

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Instituto de Ciências Exatas

Programa de Pós-Graduação em Estatística

Emilly Malveira de Lima

**Predição em análise de sobrevivência:
aplicação em estudo envolvendo óbito de
pacientes chagásicos cardiopatas**

Belo Horizonte

2019

Emilly Malveira de Lima

Predição em análise de sobrevivência: aplicação em estudo envolvendo óbito de pacientes chagásicos cardiopatas

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Minas Gerais como requisito parcial para obtenção do título de Mestre em Estatística

Orientador: Enrico Antônio Colosimo

Coorientadora: Maria do Carmo Pereira Nunes

Belo Horizonte

2019

© 2019, Emilly Malveira de Lima.
Todos os direitos reservados

Ficha catalográfica elaborada pela bibliotecária Irénquer Vismeg Lucas
Cruz - CRB 6ª Região nº 819

Lima, Emilly Malveira de.

L732p Predição em análise de sobrevivência: aplicação em estudo envolvendo óbito de pacientes chagásicos cardiopatas / Emilly Malveira de Lima. — Belo Horizonte, 2019.
57 f.: il.; 29 cm.

Dissertação (mestrado) - Universidade Federal de Minas Gerais – Departamento de Estatística

Orientador: Enrico Antônio Colosimo

1. Estatística – Teses. 2. Teoria da previsão – Medicina - Teses. 3. Análise de sobrevivência (Biometria) - Teses. 4. Doença de Chagas - Teses.
I. Orientador. II. Título.

CDU 519.2(043)



UNIVERSIDADE FEDERAL DE MINAS GERAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

UFMG

ATA DA DEFESA DA DISSERTAÇÃO DA ALUNA EMILLY MALVEIRA DE LIMA

Realizou-se, no dia 08 de outubro de 2019, às 14:00 horas, 2076 ICEX, da Universidade Federal de Minas Gerais, a 245ª defesa de dissertação, intitulada *Predição em análise de sobrevivência: Aplicação em estudo envolvendo óbito de pacientes chagásicos cardiopatas*, apresentada por EMILLY MALVEIRA DE LIMA, número de registro 2017664248, graduada no curso de ESTATÍSTICA, como requisito parcial para a obtenção do grau de Mestre em ESTATÍSTICA, à seguinte Comissão Examinadora: Prof(a). Enrico Antonio Colosimo - Orientador (DEST/UFMG), Prof(a). Antonio Carlos Pedroso de Lima (IME/USP), Prof(a). Maria do Carmo Pereira Nunes (FM/UFMG), Prof(a). Vinícius Diniz Mayrink (DEST/UFMG).

A Comissão considerou a dissertação:

Aprovada

Reprovada

Finalizados os trabalhos, lavrei a presente ata que, lida e aprovada, vai assinada por mim e pelos membros da Comissão.
Belo Horizonte, 08 de outubro de 2019.

Prof(a). Enrico Antonio Colosimo (Doutor)

Prof(a). Antonio Carlos Pedroso de Lima (Doutor)

Prof(a). Maria do Carmo Pereira Nunes (Doutora)

Prof(a). Vinícius Diniz Mayrink (Doutor)

Às minhas mães.

Resumo

Modelos de predição de risco têm sido aplicados em diversas áreas, especialmente na área médica. Em muitas situações, podem servir como uma ferramenta de auxílio clínico, ajudando a definir grupos de risco pela gravidade e na tomada de decisão acerca de um tratamento mais adequado. Há casos em que o interesse pode ser, por exemplo, prever o risco de um paciente ir a óbito em um determinado tempo, para isso são incorporadas variáveis (marcadores clínicos) registradas no início do estudo e ao longo do tempo. Este trabalho foi motivado pela necessidade de construir um escore de risco para pacientes chagásicos cardiopatas, da coorte prospectiva do projeto SaMi-Trop. Os pacientes, residentes em municípios do norte de Minas Gerais, foram acompanhados, inicialmente, por dois anos e a partir daí surgiu o interesse em construir um escore para o risco de óbito em dois anos, com informações da linha de base. O seguimento do estudo permitiu que a natureza do risco de óbito seja dinâmica, pois há marcadores clínicos que mudam ao longo do tempo, e partir disto há a necessidade de atualização do escore de risco. Neste trabalho, foi investigado o emprego de quatro abordagens: *Naive 1*, *Naive 2*, *landmark (LM)* e modelagem conjunta (*JM*). Esses métodos baseiam-se no modelo de regressão de Cox e na classe de modelos conjuntos para dados longitudinais de sobrevivência. Realizou-se a predição na linha de base e dinâmica e, avaliou-se a capacidade preditiva de cada método usando medidas de *discriminação* e *calibração*. Os resultados sugeriram que todas as abordagens possuem discriminação satisfatória, no entanto, nem todos apresentaram boa calibração e retornaram estimativas das probabilidades pouco acuradas.

Palavras-chaves: dados longitudinais e de sobrevivência, modelo de Cox, modelagem conjunta, *landmark*, predição de risco.

Abstract

Risk prediction models have been used in many areas, especially in the medical field. In many situations, they can be used as a clinical aid tool to help, for example, to define severity risk groups, and decide about the most appropriate treatment. There are circumstances where the interest may be, for example, to predict the risk of a patient being death at a given time, incorporating variables (clinical markers) recorded at baseline and over time. This work was motivated by the need to construct a risk score for patients with Chagas cardiomyopathy from the prospective SaMi-Trop cohort. Patients residing in 21 municipalities of the northern Minas Gerais were followed, initially, for two years and subsequently, it was of interest to create a two-year death risk score with baseline information. The follow-up of the study allowed the nature of the risk of death to be dynamic, whereas there are clinical markers that change over time, and therefore the risk score needs to be updated. In this article, we investigate the use of four approaches: *Naive 1*, *Naive 2*, landmark (*LM*), and joint modelling (*JM*). These methods are based on the Cox regression model and the class of models for longitudinal and survival data. It was performed static and dynamic predictions, and the performance of each method was assessed using discrimination and calibration measures. The suggested results for all approaches have satisfactory discrimination, however, not all shown have a good calibration and returned inaccurate probabilities.

Key-words: longitudinal and survival data, Cox model, joint modelling, landmarking, risk prediction.

Sumário

1	Introdução	9
1.1	Predição do risco	9
1.2	Objetivos	11
1.3	Organização da dissertação	11
2	Predição estática - linha de base	13
2.1	Conceitos básicos	13
2.2	Modelo de regressão de Cox	14
2.2.1	Inferência no modelo de Cox	14
2.2.2	Estimando $S(t \mathbf{x}_0)$	15
2.2.3	Suposição de taxas de falha proporcionais	17
2.3	Validação do modelo de predição	17
2.3.1	Métodos de validação	18
2.4	Medidas de avaliação	18
2.4.1	Discriminação	18
2.4.2	Calibração	19
2.4.3	Exemplo numérico	21
2.5	Exemplo real: Contagem de CD4	24
2.5.1	Validação do modelo: discriminação e calibração	25
2.5.1.1	Treino e teste	25
2.5.1.2	Validação cruzada	28
3	Predição dinâmica	33
3.1	Abordagem ingênua	33
3.2	Landmark	34
3.3	Modelagem conjunta	35
3.3.1	Submodelo longitudinal	35
3.3.2	Submodelo de sobrevivência	35
3.3.3	Estimação	36
3.3.4	Diagnóstico do modelo	38
3.3.5	Predição	38
3.4	Validação do modelo	39
3.5	Exemplo: contagem de CD4	39

4	Aplicação	42
4.1	Descrição dos dados	42
4.2	Modelos	43
4.2.1	Linha de base	43
4.2.2	Dinâmico	45
4.3	Predições	50
5	Conclusões	53
5.1	Considerações finais	53
	Referências	55

1 Introdução

1.1 Predição do risco

A modelagem estatística para dados clínicos, muitas vezes, é usada para construir modelos de predição para obter o risco de um indivíduo experimentar um evento em uma determinada janela de tempo. Esse risco trata-se da probabilidade da ocorrência de um evento até um tempo fixado e, o evento a ser observado pode ser: o diagnóstico/cura de uma doença, suspensão do tratamento com determinada medicação e principalmente o óbito.

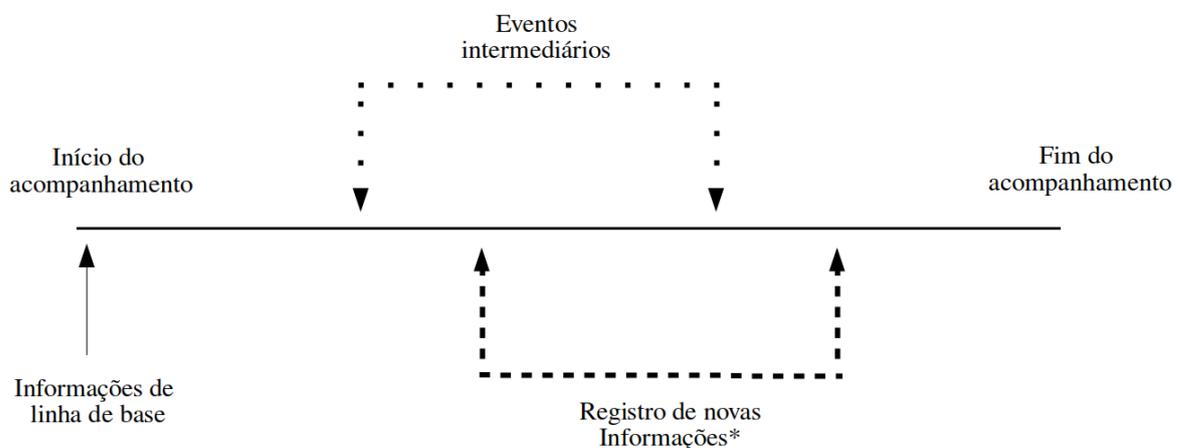
Quando a resposta de interesse for o tempo até a ocorrência do evento, pode acontecer do evento não ser observado para todos os indivíduos e se ter uma informação incompleta desse tempo. Os dados gerados são típicos de *análise de sobrevivência*, área da estatística cuja principal característica é a presença de *censura*, que é uma informação parcial da variável resposta. Alguns mecanismos de censura são diferenciados em estudos clínicos e a situação frequentemente encontrada é a censura à direita, pois o tempo de ocorrência do evento de interesse será a direita do tempo registrado (Colosimo e Giolo, 2006). Nesta dissertação serão tratados estudos nos quais ocorre esse tipo de mecanismo.

Na prática, modelos de regressão são utilizados para relacionar covariáveis (marcadores clínicos) do indivíduo com a resposta de interesse e assim, prever o risco. A técnica mais utilizada para construir esses modelos em análise de sobrevivência é o modelo de regressão de Cox (Cox, 1972), que utiliza as informações dos marcadores registradas no início do estudo (linha de base) e o tempo observado até a ocorrência evento. Com essa abordagem a estimação se baseia em informações da linha de base, tornando a predição estática.

Estudos clínicos que envolvem uma resposta temporal são frequentemente prospectivos e de longa duração (Colosimo e Giolo, 2006), gerando novas medições para marcadores que variam ao longo do tempo. Um exemplo, que na verdade é a motivação prática desse trabalho, é a coorte prospectiva do Projeto SaMi-Trop (*São Paulo-Minas Gerais Tropical Medicine Research Center*). O projeto conduz um estudo envolvendo pacientes chagásicos cardiopatas, residentes em 21 municípios do norte do Estado de Minas Gerais. Informações desses pacientes, como idade, sexo, frequência cardíaca e o exame de eletrocardiograma, foram registradas na linha de base, determinada entre os anos de 2013 e 2014; posteriormente uma nova visita foi realizada entre 2015 e 2016. A partir das informações obtidas no início do estudo, surgiu o interesse em estimar o risco desses pacientes irem a óbito em 2 anos, ou seja, dadas as informações registradas no início do estudo,

obter a probabilidade de o óbito ocorrer em 2 anos. Neste caso, tem-se exatamente o cenário de predição estática, citado anteriormente, e como aconteceu uma nova coleta de dados entre os anos de 2015 e 2016, há também a necessidade de atualizar o risco predito para aqueles pacientes que sobreviveram após os primeiros dois anos do estudo, tendo-se assim a predição dinâmica. A caracterização completa do estudo pode ser conferida em Cardoso et al. (2016).

Um exemplo desse tipo de estudo é exibido na Figura 1, em que, ao longo do acompanhamento, são observados eventos intermediários e registradas novas informações para aqueles pacientes que se encontram vivos no instante da nova visita.



* Linha de base dinâmica (sobreviventes)

Figura 1 – Exemplo de estudo clínico de longa duração com informações coletadas ao longo do acompanhamento.

Com novas informações do marcador longitudinal, o objetivo passa a ser atualizar a estimativa do risco daqueles pacientes que não experimentaram o evento de interesse (sobreviventes, como ilustrado na Figura 1). Uma forma ingênua de obter a atualização do risco é realizar a nova predição com o modelo construído na linha de base, porém, dessa maneira, obteriam-se estimativas do risco pouco eficientes, já que as estimativas para os parâmetros do modelo vêm apenas das informações da linha de base e a nova informação do marcador não é acrescentada na estimação. Portanto, torna-se necessário incorporar as novas informações durante o desenvolvimento do modelo visando obter predições mais acuradas.

Na literatura, há diversos trabalhos que abordam metodologias envolvendo esse tipo de situação quando se tem dados longitudinais e de sobrevivência. Tsiatis et al. (1995) e Wulfsohn e Tsiatis (1997), tratam da modelagem para esse tipo de dados conjuntamente, lidando tanto a parte de medidas repetidas dos marcadores longitudinais,

quanto o processo do tempo de sobrevivência. Com relação à predição baseada nesse tipo de modelo, Proust-Lima e Taylor (2009) Rizopoulos (2011), Njagi et al. (2013) trazem propostas dinâmicas, usando a modelagem conjunta, que possibilitam obter as estimativas do risco de ocorrência do evento de interesse.

Outra proposta para realizar predições dinâmicas segue o conceito de “*landmark*”. Foi introduzido por Anderson et al. (1983) com o intuito de analisar o tempo de sobrevivência com variáveis que mudam ao longo do tempo e tomando pontos de referência (linha de base dinâmica), após o início do estudo, onde foram registradas novas informações dessas variáveis. Zheng e Heagerty (2005) e Houwelingen (2007) apresentam propostas para realizar as predições a partir desse conceito. Zheng e Heagerty (2005) propuseram abordagem semiparamétrica que incorpora uma medida específica registrada para as variáveis tempo-dependente, chamando de modelagem condicional; Maziarz et al. (2017) apresentam extensões para os estimadores dos parâmetros do modelo proposto em 2005. Houwelingen (2007) desenvolveu uma metodologia na qual considera efeitos dinâmicos para as covariáveis longitudinais. Rizopoulos et al. (2017) realizam uma aplicação considerando covariáveis tempo-dependente utilizando *landmark* padrão.

Quando se trata de construir modelos de predição, a avaliação da performance desses modelos é um dos passos necessários para utilizá-los como uma ferramenta de auxílio clínico. Há diferentes meios para avaliar a capacidade de prever eventos futuros. Dois conceitos bastante comuns na avaliação de um modelo é a *discriminação* e *calibração*. Steyerberg et al. (2010), D’Agostino e Nam (2003), Pepe (2003), Harrell Jr. (2006), Cook (2007) abordam estes conceitos voltados para modelos de predição clínica e dados de sobrevivência.

1.2 Objetivos

O objetivo desta dissertação é comparar as metodologias de predição em dados longitudinais de sobrevivência e empregá-las para obter predições dinâmicas do risco de óbito para pacientes da Coorte do SaMi-Trop que sobreviveram a dois anos de acompanhamento e tiveram seus marcadores atualizados após a linha de base. Apresentando as etapas de construção de um modelo de predição tanto na linha de base, quanto dinâmica, de acordo com as abordagens: *Naive* (ingênuo), *Landmark* e Modelagem Conjunta.

1.3 Organização da dissertação

No Capítulo 2 são apresentados alguns conceitos básicos em análise de sobrevivência, o modelo de regressão de Cox para predição estática do risco, as técnicas para avaliação da performance do modelo em termos de discriminação e calibração e, final-

mente, a exemplificação com um conjunto de dados sobre pacientes com HIV no qual o marcador longitudinal é a contagem de células CD4. Esse já é bastante explorado na literatura.

No Capítulo 3 são apresentadas as metodologias com enfoque para a predição dinâmica e novamente os dados de CD4 são utilizados para exemplificar, agora tomando predições com a modelagem conjunta, *landmark* e *naive*.

No Capítulo 4 empregam-se as metodologias abordadas nos Capítulos anteriores à base de dados da coorte do projeto SaMi-Trop para realizar a predição do risco de óbito dos pacientes em até 2 e 4 anos, sendo que: para dois anos utiliza-se com as informações da linha de base e para quatro anos, considera-se até o segundo registro do marcador longitudinal.

O Capítulo 5 encerra esta dissertação com um resumo indicando as principais conclusões.

2 Predição estática - linha de base

Neste capítulo são apresentados alguns conceitos básicos em análise de sobrevivência e as etapas para construção do modelo de predição com dados de tempo de vida. O modelo de Cox é empregado com enfoque na realização da predição a partir da linha de base e uma aplicação a um conjunto de dados, já explorado na literatura, ilustra essa metodologia.

2.1 Conceitos básicos

A análise de sobrevivência destina-se à dados cuja resposta de interesse é o tempo de vida de um indivíduo (ou tempo de falha). Esse tempo é observado desde o ponto inicial do estudo até o momento da ocorrência do evento de interesse (falha). No entanto, muitas vezes, nem todos os indivíduos irão experimentar o evento, fazendo com que o tempo observado não seja o tempo de falha. Este fato define a principal característica dos dados de sobrevivência, a *censura*, que representa uma informação incompleta a respeito do tempo de falha.

O tempo até a ocorrência de um evento é representado pela variável aleatória não negativa T e o tempo de censura pela variável aleatória C , sendo T e C independentes. A observação do tempo de falha é dada por $t = \min\{T, C\}$ juntamente com uma variável indicadora $\delta = I(T \leq C)$ que vale 1 para $T \leq C$ e 0. para o contrário.

É de grande interesse fazer inferência sobre a distribuição de T , que tem como função de distribuição acumulada: $F(t) = P(T \leq t)$ e densidade: $f(t) = dF(t)/dt$. Uma outra função fundamental para interpretação do tempo de falha é a função de sobrevivência, que fornece a probabilidade de um indivíduo sobreviver até o tempo t , tal que,

$$S(t) = P(T > t) = 1 - F(t). \quad (2.1)$$

Outras duas funções fundamentais para descrever a distribuição de T são, respectivamente, a função taxa de falha e a função taxa de falha acumulada:

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

$$\Lambda(t) = \int_0^t \lambda(u) du = -\log[S(t)].$$

Essas três funções se relacionam entre si, de tal forma que:

$$\lambda(t) = f(t)/S(t) = -d \log[S(t)]/dt,$$

$$S(t) = \exp\{-\Lambda(t)\}. \quad (2.2)$$

A expressão em (2.2) é fundamental para, posteriormente, obter o risco de ocorrência do evento até o tempo t .

2.2 Modelo de regressão de Cox

O modelo de regressão de Cox é o mais popular para a análise de dados em que se deseja avaliar a relação de covariáveis com o tempo de falha. É uma abordagem semiparamétrica que tem a equação do modelo definida em termos da função taxa de falha. Sua especificação segue a forma:

$$\lambda(t \mid \mathbf{x}, y(t)) = \lambda_0(t)e^{\mathbf{x}^\top \boldsymbol{\beta} + y(t)\gamma}, \quad (2.3)$$

em que $\lambda_0(t)$ é chamada função taxa de falha basal (componente não-paramétrica e não negativa), $\boldsymbol{\beta}$ é um vetor com os coeficientes associados às covariáveis $\mathbf{x} = (x_1, \dots, x_p)^\top$ e γ quantifica o efeito da covariável tempo-dependente, $y(t)$, com seus valores medidos até t .

Quando o estudo envolver apenas covariáveis fixas, tem-se um caso particular do modelo em (2.3), que é conhecido como modelo de Cox com *taxas de falha proporcionais* (modelo TFP); ou seja,

$$\lambda(t \mid \mathbf{x}) = \lambda_0(t)e^{\mathbf{x}^\top \boldsymbol{\beta}}. \quad (2.4)$$

A proporcionalidade das taxas é a principal suposição do modelo TFP, pois ao tomar a razão das taxas de dois indivíduos i e j , resulta em uma quantidade em que não há a função taxa de falha basal, portanto, constante ao longo do tempo. Escreve-se

$$\frac{\lambda(t \mid \mathbf{x}_i)}{\lambda(t \mid \mathbf{x}_j)} = \frac{\lambda_0(t)e^{\mathbf{x}_i^\top \boldsymbol{\beta}}}{\lambda_0(t)e^{\mathbf{x}_j^\top \boldsymbol{\beta}}} = \exp\{\mathbf{x}_i^\top \boldsymbol{\beta} - \mathbf{x}_j^\top \boldsymbol{\beta}\}.$$

2.2.1 Inferência no modelo de Cox

Considere uma amostra de tamanho n em que, para o i -ésimo indivíduo, foram observadas as covariáveis fixas e tempo-dependente, representadas pelo vetor $\mathbf{x}_i(t) = (\mathbf{x}_i, y_i(t))^\top$, o tempo de falha, t_i , e a indicadora de falha, δ_i ($i = 1, \dots, n$), formando $\{t_i, \delta_i, \mathbf{x}_i(t)\}$.

Neste trabalho, estimação dos coeficientes em (2.3) será baseada no método de *máxima verossimilhança parcial*, proposto por Cox (1975). A função de verossimilhança do modelo possui uma característica especial, pois é dada em termos do produto das verossimilhanças dos k ($k \leq n$) indivíduos para os quais foi observado o evento de interesse,

em tempos distintos $t_1 < \dots < t_k$. A verossimilhança parcial para o modelo de Cox com covariável dependente do tempo é escrita como segue:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^k \frac{e^{\mathbf{x}_i^\top(t_i)\boldsymbol{\theta}}}{\sum_{j \in R(t_i)} e^{\mathbf{x}_j^\top(t_i)\boldsymbol{\theta}}} = \prod_{i=1}^n \left(\frac{e^{\mathbf{x}_i^\top(t_i)\boldsymbol{\theta}}}{\sum_{j \in R(t_i)} e^{\mathbf{x}_j^\top(t_i)\boldsymbol{\theta}}} \right)^{\delta_i}, \quad (2.5)$$

em que $\boldsymbol{\theta} = (\boldsymbol{\beta}, \gamma)$ é o vetor de parâmetros, $R(t_i)$ é o conjunto de índices dos indivíduos que não experimentaram o evento ou não foram censurados até t_i , ou seja, indivíduos cujo tempo evento/censura é maior que t_i . Cada termo da verossimilhança parcial, pode ser entendido como a probabilidade condicional de escolher, no conjunto de indivíduos sob risco, um indivíduo i que irá experimentar o evento, dado um conjunto de indivíduos sob risco no tempo t_i e dado que uma falha ocorreu em t_i . Note ainda, que para cada conjunto $R(t_i)$, a covariável tempo-dependente assumirá um único valor específico. Quando houver empates nos tempos observados $L(\boldsymbol{\theta})$ é modificada e a aproximação proposta por Breslow (1974) é usualmente empregada.

Os estimadores de $\boldsymbol{\theta}$ que maximizam $L(\boldsymbol{\theta})$ são encontrados resolvendo o sistema de equações definido pela função escore $U(\boldsymbol{\theta}) = \partial \log L(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$, que consiste em obter as raízes de:

$$U(\boldsymbol{\theta}) = \sum_{i=1}^n \delta_i \left\{ \mathbf{x}_i(t_i) - \frac{\sum_{j \in R(t_i)} \mathbf{x}_j(t_i) e^{\mathbf{x}_j^\top(t_i)\boldsymbol{\theta}}}{\sum_{j \in R(t_i)} e^{\mathbf{x}_j^\top(t_i)\boldsymbol{\theta}}} \right\} = 0. \quad (2.6)$$

Na situação em que o ajuste for feito para o modelo TFP, os estimadores serão encontrados a partir da amostra com informações da linha de base para todas as covariáveis, enquanto que para o modelo com covariável tempo-dependente, amostra é composta por todas as observações longitudinais disponíveis desta covariável. As propriedades assintóticas destes estimadores são bem estabelecidas na literatura; ver Andersen e Gill (1982). A partir disto, é possível construir intervalos de confiança e realizar testes de hipóteses a respeito dos coeficientes do modelo utilizando as estatísticas de teste de Wald e razão de verossimilhança (RV). Mais detalhes sobre o modelo de Cox e suas aplicações na área médica podem ser encontrados em Colosimo e Giolo (2006), Therneau e Grambsch (2000), Collett (2015), entre outros.

2.2.2 Estimando $S(t|\mathbf{x}_0)$

A partir da estimação dos coeficientes do modelo de Cox, o interesse é obter a função de sobrevivência para os indivíduos na amostra. Considere $\mathbf{x}_0 = \{\mathbf{x}, y(t=0)\}$ um vetor com as covariáveis de linha de base para um certo indivíduo, com isso um estimador para $S(t|\mathbf{x}_0)$ é dado por

$$\hat{S}(t | \mathbf{x}_0) = \left[\hat{S}_0(t) \right]^{\exp\{\mathbf{x}_0^\top \hat{\boldsymbol{\theta}}\}}, \quad (2.7)$$

em que $\hat{\boldsymbol{\theta}}$ é a solução de (2.6), $\hat{S}_0(t) = \exp\{-\hat{\Lambda}_0(t)\}$ é a sobrevida basal obtida por meio da relação com a função taxa de falha acumulada basal que, por sua vez, é estimada de forma não-paramétrica a partir do estimador de Nelson-Aalen. Então, seja δ_j o número de falhas em t_j , a função de risco acumulado basal do modelo de Cox é dada pela expressão a seguir:

$$\hat{\Lambda}_0(t) = \sum_{j:t_j < t} \frac{\delta_j}{\sum_{j \in R(t_j)} \exp\{\mathbf{x}_j^\top(t_j) \hat{\boldsymbol{\theta}}\}}, \quad (2.8)$$

em que $\mathbf{x}(t)$ é o vetor com covariáveis incluindo as fixas e a covariável tempo-dependente $y(t)$. Note que, novamente, é necessário usar a amostra inicial para obter tais quantidades, ou seja, há dois processos de estimação, o primeiro envolvendo a amostra completa para obter as estimativas dos coeficientes e o segundo refere-se a estimação da função de sobrevivência via relação com a função taxa de falha basal.

O caso particular desse estimador é dado quando o enfoque é o modelo de Cox TFP, em que todas as covariáveis são fixas. Nesse caso, sob certas condições de regularidade e para um t fixo, assintoticamente $\hat{S}(t|\mathbf{x}_0)$ tem uma distribuição normal com média $S(t|\mathbf{x}_0)$ e variância que pode ser estimada por (Klein e Moeschberger, 2003):

$$\widehat{Var}[\hat{S}(t|\mathbf{x}_0)] = [\hat{S}(t|\mathbf{x}_0)]^2 [q_1(t) + q_2(t | \mathbf{x}_0)], \quad (2.9)$$

em que

$$q_1(t) = \sum_{j:t_j < t} \frac{\delta_j}{\left(\sum_{j \in R(t_j)} \exp\{\mathbf{x}_j^\top \hat{\boldsymbol{\beta}}\} \right)^2} \quad (2.10)$$

é um estimador para a variância de $\hat{\Lambda}_0(t)$ se $\hat{\boldsymbol{\beta}}$ for o verdadeiro valor de $\boldsymbol{\beta}$ e

$$q_2(t | \mathbf{x}_0) = \mathbf{q}_3^\top(t | \mathbf{x}_0) \widehat{Var}(\hat{\boldsymbol{\beta}}) \mathbf{q}_3(t | \mathbf{x}_0) \quad (2.11)$$

reflete a incerteza no processo de estimação dos coeficientes, com \mathbf{q}_3 sendo um vetor p dimensional em que seu k -ésimo ($k = 1, \dots, p$) elemento é definido por:

$$q_3(t | \mathbf{x}_0)_k = \sum_{j:t_j < t} \left[\frac{\sum_{j \in R(t_j)} x_{jk} \exp\{\mathbf{x}_j^\top \hat{\boldsymbol{\beta}}\}}{\sum_{j \in R(t_j)} \exp\{\mathbf{x}_j^\top \hat{\boldsymbol{\beta}}\}} - x_{0k} \right] \left[\frac{\delta_j}{\sum_{j \in R(t_j)} \exp\{\mathbf{x}_j^\top \hat{\boldsymbol{\beta}}\}} \right]. \quad (2.12)$$

A partir dessa variância é possível calcular intervalos de confiança para $S(t|\mathbf{x}_0)$. Ainda segundo Klein e Moeschberger (2003), intervalos que consideram as transformações como logaritmo ou raiz quadrada do arco seno, de $S(t|\mathbf{x}_0)$, têm melhor performance que o intervalo sem transformação.

Obter a variância de (2.7) quando se tem o estimador (2.8) com covariável tempo-dependente não é trivial (Klein e Moeschberger, 2003; Therneau e Grambsch, 2015;

Elashoff et al., 2016). Elashoff et al. (2016) sugere que nesse caso, o método *bootstrap* (Efron e Tibshirani, 1994) deva ser usado para construir um intervalo de confiança. Portanto, neste trabalho optou-se por calcular intervalos de confiança usando o método de *bootstrap*, que consiste em re-estimar os parâmetros do modelo (TFP e covariável tempo-dependente) a partir de reamostras *bootstrap* geradas da amostra original e assim obter os erros padrões dos estimadores.

Ressalta-se que o interesse é realizar a predição do risco de ocorrência do evento em um tempo t , pré-especificado. A formulação do estimador para essa probabilidade, a partir do modelo de Cox, é dada pela expressão em (2.13):

$$\hat{r}(t | \mathbf{x}_0) = 1 - \left[\hat{S}_0(t) \right]^{\exp\{\mathbf{x}_0^\top \hat{\boldsymbol{\theta}}\}}. \quad (2.13)$$

2.2.3 Suposição de taxas de falha proporcionais

Quando se trata de construir um modelo de predição, a partir do modelo de regressão de Cox, um outro ponto importante que deve ser levado em conta, é a adequação do modelo em relação à suposição de taxas de falha proporcionais. Embora seja bastante flexível, devido à presença do componente não-paramétrico, ele não se ajusta a qualquer situação clínica (Colosimo e Giolo, 2006). Um método gráfico para avaliar essa suposição foi proposto por Grambsch e Therneau (1994), baseado na visualização dos resíduos de Schoenfeld *versus* o tempo, para cada covariável. Constatar a adequação do modelo é passo que deve ser realizado antes mesmo de verificar a validade do modelo com relação a capacidade preditiva (discriminação e calibração).

2.3 Validação do modelo de predição

Uma etapa importante após a estimação do modelo de predição é avaliar a capacidade preditiva do modelo. A performance de um modelo de predição na área médica, especialmente quando se tem resposta dicotômica e de sobrevivência, é, na maioria das vezes, avaliada por sua *discriminação* e *calibração* [Woodward et al. (2017); Steyerberg et al. (2010); Steyerberg (2008); Harrell (2015); D'Agostino e Nam (2003)]. A ideia básica desses dois conceitos é que: um modelo válido é capaz de discriminar, de forma satisfatória, casos de não casos (indivíduos que desenvolveram o evento até t , daqueles que não) e, juntamente com isto, realizar predições acuradas, ou seja, que o risco predito seja próximo do que é observado. Steyerberg et al. (2010) apresenta uma revisão de algumas medidas, que tradicionalmente são utilizadas para avaliação do modelo, como a Área sob a Curva ROC (*Area under the ROC curve*) para discriminação e a estatística de Hosmer-Lemeshow para calibração. Essas técnicas são apresentadas na Seção 2.4.

2.3.1 Métodos de validação

Em se tratando de um modelo de predição, é necessário avaliar sua validade em uma amostra diferente da que foi realizada a estimação dos parâmetros, a fim de evitar estimativas superestimadas (*overfitting*). Há dois métodos principais para validação de modelo: *interna* e *externa* (Harrell, 2015). A validação externa considera testar o modelo em uma amostra cujos indivíduos possuem características diferentes, seja por questões culturais ou temporais. Já na interna, o modelo é testado em uma parcela da amostra disponível para desenvolvimento e há basicamente três maneiras de implementá-la:

1) **treino e teste**: a amostra original é dividida de maneira que uma parte seja utilizada para desenvolver o modelo (amostra de treino) e outra para testar (amostra de teste). Geralmente usa-se uma proporção de 70%-30% ou 80%-20% para treino-teste;

2) **validação cruzada**: que pode ser entendida como uma generalização do treino e teste e consiste em dividir a amostra em k subamostras (*k-folds*), usualmente $k = 10$, e permutar entre treinamento e teste. Para cada k realização é calculada a medida de validação de interesse e, ao final, retirada a média entre as k realizações. Este procedimento pode ser repetido um certo número de vezes, visando ter resultados mais consistentes do que quando se valida o modelo em apenas uma amostra de teste.

3) **bootstrap**: neste método a amostra de desenvolvimento é reamostrada, com reposição, B vezes e o modelo é validado calculando as medidas de desempenho nas B amostras geradas.

Logo, ao validar o modelo utilizando esses métodos, busca-se encontrar resultados mais confiáveis das medidas de avaliação e evitar a escolha de um modelo de predição que produza resultados super ou subestimados, ou seja, que irá predizer um risco alto para um paciente que tem risco baixo e vice-versa. Mais detalhes sobre esses métodos, suas vantagens e desvantagens podem ser encontrados em Harrell (2015).

2.4 Medidas de avaliação

2.4.1 Discriminação

A Área sob a Curva ROC (AUC), também conhecida como Estatística C, é a medida mais utilizada quando se quer avaliar a discriminação do modelo de predição (Woodward et al., 2017). Tal medida é obtida em função da Curva ROC (*Receiver Operating Characteristic Curve*), que reúne o conjunto de pares de *sensibilidade*: taxa de verdadeiros positivos, ou seja, eventos classificados corretamente como eventos; e *especificidade*: taxa de verdadeiros negativos, não-eventos classificados corretamente como não-eventos, para diferentes pontos de corte (valores do risco) e, a área sob esta curva é justamente o que dá a interpretabilidade sobre o desempenho do modelo ao discriminar evento de não-evento.

A AUC é definida por Hanley e McNeil (1982) seguinte forma:

$$\text{AUC}(t) = \Pr[r_i(t | \mathbf{x}_i^*) > r_j(t | \mathbf{x}_j^*) | \delta_i = 1, \delta_j = 0], \quad i, j = 1, \dots, n \quad (2.14)$$

em que $r_i(t | \mathbf{x}_i^*)$ e $r_j(t | \mathbf{x}_j^*)$ são, respectivamente, os riscos do i -ésimo indivíduo, que teve o evento observado, e do j -ésimo indivíduo, para qual o evento não foi observado, dadas todas suas informações avaliadas na linha de base \mathbf{x}_i^* e \mathbf{x}_j^* , respectivamente. Com isso, entende-se que esta probabilidade (2.14) é exatamente o que vai quantificar a “chance” do modelo discernir se, um paciente i , que experimentou o evento em t , tem um risco maior que um paciente j que não experimentou o evento.

Segundo Houwelingen e Putter (2012), uma estimativa para a área sob a Curva ROC é construída em termos da subamostra de pacientes que não tiveram o evento observado, n_0 , e daqueles que sim, n_1 , e também, de acordo com seus respectivos riscos preditos, $\hat{r}_{0i}(t | \mathbf{x}_i^*)$, $\hat{r}_{1j}(t | \mathbf{x}_j^*)$. Sendo assim, um estimador para a AUC é:

$$\widehat{\text{AUC}}(t) = \frac{\sum_{j=1}^{n_1} \sum_{i=1}^{n_0} (I\{\hat{r}_{1j}(t | \mathbf{x}_j^*) > \hat{r}_{0i}(t | \mathbf{x}_i^*)\} + \frac{1}{2} I\{\hat{r}_{1j}(t | \mathbf{x}_j^*) = \hat{r}_{0i}(t | \mathbf{x}_i^*)\})}{n_0 n_1}. \quad (2.15)$$

Na primeira parcela das somas, há uma função indicadora que vale 1, quando o risco de um paciente j , que experimentou o evento, for maior que o risco de um paciente i , que não experimentou o evento, e quando houver empate entre os riscos desses indivíduos, é somado o valor de 1/2. Note que a Curva ROC possuirá uma forma em que, abaixo dela, haverá áreas retangulares e triangulares entre cada ponto de corte, e cada um gera pares de sensibilidade e especificidade diferentes. Sendo assim, o estimador da área sob a curva pode ser entendido como a soma das áreas retangulares (quando não há empate nos riscos comparados: primeira parcela da soma) e triangulares (quando há empate no riscos comparados: segunda parcela da soma). Mais detalhes sobre a Curva ROC na avaliação de testes para predição clínica podem ser encontrados em Pepe (2003).

O modelo de predição, segundo a medida AUC, tem discriminação perfeita se $\text{AUC} = 1$, o que na prática é improvável que aconteça. Quando $\text{AUC} = 0.5$, o modelo é dito sem discriminação, ou seja, identifica ao acaso um indivíduo que irá ou não experimentar o evento.

2.4.2 Calibração

Uma medida de calibração é utilizada para avaliar, em termos gerais, a acurácia do modelo de predição, ou seja, se os riscos estimados estão próximos dos riscos observados. Enquanto na discriminação o interesse é verificar se o modelo é capaz de distinguir os eventos futuros dos não-eventos, agora deseja-se saber se, quando comparadas, as probabilidades preditas para indivíduos aproximam-se do que foi observado.

Woodward (2013) identifica quatro tipos de calibração: *perfeita*, *global*, *média* e *por grupos* e, ainda ressalta que, entre esses quatro tipos, o teste mais efetivo para avaliar a calibração é *por grupos*. Esses grupos são divididos a partir da ordenação dos riscos e comparados com os respectivos valores observados nos grupos. O teste de Hosmer-Lemeshow (Hosmer e Lemeshow, 1980) é a forma mais comum de avaliar se um modelo é adequado, sob a hipótese nula de que se tem um modelo calibrado. A princípio, esse teste foi proposto para avaliar a adequação do modelo de regressão logística, mas há adaptações quando se trata de modelos para dados de sobrevivência. Uma extensão da sua estatística de teste proposta por D'Agostino e Nam (2003), que incorpora estimativas do estimador de Kaplan-Meier, é

$$HL_{DN}(t) = \sum_{j=1}^g n_j \frac{[(1 - \widehat{KM}_j(t)) - \hat{r}_j(t | \mathbf{x}^*)]^2}{\hat{r}_j(t | \mathbf{x}^*)(1 - \hat{r}_j(t | \mathbf{x}^*))} \stackrel{H_0}{\sim} \chi_{g-1}^2,$$

em que g é o número de grupos (usualmente $g=10$) e para o j -ésimo grupo, $\widehat{KM}_j(t)$ que é o estimador de Kaplan-Meier, avaliado no período de tempo especificado, e a média do risco predito, $\hat{r}_j(t | \mathbf{x}^*)$. A estatística HL_{DN} , sob H_0 , tem distribuição *qui-quadrado* com $g - 1$ graus de liberdade. Outra variação dessa estatística de teste é apresentada por (Demler et al., 2015).

Uma forma gráfica para avaliar a calibração, é por meio de um *gráfico de calibração*, no qual também são comparados os riscos esperados (preditos) com os observados. A construção do gráfico é bem similar à ideia do teste de Hosmer-Lemeshow, os riscos são divididos de acordo com os decis (grupos) e para cada decil é calculada a média dos riscos preditos e a frequência dos eventos observados (Woodward et al., 2017). No exemplo da próxima seção é possível apreciar os gráficos de calibração (Figura 6).

Para comparar a acurácia dos modelos de predição, também é considerado o *erro esperado* da predição de eventos futuros, como feito em Rizopoulos et al. (2017). A definição desta medida é dada em termos de uma média que leva em conta a soma de funções indicadoras e uma função de perda (quadrática ou absoluta), como é mostrado a seguir:

$$\begin{aligned} \widehat{PE}(t) &= \frac{1}{n(t)} \sum_{i:T_i \geq t} I(T_i \geq t)L[0 - \hat{r}_i(t | \mathbf{x}_i^*)] + \delta_i I(T_i < t)L[1 - \hat{r}_i(t | \mathbf{x}_i^*)] \\ &+ (1 - \delta_i)I(T_i < t)\{(1 - \hat{r}_i(t | \mathbf{x}_i^*))L[0 - \hat{r}_i(t | \mathbf{x}_i^*)] \\ &+ [\hat{r}_i(t | \mathbf{x}_i^*)]L[1 - \hat{r}_i(t | \mathbf{x}_i^*)]\}, \end{aligned} \quad (2.16)$$

em que $n(t)$ é o número de indivíduos sob risco em t e $L(\cdot)$ denota a função de perda. Este estimador do erro de predição foi uma proposta de Henderson et al. (2002) que acomoda as censuras presentes na amostra. Observe que nos dois primeiros termos da soma são considerados os indivíduos cujo tempo observado é maior que o tempo de predição (censurados) e os que experimentam o evento antes de t . Já o terceiro termo corresponde aos indivíduos que foram censurados antes de t .

2.4.3 Exemplo numérico

Considere uma amostra de 25 indivíduos que tiveram seus marcadores registrados e a partir de um certo tempo de acompanhamento, observou-se que 8 experimentaram o evento. A partir do modelo de Cox, foram obtidos os valores preditos do risco em $t_{\text{pred}} = 5$ meses. Ilustram-se a construção da curva ROC, o cálculo da estimativa da AUC e do erro de predição esperado (PE - com a função de perda quadrática). Na Tabela 1, observe que para um tempo de predição $t_{\text{pred}} = 5$ meses, cinco deles tiveram seus tempos censurados (do indivíduo 21 ao 25). Esses 5 indivíduos são incorporados apenas no cálculo do erro de predição esperado.

Tabela 1 – Riscos preditos ($\hat{r}_i(t_{\text{pred}}=5 \mid \mathbf{x}_i^*)$), indicadora de falha (δ_i) e tempos observados (t_i) para os 25 indivíduos do exemplo.

Indivíduo	$\hat{r}_i(t_{\text{pred}}=5 \mid \mathbf{x}_i^*)$	δ_i	t_i
1	0.31	1	1.54
2	0.26	1	1.29
3	0.55	1	2.76
4	0.05	0	6.09
5	0.47	0	5.34
6	0.48	0	6.25
7	0.82	1	0.28
8	0.37	0	6.76
9	0.55	0	5.56
10	0.17	0	5.80
11	0.63	0	6.53
12	0.89	1	2.34
13	0.28	0	6.34
14	0.40	1	2.42
15	0.77	1	4.06
16	0.67	1	1.85
17	0.20	0	5.41
18	0.36	0	5.72
19	0.36	0	5.72
20	0.69	0	6.38
21	0.11	0	1.51
22	0.14	0	3.31
23	0.11	0	3.65
24	0.15	0	2.65
25	0.08	0	1.83

Para construir a curva ROC e calcular a AUC é necessário excluir os indivíduos censurados cujo tempo observado é menor que o $t_{\text{pred}} = 5$, pois não se sabe o tempo do desfecho, ou seja, se experimentou ou não o evento. Considere os dois passos seguintes:

- (1) ordenar, de forma decrescente, os indivíduos de acordo com os riscos (Tabela 2).

(2) separar eventos de não eventos para comparar os riscos (AUC).

Tabela 2 – Riscos estimados ordenados do maior para o menor.

Indivíduo	$\hat{r}_i(t_{\text{pred}} = 5 \mid \mathbf{x}_i^*)$	δ_i	Indivíduo	$\hat{r}_i(t_{\text{pred}} = 5 \mid \mathbf{x}_i^*)$	δ_i
12	0.89	1	14	0.40	1
7	0.82	1	8	0.37	0
15	0.77	1	18	0.36	0
20	0.69	0	19	0.36	0
16	0.67	1	1	0.31	1
11	0.63	0	13	0.28	0
3	0.55	1	2	0.26	1
9	0.55	0	17	0.20	0
6	0.48	0	10	0.17	0
5	0.47	0	4	0.05	0

Diante disso, para calcular a AUC, a equação (2.15) é simplificada de tal forma que os elementos da soma são denotados por

$$a_{ij} = I\{\hat{r}_{1j}(t_{\text{pred}} \mid \mathbf{x}_j^*) > \hat{r}_{0i}(t_{\text{pred}} \mid \mathbf{x}_i^*)\} + \frac{1}{2} I\{\hat{r}_{1j}(t_{\text{pred}} \mid \mathbf{x}_j^*) = \hat{r}_{0i}(t_{\text{pred}} \mid \mathbf{x}_i^*)\},$$

sendo assim, a forma resumida para para o estimador da AUC fica sendo:

$$\widehat{AUC}(t_{\text{pred}}) = \frac{\sum_{j=1}^{n_1} \sum_{i=1}^{n_0} a_{ij}}{n_0 n_1} = \frac{\sum_{j=1}^{n_1} a_j}{n_0 n_1},$$

em que $n_0 = 12$ indivíduos na subamostra de não-eventos e $n_1 = 8$ na subamostra de eventos. De acordo com a subamostra, os subconjuntos de riscos estimados para não-eventos e eventos são, respectivamente,

- $\hat{r}_{0i} = \{0.69, 0.63, 0.55, 0.48, 0.47, 0.37, 0.36, 0.36, 0.28, 0.20, 0.17, 0.05\}$,
- $\hat{r}_{1j} = \{0.89, 0.82, 0.77, 0.67, 0.55, 0.40, 0.31, 0.26\}$.

Cada risco em \hat{r}_{1j} é comparado com os riscos em \hat{r}_{0i} , resultando nas seguintes parcelas da soma:

- para $j = 1, 2$ e $3 \implies a_j = 12$,
- para $j = 4 \implies a_j = 11$,
- para $j = 5 \implies a_j = 9 + 1/2$ (empate do risco de um evento com um não-evento),
- para $j = 6 \implies a_j = 7$,

- para $j = 7 \implies a_j = 4$,
- para $j = 8 \implies a_j = 3$.

Logo, $\sum_{j=1}^8 a_j = 70.5$ e a estimativa para a Área sob a Curva ROC é

$$\widehat{AUC}(t_{\text{pred}}) = \frac{70.5}{8 \times 12} = 0.734.$$

A curva ROC, ilustrada na Figura 2, tem sua construção feita primeiramente, ordenando as estimativas das probabilidades, da maior para a menor. A partir do par (0.0), a curva toma forma avaliando cada probabilidade (pronto de corte) e realizando as seguintes transições, de acordo com o número de eventos e não-eventos, #eventos e #não-eventos, respectivamente:

- 1) $\frac{\text{\#eventos}}{n_1}$, na vertical, quando for evento;
- 2) $\frac{\text{\#não-eventos}}{n_0}$, na horizontal, quando não for evento;
- 3) na vertical e na horizontal, quando houver empate, sendo a soma das duas razões supracitadas.

Nos dois primeiros casos, as áreas formadas sob a curva são retângulos e no terceiro caso, trapézios. No gráfico da Figura 2, são respectivamente, as áreas hachuradas com a linha contínua e tracejada.

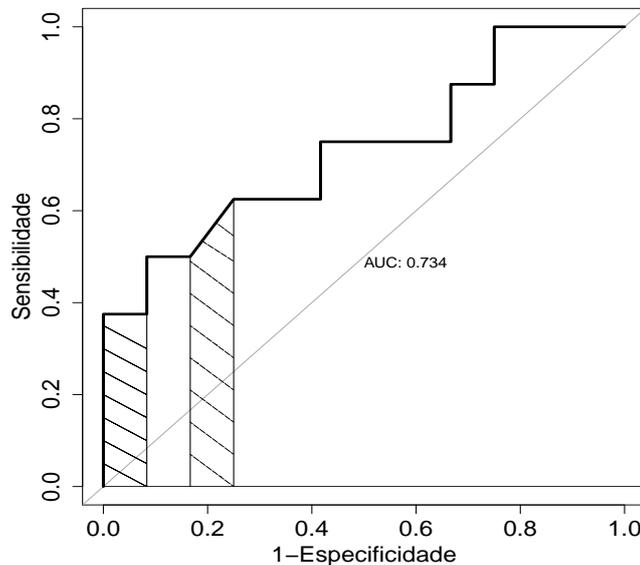


Figura 2 – Exemplo Curva ROC. Retângulo indicado pela área hachurada com linha contínua e trapézio pela linha tracejada.

Com relação ao erro de predição esperado, baseado no estimador em (2.16), tem-se uma estimativa igual a 0.2241. Supondo que essa medida não incorpora as observações censuradas com $t_i < t_{pred}$, a terceira parcela da soma não existiria e a estimativa seria 0.1983, tornando-se subestimada.

2.5 Exemplo real: Contagem de CD4

Para ilustrar as etapas de construção de um modelo de predição, ambos métodos foram empregados com a finalidade de realizar predições do risco de óbito de pacientes em um estudo comparativo, apresentado por (Abrams et al., 1994). O estudo envolvia a contagem de células CD4 em uma coorte de 467 pacientes com infecção avançada de HIV que estavam sob tratamento antirretroviral. Na linha de base, foram registradas as informações de cada paciente, tendo como marcadores de interesse: tipo de tratamento antirretroviral (Didanosina (ddI) e Zalcitabina (ddC)) e o número de células CD4, variável tempo-dependente que foi registrada em 5 instantes durante o estudo: na entrada (linha de base) e em 2, 6, 12 e 18 meses após o início do estudo.

A contagem de CD4 não é inserida no modelo diretamente, usa-se a raiz quadrada, assim como é feito em (Rizopoulos, 2011), mas também poderia ser considerado o logaritmo da contagem como em (Tsiatis et al., 1995), isso é justificado devido ao comportamento dessa variável ser assimétrico. Os tratamentos foram definidos aleatoriamente (estudo clínico aleatorizado) para os pacientes, de tal forma que 237 foram alocados ao grupo ddC e 230 ao ddI. O evento de interesse é o óbito e dos 467 pacientes, 188 experimentaram o evento. Desta forma, a variável resposta de interesse é o tempo desde a linha de base até a ocorrência do óbito.

Foi ajustado tanto o modelo de Cox de taxas de falhas proporcionais (Modelo B) quanto o modelo com covariável tempo-dependente (Modelo TP). O ajuste do Modelo B foi realizado na amostra com 467 pacientes que tinham a medida de CD4 na linha de base. Para o Modelo TP, o ajuste envolveu 1405 observações disponíveis para contagem de CD4 de todos os pacientes com uma ou mais medidas. Os ajustes dos modelos são mostrados nas Tabelas 3 e 4.

Tabela 3 – Ajuste Modelo B: estimativas obtidas a partir do modelo de Cox de taxas de falha proporcionais. (RTF: Razão de Taxas de Falha).

	Coeficientes	\widehat{RTF}	Erro Padrão	Valor z	P-valor
ddI	0.2642	1.3024	0.1464	1.8044	0.0712
$\sqrt{CD4}$	-0.1832	0.8326	0.0223	-8.2241	<0.001

Tabela 4 – Ajuste Modelo TP: estimativas obtidas a partir do modelo de Cox com covariável tempo-dependente.

	Coeficientes	\widehat{RTF}	Erro Padrão	Valor z	P-valor
ddI	0.3095	1.3627	0.1465	2.1120	0.0347
$\sqrt{CD4}$	-0.1934	0.8241	0.0244	-7.9367	<0.001

Como visto anteriormente, o risco estimado (predito) é formulado em termos da função de sobrevivência basal, $S_0(t)$, e as estimativas para esta quantidade em 6 e 12 meses (Tabela 5). Nota-se que a sobrevida basal em 6 meses é muito semelhante para os dois modelos, no entanto, em 12 meses o Modelo TP possui uma estimativa maior.

Tabela 5 – Estimativas para a sobrevida basal em 6 e 12 meses.

	$\hat{S}_0(6)$	$\hat{S}_0(12)$
Modelo B	0.8958	0.7289
Modelo TP	0.8953	0.7452

A partir dos modelos ajustados, o objetivo é validá-los usando as técnicas de treino e teste e validação cruzada e após a validação, usá-los para prever o risco de óbito em 6 meses e 12 meses a partir das informações do tipo de droga e a contagem de CD4.

2.5.1 Validação do modelo: discriminação e calibração

A avaliação da discriminação e da calibração dos modelos foi feita considerando tanto o método de treino e teste quanto o de validação cruzada. Ambos os modelos foram treinados com 70% dos dados disponíveis e testados em 30% da amostra e o procedimento de validação cruzada foi feito repetido 20 vezes. Para predição em 12 meses, foram excluídos aqueles pacientes que experimentaram o evento entre 6 e 12, logo o conjunto de teste avaliado, em 6 meses, possuía 140 pacientes e para 12 meses, 121 pacientes.

2.5.1.1 Treino e teste

Na validação com o método de treino e teste, os valores resultantes para a $AUC(t = 6 \text{ meses})$ para os Modelos B e TP foram, respectivamente, 0.746 (IC 95%: 0.632;0.860) e 0.744 (IC 95%: 0.630;0.858), mostrando que a capacidade de discriminar, entre os indivíduos que irão à óbito em até 6 meses e os não-eventos é razoável. Ainda é possível notar que a discriminação de ambos os modelos acabou sendo melhor ao aumentar a janela de predição de 6 para 12 meses, com estimativas da AUC superiores a 0.80.

Em termos da calibração, as estimativas do erro de predição (\widehat{PE}) são semelhantes para os dois horizontes de predição e não apontam qual modelo seria melhor. No entanto, quando avaliada pelo teste de Hosmer-Lemeshow, a um nível de 5% de significância, há evidências de que os dois modelos são adequados para predizer em 6 meses (p-valor > 0.05), por outro lado, para predizer em 12 meses, apenas o Modelo B mostrou ser capaz de retornar estimativas acuradas (p-valor=0.1355). Os resultados da validação são mostrados a seguir:

Tabela 6 – Medidas de avaliação calculadas na amostra de teste: n=140 na predição em 6 meses e n=121 na predição em 12 meses.

	AUC (IC 95%)	P-valor*	\widehat{PE}
<i>t</i> = 6 meses			
Modelo B	0.746 (0.632;0.860)	0.0708	0.1071
Modelo TP	0.744 (0.630;0.858)	0.0736	0.1077
<i>t</i> = 12 meses			
Modelo B	0.811 (0.715;0.907)	0.1355	0.1640
Modelo TP	0.812 (0.718;0.906)	0.0522	0.1661

*Teste de Hosmer-Lemeshow

As curvas ROC são apresentadas nas Figuras 3 e 4 e os gráficos de calibração nas Figuras 5 e 6. Avaliando os gráficos de calibração, é possível identificar, para quais grupos das probabilidades estimadas, os modelos sub ou superestima o risco. Nota-se uma pequena diferença entre cada um e, visualmente, o Modelo TP tem as probabilidades preditas mais próximas das verdadeiras, com alguns pontos ligeiramente mais próximos a linha tracejada. No entanto, avaliar se o modelo está calibrado por meio pelo método gráfico é um tanto quanto subjetivo, por isso torna-se necessário ter um teste ou uma medida de avaliação como complemento para conclusão.

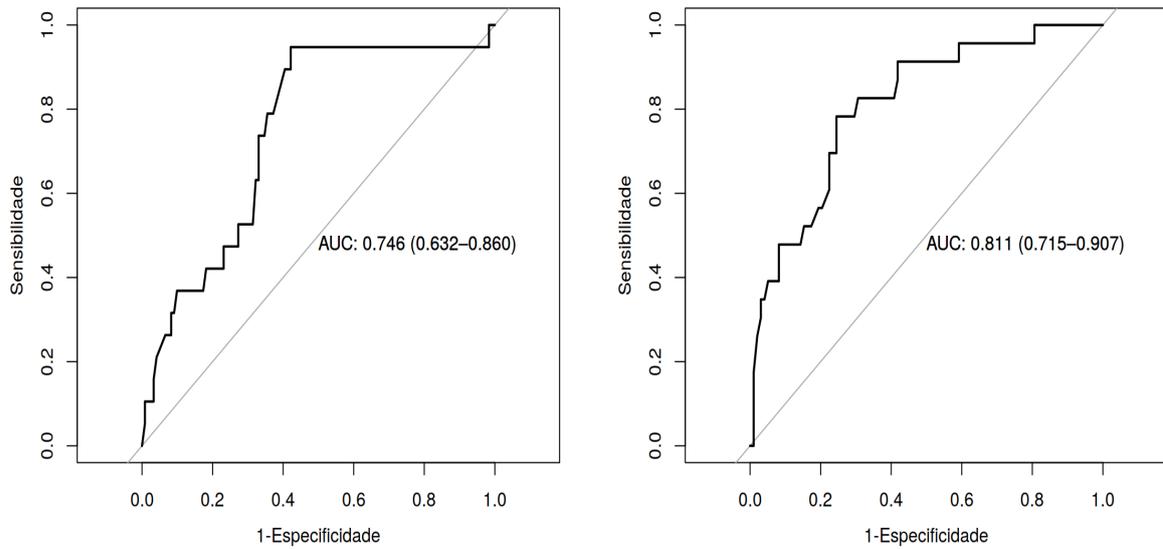


Figura 3 – Curvas ROC para o Modelo B: predição para 6 meses (esquerda) e 12 meses (direita).

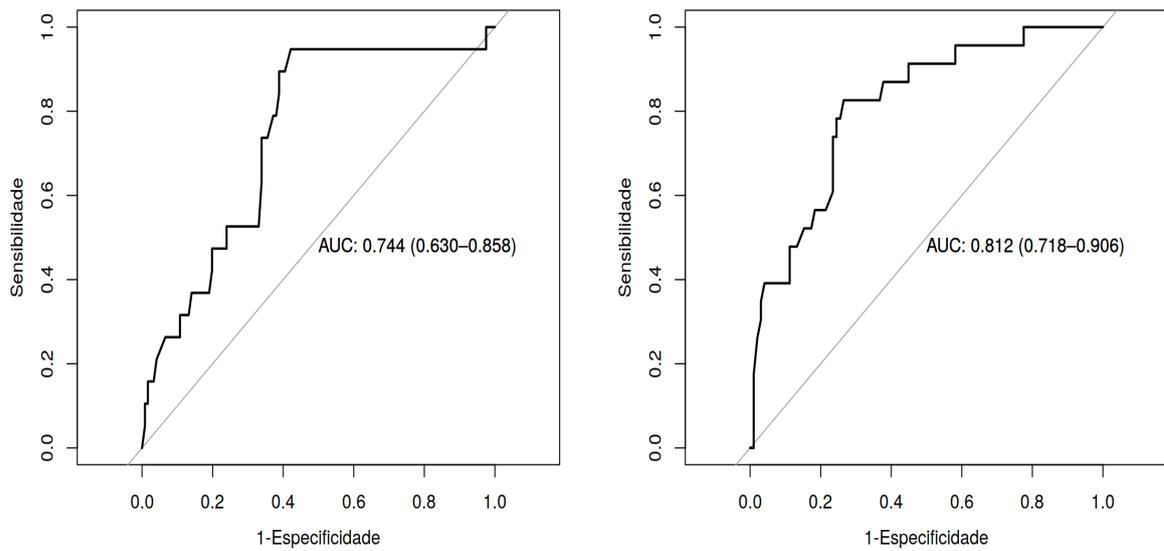


Figura 4 – Curvas ROC para o Modelo TP: predição para 6 meses (esquerda) e 12 meses (direita).

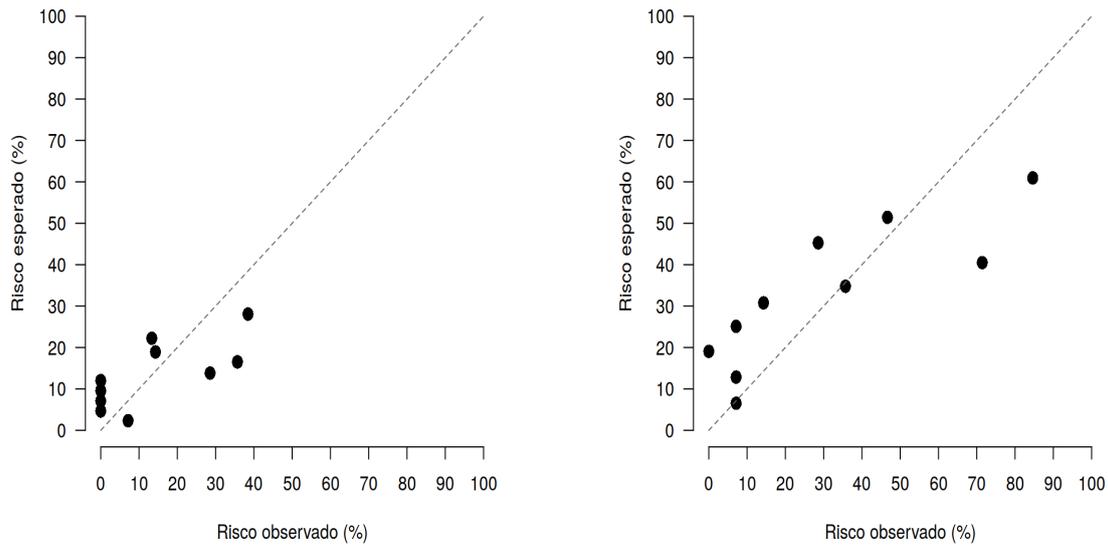


Figura 5 – Gráficos de calibração para o Modelo B: predição para 6 meses (esquerda) e 12 meses (direita).

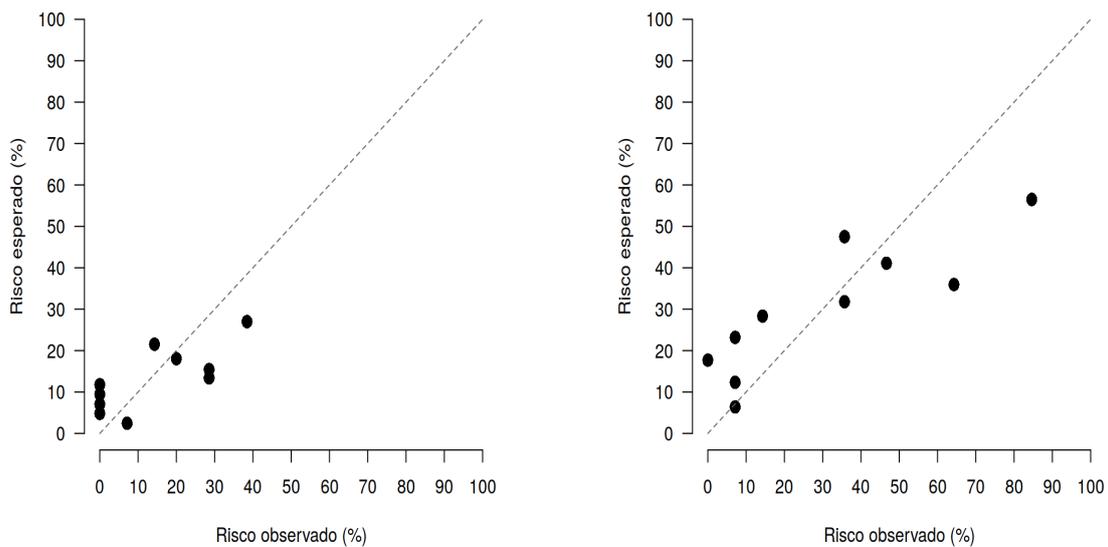


Figura 6 – Gráficos de calibração para o Modelo TP: predição para 6 meses (esquerda) e 12 meses (direita).

2.5.1.2 Validação cruzada

Os resultados obtidos com o método de validação cruzada são apresentados na Tabela 7. Nota-se que, quando se trata da capacidade de discriminar os eventos de não-eventos que ocorrem em 6 ou 12 meses, praticamente não há diferença entre as estimativas da AUC. Observa-se também que, aumentando a janela de predição a área sob a curva ROC tem valores maiores (0.743 para os dois modelos). Os intervalos de confiança possuem maior comprimento para a janela de predição em 12 meses e, para esse instante de predição, o Modelo TP tem apresentado o intervalo mais curto.

Com relação a calibração, o Modelo B apresentou resultados favoráveis para o Teste de Hosmer-Lemeshow e para o erro de predição (PE). Nos dois horizontes de predição, as estimativas de 0.107 e 0.183 para 6 e 12 meses, respectivamente, foram menores e o p-valor > 0.05 atestou que apenas o Modelo B é adequado. Por outro lado, nota-se que o modelo ajustado considerando CD4 como covariável tempo-dependente, não é um modelo calibrado e esse resultado contradiz o que foi observado no método de validação por treino e teste. A partir desses resultados, surge a pergunta sobre qual método é melhor para validar um modelo de predição de risco, pois realizar a validação em apenas uma amostra de teste, pode ser uma questão de sorte: tirar uma amostra “adequada”, para qual o modelo terá um desempenho de predição bom, ou uma “inadequada” na qual as predições não serão boas. Sendo assim, o método de validação cruzada seria uma forma de contornar esse problema gerando várias amostras de testes para serem usadas para o cálculo das medidas de validação.

Tabela 7 – Resultados validação cruzada. Médias das medidas obtidas nas 20 repetições.

	AUC (IC 95%)	P-valor*	\widehat{PE}
Predição para 6 meses			
Modelo B	0.715 (0.499;0.923)	0.389	0.107
Modelo TP	0.710 (0.578;0.830)	0.051	0.133
Predição para 12 meses			
Modelo B	0.743 (0.594;0.896)	0.383	0.183
Modelo TP	0.743 (0.656;0.830)	0.026	0.224

*Teste de Hosmer-Lemeshow

Como visto acima, apenas o Modelo B se mostrou calibrado para os dois horizontes de predição e o fato do Modelo TP não ter resultados favoráveis à essa condição, não o invalida de ser utilizado para realizar predições futuras. D’Agostino e Nam (2003) ressaltam que um modelo poder ter uma performance boa segundo uma medida e ruim por outra. Harrell et al. (1996) recomendam que um modelo com uma *boa discriminação* e *fraca calibração* seja preferível a um modelo com *boa calibração* e *fraca discriminação*, uma vez que é possível recalibrar o modelo reordenando as probabilidades e, no entanto, estas não sofrem alteração para converter em uma discriminação diferente. Sendo assim, em um cenário ideal, seria necessário recalibrar o Modelo TP, mas como foi explicado, esse modelo ainda tem uma boa discriminação, assim como o Modelo B, e será empregado para fazer predições.

Predição do risco de óbito

Com resultados obtidos na validação dos modelos, é possível obter a predição do risco de óbito a partir de cada modelo e analisar as estimativas pontuais e intervalares obtidas via intervalos de confiança *bootstrap* percentílicos. Para ilustrar o desempenho dos modelos, foram selecionados 10 pacientes da amostra, na Tabela (8), para apresentar as estimativas do risco nos dois tempos de predição (6 e 12 meses). Para avaliar o comportamento das estimativas, buscou-se selecionar pacientes com o risco alto e baixo seguindo os valores da raiz quadrada da contagem CD4, ou seja, $CD4 < 200$ e $CD4 \geq 200$. Além disso, seguindo tratamentos antirretrovirais diferentes. Com esse cenário, espera-se que aumentando a janela de predição, de 6 para 12 meses, as estimativas pontuais sejam maiores e a incerteza também aumente.

Tabela 8 – Valores na linha de base da raiz quadrada da contagem de CD4 e tipo de droga para 10 pacientes da amostra.

Paciente	1	7	20	59	80	86	183	217	348	353
CD4	114.06	45.97	362.90	129.96	119.90	242.11	70.06	165.12	247.12	107.95
Droga	ddC	ddC	ddC	ddC	ddC	ddI	ddC	ddI	ddI	ddI

Na Figura 7 são mostradas as estimativas do risco de óbito em 6 e 12 meses para os 10 pacientes. É constatado que as estimativas mais baixas do risco de óbito são para aqueles pacientes que têm a contagem de CD4 maior. Outro ponto que se observa é que, em todos os casos, as estimativas geradas pelo Modelo TP são menores que as do Modelo B, veja as estimativas para os pacientes 7 e 183, por exemplo. Além disto, a incerteza das estimativas, avaliada pelo comprimento dos intervalos de confiança, também é maior, já que os intervalos de predição para o Modelo B são maiores que o intervalos obtidos pelo Modelo TP.

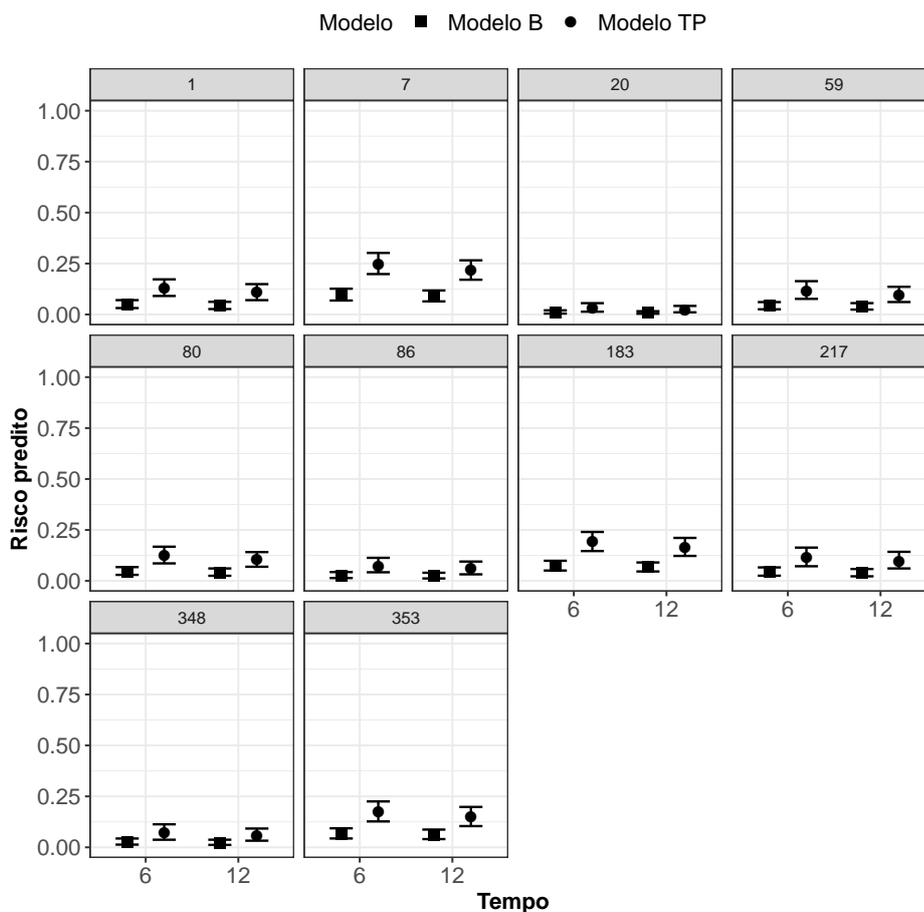


Figura 7 – Estimativas para a probabilidade dos pacientes, com seus respectivos valores de CD4 e tipo droga, irem à óbito em até 6 e 12 meses e intervalos de predição *bootstrap* percentílico de 95% obtidos nos Modelos B e TP.

Por fim, avaliando o efeito de cada droga e a contagem de CD4, apresenta-se o gráfico na Figura 8, pois traduz melhor a relação do risco, estimado em 6 meses, com a raiz quadrada da contagem de CD4 e o tipo de antirretroviral. É possível identificar o efeito na diminuição do risco para valores maiores de CD4 e também que, visualmente, o risco do paciente ir à óbito em seis meses, quando se tem uma contagem baixa, é maior para aqueles que fizeram o uso da droga ddI.

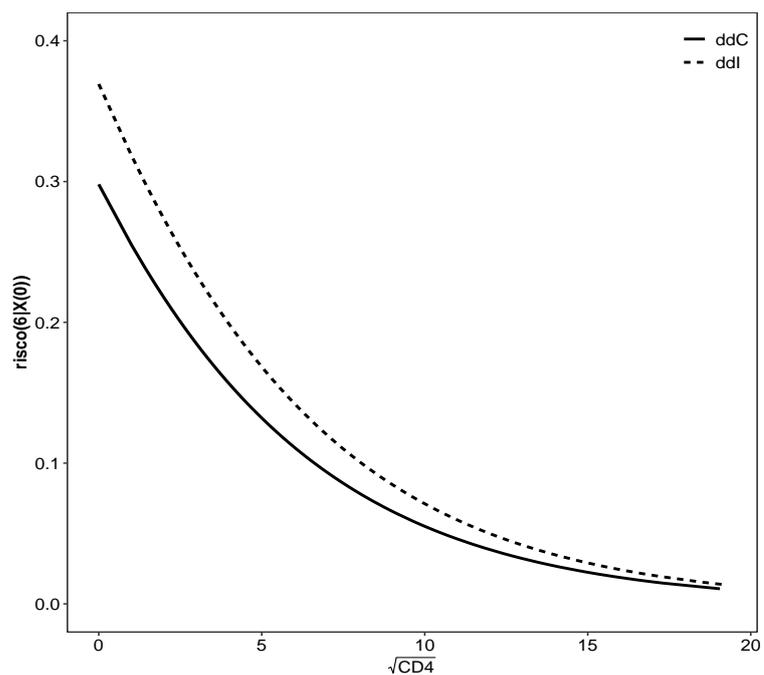


Figura 8 – Valores para o risco de óbito em 6 meses dada a informação da linha de base e de acordo com o tipo de droga e variação da raiz quadrada da contagem de células de CD4.

3 Predição dinâmica

Neste capítulo são apresentadas as metodologias que possibilitam realizar a predição dinâmica. Uma abordagem ingênua é considerada na Seção 3.1, utilizando o modelo de Cox ajustado na linha de base. Na Seção 3.2, o modelo de Cox é estimado em um novo instante de referência ou *landmark* (*LM*). Na Seção 3.3 é apresentada a modelagem conjunta (*JM*) para tratar tanto do processo da resposta de sobrevivência quanto do marcador longitudinal, bem como a obtenção do risco predito a partir deste modelo. Na Seção 3.5, retoma-se o exemplo da contagem de CD4, agora realizando predição dinâmica para o risco de óbito.

3.1 Abordagem ingênua

Considere que, ao dar seguimento no estudo, os pacientes que sobreviveram ao tempo t tiveram novas medições de seus marcadores longitudinais, as quais serão incluídas na amostra. A partir desse novo registro, há o interesse em obter uma nova estimativa para a probabilidade de um indivíduo experimentar o evento de interesse até um novo tempo u , $u > t$. Uma forma de atualizar o risco é usar o modelo construído na linha de base e “plugar”, no estimador em (2.13), a nova informação obtida em um tempo t , $\mathbf{x}_t = \{\mathbf{x}, y(t)\}$, que contém tanto as informações de linha de base, quanto o novo valor medido para o marcador longitudinal. É considerada uma maneira *ingênua* para obter o risco dinamicamente, pois não há atualização da estimação dos parâmetros do modelo. A partir desse enfoque, são apresentadas duas abordagens, que serão chamadas de *Naive 1* (\hat{r}_{N_1}) e *Naive 2* (\hat{r}_{N_2}).

Naive 1: a predição para o indivíduo que sobreviveu até t é feita inserindo a nova informação do marcador longitudinal na fórmula (2.13) do risco, usando a sobrevida basal para o novo tempo de predição. Logo, o estimador que se tem é:

$$\begin{aligned} r_{N_1}(u|t) &= 1 - P(T - t > u \mid T > t, \mathbf{x}_t) \\ \hat{r}_{N_1}(u|t) &= 1 - \left[\hat{S}_0(u) \right]^{\exp\{\mathbf{x}^\top \hat{\beta} + y(t) \hat{\gamma}\}}. \end{aligned} \tag{3.1}$$

Naive 2: o segundo método para atualização do risco predito, leva em consideração a probabilidade condicional de que o indivíduo sobreviveu até t e a informação nova

da covariável tempo-dependente, tal que

$$\begin{aligned}
 r_{N_2}(u|t) &= 1 - P(T > u \mid T > t, \mathbf{x}_t) \\
 &= 1 - \frac{P(T > u, T > t \mid \mathbf{x}_t)}{P(T > t \mid \mathbf{x}_t)} \\
 &= 1 - \frac{P(T > u \mid \mathbf{x}_t)}{P(T > t \mid \mathbf{x}_t)} \\
 \hat{r}_{N_2}(u|t) &= 1 - \frac{\hat{S}(u \mid \mathbf{x}_t)}{\hat{S}(t \mid \mathbf{x}_t)}. \tag{3.2}
 \end{aligned}$$

Observe que nessa segunda abordagem para fazer a atualização do risco, a quantidade (3.2) dispõe do estimador da função de sobrevivência avaliado em t e u e, para obter as estimativas para essas quantidades, são utilizados as estimativas para os parâmetros do modelo construído na linha de base.

3.2 Landmark

As ideias iniciais dessa metodologia foram apresentadas por Anderson et al. (1983), que tinham como finalidade lidar com covariável tempo-dependente na análise do tempo de sobrevivência de pacientes com câncer. O intuito era tomar pontos de referência, após o início do acompanhamento, e dar seguimento na análise para obter uma nova probabilidade do paciente sobreviver a mais um tempo u dado que sobreviveu ao tempo t . O propósito desse tipo de análise é fixar um tempo s ($s \geq t$) durante o estudo e realizar predições a partir das informações medidas neste ponto, ou seja, a linha de base considerada não seria mais formada pelas medidas no instante inicial do estudo e sim no ponto s . Os indivíduos que se encontram vivos neste tempo s fazem parte de um “novo” conjunto da amostra que será utilizado para estimação dos parâmetros de um modelo de Cox.

A estimação dos parâmetros do modelo é realizada como no Capítulo 2, agora com a verossimilhança contém dados apenas dos pacientes sob risco no ponto de referência s . Logo, um indivíduo i que sobreviveu ao tempo t e teve uma nova medição do marcador longitudinal em s , $y(s)$, tem uma estimativa para o risco em um determinado tempo u ($u > t$) dada por

$$\hat{r}_i(u \mid y(s)) = 1 - \left[\hat{S}_0(u) \right]^{\exp\{\mathbf{x}_i^\top \hat{\boldsymbol{\beta}}^* + \hat{\gamma}^* y(s)\}} \tag{3.3}$$

em que, assim como mostrado anteriormente, $\hat{\boldsymbol{\beta}}^*$ e $\hat{\gamma}^*$ são os estimadores de máxima verossimilhança parcial dos coeficientes do modelo de Cox e $\hat{S}_0(u)$ é a sobrevida basal estimada para a nova linha de base avaliada no tempo u .

Como o modelo é reajustado com as novas medidas do marcador longitudinal, espera-se que as estimativas do risco sejam mais acuradas do que as obtidas com as abor-

dagens *Naive 1* e *Naive 2*. Isso é motivado pelo fato de que os marcadores longitudinais podem mudar seu efeito durante o estudo e re-estimando os coeficientes do modelo, esse efeito pode ser captado, fazendo com que as predições do risco sejam melhores. No entanto, o processo longitudinal completo do marcador que depende do tempo não é incorporado no modelo e também não se utiliza a amostra completa, visto que aqueles pacientes que experimentaram o evento antes de t não terão um novo registro do marcador.

3.3 Modelagem conjunta

Para as especificações da modelagem conjunta, considere t_i como o tempo observado até o evento para o i -ésimo indivíduo ($i = 1, \dots, n$), $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})^\top$ as covariáveis fixas relacionadas ao tempo até a ocorrência do evento e a covariável tempo-dependente, chamada e denotada nesta seção por covariável/marcador longitudinal e $y_i(t)$, respectivamente. Na modelagem conjunta, também deseja-se examinar a associação entre a covariável tempo-dependente e o risco de ocorrência do evento, para isso é adotado um submodelo para explicar o processo longitudinal de $y_i(t)$ e um submodelo que associa esse processo ao tempo de ocorrência do evento. Ambos os submodelos são apresentados nas Subseções 3.3.1 e 3.3.2, respectivamente

3.3.1 Submodelo longitudinal

Seguindo Rizopoulos (2012), que foca em uma resposta longitudinal contínua, o processo longitudinal do marcador $y(t)$ é aproximado seguindo a abordagem do modelo linear de efeitos mistos, tal que, para o i -ésimo indivíduo, sua formulação tem a seguinte estrutura:

$$y_i(t) = y_i^*(t) + \varepsilon_i(t) = \mathbf{w}'_i \boldsymbol{\alpha} + \mathbf{z}'_i(t) \mathbf{b}_i + \varepsilon_i(t), \quad \varepsilon_i(t) \sim N(0, \sigma^2), \quad (3.4)$$

em que $y_i^*(t)$ denota o verdadeiro valor não observado da resposta longitudinal no instante t , \mathbf{w}'_i e $\mathbf{z}'_i(t)$ são os vetores de covariáveis com seus respectivos efeitos fixos e aleatórios $\boldsymbol{\alpha}$ e \mathbf{b}_i . Para acomodar os efeitos aleatórios, é assumido que $\mathbf{b}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{b_i})$ e ainda, assume-se que \mathbf{b}_i é independente do erro $\varepsilon_i(t)$.

3.3.2 Submodelo de sobrevivência

Um modelo de risco relativo quantifica a associação da resposta longitudinal $y_i^*(t)$ com o risco até a ocorrência de um evento em termos da função taxa de falha na expressão (3.5)

$$\lambda(t | y_i^*(t), \mathbf{x}_i) = \lambda_0(t) \exp\{\gamma^* y_i^*(t) + \mathbf{x}_i^\top \boldsymbol{\beta}^*\}, \quad t > 0, \quad (3.5)$$

em que $\lambda_0(t)$ é a função de risco basal, podendo ser não especificada, como no modelo de Cox ou proceder utilizando um modelo paramétrico ou alguma aproximação como, por

exemplo, funções definidas por partes do tipo constante por partes ou *splines* (Rizopoulos, 2010). No entanto, no contexto da modelagem conjunta, ela precisa ser definida, pois a não especificação de $\lambda_0(t)$ acarreta na subestimação dos parâmetros do modelo (Rizopoulos, 2012; Hsieh et al., 2006).

A função de sobrevivência, baseada nesse modelo, é obtida a partir da relação com a função de risco acumulado (2.2) e é escrita da seguinte forma:

$$\begin{aligned} S(t | y_i^*(t), \mathbf{x}_i) &= P(T_i > t | y_i^*(t), \mathbf{x}_i) \\ &= \exp \left\{ - \int_0^t \lambda_0(s) \exp \{ \gamma^* y_i^*(s) + \mathbf{x}_i^\top \beta^* \} ds \right\}. \end{aligned} \quad (3.6)$$

A distribuição conjunta de ambas as respostas dos modelos é definida em termos da independência condicionada aos efeitos aleatórios \mathbf{b}_i . A seguir é abordada a estimação do modelo considerando essa distribuição conjunta.

3.3.3 Estimação

A estimação dos parâmetros do modelo conjunto, considerando os dois submodelos apresentados anteriormente, pode ser feita tanto sob abordagem frequentista [Tsiatis et al. (1995); Henderson et al. (2002); Tsiatis e Davidian (2004)] ou sob a abordagem Bayesiana [Wang e Taylor (2001); Brown e Ibrahim (2003)]. O método clássico será tratado neste trabalho e consiste em maximizar uma função de verossimilhança com respeito aos parâmetros a serem estimados para a distribuição conjunta das respostas de sobrevivência e longitudinal $\{t_i, \delta_i, y_i\}$. Para definir essa distribuição conjunta, assume-se que o vetor de efeitos aleatórios explicam a associação entre os dois processos e a correlação entre as medidas longitudinais de cada indivíduo. Sendo assim, dado os efeitos aleatórios, as medidas longitudinais do i -ésimo indivíduo são independentes, tal que,

$$p(t_i, \delta_i, y_i | \mathbf{b}_i, \boldsymbol{\theta}^*) = p(t_i, \delta_i | y_i, \mathbf{b}_i; \boldsymbol{\theta}^*) p(y_i | \mathbf{b}_i; \boldsymbol{\theta}^*), \quad (3.7)$$

em que $\boldsymbol{\theta}^*$ é o vetor com os parâmetros do submodelo para resposta de sobrevivência, do processo longitudinal e dos efeitos aleatórios. Também assume-se que há independência entre as medias longitudinais de um indivíduo dado os efeito aleatórios. Tem-se

$$p(y_i | \mathbf{b}_i; \boldsymbol{\theta}^*) = \prod_{j=1}^{n_i} p(y_i(t_{ij}) | \mathbf{b}_i; \boldsymbol{\theta}^*), \quad (3.8)$$

em que n_i é o número de medidas do i -ésimo indivíduo. A partir de (3.7), a contribuição do i -ésimo indivíduo para a função de verossimilhança é descrita por

$$\begin{aligned}
 L_i(\boldsymbol{\theta}^*) &= p(t_i, \delta_i, y_i; \boldsymbol{\theta}^*) \\
 &= \int p(t_i, \delta_i, y_i, \mathbf{b}_i; \boldsymbol{\theta}^*) d\mathbf{b}_i \\
 &= \int p(t_i, \delta_i | y_i, \mathbf{b}_i; \boldsymbol{\theta}^*) p(y_i | \mathbf{b}_i; \boldsymbol{\theta}^*) p(\mathbf{b}_i; \boldsymbol{\theta}^*) d\mathbf{b}_i \\
 &= \int p(t_i, \delta_i | y_i, \mathbf{b}_i; \boldsymbol{\theta}^*) \left[\prod_{j=1}^{n_i} p(y_i(t_{ij}) | \mathbf{b}_i; \boldsymbol{\theta}^*) \right] p(\mathbf{b}_i; \boldsymbol{\theta}^*) d\mathbf{b}_i. \quad (3.9)
 \end{aligned}$$

O primeiro termo da integral, que é a verossimilhança para o submodelo de sobrevivência, pode ser escrito da seguinte forma:

$$\begin{aligned}
 p(t_i, \delta_i | \mathbf{b}_i; \boldsymbol{\theta}^*) &= [\lambda_i(t_i | y_i^*; \boldsymbol{\theta}^*)]^{\delta_i} S(t_i | y_i^*; \boldsymbol{\theta}^*) \\
 &= [\lambda_0(t_i) \exp\{\eta y_i^*(t_i) + \gamma' \mathbf{x}_i(t_i)\}]^{\delta_i} \\
 &\quad \times \exp \left\{ - \int_0^{t_i} \lambda_0(s) \exp\{\eta y_i^*(s) + \gamma' \mathbf{x}_i(s)\} ds \right\}, \quad (3.10)
 \end{aligned}$$

em que $\lambda_0(\cdot)$ pode ser qualquer função positiva no tempo, seguindo a definição de uma aproximação não paramétrica em (3.5), $\lambda_i(\cdot)$ é dada pela função de risco apresentada em (3.5) referente ao submodelo para o tempo até o evento e a função de sobrevivência,

$$\begin{aligned}
 S_i(t_i | y_i^*; \boldsymbol{\theta}^*) &= P(T > t_i | y_i^*, \mathbf{x}_i) \\
 &= \exp \left\{ - \int_0^{t_i} \lambda_i(s | y_i^*; \boldsymbol{\theta}^*) ds \right\}. \quad (3.11)
 \end{aligned}$$

E ainda, $p(y_i(t_{ij}) | \mathbf{b}_i; \boldsymbol{\theta}^*)$ é a densidade de uma distribuição normal univariada para a resposta longitudinal e $p(\mathbf{b}_i; \boldsymbol{\theta}^*)$ é densidade de uma normal multivariada para os efeitos aleatórios.

Os estimadores de máxima verossimilhança são obtidos realizando a maximização da log-verossimilhança, $\log L(\boldsymbol{\theta}^*)$, e requerem o emprego de métodos de integração numérica para aproximar as integrais presentes na equação. Sob a abordagem clássica, a solução se dá utilizando algoritmos como *expectation-maximization* (algoritmo EM) e Newton-Raphson. Rizopoulos (2012) apresenta a estrutura completa do algoritmo EM empregado para estimação no cenário de modelagem conjunta, tratando os efeitos aleatórios como dados omissos. A aplicação do método pode ser realizada com o auxílio de bibliotecas do R como, por exemplo, o pacote JM (Rizopoulos, 2010) que ajusta o modelo considerando a estimação por máxima verossimilhança ou a implementação usando abordagem Bayesiana com método de Monte Carlo via Cadeia de Markov, encontrada no pacote JMBayes.

3.3.4 Diagnóstico do modelo

Assim como visto no cenário de linha de base, a utilização de um modelo na prática requer que suas suposições sejam verificadas. Na modelagem conjunta, essa avaliação é feita sob as suposições do modelo linear de efeitos mistos para a resposta longitudinal e o modelo risco relativo para a resposta de sobrevivência que são, respectivamente: homocedasticidade vinda da estrutura da matriz de covariância e normalidade atribuída ao erro; e a forma funcional da covariável tempo-dependente e a adequação do modelo no ajuste dos dados. Rizopoulos (2012) apresenta diferentes tipos de resíduos para ambas as respostas, que podem ser analisados graficamente.

3.3.5 Predição

Após a construção do modelo (estimação) o enfoque é prever a sobrevida esperada de um indivíduo, dado que ele já sobreviveu a um certo tempo t e que tem um histórico de informações até esse tempo, denotado por $\mathcal{Y}_i(t)$. Considera-se a probabilidade condicional de um indivíduo sobreviver até $u > t$ dado que ele sobreviveu até t , sendo expressa por

$$S_i(u | t, \mathcal{Y}_i(t)) = P(T_i \geq u | T_i > t, \mathcal{Y}_i(t), \boldsymbol{\theta}^*), \quad (3.12)$$

Essa probabilidade condicional pode ser reescrita como

$$\begin{aligned} P(T \geq u | T > t, \mathcal{Y}_i(t); \boldsymbol{\theta}^*) &= \int P(T \geq u | T > t, \mathcal{Y}_i(t), \mathbf{b}_i; \boldsymbol{\theta}^*) \\ &\quad \times p(\mathbf{b}_i | T > t, \mathcal{Y}_i(t); \boldsymbol{\theta}^*) d\mathbf{b}_i \\ &= \int P(T_i \geq u | T > t, \mathbf{b}_i; \boldsymbol{\theta}^*) p(\mathbf{b}_i | T > t, \mathcal{Y}_i(t); \boldsymbol{\theta}^*) d\mathbf{b}_i \\ &= \int \frac{S_i[u | y_i^*(u, \boldsymbol{\theta}^*); \boldsymbol{\theta}^*]}{S_i[t | y_i^*(t, \boldsymbol{\theta}^*); \boldsymbol{\theta}^*]} p(\mathbf{b}_i | T > t, \mathcal{Y}_i(t); \boldsymbol{\theta}^*) d\mathbf{b}_i, \end{aligned} \quad (3.13)$$

Sendo que $S_i(\cdot)$ é a função de sobrevivência apresentada em (3.11) e $y_i^*(\cdot)$ é a componente que carrega as informações longitudinais aproximada no modelo de efeitos mistos. A partir de (3.13), deriva-se um estimador para a sobrevida esperada, denominado estimador de primeira ordem (Rizopoulos, 2011), escrito como:

$$\hat{S}_i(u | t, \mathcal{Y}_i(t)) = \frac{S_i \left[u | y_i^*(u, \hat{\boldsymbol{\theta}}^*); \hat{\boldsymbol{\theta}}^* \right]}{S_i \left[t | y_i^*(t, \hat{\boldsymbol{\theta}}^*); \hat{\boldsymbol{\theta}}^* \right]}, \quad (3.14)$$

em (3.14) tem-se que $\hat{\boldsymbol{\theta}}^*$ é o estimador de máxima verossimilhança para o modelo conjunto e n_i é o número de medidas do marcador longitudinal para o indivíduo i .

Sobre esse estimador de primeira ordem em (3.14), Rizopoulos (2011) relata que funciona bem na prática, porém tem seus impasses, como a dificuldade de se obter o erro padrão e/ou intervalos de confiança para $S_i(u | t, \mathcal{Y}_i(t))$. A solução encontrada por

Rizopoulos (2011) e Proust-Lima e Taylor (2009) é a realização de métodos de Monte Carlo motivada por uma formulação Bayesiana assintótica do modelo conjunto. Novamente, uma alternativa é obter estimativas para os erros padrões pelo método de *bootstrap*. Elashoff et al. (2016) sugerem que B amostras *bootstrap* ($B=200-500$) sejam usadas para aproximar o erro padrão dos parâmetros estimados em cada uma das B amostras.

A obtenção dessas estimativas também pode ser feita a partir do pacote JM (Rizopoulos, 2010). Como nota do pacote, tem-se que as probabilidades preditas não são calculadas sob a não especificação da função taxa de falha basal no submodelo de sobrevivência. Como citado anteriormente, ter $\lambda_0(\cdot)$ não especificada acarreta na subestimação do erro padrão dos parâmetros na modelagem conjunta e as opções para evitar esse problema é admitir para a taxa de falha basal uma distribuição paramétrica, como Gama ou Weibull, ou uma aproximação com funções por partes ou *splines*. Neste trabalho, optou-se por empregar o modelo conjunto usando a distribuição Weibull na especificação de $\lambda_0(\cdot)$.

3.4 Validação do modelo

A avaliação da predição baseada nas abordagens apresentadas neste capítulo também é feita em termos dos conceitos de discriminação e calibração. Por exemplo, Proust-Lima e Taylor (2009) avaliam a acurácia das estimativas, obtidas pela modelagem conjunta, seguindo duas medidas: erro de predição absoluto e uma média ponderada do erro de predição, Rizopoulos (2011; 2012) avalia a capacidade preditiva, também baseada na modelagem conjunta, apenas em termos da área sob a curva ROC e Rizopoulos et al. (2017) englobam tanto discriminação (AUC) quanto a calibração (erro de predição esperado: PE) para verificar a capacidade preditiva com *landmark* e modelagem conjunta. Portanto, a validade do modelo também será avaliada seguindo as medidas AUC e PE apresentadas em (2.15) e (2.16), respectivamente.

3.5 Exemplo: contagem de CD4

Voltando ao exemplo da contagem de células CD4 em pacientes com HIV, agora o interesse é realizar predições em 12 e 18 meses, para os pacientes que sobreviveram a 6 meses. A análise foi conduzida ajustando modelos para técnica de *landmark*, em que foi considerado um conjunto de dados com pacientes sob risco em 6 meses ($n = 403$) e 12 meses ($n = 318$). O modelo conjunto é ajustado com toda a amostra e todas as informações disponíveis da contagem de CD4.

As estimativas para o efeito da raiz quadrada de CD4, para cada abordagem, estão sumarizados na Tabela 9. Observa-se que o efeito do marcador é negativo, indicando que o aumento da contagem de CD4 implica na redução do risco de óbito. Note também que

as estimativas dos ajustes com o modelo de Cox, sem modelar o efeito de CD4, são todas menores do que as obtidas com a modelagem conjunta.

Tabela 9 – Estimativas do efeito da raiz quadrada da contagem de CD4 obtidas com todas as abordagens.

	Estimativa	Erro Padrão	z	p-valor
Cox B	-0.791	0.084	-9.376	< 0.0001
Cox TP	-0.723	0.080	-9.041	< 0.0001
$LM(s = 6)$	-0.720	0.100	-7.195	< 0.0001
$LM(s = 12)$	-0.615	0.175	-3.505	0.0005
JM_{wb}	-0.280	0.035	-7.865	< 0.0001

LM : landmark

JM_{wb} : modelo conjunto assumindo um modelo Weibull para $\lambda_0(\cdot)$

Validação do modelos

Como feito no capítulo anterior, empregou-se o método de validação cruzada para extrair as medidas de validação e avaliar a capacidade preditiva de todos os métodos. Foi necessário realizar algumas adaptações nos algoritmos na predição ingênua, de modo que ao treinar o modelo eram incluídos os pacientes que tinham tempo observado inferior a 6 meses. Novamente as medidas de performance foram avaliadas e o segundo método da predição ingênua, *Naive 2*, foi o que melhor discrimina pacientes que irão experimentar o evento em 12 meses daqueles que não, $AUC = 0.795$. O método de *landmark* se apresenta como o mais calibrado, $\widehat{PE} = 0.151$, gerando probabilidades preditas mais próximas do que é observado.

Avaliando ambas as medidas de performance do modelo, todos os métodos tiveram desempenhos muito próximos segundo os resultados da validação cruzada. Nota-se que os métodos da predição ingênua apresentaram intervalos de confiança da AUC mais largos que os outros métodos. Quanto a calibração não é observada nenhuma grande diferença entre os métodos. A partir desses resultados, as quatro abordagens foram utilizadas para fazer predições dinâmicas do risco de óbito individuais para 10 pacientes do conjunto de dados.

Tabela 10 – Resultado da validação cruzada para os métodos de predição de acordo com as medidas de avaliação (AUC e Erro predito esperado).

$t = 12$ meses	AUC (IC 95%)	\widehat{PE}
<i>Naive 1</i>	0.727 (0.555 ; 0.899)	0.184
<i>Naive 2</i>	0.795 (0.633 ; 0.954)	0.162
<i>Landmark</i>	0.715 (0.633 ; 0.896)	0.151
JM_{wb}	0.753 (0.608 ; 0.898)	0.183

Predição dinâmica do risco de óbito

O esquema para estimar as probabilidades do risco de óbito do paciente, dado que ele já sobreviveu um certo tempo, baseou-se em fixar dois horizontes de predição, em 12 meses e 18 meses. Para comparar as predições geradas por cada método, optou-se por escolher 10 pacientes sob risco em cada ponto de referência e para estes obter as estimativas do risco de óbito no horizonte de tempo especificado.

Na Figura 9, são expressadas as estimativas pontuais e intervalos de predição *bootstrap* percentílico do risco predito para os 10 pacientes em 12 e 18 meses. Observa-se os métodos *Naive 1* e *Naive 2* (Modelo B e Modelo TP) retornam valores maiores para a estimativa do risco, comparados aos outros métodos. A predição resultante dos métodos de *JM*, *LM* e *Naive 2* tendem a apresentar estimativas mais próximas, com interseção dos intervalos de predição. Observa-se também que há o aumento do erro padrão ao ampliar a janela de predição.

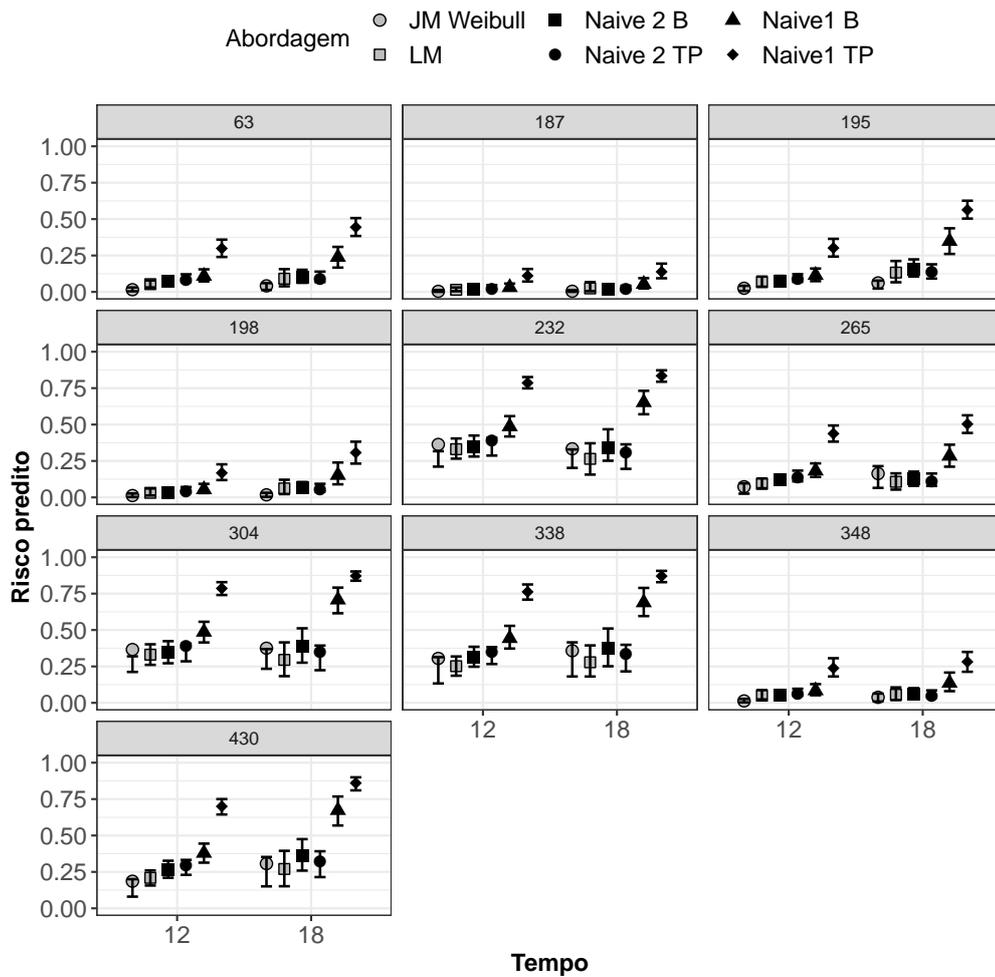


Figura 9 – Predições pontuais e intervalos de predição do risco, obtidas por cada um dos métodos, para os pacientes 63, 187, 195, 198, 232, 262, 304, 338, 348, 430 em 12 e 18 meses.

4 Aplicação

Neste capítulo, apresenta-se uma aplicação aos dados da coorte do projeto SaMi-Trop, empregando as metodologias discutidas anteriormente. O foco é realizar predições do risco de óbito dos paciente em 2 e 4 anos após o início do estudo.

4.1 Descrição dos dados

Para o estudo do projeto SaMi-Trop foi estabelecida uma coorte prospectiva com 1959 pacientes com cardiomiopatia chagásica, que residem em 21 municípios do norte de Minas Gerais atendidos pela Rede de Telessaúde de Minas Gerais. Os pacientes foram acompanhados por pelo menos dois anos, com uma visita na linha de base (2013-2014) e outra em 24 meses. A amostra considerada para desenvolver a análise contém informações de 1544 pacientes para quais se tinham os dados completos das 5 covariáveis relacionadas ao tempo até a ocorrência do óbito. Um modelo de predição na linha de base foi construído considerando os seguintes marcadores:

- *idade* do paciente (em anos).
- *CF*: classe funcional segundo a classificação da NYHA (*New York Heart Association*). São 4 classes e cada classe diz respeito a capacidade do paciente realizar as atividades cotidianas. Por exemplo, um paciente pertencente à Classe I, executa as atividades normalmente e um paciente na Classe 4 não têm condições de realizar as atividades de rotina. Essa variável foi dicotomizada e assumiu valor 0 se o paciente estava na Classe I e valor 1 se estava na classe II, III ou IV.
- *frequência cardíaca* (em batimentos por minuto).
- *ECG*: variável dicotômica que indica alteração do exame de eletrocardiograma por meio da medida de duração do complexo *QRS* ≥ 150 ms (1=*QRS* ≥ 150 ms, 0=*QRS* < 150 ms).
- *IC*: insuficiência cardíaca, dicotômica (1=sim, 0=não), diagnosticada pelo marcador NT-ProBNP (N-Terminal do Pró-hormônio do Peptídeo Natriurético do tipo B).

Na Tabela 11, são mostradas estatísticas descritivas para as 5 covariáveis na linha de base e na segunda visita apenas para frequência cardíaca (*Freq. Card. 2*). Há informação da média e desvio padrão para as variáveis contínuas e a frequência e porcentagem para as categóricas (n (%)). A idade média dos pacientes, na linha de base, era de 59.4 anos, 13.9%

tinham insuficiência cardíaca diagnosticada pelo marcador NT-ProBNP, 17.3% alteração no ECG e 46.2% eram pacientes que realizavam suas atividades cotidianas normalmente.

Tabela 11 – Medidas descritivas das covariáveis.

	<i>Idade</i>	<i>Freq. Card. 1</i>	<i>Freq. Card. 2</i>
Média (Desvio Padrão)	59.4 (12.7)	66.2 (13.9)	65.0 (13.8)
	<i>IC</i>	<i>ECG</i>	<i>CF</i>
n (%)	214 (13.9%)	267 (17.3%)	714 (46.2%)

Após a segunda visita, o interesse é estimar o risco de óbito em mais dois anos, dado que o paciente sobreviveu aos primeiros dois anos e suas informações clínicas foram atualizadas na segunda visita. O marcador longitudinal que será analisado é a *frequência cardíaca*. Até o momento da finalização deste trabalho, não foi possível obter a informação do desfecho após a 2ª visita para todos aqueles que sobreviveram aos 2 anos, com isso as análises foram feitas com os dados disponíveis até o momento. Obteve-se a informação completa do desfecho e da segunda medida da *frequência cardíaca* somente para 283 pacientes. O tempo mediano de acompanhamento desses pacientes foi de 63 meses. No primeiro momento foi observado que 110 pacientes experimentaram o evento antes de 2 anos e, após a segunda visita, 80 pacientes foram à óbito.

4.2 Modelos

Os modelos de Cox (linha de base e *landmark*) e conjunto foram ajustados aos dados do SaMi-Trop, para posteriormente estimar os riscos de óbito em 2 e 4 anos de cada paciente. O ajuste do modelo de Cox na linha de base foi feito na amostra com os 1544 pacientes, sendo possível obter a sobrevida basal, $S_0(t)$, para $t = 2, 4$ anos e consequentemente ter a predição na linha de base e dinâmica com os estimadores *Naive 1* e *Naive 2*. Na abordagem *landmark* o modelo de Cox foi ajustado numa amostra menor, com 283 pacientes, para os quais foi possível obter o registro de último contato ou evento após a segunda visita do acompanhamento. E ainda, o conjunto de dados para o ajuste com a modelagem conjunta possui 1824 observações.

4.2.1 Linha de base

Os coeficientes estimados, a razão das taxas de falha (RTF), o erro padrão e o p -valor associado aos testes de significância de cada marcador são apresentados na Tabela 12. Observa-se que todos os marcadores possuem efeito positivo na taxa de falha, ou seja, aumentam o risco de óbito e o marcador de maior efeito no modelo foi *IC*.

Tabela 12 – Ajuste do modelo de Cox na amostra com $n = 1544$.

	Estimativa	\widehat{RTF} (IC 95%)	Erro padrão	p -valor
<i>IC</i>	1.756	5.787 (3.865;8.665)	0.206	< 0.001
<i>ECG</i>	0.521	1.684 (1.106;2.563)	0.214	0.015
<i>CF</i>	0.479	1.615 (1.079;2.415)	0.205	0.019
<i>Idade</i>	0.037	1.038 (1.022;1.053)	0.008	< 0.001
<i>Freq. Cardíaca</i>	0.017	1.017 (1.005;1.028)	0.006	0.004

A partir das estimativas na Tabela 12 e da função e sobrevivência basal nos tempos $t = 2$ e 4 anos ($\hat{S}_0(2) = 0.9602$ e $\hat{S}_0(4) = 0.9513$), calcula-se o risco de óbito para estes instantes de tempo, respectivamente. Para a predição a partir da linha de base, a performance da capacidade preditiva do modelo foi validada usando as técnicas de treino e teste e validação cruzada. Na primeira técnica, a amostra foi separada com 1082 pacientes para treino e 463 teste. Na validação cruzada, os resultados foram gerados a partir da validação em 5-*folds*.

Na Tabela 13, são exibidos os resultados da validação com as estimativas para a AUC, o PE e o p -valor do teste de calibração. Nota-se que o modelo de linha de base mostrou-se capaz de discriminar de forma satisfatória ($AUC \geq 0.8$) aqueles pacientes que irão a óbito em 2 anos, daqueles que não. Tanto o teste de Hosmer-Lemeshow quanto o erro de predição indicam que o modelo tem boa acurácia e retorna probabilidades preditas próximas do que foi observado.

Tabela 13 – Medidas de avaliação da performance para predição do risco de óbito em $t = 2$ anos.

	\widehat{AUC}	p -valor*	\widehat{PE}
Treino e teste	0.839 (0.768;0.910)	0.823	0.058
Validação cruzada	0.803 (0.661;0.942)	0.466	0.057

*Teste de Hosmer-Lemeshow

Na Figura 10, tem-se a representação da curva ROC para o resultado da área sob a curva mostrado na Tabela 13 e, o gráfico de calibração exibindo os valores observados e esperados com pontos próximos à reta, dando a indicação que o modelo é calibrado. Com base nesses resultados, pode-se concluir que o modelo desenvolvido para predição estática do risco de óbito em 2 anos é capaz de discriminar, satisfatoriamente, indivíduos que irão a óbito em 2 anos daqueles que não irão e, retornar predições acuradas do risco predito.

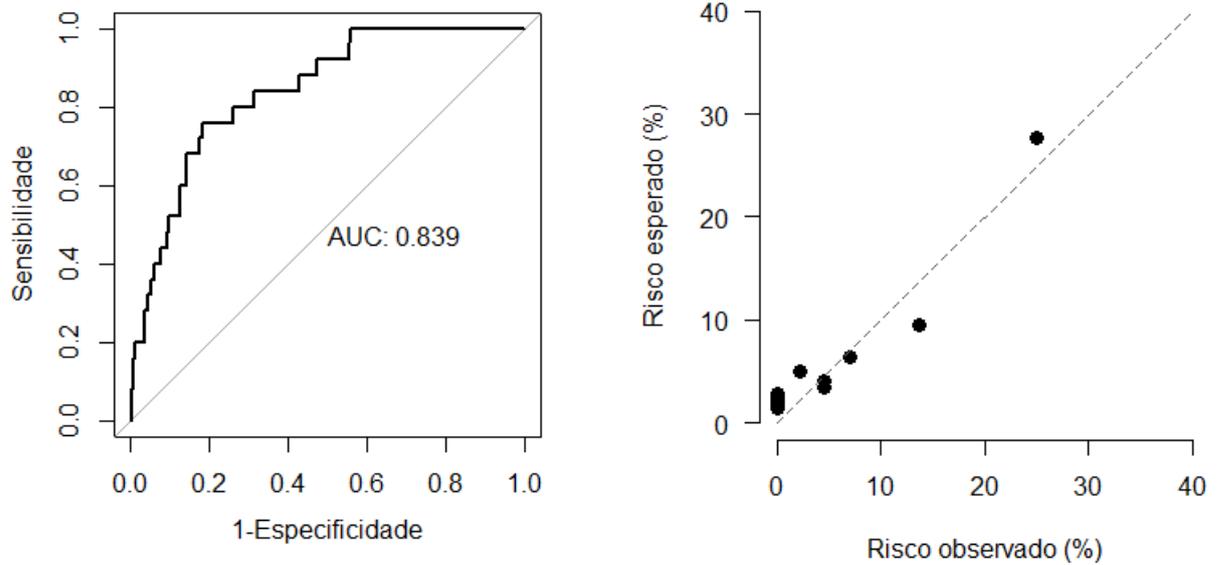


Figura 10 – Avaliação do desempenho do modelo de predição no conjunto de teste. Curva ROC (esquerda) e gráfico de valores observados *vs.* esperados (direita).

4.2.2 Dinâmico

Para a predição dinâmica foram ajustados os modelo de Cox na nova linha de base com $n = 283$ pacientes e o modelo conjunto em $n = 1824$ observações com *frequência cardíaca* avaliada nos dois instantes de medição. As estimativas para os coeficientes das 5 covariáveis no modelo *landmark* encontram-se na Tabela 14. Nota-se que no ajuste feito na amostra menor, os efeitos de *ECG*, *classe funcional* e *frequência cardíaca* passaram a ser não significativos. A sobrevida basal em 4 anos, tendo este ajuste como base, é $\hat{S}_0(4) = 0.935$, note que essa estimativa é um pouco menor que a obtida pelo modelo de linha de base para este mesmo instante de tempo.

Tabela 14 – Ajuste do modelo de Cox na nova linha de base (*landmark*).

	Estimativa	\widehat{RTF} (IC 95%)	Erro padrão	<i>p</i> -valor
<i>IC</i>	0.954	2.596 (1.633;4.129)	0.236	<0.001
<i>ECG</i>	0.242	1.274 (0.803;2.019)	0.235	0.304
<i>CF</i>	0.160	1.174 (0.743;1.854)	0.233	0.493
<i>Idade</i>	0.016	1.016 (0.999;1.033)	0.008	0.069
<i>Freq. Cardíaca</i>	0.001	1.001 (0.990;1.013)	0.005	0.810

Para a modelagem conjunta, todas as covariáveis, exceto *idade*, foram usadas para explicar a *frequência cardíaca* como resposta longitudinal. No submodelo de sobrevivência atribuiu-se à $\lambda_0(t)$ a distribuição o Weibull e incluíram-se todas as 5 covariáveis, com *Freq. Cardíaca 2* sendo o valor obtido pelo submodelo longitudinal. Na Tabela 15, são mostradas as estimativas do modelo ajustado.

Tabela 15 – Ajuste do submodelo de sobrevivência dentro da modelagem conjunta com a especificação da distribuição Weibull para $\lambda_0(t)$.

	Estimativa	Erro padrão	Valor z	p -valor
<i>IC</i>	1.441	0.157	9.149	<0.001
<i>ECG</i>	0.449	0.159	2.821	0.005
<i>CF</i>	0.341	0.156	2.188	0.029
<i>Freq. Cardíaca 2</i>	0.006	0.003	1.742	0.082
<i>Idade</i>	0.032	0.006	5.389	<0.001

Os resultados na tabela acima indicam todas que as covariáveis estão associadas com o tempo até o óbito, embora o efeito de *Freq. Cardíaca 2* não seja tão expressivo como foi no modelo de linha de base. Os ajustes para o *landmark* e para a modelagem conjunta viabilizam a obtenção das estimativas para o risco de óbito no novo horizonte de predição.

Os modelos para predição dinâmica foram submetidos a uma validação a partir das medidas de discriminação e calibração usando a técnica de validação cruzada com *5-folds*. Na Tabela 16 encontram-se os resultados com as estimativas para a AUC e o PE. Com base na estimativa pontual da AUC, a abordagem com melhor desempenho na discriminação foi o *Naive 2* ($\widehat{AUC} = 0.775$) e a pior, o *Naive 1* ($\widehat{AUC} = 0.551$). No entanto, observa-se que o *Naive 2* possui um intervalo de confiança mais largo, comparado com o modelo conjunto ($\widehat{AUC} = 0.737$), que teve a segunda melhor estimativa da AUC. A partir do erro de predição esperado, o modelo de Cox na nova linha de base, se mostrou mais calibrado ($\widehat{PE} = 0.069$), gerando probabilidades preditas mais próximas do observado.

Tabela 16 – Medidas de avaliação da performance por validação cruzada com $k=5$ *folds* para predição do risco de óbito em 4 anos.

	\widehat{AUC} (IC 95%)	\widehat{PE}
<i>Naive 1</i>	0.551 (0.259;0.843)	0.388
<i>Naive 2</i>	0.775 (0.531;0.992)	0.228
<i>LM</i>	0.649 (0.379;0.906)	0.069
<i>JM</i>	0.737 (0.621;0.852)	0.462

Como o modelo *LM* foi ajustado para uma amostra menor, as medidas de validação acabam não sendo totalmente comparáveis. Na intenção ter uma forma a qual fosse possível comparar as quatro abordagens, foram geradas curvas de sobrevivência (Kaplan-Meier) de acordo com as seguintes categorias de risco: baixo, intermediário e alto. Cada categoria de risco foi definida de acordo com um limiar de risco específico, estes limiares são apresentados na Tabela 17.

O Limiar 1 foi definido para o modelo de predição na linha de base conjuntamente com os especialistas e, a partir destes limiares foram obtidos bons resultados no estudo feito para o escore de risco de óbito em 2 anos. Como o comportamento do risco tende a mudar na predição dinâmica, foram especificados outros pontos para definir as categorias de risco, o Limiar 2. É esperado que um seja modelo bom se conseguir segregar os grupos de risco de tal forma que a sobrevida seja maior para a categoria “baixo risco”, que seja menor para “alto risco” e que a sobrevida da categoria “risco intermediário” esteja entre essas duas.

Tabela 17 – Limiares para definir as categorias do risco predito.

Risco	Limiar 1	Limiar 2
Baixo	$< 2\%$	$< 5\%$
Intermediário	$\geq 2\%$ e $< 10\%$	$\geq 5\%$ e $< 20\%$
Alto	$\geq 10\%$	$\geq 20\%$

Para as categorias obtidas com o Limiar 1 (Figura 11), o método LM não atribuiu, para nenhum dos 283 pacientes, estimativas do risco menores que 2%, ou seja, nenhum paciente foi alocado na categoria de alto risco, por essa razão não há a curva de sobrevivência para essa categoria. Observa-se um comportamento bastante similar entre as abordagens *Naive 1* e *JM*, de acordo com esses métodos, nenhum paciente presente na categoria de risco baixo sofreu o evento e ainda, observa-se que há uma separação das curvas de risco intermediário e risco alto (diferença significativa: Tabela 18). No método *Naive 2*, verifica-se que pacientes do grupo de baixo risco têm a sobrevida maior, no entanto as curvas para as outras duas categorias de risco não diferem (p -valor=0.827).

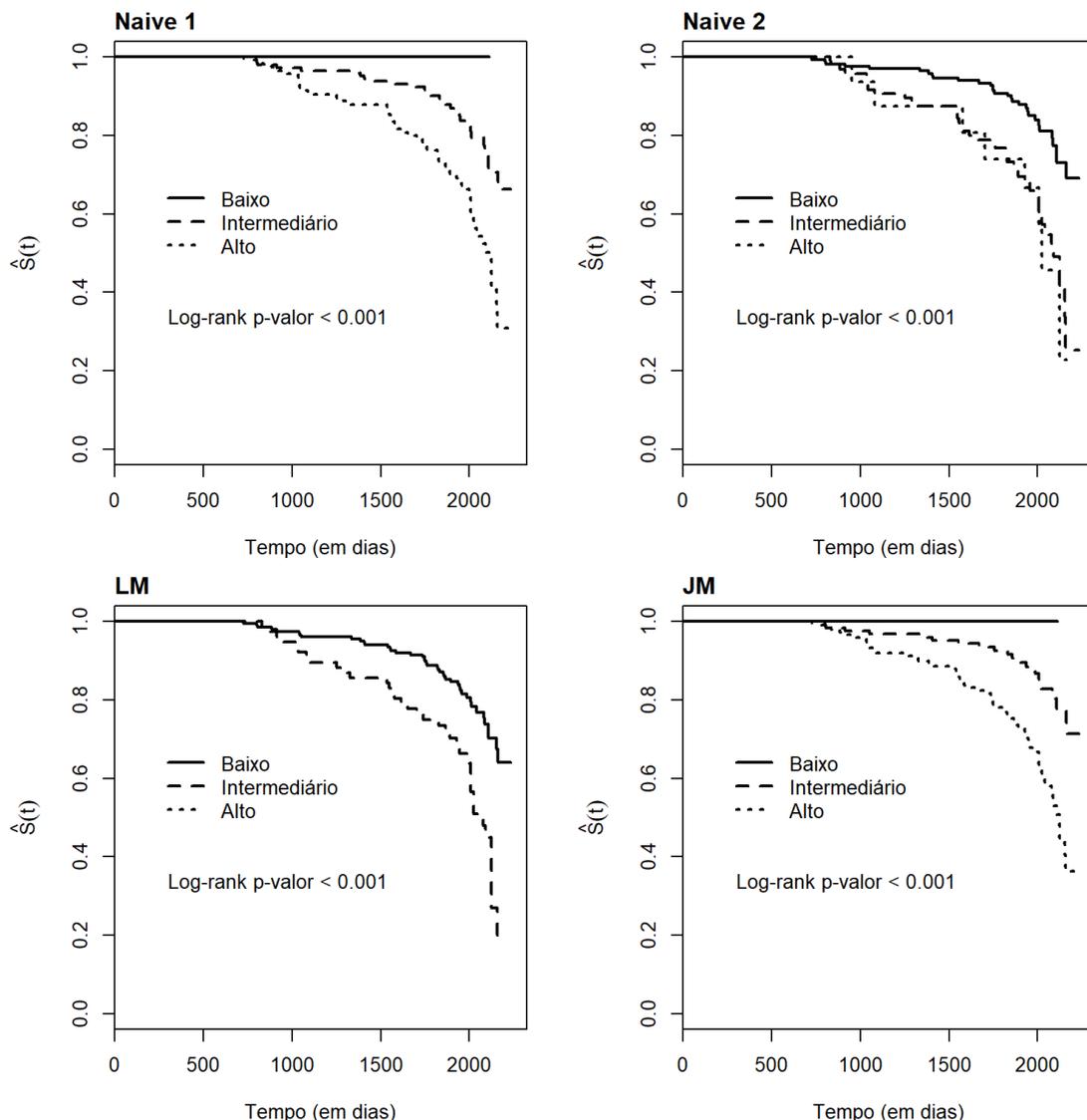


Figura 11 – Curvas de Kaplan-Meier de acordo com os grupos de risco baixo, intermediário e alto para o Limiar 1.

Na Figura 12, são exibidas as curvas de sobrevivência para os grupos gerados a partir do Limiar 2. Nota-se que, para o *Naive 2*, as curvas têm um padrão que difere do anterior e, em parte do acompanhamento, pacientes alocados no grupo de maior risco apresentam sobrevida maior. No caso do *LM*, observa-se que ao aumentar o limiar do grupo de alto risco, poucos pacientes que experimentaram o evento em 4 anos e, que deveriam ter o risco mais alto, foram para essa categoria. Para o *Naive 1* agora não há quase nenhuma diferença entre as categorias de risco, embora o teste tenha apontado diferença significativa entre as categorias “baixo” e “intermediário”. A abordagem *JM* é a única em que as curvas se comportam aproximadamente como o esperado (curvas separadas para as três categorias), sendo constatada diferença significativa entre os grupos “intermediário” e “alto” (ver Tabela 19).

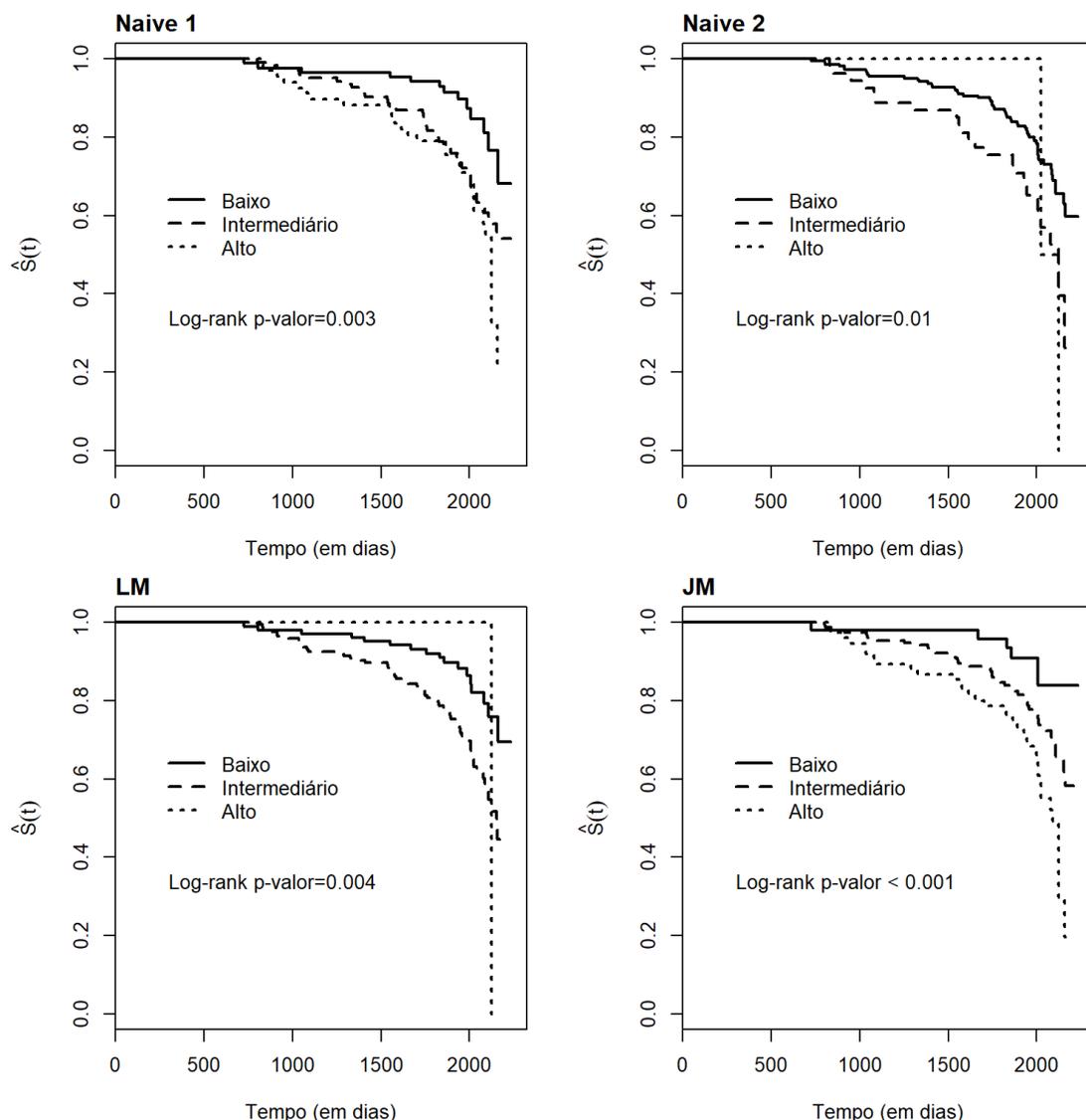


Figura 12 – Curvas de Kaplan-Meier para os grupos de risco baixo, intermediário e alto para o Limiar 2.

Como foi apontado na avaliação dos gráficos, o teste de Log-rank atestou diferença significativa entre alguns grupos de risco de acordo com a abordagem. Para identificar quais grupos diferem, foi feita uma comparação dois-a-dois das curvas entre as categorias de riscos “baixo contra intermediário” e “intermediário contra alto” usando o método de Bonferroni. O nível de significância corrigido para os 3 grupos, é $\alpha^* = 0.05/3 = 0.017$. As tabelas de comparação (Tabelas 18 e 19) apresentam a estatística de teste e o p -valor associados à cada comparação.

Tabela 18 – Comparação dois-a-dois das curvas de sobrevivência pelo teste de Log-rank, Limiar 1 (χ^2 : estatística de teste).

	Baixo vs. intermediário		Intermediário vs. alto	
	χ^2	p -valor	χ^2	p -valor
<i>Naive 1</i>	2.733	0.098	16.122	< 0.001
<i>Naive 2</i>	20.054	< 0.001	0.047	0.827
<i>LM</i>	-	-	-	-
<i>JM</i>	0.635	0.425	18.356	< 0.001

Com base nos resultados na Tabela 18, pode-se concluir que há diferença significativa entre as categorias “baixo” e “intermediário” nas abordagens *Naive 1* e *Naive 2*. Na comparação Intermediário vs. alto constata-se diferença entre os grupos do *Naive 1* e *JM*.

Tabela 19 – Comparação dois-a-dois das curvas de sobrevivência pelo teste de Long-rank, Limiar 2. (χ^2 : estatística de teste)

	Baixo vs. intermediário		Intermediário vs. alto	
	χ^2	p -valor	χ^2	p -valor
<i>Naive 1</i>	7.296	0.007	0.989	0.319
<i>Naive 2</i>	6.936	< 0.008	0.070	0.791
<i>LM</i>	10.708	0.001	0.136	0.711
<i>JM</i>	3.683	0.055	7.148	0.007

Na comparação apresentada na Tabela 19 observa-se que há evidências de diferença entre a sobrevida das categorias “baixo” e “intermediário” em todas as abordagens. Apenas a modelagem conjunta apresenta diferença entre os grupos de risco intermediário e alto, o que, sob uma visão clínica, a torna uma abordagem interessante por distinguir os três grupos.

4.3 Predições

Um cenário de predição foi construído para reproduzir as estimativas do risco de óbito de acordo com algumas possíveis combinações das covariáveis. Supondo que são 18 pacientes com idades de 25 e 65 anos e com três diferentes configurações para as variáveis dicotômicas: sendo alguns indivíduos com todas marcadas como “1” (caso mais grave), apenas *IC* (covariável com maior efeito nos modelos) ou com “0” em todas (caso menos grave). As duas medidas da frequência cardíaca estão de acordo com o seguinte padrão: estável, aumento ou diminuição comparando a primeira com a segunda medida. Os padrões completos são apresentados na Tabela 20.

Tabela 20 – Padrões das covariáveis para os 18 pacientes do cenário fictício.

Paciente	Idade	Freq. Card 1	Freq. Card 2	IC	ECG	CF
1		60	70	0	0	0
2		80	40	0	0	0
3		70	120	0	0	0
4	25	60	70	1	0	0
5		80	40	1	0	0
6		70	120	1	0	0
7		60	70	1	1	1
8		80	40	1	1	1
9		70	120	1	1	1
10		60	70	0	0	0
11		80	40	0	0	0
12		70	120	0	0	0
13	65	60	70	1	0	0
14		80	40	1	0	0
15		70	120	1	0	0
16		60	70	1	1	1
17		80	40	1	1	1
18		70	120	1	1	1

Para cada um dos 18 pacientes foram obtidas as estimativas do risco de óbito em 4 anos e os intervalos de predição considerando as covariáveis fixas e o histórico das duas medidas de *frequência cardíaca*. Os resultados são apresentados na Figura 13, em que, para cada paciente, há 4 estimativas, que se referem a cada uma das abordagens.

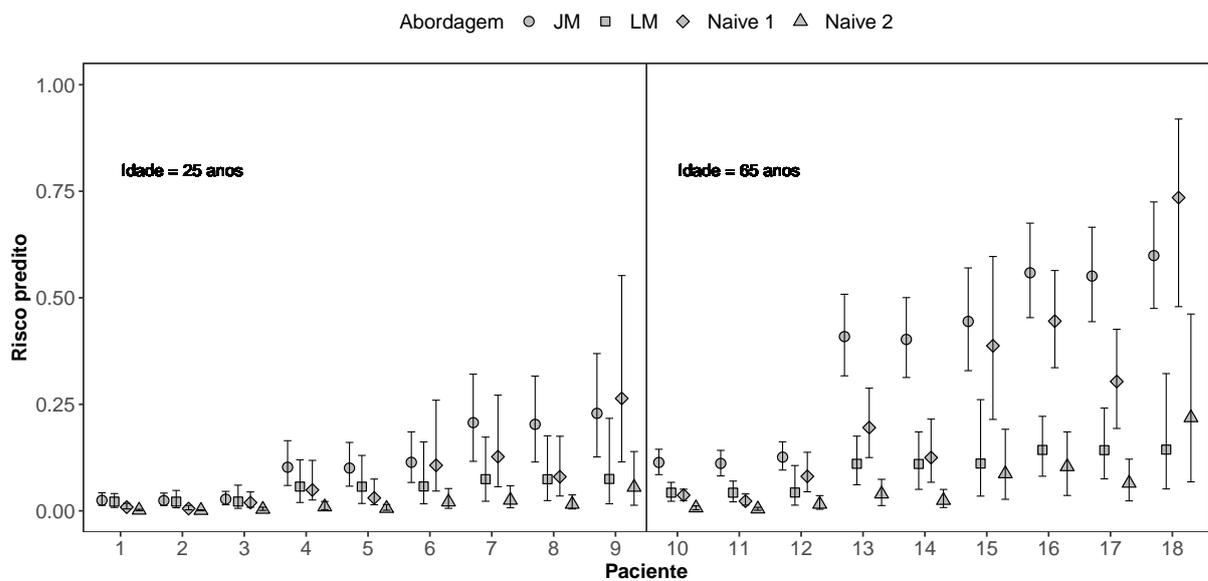


Figura 13 – Estimativas do risco de óbito em 4 anos e intervalo de predição *bootstrap* percentílico de 95% para os indivíduos exibidos na Tabela 20.

A partir dos resultados acima é possível identificar o efeito que *idade* tem de aumentar o risco, principalmente ao comparar padrões daqueles pacientes considerados

mais graves (que assumem valor 1 para *IC*, *ECG* e *CF*). Quanto ao marcador *frequência cardíaca*, aparentemente, também tem um efeito de aumentar o risco quando cresce ou permanece estável (com medidas consideradas de boa a normal) e quando o paciente tem uma segunda medida menor que a primeira, o risco parece ser sutilmente menor.

Comparando estimativas do risco de acordo com as abordagens, nota-se que, ao contrário do que foi visto no exemplo de CD4, a modelagem conjunta e o *landmark* resultam em estimativas maiores que *Naive 1* e *Naive 2*. Isso é fortemente motivado pela limitação do estudo de haver poucas medidas para ajustar o modelo conjunto e assim captar da melhor forma o verdadeiro efeito do marcador longitudinal e, também pouca informação sobre o tempo de acompanhamento do estudo para aqueles que sobreviveram aos dois primeiros anos e compõem a amostra na qual se estima o modelo de Cox na nova linha de base.

5 Conclusões

5.1 Considerações finais

O presente trabalho abordou métodos para realizar predição dinâmica para dados longitudinais e de sobrevivência. Foi visto que o modelo de regressão de Cox, já bem estabelecido na literatura e frequentemente usado em aplicações de dados clínicos, é uma ferramenta que pode ser utilizada tanto para criar um modelo de predição estática a partir da linha de base, quanto dinâmico, seguindo as abordagens *Naive 1*, *Naive 2* e *Landmark*. Como alternativa para predição dinâmica, tem-se a modelagem conjunta, que vem se estabelecendo para o enfoque de predição.

Na linha de base constatou-se, com o exemplo da contagem de CD4, que todas as abordagens são capazes de discriminar eventos de não eventos e também têm boa calibração. Para esse exemplo, concluiu-se que o modelo de Cox com covariável tempo-dependente, apesar de ter mais informação na estimação, não se sai bem nas predições e tem fraca calibração. Já o modelo de Cox ajustado somente com informações na linha de base, consegue se sair ligeiramente melhor, por apresentar resultados com boa discriminação, ser calibrado e ter menor erro de predição. Para os dados do SaMi-Trop, o modelo de Cox na linha de base também teve uma boa performance de predição para o risco de óbito em 2 anos, com medidas de validação muito boas.

No enfoque de predição dinâmica, todas as abordagens apresentaram medidas de performance muito próximas, sendo que com o *Naive 1* e o *Naive 2* com um pouco mais de incerteza para a medida de discriminação. As abordagens *Naive 1* e JM, se mostraram parecidas em termos de calibração, apesar das estimativas do erro, para as 4 abordagens, não diferirem tanto. O *Naive 2* tanto no exemplo, quanto na aplicação teve AUC médio melhor. Entretanto, na aplicação real mostrou ter uma incerteza maior para esta medida. Apesar da limitação dos dados na aplicação, foi visto que o modelo de Cox para a abordagem de *landmark* também é uma boa e simples alternativa para realizar predições dinâmicas.

A modelagem conjunta não apresentou resultados diferentes das outras abordagens para o exemplo de CD4. Na aplicação, claramente teve seu desempenho afetado pela escassez do número de medidas de frequência cardíaca, o que fez com que tivesse a performance similar ao *Naive 1*. No entanto, ainda assim se mostrou possível de aplicação nesse cenário.

Ao realizar predições no cenário fictício, as abordagens *Naive 1* e *Naive 2* retornam estimativas mais baixas do risco de óbito comparados ao *LM* e *JM*, com modelagem

conjunta gerando estimativas mais altas do que todos os métodos.

A obtenção dos intervalos de predição é um ponto que difere bastante as abordagens com o modelo de Cox para a modelagem conjunta. Para gerar os intervalos foram utilizadas 200 reamostras *bootstrap* e as estimativas intervalares para o modelo de Cox foram obtidas em menos de 5 segundos, enquanto que esse processo demorou em torno de 1 hora para a modelagem conjunta

Embora, a modelagem conjunta seja mais completa em termos de modelar tanto a parte de sobrevivência quando a parte longitudinal dos dados, ela acabará sendo afetada quando houver poucas informações do marcador longitudinal. Ao optar por trabalhar com o modelo de Cox tem-se uma ferramenta simples e eficiente.

Como trabalhos futuros, ressalta-se que o cenário de dados longitudinais e de sobrevivência pode ser abordado com extensões mais flexíveis do método de *landmark*. Há propostas feitas por Zheng e Heagerty (2005), que tratam da predição dinâmica da probabilidade de sobrevivência utilizando *modelos parcialmente condicionais*. Houwelingen (2007) também apresenta técnicas com esse objetivo, mas mantém o intuito inicial dado para essa abordagem. Zheng e Heagerty (2005), Houwelingen (2007) e Maziarz et al. (2017) trabalham em aproximações que permitem considerar os efeitos dinâmicos das covariáveis longitudinais ao longo do tempo.

Referências

- ABRAMS, D. I.; GOLDMAN, A. I.; LAUNER, C.; KORVICK, J. A.; NEATON, J. D.; CRANE, L. R.; GRODESKY, M.; WAKEFIELD, S.; MUTH, K.; KORNEGAY, S. et al. A comparative trial of didanosine or zalcitabine after treatment with zidovudine in patients with human immunodeficiency virus infection. *New England Journal of Medicine*, v. 330, n. 10, p. 657–662, 1994.
- ANDERSEN, P. K.; GILL, R. D. Cox’s regression model for counting processes: a large sample study. *The Annals of Statistics*, n. 4, p. 1100–1120, 1982.
- ANDERSON, J. R.; CAIN, K. C.; GELBER, R. D. Analysis of survival by tumor response. *Journal of Clinical Oncology*, v. 1, n. 11, p. 710–719, 1983.
- BRESLOW, N. Covariance analysis of censored survival data. *Biometrics*, p. 89–99, 1974.
- BROWN, E. R.; IBRAHIM, J. G. A bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics*, v. 59, n. 2, p. 221–228, 2003.
- CARDOSO, C. S.; SABINO, E. C.; OLIVEIRA, C. D. L.; OLIVEIRA, L. C. de; FERREIRA, A. M.; CUNHA-NETO, E.; BIERRENBACH, A. L.; FERREIRA, J. E.; HAIKAL, D. S.; REINGOLD, A. L.; RIBEIRO, A. L. P. Longitudinal study of patients with chronic Chagas cardiomyopathy in Brazil (SaMi-Trop project): a cohort profile. *British Medical Journal Publishing Group*, v. 6, n. 5, 2016.
- COLLETT, D. *Modelling survival data in medical research*. New York: Chapman & Hall/CRC, 2015.
- COLOSIMO, E. A.; GIOLO, S. R. *Análise de sobrevivência aplicada*. São Paulo: Edgard Blucher, 2006.
- COOK, N. R. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*, v. 115, n. 7, p. 928–935, 2007.
- COX, D. R. Regression models and life-tables. *Journal of the Royal Statistical society. Series B (Methodological)*, [Royal Statistical Society, Wiley], v. 34, n. 2, p. 187–220, 1972.
- COX, D. R. Partial likelihood. *Biometrika*, v. 62, n. 2, p. 269–276, 1975.
- D’AGOSTINO, R.; NAM, B.-H. Evaluation of the performance of survival analysis models: Discrimination and calibration measures. In: *Advances in Survival Analysis*. [S.l.]: Elsevier, 2003. v. 23, p. 1 – 25.
- DEMLER, O. V.; PAYNTER, N. P.; COOK, N. R. Tests of calibration and goodness-of-fit in the survival setting. *Statistics in Medicine*, v. 34, n. 10, p. 1659–1680, feb 2015.
- EFRON, B.; TIBSHIRANI, R. *An Introduction to the Bootstrap*. New York: Chapman and Hall, 1994.
- ELASHOFF, R.; LI, N. et al. *Joint modeling of longitudinal and time-to-event data*. New York: Chapman and Hall/CRC, 2016.
- GRAMBSCH, P. M.; THERNEAU, T. M. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, v. 81, n. 3, p. 515–526, 1994.

- HANLEY, J. A.; MCNEIL, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, v. 143, n. 1, p. 29–36, 1982.
- HARRELL, F. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. [S.l.]: Springer International Publishing, 2015. (Springer Series in Statistics).
- HARRELL, F. E.; LEE, K. L.; MARK, D. B. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, v. 15, n. 4, p. 361–387, 1996.
- HARRELL JR., F. E. *Regression Modeling Strategies*. New York: Springer, 2006.
- HENDERSON, R.; DIGGLE, P.; DOBSON, A. Identification and efficacy of longitudinal markers for survival. *Biostatistics*, v. 3, n. 1, p. 33–50, 2002.
- HOSMER, D. W.; LEMESHOW, S. Goodness of fit tests for the multiple logistic regression model. *Communications in Statistics-Theory and Methods*, v. 9, n. 10, p. 1043–1069, 1980.
- HOUWELINGEN, H. C. V. Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics*, v. 34, n. 1, p. 70–85, 2007.
- HOUWELINGEN, H. van; PUTTER, H. *Dynamic prediction in clinical survival analysis*. Boca Raton: CRC Press, 2012.
- HSIEH, F.; TSENG, Y.-K.; WANG, J.-L. Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics*, v. 62, n. 4, p. 1037–1043, 2006.
- KLEIN, J. P.; MOESCHBERGER, M. L. *Survival analysis: techniques for censored and truncated data*. New York: Springer, 2003.
- MAZIARZ, M.; HEAGERTY, P.; CAI, T.; ZHENG, Y. On longitudinal prediction with time-to-event outcome: Comparison of modeling options. *Biometrics*, v. 73, n. 1, p. 83–93, 2017.
- NJAGI, E. N.; RIZOPOULOS, D.; MOLENBERGHS, G.; DENDALE, P.; WILLEKENS, K. A joint survival-longitudinal modelling approach for the dynamic prediction of rehospitalization in telemonitored chronic heart failure patients. *Statistical Modelling*, v. 13, n. 3, p. 179–198, 2013.
- PEPE, M. S. *The statistical evaluation of medical tests for classification and prediction*. Oxford: Oxford University Press, 2003.
- PROUST-LIMA, C.; TAYLOR, J. M. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment psa: a joint modeling approach. *Biostatistics*, v. 10, n. 3, p. 535–549, 2009.
- RIZOPOULOS, D. JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software*, v. 35, n. 9, p. 1–33, 2010.
- RIZOPOULOS, D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, Wiley, v. 67, n. 3, p. 819–829, 2011.
- RIZOPOULOS, D. *Joint models for longitudinal and time-to-event data: With applications in R*. New York: Chapman and Hall/CRC, 2012.
- RIZOPOULOS, D.; MOLENBERGHS, G.; LESAFFRE, E. M. Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. *Biometrical Journal*, v. 59, n. 6, p. 1261–1276, 2017.

- STEYERBERG, E. W. *Clinical prediction models: a practical approach to development, validation, and updating*. New York: Springer, 2008.
- STEYERBERG, E. W.; VICKERS, A. J.; COOK, N. R.; GERDS, T.; GONEN, M.; OBUCHOWSKI, N.; PENCINA, M. J.; KATTAN, M. W. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology*, v. 21, n. 1, p. 128–38, 2010.
- THERNEAU, T. M. *A Package for Survival Analysis in S*. [S.l.], 2015. Version 2.38. Disponível em: <<https://CRAN.R-project.org/package=survival>>.
- THERNEAU, T. M.; GRAMBSCH, P. M. *Modeling survival data: extending the Cox model*. New York: Springer, 2000.
- TSIATIS, A.; DEGRUTTOLA, V.; WULFSOHN, M. Modeling the relationship of survival to longitudinal data measured with error. applications to survival and cd4 counts in patients with aids. *Journal of the American Statistical Association*, v. 90, n. 429, p. 27–37, 1995.
- TSIATIS, A. A.; DAVIDIAN, M. Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica*, v. 14, n. 3, p. 809–834, 2004.
- WANG, Y.; TAYLOR, J. M. G. Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association*, v. 96, n. 455, p. 895–905, 2001.
- WOODWARD, M. *Epidemiology: study design and data analysis*. New York: Chapman and Hall/CRC, 2013.
- WOODWARD, M.; TUNSTALL-PEDOE, H.; PETERS, S. A. Graphics and statistics for cardiology: clinical prediction rules. *Heart*, v. 103, n. 7, p. 538–545, 2017.
- WULFSOHN, M. S.; TSIATIS, A. A. A joint model for survival and longitudinal data measured with error. *Biometrics*, v. 53, n. 1, p. 330–339, 1997.
- ZHENG, Y.; HEAGERTY, P. J. Partly conditional survival models for longitudinal data. *Biometrics*, v. 61, n. 2, p. 379–391, 2005.