

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

PEDRO VITOR BERNARDES BRANDÃO

AVALIAÇÃO DE DIFERENTES MÉTODOS DE PREDIÇÃO
DINÂMICA EM ANÁLISE DE SOBREVIVÊNCIA

Belo Horizonte
2025

PEDRO VITOR BERNARDES BRANDÃO

AVALIAÇÃO DE DIFERENTES MÉTODOS DE PREDIÇÃO
DINÂMICA EM ANÁLISE DE SOBREVIVÊNCIA

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.

Belo Horizonte
2025

Brandão, Pedro Vítor Bernardes.

B817a Avaliação de diferentes métodos de predição dinâmica em análise de sobrevivência [recurso eletrônico] / Pedro Vítor Bernardes Brandão – 2025.
1 recurso online (102 f. il., color.) : pdf.

Orientador: Enrico Antônio Colosimo.
Dissertação (mestrado) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística.
Referências: f. 99-102

1. Estatística – Teses. 2. Análise de sobrevivência (Biometria) algoritmos de predição – Teses. 3. Unidade de tratamento intensivo – Recém-nascidos - Teses. 4. Cardiopatias – Doença de Chagas - São Paulo – Minas Gerais – Teses. I. Colosimo, Enrico Antônio. I. Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística. III. Título.

CDU 519.2(043)



UNIVERSIDADE FEDERAL DE MINAS GERAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA



FOLHA DE APROVAÇÃO

Avaliação de diferentes métodos de predição dinâmica em análise de sobrevivência

PEDRO VÍTOR BERNARDES BRANDÃO

Dissertação submetida à Banca Examinadora, designada pelo Colegiado do Programa de Pós-Graduação em ESTATÍSTICA, como requisito para obtenção do grau de Mestre em ESTATÍSTICA, área de concentração ESTATÍSTICA E PROBABILIDADE.

Aprovada em 11 de abril de 2025, pela banca constituída pelos membros:

Documento assinado digitalmente
gov.br ENRICO ANTONIO COLOSIMO
Data: 11/04/2025 17:06:24-0300
Verifique em <https://validar.iti.gov.br>

Prof. Enrico Antonio Colosimo
Orientador – DEST/UFMG

Documento assinado digitalmente
gov.br SILVANA MARCIA BRUSCHI KELLES
Data: 15/04/2025 13:28:50-0300
Verifique em <https://validar.iti.gov.br>

Profa. Silvana Márcia Bruschi Kelles
Faculdade de Medicina/PUC Betim

Documento assinado digitalmente
gov.br VINICIUS DINIZ MAYRINK
Data: 16/04/2025 08:32:32-0300
Verifique em <https://validar.iti.gov.br>

Prof. Vinícius Diniz Mayrink
DEST/UFMG

Documento assinado digitalmente
gov.br TARCIANA LIBERAL PEREIRA DE ARAUJO
Data: 15/04/2025 01:02:34-0300
Verifique em <https://validar.iti.gov.br>

Profa. Tarciana Liberal Pereira
DEST/UFMG

Belo Horizonte, 11 de abril de 2025.

À Deus, minha esposa e minha família.

Agradecimentos

Primeiramente, agradeço a Deus, pois sem Ele eu não teria chegado até o fim desse percurso. Sua força e sabedoria me sustentaram em cada etapa desta jornada.

À minha esposa, Lais, meu mais profundo agradecimento. Seu apoio incondicional e sua presença constante foram meu maior suporte durante todo o processo de construção desta dissertação. Sem você, este caminho teria sido muito mais árduo.

Agradeço à minha mãe, Vânia, ao meu pai, Paulo, e ao meu irmão, Paulo Júnior, pelo incentivo contínuo e pelo apoio inestimável ao longo de todo o mestrado. Vocês foram meu alicerce, proporcionando-me força e motivação para seguir em frente.

Ao meu orientador, Enrico Antônio Colosimo, minha gratidão pela paciência, incentivo e por compartilhar seu conhecimento de forma generosa. Seu apoio foi fundamental para minha formação e crescimento acadêmico, e cada troca de experiência foi enriquecedora.

Também expressei minha gratidão à equipe atuarial e a todos que colaboraram da empresa na qual trabalho, que estiveram ativamente envolvidos na construção deste conhecimento e dos resultados. A contribuição de todos foi essencial para a concretização deste estudo.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001. Agradeço também ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pelo suporte concedido. O suporte dessas entidades foi fundamental para o desenvolvimento deste estudo, proporcionando acesso a estruturas acadêmicas adequadas, recursos bibliográficos essenciais e equipamentos que viabilizaram a realização da pesquisa.

A todos que, de alguma forma, estiveram ao meu lado e me ajudaram a chegar até aqui, o meu muito obrigado!

Resumo

Este trabalho explorou a aplicação de técnicas de predição dinâmica em análise de sobrevivência, avaliando sua eficácia em dois cenários clínicos: a utilização de CTI/UTI neonatal e risco de óbito em pacientes com cardiopatia chagásica. Quatro métodos principais foram utilizados: Naive 1, Naive 2, Landmark e Modelagem Conjunta (JM). No contexto da Modelagem Conjunta, foram avaliadas três distribuições para a linha de base (Weibull, Piecewise e Spline), permitindo uma análise detalhada das vantagens e limitações de cada abordagem. Os resultados indicaram que, nos estudos analisados, a técnica JM apresentou um bom desempenho em termos de discriminação e calibração, equiparando-se e, em alguns casos, superando outras abordagens de predição dinâmica estudadas neste trabalho, especialmente devido à sua capacidade de integrar variáveis longitudinais. Além disso, observou-se que o envelhecimento das variáveis dependentes do tempo influencia a performance dos modelos, ressaltando a importância de atualizações contínuas. Este estudo reforça a relevância de métodos dinâmicos para cenários clínicos complexos, com impacto direto na prática clínica e na tomada de decisões baseadas em evidências.

Palavras-Chave: dados longitudinais e de sobrevivência; modelo de Cox; landmark; modelagem conjunta; Weibull; Piecewise; Spline; predição de risco.

Abstract

This study explored the application of dynamic prediction techniques in survival analysis, assessing their effectiveness in two clinical scenarios: the use of neonatal ICU and the risk of death in patients with Chagas cardiomyopathy. Four main methods were employed: Naive 1, Naive 2, Landmark, and Joint Modeling (JM). Within the Joint Modeling framework, three baseline distributions were evaluated (Weibull, Piecewise, and Spline), allowing for a detailed analysis of the advantages and limitations of each approach. The results indicated that, in the analyzed studies, the JM technique demonstrated strong performance in terms of discrimination and calibration, matching or even surpassing other dynamic prediction approaches studied in this work, particularly due to its ability to integrate longitudinal variables. Additionally, it was observed that the aging of time-dependent variables influences model performance, highlighting the importance of continuous updates. This study reinforces the relevance of dynamic methods for complex clinical scenarios, with a direct impact on clinical practice and evidence-based decision-making.

Keywords: longitudinal and survival data; Cox model; landmark; joint modeling; Weibull; Piecewise; Spline; risk prediction.

Lista de Figuras

1	Ilustração da Linha do Tempo do acompanhamento clínico dinâmico.	27
2	Acompanhamento clínico do estudo de caso de CTI/UTI neonatal.	34
3	Box-Plot da variável percentil do peso.	42
4	Acompanhamento clínico do estudo de caso de cardiopatas chagásicos.	42
5	Resíduos de Schoenfeld versus tempo para os marcadores.	51
6	Resíduos de Schoenfeld versus tempo para os marcadores.	52
7	Resíduos de Schoenfeld versus tempo para os marcadores.	53
8	Resíduos de Schoenfeld versus tempo para os marcadores.	54
9	Curva ROC para predição de 28 dias.	55
10	Risco esperados vs. risco observados para predição de 28 dias.	55
11	Curvas de Kaplan-Meier para os modelos Naive 1 e Naive 2 com o limiar 1.	64
12	Curvas de Kaplan-Meier para os modelos LandMark e Modelos Con-juntos com o limiar 1.	65
13	Curvas de Kaplan-Meier para os modelos Naive 1, Naive 2, LandMark e JM - Weibull com o limiar 2.	66
14	Curvas de Kaplan-Meier para os modelos JM - Piecewise e JM - Spline com o limiar 2.	67
15	Curvas de Kaplan-Meier para os quatro primeiros modelos.	75
16	Curvas de Kaplan-Meier para os dois modelos restantes.	76
17	Curvas de Kaplan-Meier para os dois primeiros modelos.	76
18	Curvas de Kaplan-Meier para os quatro modelos restantes.	77
19	Resíduos de Schoenfeld versus tempo para os marcadores.	81
20	Curva ROC.	84
21	Risco esperados vs. risco observados.	84
22	Curvas de Kaplan-Meier segundo os grupos de risco (baixo, inter-mediário e alto) conforme o limiar 1 – abordagens: Naive 1, Naive 2, LandMark e JM - Weibull.	90
23	Curvas de Kaplan-Meier segundo os grupos de risco (baixo, inter-mediário e alto) conforme o limiar 1 – abordagens: JM - Piecewise e JM - Spline.	91
24	Curvas de Kaplan-Meier segundo os grupos de risco (baixo, inter-mediário e alto) conforme o limiar 2 – abordagens: Naive 1, Naive 2, LandMark e JM - Weibull.	92

25	Curvas de Kaplan-Meier segundo os grupos de risco (baixo, inter-	
	mediário e alto) conforme o limiar 2 – abordagens: JM - Piecewise e	
	JM - Spline.	93

Lista de Tabelas

1	Distribuição de Frequências (Absoluta e Proporcional) das Variáveis relacionadas as mães.	37
2	Distribuição de Frequências (Absoluta e Proporcional) do número de semanas da gestação.	38
3	Distribuição de Frequências (Absoluta e Proporcional) da Idade no momento do parto	38
4	Distribuição de Frequências (Absoluta e Proporcional) do Tempo de internação até o parto.	39
5	Distribuição de Frequências (Absoluta e Proporcional) das Variáveis relacionadas ao estado de saúde das mães.	40
6	Frequência relativa de Apgar do Primeiro Minuto por indivíduos que utilizaram/não utilizaram CTI/UTI neonatal.	41
7	Distribuição de Frequências (Absoluta e Proporcional) do estado de saúde das crianças.	41
8	Descrição da Idade (2013-2014).	43
9	Descrição da Frequência cardíaca (2013-2014).	44
10	Média e mediana das covariáveis quantitativas e frequência relativa das covariáveis categóricas ao longo do tempo.	45
11	Estimativas, risco relativo, EP e p-valor - Modelo de Cox.	49
12	Performance do modelo (discriminação (AUC) e calibração (Teste de Hosmer Lemenshow e Erro de Predição)) de linha de base para prever a utilização de CTI/UTI neonatal em 28 dias.	54
13	Estimativas, razão de taxa de falha, erro padrão e p-valor do modelo LandMark para linha de base em 28 dias para o risco de internação em CTI.	56
14	Estimativas, $\hat{H}R$ e p-valor do submodelo de sobrevivência com especificação da distribuição Weibull na linha de base.	57
15	Estimativas, $\hat{H}R$ e p-valor do submodelo de sobrevivência com especificação da distribuição Piecewise na linha de base.	58
16	Estimativas, $\hat{H}R$ e p-valor do submodelo de sobrevivência com especificação da distribuição Spline na linha de base.	59
17	Resultados de validação pelas métricas de discriminação (AUC) e calibração (Teste de Hosmer Lemenshow e Erro de Predição) por modelo e método.	60
18	Limiares para definir as categorias do risco predito.	62

19	Classificação dos limiares, evento e risco para os diferentes modelos. Sendo o evento a internação em CTI/UTI neonatal.	63
20	Comparação dois-a-dois das curvas de sobrevivência pelo teste de log- rank, Limiar 1.	68
21	Comparação dois-a-dois das curvas de sobrevivência pelo teste de log- rank, Limiar 2.	69
22	Estimativas, razão de taxa de falha, erro padrão e p-valor do modelo LandMark – 1 Ano (365 dias).	70
23	Resultados de validação pelas métricas de discriminação (AUC) e calibração (Teste de Hosmer Lemenshow e Erro de Predição) por modelo e método em 1 ano.	71
24	Limiares para definir as categorias do risco predito.	72
25	Classificação dos limiares, evento e risco para os diferentes modelos. Sendo o evento a internação em CTI/UTI neonatal.	73
26	Comparação dois-a-dois das curvas de sobrevivência pelo teste de log- rank, Limiar 1.	78
27	Comparação dois-a-dois das curvas de sobrevivência pelo teste de log- rank, Limiar 2.	79
28	Estimativas, razão das taxas de falha, erro padrão e p-valor do modelo de Cox	80
29	Resultados de validação pelas métricas de discriminação (AUC) e ca- libração (Teste de Hosmer Lemenshow e Erro de Predição) do modelo de linha de base.	82
30	Resultados de validação pelas métricas de discriminação (AUC) e calibração (Teste de Hosmer Lemenshow e Erro de Predição) por modelo e método em 10 repetições, utilizando 5 folds e 200 reamostras bootstrap	83
31	Estimativas, razão das raxas de falha, erro padrão e p-valor do modelo de Cox em uma nova linha de base (Landmark)	84
32	Estimativas, $\hat{H}R$ e p-valor do submodelo de sobrevivência com espe- cificação da distribuição de Weibull na linha de base.	85
33	Estimativas, $\hat{H}R$ e p-valor do submodelo de sobrevivência com espe- cificação da distribuição Piecewise na linha de base (5 nós).	85
34	Estimativas, $\hat{H}R$ e p-valor do submodelo de sobrevivência com espe- cificação da distribuição Spline na linha de base (5 nós)	86

35	Resultados de validação pelas métricas de discriminação (AUC) e calibração (Teste de Hosmer Lemenshow e Erro de Predição) por modelo e método.	87
36	Limiares para definir as categorias do risco predito.	88
37	Classificação de risco, óbitos e probabilidade de morte para os diferentes modelos.	89
38	Comparação dois-a-dois das curvas de sobrevivência pelo teste de log-rank, Limiar 1.	94
39	Comparação dois-a-dois das curvas de sobrevivência pelo teste de log-rank, Limiar 2.	94

Sumário

1	Introdução	15
1.1	Predições Clínicas	15
1.2	Objetivos	18
1.3	Organização do Trabalho	18
2	Análise de Sobrevivência para Predição de Riscos	19
2.1	Modelo de Regressão de Cox	20
2.2	Predição Estática	22
2.2.1	Suposição de Taxas de Falha Proporcionais	22
2.3	Validação do Modelo de Predição	23
2.3.1	Discriminação	23
2.3.2	Calibração	24
3	Predição Dinâmica	27
3.1	Abordagem Ingênua	27
3.2	Landmark	28
3.3	Modelagem Conjunta	29
3.3.1	Modelo Longitudinal	29
3.3.2	Modelo de Sobrevivência	29
3.3.3	Função de Taxa de Falha	30
3.3.4	Estimação	32
3.3.5	Predição	32
3.4	Avaliação dos Modelos	33
4	Bases de Dados	34
4.1	Caracterização da Base de Dados: Utilização de CTI/UTI Neonatal	34
4.1.1	Breve Descrição dos Dados	37
4.2	Caracterização da Base de Dados: Coorte de Cardiopatas Chagásicos	42
	- SamiTrop	42
4.2.1	Breve Descrição dos Dados	43
5	Métodos Computacionais	46
6	Resultados	49
6.1	Modelos de Predição - Utilização CTI/UTI Neonatal	49
6.1.1	Modelo de Cox na Linha de Base	49
6.1.2	Predição Estática em 28 dias	54

6.1.3	Modelo de Predição Dinâmica - 6 Meses	56
6.1.4	Estudo dos Limiares de Risco - 6 Meses	61
6.1.5	Modelo de Predição Dinâmica - 1 Ano	69
6.1.6	Estudo dos Limiares de Risco - 1 Ano	72
6.2	Modelos de Predição - Coorte Prospectiva SamiTrop	80
6.2.1	Modelo de Cox na Linha de Base	80
6.2.2	Modelo de Predição Dinâmica	84
6.2.3	Estudo dos Limiares de Risco	88
7	Conclusão e Próximos Passos	96
8	Referências	99

1 Introdução

1.1 Predições Clínicas

O avanço das tecnologias na saúde tem intensificado a busca por estratégias de promoção e prevenção de riscos, impulsionando o desenvolvimento de modelos probabilísticos para estimar a ocorrência de eventos em uma janela temporal, como morte por uma causa específica, manifestação ou cura de uma doença, entre outros.

Conforme apresenta Rizopoulos et al. (2017) em seu estudo, atualmente existe um grande interesse na avaliação precisa de risco para prevenção e tratamento de doenças. Os médicos usam escores de risco para tomar decisões apropriadas, como prescrever tratamento ou exames médicos extras ou sugerir terapias alternativas. Pessoas que apresentam doenças crônicas, em geral, possuem tempo de vida curto, sendo essencial no acompanhamento dos pacientes as perspectivas de sobrevivência no curto e longo prazo, aumentando a qualidade de vida e a assertividade no tratamento.

Existem diversas abordagens para a construção de algoritmos de predição, dentre elas os modelos de análise de sobrevivência. Esses que são corriqueiramente utilizados na área da saúde, se baseiam nas estimativas de probabilidade do tempo de sobrevivência dos indivíduos. Duas grandes vertentes na análise da sobrevivência, são os modelos de previsão estático e dinâmico.

Para a construção dos modelos, são utilizadas covariáveis que dizem respeito aos indivíduos e ao seu estado de saúde, tais como: idade, sexo, estado civil, peso, Eletrocardiograma (ECG), dentre outras. Essas variáveis serão essenciais para estimação da probabilidade do tempo até a ocorrência do evento de interesse, sendo que elas podem ser divididas em dois grandes grupos: estáticas e dinâmicas. As variáveis estáticas são aquelas que não se alteram com o passar do tempo, ou seja, após a primeira coleta de informações, elas não serão registradas novamente já que sua natureza é imutável, ou então, pode demorar muito tempo para ser alterada. Já as variáveis dinâmicas ou dependentes do tempo, são aquelas que estão em constante mudança, e por isso nas novas medições precisarão ser atualizadas. Algumas dessas variáveis, denominadas biomarcadores, são medidas biológicas que fornecem informações sobre o estado de saúde e podem refletir a evolução da doença, como por exemplo a frequência cardíaca, pressão sanguínea e peso.

Na modelagem estatística, é possível mensurar a probabilidade de um evento ocorrer até um determinado tempo. Para realizar a predição, as informações são extraídas no início do acompanhamento dos indivíduos, na qual é conhecido como linhas de base. Neste momento inicial, podem ser extraídas tanto variáveis estáticas

como dinâmicas, porém, as medições serão realizadas uma única vez, ou seja, os biomarcadores não serão medidos novamente dentro do intervalo de predição.

Estudos que envolvem respostas temporais são em geral prospectivos e de longa duração (Colosimo e Giolo, 2024), desta forma, a utilização dos biomarcadores em modelos estáticos pode causar um certo viés nos resultados, pois o comportamento dessas variáveis podem não ser adequadamente capturado por modelos que realizam medições uma única vez na linha temporal. Segundo Houwelingen e Putter (2022) a preocupação em utilizar variáveis dinâmicas em modelos estáticos é que o “envelhecimento” dos preditores a partir da linha de base, pode levar a uma incompatibilidade entre o verdadeiro valor subjacente do biomarcador e a última observação. Sendo que esse erro de medição pode inclusive ocasionar a violação de riscos proporcionais.

Além disso, quando se trata de estudos clínicos, os sintomas ou prognósticos podem variar substancialmente ao longo do tempo, desta forma, são comumente utilizados os modelos de predições dinâmicos. Em termos gerais, esse tipo de predição avalia no período de seguimento os pacientes que sobreviveram até um determinado tempo e registraram novas medidas longitudinais de seus marcadores dinâmicos. Além de coletar novos registros, há interesse em obter uma nova estimativa para a probabilidade de um indivíduo vivenciar o evento até um novo tempo pré-determinado.

A predição dinâmica é considerada para pacientes que sobreviveram até t anos e que possuem novas medições dos biomarcadores e deseja-se estimar a probabilidade de sobreviverem a um tempo $u > t$. Para a construção do modelo probabilístico podem existir diversos biomarcadores, contudo, os modelos construídos neste trabalho serão baseados em apenas um marcador longitudinal. A inclusão de múltiplos marcadores, embora relevante, será reservada para investigações futuras, dada a sua natureza metodológica mais desafiadora.

Segundo Gould et al., (2015), a predição dinâmica por meio da modelagem conjunta pode beneficiar as análises dos resultados longitudinais e de sobrevivência. O uso de modelos mistos longitudinais para incorporar os efeitos das variáveis no tempo na avaliação dos desfechos de sobrevivência representa com mais precisão a influência quantitativa desses fatores nas estimativas de sobrevida, em comparação com a inclusão direta do fator no modelo de regressão de sobrevivência.

Dois estudos de casos reais são utilizados nas modelagens do presente estudo. O primeiro deles trata de uma coorte prospectiva de uma cooperativa médica, que busca estimar o risco dos recém nascidos virem a utilizar a internação em CTI/UTI neonatal no primeiro ano de vida. As probabilidades serão mensuradas em três pontos estratégicos: os primeiros 28 dias, tempo em que uma criança é conside-

rada recém nascida, em seguida para as crianças que permaneceram no estudo e que não foram internadas até o vigésimo oitavo dia, será calculada a probabilidade de internação no sexto mês (182 dias). Por fim, para aqueles que não foram internados em CTI/UTI neonatal dentro dos primeiros seis meses, será calculada a probabilidade das crianças utilizarem o CTI/UTI ao fim do seu primeiro ano de vida.

De acordo com Surita et al. (2020), os comportamentos de saúde da mãe durante a gravidez exercem uma influência significativa sobre a saúde do filho nos primeiros anos de vida. Partindo dessa premissa, o primeiro estudo preditivo será desenvolvido com o objetivo de estimar a probabilidade do tempo até que a criança necessite de cuidados em CTI/UTI neonatal, partindo das informações de saúde e as características pessoais das mães, buscando identificar fatores de risco e padrões que possam antecipar a necessidade de intervenção neonatal e, assim, melhorar a capacidade de planejamento e resposta do sistema de saúde. Para o primeiro estudo, o percentil do peso da criança é o biomarcador de interesse que muda com o tempo.

O segundo estudo se trata da coorte prospectiva do Projeto SaMi-Trop (Centro de Pesquisas em Medicina Tropical São Paulo-Minas Gerais). O estudo conta com uma coorte multicêntrico de portadores de cardiopatia chagásica residente em vinte e uma cidades mineiras, acompanhadas a partir da linha de base no período de 2013 a 2014. Após a coleta dos dados na linha de base, foram registradas novas entrevistas com os pacientes que sobreviveram aos primeiros dois anos, e que não saíram da pesquisa nos anos de 2015 a 2016. Os participantes da pesquisa responderam a um questionário e foram examinados, além de coletarem amostras de sangue e exame de ECG, conforme aponta Cardoso et al., (2016).

Dentre os modelos que podem ser utilizados para realizar previsões na linha de base, o modelo de riscos proporcionais de Cox é amplamente aplicado na área da saúde. Neste modelo, o autor propõe a estimação dos coeficientes de regressão por meio da verossimilhança parcial (Cox, 1975).

Para a modelagem dinâmica, este trabalho explorará inicialmente uma abordagem 'ingênua' de previsão, que consiste em incorporar novas informações em diferentes pontos temporais, utilizando os modelos ajustados na linha de base. Por ser uma maneira simples de realizar previsões dinâmicas, ela pode introduzir viés nos resultados devido ao 'envelhecimento' da coorte. Como alternativa, será considerada a abordagem de 'Landmarking', introduzida por Anderson et al. (1983), na qual os marcadores são atualizados ao longo do tempo, permitindo uma análise mais precisa das variações temporais. Por fim, este trabalho adotará também a modelagem conjunta de marcadores longitudinais e do processo de sobrevivência como método

para predição dinâmica de risco, buscando capturar a evolução do risco de forma contínua e integrada.

1.2 Objetivos

Este trabalho tem como objetivo comparar diferentes métodos de predição em análise de sobrevivência para dados dinâmicos, utilizando medidas de discriminação e calibração. Ao considerar modelos que incorporam informações longitudinais ao longo do tempo, busca-se identificar aquele com melhor desempenho para contextos em que as características dos indivíduos mudam no tempo, como em coortes prospectivas.

Além disso, o estudo propõe a construção de dois modelos preditivos: um para estimar a probabilidade de recém-nascidos necessitarem de internação em CTI/UTI neonatal no primeiro ano de vida, e outro para prever o risco de óbito em indivíduos com cardiopatia chagásica. Esses modelos visam identificar precocemente fatores de risco e padrões associados a desfechos adversos, contribuindo para ações clínicas preventivas mais eficazes.

1.3 Organização do Trabalho

No Capítulo 2 deste trabalho são introduzidos os conceitos básicos de análise de sobrevivência, o modelo de regressão de Cox para predição estática, e, por fim, as métricas utilizadas para avaliar a discriminação e calibração dos modelos de predição. No Capítulo 3, são apresentadas as quatro abordagens de predição dinâmica que fundamentam a análise deste estudo. No Capítulo 4 são descritas as bases de dados dos dois estudos de caso analisados ao longo do trabalho: a coorte Sa-MiTrop e a coorte das mães de recém nascidos. O Capítulo 5 abordará as principais técnicas computacionais utilizadas na predição do risco. No Capítulo 6, serão discutidos os resultados obtidos para ambos os estudos de caso. Finalmente, o Capítulo 7 encerra as discussões com um resumo indicando as principais conclusões obtidas até o momento, além de elencar próximos passos a serem trilhados na pesquisa em predição dinâmica.

2 Análise de Sobrevivência para Predição de Riscos

O termo análise de sobrevivência surgiu a partir de pesquisas biomédicas que estudavam as causas de morte dos indivíduos, levando em conta o tempo transcorrido entre o diagnóstico (linha de base) de uma dada doença e a morte do indivíduo. Desta forma, a “análise de sobrevivência” lida com a modelagem estatística e a análise do “tempo de sobrevivência”.

O tempo de sobrevivência, reflete o período entre o ponto de partida do estudo e a ocorrência do evento pré-especificado, não existindo restrições para a definição do evento de interesse, já que as técnicas de sobrevivência podem ser observadas em diversas áreas. Em aplicações industriais o evento de interesse geralmente é a falha de um componente, já na área da saúde esse evento de interesse pode ser a morte, o aparecimento ou a cura de uma doença, conforme destaca Houwelingen e Putter (2012).

Na modelagem dos dados em análise de sobrevivência, pode ocorrer que alguns eventos não sejam observados em alguma instância, sendo essa a principal característica dos dados de sobrevivência. A não ocorrência do evento pode estar associado a janela temporal limitada ou ao aparecimento de outros eventos que não estejam associados ao estudo e que impeçam a ocorrência do desfecho, exemplificando: se o estudo se dedica a estudar a morte por alguma doença crônica e alguns dos pacientes vierem a óbito por outras causas, esses nunca experimentarão o evento de interesse da modelagem. Esse fenômeno é conhecido na literatura como censura.

O tempo de sobrevivência real corresponde ao momento exato em que o evento de interesse ocorre, enquanto o tempo de sobrevivência observado é aquele registrado durante o acompanhamento do indivíduo, podendo ser incompleto. Segundo Colosimo e Giolo (2024), as censuras podem ser classificadas como: (i) censura à direita, quando o tempo observado é menor que o tempo real; (ii) censura à esquerda, quando o tempo observado é maior que o tempo real; e (iii) censura intervalar, quando se sabe apenas que o evento ocorreu dentro de um intervalo de tempo.

Resumindo, pode-se dizer que a censura à direita é quando o evento de interesse não é observado até o término do estudo, censura à esquerda quando o evento de interesse ocorreu antes do sujeito ser observado no estudo e censura intervalar, quando se sabe apenas que o evento de interesse ocorreu dentro de algum intervalo. Observe que o tempo real de ocorrência do evento é desconhecido em todos os três casos. Dos três tipos de censura apresentados, o mais encontrado na literatura é a censura à direita, e que será tratada neste trabalho.

De modo geral os métodos de análise de sobrevivência são amplamente utilizados na área da saúde, e buscam estimar a distribuição do tempo entre o início do estudo e a ocorrência do evento de interesse, associado as covariáveis de interesse. Estas covariáveis podem ser estáticas ou dinâmicas (tempo dependente). Tanto nos modelos estáticos como nos dinâmicos, as covariáveis possuem os mesmos objetivos preditivos, porém se diferem no momento e na quantidade de vezes em que são coletadas.

2.1 Modelo de Regressão de Cox

Na construção da modelagem estática e dinâmica, o modelo de riscos proporcionais de Cox (1972) é amplamente conhecido e utilizado. O modelo tem o objetivo de avaliar a relação entre as covariáveis com o tempo de falha. O modelo proposto é amplamente utilizado na área da saúde devido a sua flexibilidade e robustez. Esse modelo será utilizado neste trabalho para prever a probabilidade do tempo de sobrevivência ser inferior a um tempo específico t , ou seja, $P[T < t]$, em que T é a variável resposta.

O modelo de Cox é uma abordagem semi-paramétrica, que leva em consideração em sua construção a suposição de proporcionalidade dos riscos e utiliza a verossimilhança parcial para a estimação dos parâmetros (Colosimo e Giolo, 2024).

A função de taxa de falha, denotada por $\lambda(t)$, é central para a modelagem em análise de sobrevivência e confiabilidade. Ela descreve a taxa instantânea ($\lambda(t) > 0$) com que os eventos ocorrem em um intervalo de tempo $[t, t + \Delta(t)]$, condicionado ao fato de que o evento ainda não ocorreu antes do tempo t . A expressão que define a função da taxa de falha pode ser escrita como:

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

A expressão do modelo de Cox é definida em termos da função de taxa de falha, em que é composto pelo produto de duas componentes: uma paramétrica e outra não paramétrica. Sua especificação é definida da seguinte forma:

$$\lambda(t|x, y(t)) = \lambda_0(t)e^{x^T \beta + y(t)\gamma}$$

em que, $\lambda_0(t)$ é chamada de taxa de falha basal, sendo esse o componente não paramétrico, β é um vetor $p \times 1$ de coeficientes da regressão associado às covariáveis fixas $x = (x_1, x_2, \dots, x_p)^T$, e γ é o coeficiente associado ao biomarcador dependente do tempo $y(t)$. Foi considerada somente uma biomarcador neste trabalho, porém, a

generalização é trivial para a análise com mais de um biomarcador.

Vale destacar que o vetor $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$ não considera o intercepto β_0 , uma vez que ele foi absorvido pela taxa de falha basal, ou seja, quando centralizamos as covariáveis substituindo x por $x - \bar{x}$, mudamos apenas a linha de base, mas não os coeficientes da regressão, sendo essa característica válida devido ao modelo de Cox pertencer a família de escala (Houwelingen e Putter 2012).

Quando o estudo leva em consideração apenas variáveis fixas, tem-se que as taxas de falhas são proporcionais, isso, por que ao tomar a razão da taxa de falha entre dois indivíduos distintos, as funções na linha de base se anulam. A formulação do modelo de Cox, com taxas de falhas proporcionais, segue:

$$\lambda(t | x) = \lambda_0(t)e^{x^T\beta}$$

Cox, em seu artigo (Cox, 1975), propôs a estimação dos coeficientes por meio da verossimilhança parcial, uma vez que, devido à presença do componente não paramétrico, a estimação por meio do método de máxima verossimilhança se torna inapropriada. A função de verossimilhança parcial é dada em termos do produto das verossimilhanças dos k eventos, considerando uma amostra de tamanho n , com $k \leq n$, em que foi observado o evento de interesse para diferentes tempos. A formulação da verossimilhança parcial é dada por:

$$L(\theta) = \prod_{i=1}^k \left(\frac{\exp(x_i^T(t_i)\theta)}{\sum_{j \in R(t_i)} \exp(x_j^T(t_i)\theta)} \right) = \prod_{i=1}^n \left(\frac{\exp(x_i^T(t_i)\theta)}{\sum_{j \in R(t_i)} \exp(x_j^T(t_i)\theta)} \right)^{\delta_i}$$

em que δ_i é o indicador de falha, $\theta = (\beta, \gamma)$ é o vetor de parâmetros, $R(t_i)$ é o conjunto de índices dos indivíduos que não passaram pelo desfecho estudado ou não foram censurados até o tempo t_i , e $x_i(t) = (x_i, y_i(t))^T$ é o vetor de covariáveis para o i -ésimo indivíduo no tempo t . Os valores de θ que maximizam $L(\theta)$ são os estimadores de máxima verossimilhança parcial.

A função de verossimilhança parcial reflete a probabilidade condicional de se escolher um indivíduo que irá experimentar o evento de interesse no tempo t , dado um conjunto de indivíduos sob risco no mesmo momento.

Andersen e Gill (1982) apresentam as provas gerais das propriedades assintóticas dos estimadores, sendo eles consistentes e assintoticamente normais. A partir da máxima verossimilhança parcial, é possível construir intervalos de confiança e realizar testes de hipóteses (Andersen e Gill, 1982).

Um dos componentes importantes no estudo da regressão de Cox e fundamental na construção da função de sobrevivência é a função de risco acumulado na linha de

base, que, por sua vez, é estimada de forma não paramétrica a partir do estimador de Nelson-Aalen-Breslow (1972). A função de risco acumulado basal do modelo de Cox pode ser estimada pela seguinte função:

$$\hat{\Lambda}_0(t) = \sum_{j:t_j < t} \frac{\delta_j}{\sum_{i \in R(t_j)} \exp\left(x_i^T(t_j) \hat{\theta}\right)},$$

em que $x_i^T(t_j)$ é o vetor de covariáveis, sejam elas estáticas ou dependentes do tempo avaliadas em t_j .

2.2 Predição Estática

Após a estimação dos coeficientes do modelo, é de interesse obter a função de sobrevivência para cada indivíduo pertencente à amostra. Considerando um vetor de covariáveis na linha de base para certo indivíduo ($x_0 = \{x, y(t = 0)\}$), será possível estimar a função de sobrevivência $S(t | x_0)$ por:

$$\hat{S}(t | x_0) = \left(\hat{S}_0(t)\right)^{\exp(x_0^T \hat{\theta})},$$

em que $\hat{\theta}$ é o estimador de máxima verossimilhança parcial, e $\hat{S}_0(t)$ é o estimador da função de sobrevivência na linha de base, dada por:

$$\hat{S}_0(t) = \exp\left(-\hat{\Lambda}_0(t)\right).$$

Dentro da modelagem de Cox, o interesse é encontrar o risco de ocorrência de um evento em um tempo t . Assim, a partir do modelo de Cox, o risco estimado para um novo indivíduo com vetor de covariáveis x_0 , é dado por:

$$\hat{r}(t|x_0) = \hat{P}(T < t | x_0) = 1 - \left(\hat{S}_0(t)\right)^{\exp(x_0^T \hat{\theta})}.$$

2.2.1 Suposição de Taxas de Falha Proporcionais

Após a construção do modelo, antes de realizar a validação quanto à sua capacidade de predição (discriminação e calibração), é importante verificar sua adequação quanto à suposição de taxa de falhas proporcionais. O método utilizado neste trabalho foi proposto por Grambsch e Therneau (1994), onde é possível visualizar os resíduos de Schoenfeld versus o tempo para cada covariável, a fim de verificar a adequação do modelo, uma vez que o modelo de Cox por mais flexível que seja, devido à presença do componente não paramétrico, ele não se ajusta a qualquer situação

clínica (Colosimo e Giolo, 2024).

2.3 Validação do Modelo de Predição

Conforme discutido em seu livro, Houwelingen e Putter (2012), após o desenvolvimento de um modelo de predição, diferentes questões podem surgir nas avaliações do poder preditivo do modelo. As avaliações consistem no poder discriminatório do modelo, ou seja, em conseguir distinguir casos e não casos do desfecho estudado, e calibração, modelos que realizam predições com alta acurácia, onde o risco predito é próximo ao observado.

Para a investigação da capacidade preditiva dos modelos, é comumente utilizado pela literatura a partição da amostra. Esse procedimento tem como objetivo resguardar as análises de estimativas sub ou superestimadas ('*Underfitting*' e '*Overfitting*'). Harrell (2015) em seu trabalho destaca três principais métodos de partição, que podem ser utilizados: treino e teste, método que divide a base de dados em diferentes proporções, onde o treino é a parte utilizada para o ajuste do modelo e o teste a proporção utilizada para realizar as validações; validação cruzada, método que divide a base de dados em k partes de iguais proporções e utiliza cada uma dessas proporções na construção do modelo e validação em diferentes combinações. E por fim, o bootstrap, método em que a amostra passa por um processo de reamostragem com reposição N vezes, e o modelo é validado na base original a partir das medidas de desempenho usando as N reamostras. Neste trabalho será utilizado a validação cruzada com 5 folds e o bootstrap utilizando 200 reamostras.

2.3.1 Discriminação

A avaliação da discriminação do modelo será obtida por meio da Área sob a Curva ROC (AUC), sendo essa uma das medidas mais utilizadas quando se deseja avaliar o poder de discriminação dos modelos estatísticos estimados (Woodward et al., 2017). A AUC é calculada a partir da curva ROC, que cruza os pares ordenados de sensibilidade e um menos a especificidade nos diferentes pontos de corte.

Segundo Houwelingen e Putter (2012), uma possível construção para a AUC é dada em termos de uma subamostra de pacientes que não experimentaram o desfecho estudado (n_0) e aqueles que experimentaram (n_1), de acordo com seus

valores preditos ($\hat{P}_{0i}(T < t | x_i), \hat{P}_{1j}(T < t | x_j)$), podendo ser escrita como:

$$\begin{aligned} \widehat{\text{AUC}}(t) &= \frac{1}{n_0 n_1} \sum_{j=1}^{n_1} \sum_{i=1}^{n_0} \left[I \left(\hat{P}_{1j}(T < t | x_i) > \hat{P}_{0i}(T < t | x_j) \right) \right. \\ &\quad \left. + \frac{1}{2} I \left(\hat{P}_{1j}(T < t | x_i) = \hat{P}_{0i}(T < t | x_j) \right) \right] \\ &= \frac{1}{n_0 n_1} \sum_{j=1}^{n_1} \sum_{i=1}^{n_0} \left[I \left(\hat{r}_{1j}(t|x_j) > \hat{r}_{0i}(t|x_i) \right) \right. \\ &\quad \left. + \frac{1}{2} I \left(\hat{r}_{1j}(t|x_j) = \hat{r}_{0i}(t|x_i) \right) \right]. \end{aligned}$$

A interpretação da AUC baseia-se na ideia de uma média ponderada. Soma-se 1 sempre que o risco atribuído a um indivíduo que vivenciou o desfecho é maior que o risco de um indivíduo que não o vivenciou. Isso ocorre porque, ao ter o evento materializado, espera-se que esse indivíduo tenha, de fato, um risco maior. Em casos de empate entre os riscos atribuídos aos dois grupos, soma-se 0,5.

Para a avaliação dos resultados do AUC, temos que um modelo com 100% de discriminação, o resultado será igual a 1, caso que beira a impossibilidade, visto a complexidade que envolve um modelo de predição. Já um modelo que não possui nenhuma discriminação, o AUC será menor ou igual a 0,5.

2.3.2 Calibração

Já a calibração, medida que avalia a acurácia dos modelos de predição, é comumente estimada por meio do teste de Hosmer e Lemeshow (1980). O teste de hipótese, que tem como hipótese nula a proposição de que o modelo está calibrado, foi originalmente formulado para testar modelos de regressão logística, mas que teve adaptações para dados de sobrevivência. Um modelo é considerado calibrado quando as probabilidades previstas se aproximam das frequências observadas de ocorrência do evento em grupos com risco semelhante. A estatística de teste é dada por:

$$\text{HL} = \sum_{j=1}^g \frac{(d_j - n_j \bar{r}_j)^2}{n_j \bar{r}_j},$$

em que g é o número de grupos que as probabilidades preditas e observadas pelo modelo serão divididas (o usado neste trabalho será de $g = 10$), d_j é o número de eventos observados no grupo g , n_j é o número de indivíduos no grupo g e \bar{r}_j é a

média dos riscos esperados no grupo g , que pode ser denotado por:

$$\overline{\hat{r}}_j(t) = \sum_{k=1}^{n_k} \frac{\hat{r}_{jk}(t|x_{jk})}{n_j}, i = 1, \dots, g; k = 1, \dots, n_j,$$

em que $\hat{r}_{jk}(t)$ é o risco estimado para o k -ésimo indivíduo do grupo g .

A estatística de Hosmer e Lemeshow, sob H_0 , tem distribuição qui-quadrado com $S - 1$ graus de liberdade.

Uma extensão da estatística do teste de Hosmer e Lemeshow proposto por D'Agostino e Nam (2003), que incorpora estimativas do estimador de Kaplan-Meier, é dado por:

$$\text{HL}_{\text{ND}}(t) = \sum_{j=1}^g n_j \frac{\left[(1 - \hat{K}M_j - \overline{\hat{r}}_j(t)) \right]^2}{\overline{\hat{r}}_j(t)(1 - \overline{\hat{r}}_j(t))}$$

em que $\hat{K}M_j(t)$ é o estimador de Kaplan-Meier, avaliado no período de tempo especificado, e o risco predito, $\hat{r}_j(t)$. A estatística HL_{ND} , sob H_0 , tem distribuição qui-quadrado com $g - 1$ graus de liberdade.

Uma abordagem alternativa para o teste de Hosmer-Lemeshow, proposta por Demler (2015), incorpora a variância do estimador de Kaplan-Meier no denominador do teste. Os autores argumentam que a utilização da variância do estimador de Kaplan-Meier favorece a análise da calibração do modelo, tornando-o ainda mais robusto e preciso. O teste é, então, expresso por:

$$\text{HL}_{\text{ND-VAR}}(t) = \sum_{j=1}^g \frac{\left[(1 - \hat{K}M_j(t)) - \overline{\hat{r}}_j(t) \right]^2}{\hat{V}ar(\hat{K}M_j(t))}$$

em que $\hat{V}ar(\hat{K}M_j(t))$ é o estimador de Greenwood (Colossimo e Giolo, 2024)

Outra forma de avaliar a calibração dos modelos, é o gráfico associado ao teste de Hosmer e Lemeshow. Os riscos são divididos em g grupos, como nos testes apresentados anteriormente, e para cada um desses grupos é calculada a média dos riscos preditos e a frequência dos eventos observados (Woodward et al., 2017).

Além dos testes apresentados, é possível comparar a acurácia dos diferentes modelos de predição através do erro de predição (Rizopoulos et al. 2017). A medida utilizada leva em conta a soma das funções indicadoras e uma função de perda quadrática, tendo a seguinte formulação:

$$\begin{aligned} \widehat{PE}(t) = \frac{1}{n} \sum_{i=1}^n & \left[I(T_i \geq t) L[0 - \hat{P}_i(T < t | x_i)] + \delta_i I(T_i < t) L[1 - \hat{P}_i(T < t | x_i)] \right] \\ & + (1 - \delta_i) I(T_i < t) \left[(1 - \hat{P}_i(T < t | x_i)) L[0 - \hat{P}_i(T < t | x_i)] \right. \\ & \left. + \hat{P}_i(T < t | x_i) L[1 - \hat{P}_i(T < t | x_i)] \right] \end{aligned}$$

em que, $L(\cdot)$ denota a função de perda quadrática. O estimador foi proposto por Henderson, Diggle e Dobson (2002), que incorpora na formulação as censuras presentes na amostra.

3 Predição Dinâmica

A predição dinâmica atualiza as informações dos biomarcadores, para os pacientes que sobreviveram após o primeiro período de pesquisa. Em linhas gerais, no período de acompanhamento, os pacientes que sobreviveram até um determinado período t registraram novas medidas de seus marcadores longitudinais, podendo ser esses marcadores intermediários a realização de novos exames, a ocorrência de outras doenças que não seja a estudada, dentre outros. Desta forma, após t período de tempo, os biomarcadores são atualizados e a predição busca obter uma nova estimativa para a probabilidade de um indivíduo sobreviver a um novo momento u , tal que $u > t$, conforme exemplifica a Figura 1.

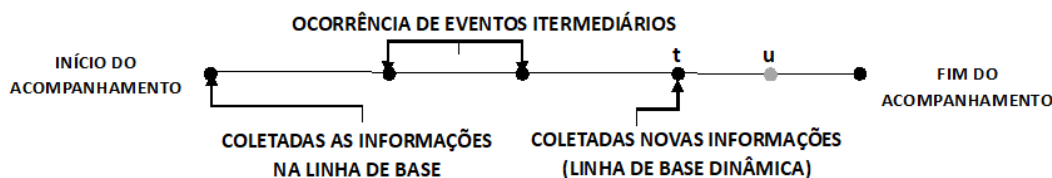


Figura 1: Ilustração da Linha do Tempo do acompanhamento clínico dinâmico.

Para o estudo de Utilização do CTI/UTI Neonatal e para a coorte SaMi-Trop, os modelos construídos serão baseados em apenas um marcador dependente do tempo. Diversas abordagens podem ser utilizadas para a construção de predições dinâmicas, dentre eles: abordagem ingênua, que utiliza o modelo de Cox da linha de base apenas inserindo as novas informações colhidas no tempo t ; Landmark, que ajusta o modelo de Cox para uma nova linha de base no tempo t , ignorando os eventos que ocorreram anteriormente; e a modelagem conjunta que associa as variáveis dependentes do tempo com o risco de ocorrência do desfecho.

3.1 Abordagem Ingênua

Os pacientes que sobreviveram ao período t , após a primeira medição, tiveram seus marcadores longitudinais atualizados. Após as novas coletas dos dados, deseja-se estimar a probabilidade de os sobreviventes ao tempo t experimentarem o evento de interesse até um novo tempo u , em que $u > t$.

Uma forma de se atualizar o risco, é utilizar o modelo construído na linha de base e adicionar ao estimador, as novas observações. Desta forma, o vetor de covariáveis contém tanto as informações de linha de base, quanto os novos valores medidos para os marcadores longitudinal (tempo dependente).

O método é considerado ingênuo, por não atualizar as estimativas dos parâmetros dos modelos, provocando certo viés nas análises, uma vez que a utilização de diferentes dados gera novos estimadores. A partir do apresentado, serão utilizados duas abordagens: Naive 1 e Naive 2.

Na predição Naive 1, utilizam-se os estimadores obtidos a partir do modelo ajustado na linha de base, realizando-se a predição do risco considerando as novas informações do marcador longitudinal, nota-se, portanto, que essa abordagem equivale a criar uma nova linha de base no tempo t , utilizando os mesmos estimadores do tempo zero. O método pode ser escrito como:

$$\begin{aligned}\widehat{r}_{N1}(u | t, x_t) &= \widehat{P}(T < u | x_t, T > t) \\ &= 1 - \widehat{P}(T - t > u - t | T > t, x_t) \\ &= 1 - \left(\widehat{S}_0(u - t)^{\exp(x^T \widehat{\beta} + y(t) \widehat{\gamma})} \right),\end{aligned}$$

em que $x_t = \{x, y(t)\}$ são as novas informações obtidas no tempo t . O vetor contém as estimativas da linha de base, além das novas medidas para os marcadores longitudinais. É possível observar que o $r_{N1}(u | t, x_t)$ assume uma nova linha de base em t utilizando das estimativas obtidas no modelo estático.

A predição Naive 2 corresponde à estimativa da probabilidade condicional de ocorrência do evento até u , dado que o indivíduo sobreviveu até t ($u > t$), com base nos estimadores do modelo da linha de base. Sua formulação é dada por:

$$\begin{aligned}\widehat{r}_{N2}(u | t, x_0) &= \widehat{P}(T < u | x_0, T > t) = 1 - \widehat{P}(T > u | T > t, x_0) \\ &= 1 - \frac{\widehat{P}(T > u, T > t | x_0)}{\widehat{P}(T > t | x_0)} \\ &= 1 - \frac{\widehat{P}(T > u | x_0)}{\widehat{P}(T > t | x_0)} \\ &= 1 - \frac{\widehat{S}_0(u) e^{x^T \widehat{\beta} + y(0) \widehat{\gamma}}}{\widehat{S}_0(t) e^{x^T \widehat{\beta} + y(0) \widehat{\gamma}}}.\end{aligned}\tag{1}$$

3.2 Landmark

Para a construção do Landmark, o ponto de partida será o modelo de Cox que permite covariáveis dependentes do tempo (Houwelingen e Putter, 2012). As ideias iniciais desse modelo foram apresentadas por Anderson et al. (1983), que estudava a sobrevivência de pacientes diagnosticados com câncer, utilizando variáveis tempo-dependentes.

O princípio da modelagem parte em fixar um tempo t , e realizar as predições

a partir das informações extraídas neste ponto. Desta forma, a linha de base não estaria mais no tempo zero, mas no ponto t . Os indivíduos que sobreviveram até t formam uma nova coorte, em que seus dados serão utilizados para a estimação dos parâmetros do modelo de Cox. A formulação do risco, para os indivíduos que sobreviveram até t , e que se deseja estimar o risco até o tempo u , para um tempo $u > t$, baseado em suas novas informações, será:

$$\widehat{r}_{LM}(u | t, x_t) = \widehat{P}(T < u - t | T > t, x_t) = 1 - \left(\widehat{S}_0^*(u - t) \right)^{\exp(x_t^T \widehat{\beta}^* + y(t) \widehat{\gamma}^*)},$$

sendo que $\widehat{\beta}^*$ e $\widehat{\gamma}^*$ são os estimadores de máxima verossimilhança parcial dos coeficientes do modelo de Cox, e $\widehat{S}_0^*(u - t)$ é a sobrevida basal estimada para a nova linha de base no tempo t .

3.3 Modelagem Conjunta

Na modelagem conjunta deseja-se avaliar a associação entre as covariáveis dependentes do tempo, e o risco da ocorrência do desfecho estudado. Para investigar o efeito de associação e estimação do risco, são construídos dois submodelos, um longitudinal e outro de sobrevivência, ligados por um efeito aleatório comum a ambos os modelos.

3.3.1 Modelo Longitudinal

Conforme apresenta Rizopoulos (2012), o modelo longitudinal se concentra em respostas longitudinais contínuas, em que o processo do marcador dependente do tempo segue a abordagem dos modelos lineares de efeitos mistos. Uma variável dependente do tempo é medida com erro, sendo sua expressão dada por:

$$y_i(t) = y_i^*(t) + \epsilon_i(t) = w_i^T \alpha + z_i^T(t) b_i + \epsilon_i(t), \quad \epsilon_i(t) \sim N(0, \sigma^2), \quad i = 1, \dots, n,$$

em que, $y_i^*(t)$ representa o verdadeiro valor não observado da resposta longitudinal no tempo t , w_i' e δ_i' são os vetores de covariáveis fixas e dependentes do tempo, e α e b_i os respectivos efeitos fixos e aleatórios. Assume-se que b_i é independente do erro e tem distribuição $N(0, \Sigma)$.

3.3.2 Modelo de Sobrevivência

O submodelo de sobrevivência quantifica a associação entre a resposta longitudinal $y_i^*(t)$ com o risco da ocorrência do evento em termos da função de taxa de falha.

Podendo ser escrito como:

$$\lambda(t | y_i^*(t), x_i) = \lambda_0(t) \exp(\gamma y_i^*(t) + x_i^T \beta), t > 0,$$

em que, $\lambda_0(t)$ é a função de risco basal.

Conforme aponta Rizopoulos (2012), nos modelos de predição de linha de base é habitual que $\lambda_0(\cdot)$ esteja no modelo completamente sem especificações, a fim de evitar o impacto de escolha incorreta para a distribuição dos tempos de sobrevivência. Contudo, na modelagem dinâmica, verifica-se que seguir um modelo sem especificações, pode levar a uma subestimação dos erros padrões das estimativas dos parâmetros (Hsieh et al., 2006).

3.3.3 Função de Taxa de Falha

A função de taxa de falha basal $\lambda_0(\cdot)$, pode ser tratada de diferentes formas. Uma opção padrão é a utilização de uma função de risco correspondente a uma distribuição paramétrica conhecida. No contexto da análise de sobrevivência, as distribuições normalmente utilizadas incluem a Weibull, a log-normal e a Gamma (Rizopoulos 2012). A função de risco correspondente à distribuição Weibull pode ser definida como:

$$\lambda_0(t) = \sigma t^{\sigma-1}.$$

A distribuição de Weibull foi escolhida para este trabalho por ser a opção predefinida no pacote do software utilizado para as modelagens. Além de sua disponibilidade direta na ferramenta, a distribuição de Weibull é amplamente empregada em análises de sobrevivência devido à sua flexibilidade para representar diferentes formas de risco ao longo do tempo, o que a torna adequada para diversos contextos clínicos e epidemiológicos.

Como alternativa as distribuições normalmente utilizadas, a literatura em análise de sobrevivência vem sugerindo a utilização de uma especificação mais flexível da função de risco de base, que são os modelos baseados em nós. Esses modelos oferecem uma abordagem versátil para modelar a função de risco de base em análise de sobrevivência. Ao contrário de métodos paramétricos tradicionais, que impõem formas específicas à função de risco, esses modelos permitem que sua forma seja determinada diretamente pelos dados, conferindo maior capacidade de adaptação à padrões complexos. Essa flexibilidade é alcançada por meio da segmentação do domínio temporal em intervalos controlados por nós, onde em cada "pedaço" da função será ajustado um polinômio com parametrizações distintas, forçando então o encontro das funções adjacentes nos pontos de corte (Wegman et al., 1983).

Dentre as abordagens propostas na literatura para modelar a função de risco de linha de base, estão: Whittemore e Keller (1986) que usaram funções escalonadas e splines lineares para obter uma estimativa não paramétrica da função de risco, Rosenberg (1995) utilizou uma aproximação de splines B, e Herndon e Harrell (1995) que usaram splines cúbicas restritas. Duas opções levantados por Rizopoulos (2012) que em geral funcionam de forma satisfatória são as abordagens: exponencial piecewise e splines. No modelo exponencial piecewise, a função de risco da linha de base assume a seguinte forma:

$$\lambda_0(t) = \sum_{q=1}^Q \xi_q I(v_{q-1} < t < v_q),$$

em que, $0 = v_0 < v_1 < \dots < v_Q$ denota uma divisão da escala de tempo, com v_q sendo maior que o tempo máximo observado, e ξ_q denota o coeficiente da função de risco no intervalo (v_{q-1}, v_q) . À medida que o número de nós aumenta, a especificação do risco na linha de base torna-se mais flexível.

Para a modelagem via splines, a função de risco de linha de base $\log \lambda_0(t)$ é expandida em funções da base B-spline para splines cúbicos, como segue:

$$\log \lambda_0(t) = k_0 + \sum_{d=1}^{m^*} k_d B_d(t, q)$$

em que $k^T = (k_0, k_1, \dots, k_m)$ são os coeficientes spline, q denota o grau das funções de base B-splines $B(\cdot)$, e $m = m^* + q - 1$ é o número de coeficientes do modelo, com m^* denotando o número de nós interiores.

A quantidade de nós em cada uma das funções pode variar e está intrinsecamente associado a características da base de dados que compõe a pesquisa, como: colinearidade entre as covariáveis, escala das covariáveis, dentre outras razões. E também está relacionado a quantidade de dados disponíveis para o ajuste da função. De acordo com Harrell (2001), para amostras menores que 100, o uso de quatro nós internos em geral produz um ajuste adequado, equilibrando flexibilidade e precisão. Para amostras maiores, o número de nós pode variar entre quatro e sete nós, sendo que, a partir desse limite a análise tende a se tornar excessivamente subjetiva.

Desta forma nos dois modelos apresentados, aumentar o número de nós provoca uma maior flexibilidade na aproximação de $\lambda_0(\cdot)$. No entanto, deve-se manter um equilíbrio entre viés e variância a fim de evitar a ocorrência de super dispersões (Rizopolus 2012).

3.3.4 Estimação

Especificada a função de risco, podemos escrever a função de sobrevivência como:

$$\bar{S}(t | y_i^*(u), x_i) = P(T_i > t | y_i^*(t), x_i) = \exp \left\{ - \int_0^t \lambda_0(s) \exp(\gamma y_i^*(s) + x_i^T \beta) ds \right\}.$$

A estimação dos parâmetros na modelagem conjunta utiliza os dois submodelos apresentados em 3.3.1 e 3.3.2, e consiste em maximizar a função de verossimilhança com respeito aos parâmetros a serem estimados $\theta^* \{ \alpha, \sigma^2, \Sigma, \gamma, \beta, \lambda_0 \}$. No caso do modelo Weibull, consideram-se todos os parâmetros, exceto λ_0 : $\theta_{\text{Weibull}}^* = \{ \alpha, \sigma^2, \Sigma, \gamma, \beta \}$, para os modelos Spline e Piecewise, consideram-se todos os parâmetros, exceto σ^2 : $\theta_{\text{Spline, Piecewise}}^* = \{ \alpha, \Sigma, \gamma, \beta, \lambda_0 \}$. Para definir a distribuição conjunta dos estimadores (t, δ, y, b) , assume-se que o vetor de efeitos aleatórios explica a associação entre os dois processos e a correlação entre as medidas longitudinais de cada indivíduo. Desta forma, as medidas longitudinais de cada indivíduo são independentes, tal que

$$p(t_i, \delta_i, y_i | \theta^*) = \int p(t_i, \delta_i | y_i, b_i; \theta^*) p(y_i | b_i; \theta^*) p(b_i) db_i.$$

Desta forma, a função de verossimilhança é obtida usando a independência condicional dos marcadores longitudinais para o mesmo indivíduo, dado o vetor de efeitos aleatórios b_i :

$$\begin{aligned} L(\theta^*) &= \prod_{i=1}^n p(t_i, \delta_i, y_i) \\ &= \prod_{i=1}^n \int p(t_i, \delta_i, y_i, b_i) db_i \\ &= \prod_{i=1}^n \int \left[\prod_{j=1}^{j_i} p(y_i(t_{ij}) | b_i) \right] p(t_i, \delta_i | y_i(t_i), b_i) p(b_i) db_i, \end{aligned}$$

em que, b_i é o número de medidas longitudinais $y(t_{i1}), i = 1, \dots, j_i$, para o i -ésimo indivíduo.

Os estimadores de máxima verossimilhança serão obtidos realizando a maximização da log-verossimilhança.

3.3.5 Predição

Após a estimação dos parâmetros, assim como nos demais modelos, será calculada a sobrevida esperada do indivíduo, dado que sobreviveu a um determinado tempo

t e possui informações históricas até esse tempo ($u(t) = \{u_i(u) : 0 < u < t\}$). A probabilidade de um indivíduo sobreviver até u , dado que ele sobreviveu até t , com $u > t$, pode ser expressa como:

$$S(u | t, y(t)) = P(T \geq u | T > t, y(t), \theta^*) = \int \frac{S(u | y^*(u, \theta^*); \theta^*)}{S(t | y^*(t, \theta^*); \theta^*)} p(b | T > t, y(t); \theta^*) db$$

em que, $S(\cdot)$ é a função de sobrevivência e $y^*(\cdot)$ é o vetor com as informações longitudinais aproximada pelo modelo de efeitos mistos. Conforme apresenta Rizopoulos (2011), a derivada da função de sobrevivência gera o estimador de primeira ordem para a sobrevida esperada:

$$\hat{S}(u | t, y(t)) = \frac{S(u | y^*(u, \hat{\theta}^*); \hat{\theta}^*; \hat{b})}{S(t | y^*(t, \hat{\theta}^*); \hat{\theta}^*; \hat{b})}, \quad (2)$$

em que, $\hat{\theta}^*$ é o estimador de máxima verossimilhança para a modelagem conjunta. Observe que a forma da predição (2) usando o JM é similar àquela do Naive 2 (1).

3.4 Avaliação dos Modelos

A avaliação dos modelos se dará pelas mesmas métricas apresentadas para o modelo de Cox de discriminação e calibração.

4 Bases de Dados

Nesta seção serão apresentadas as bases de dados utilizadas nos estudos de caso que sustentam tanto a abordagem estática quanto a abordagem dinâmica. Compreender as características dos dados antes da modelagem é fundamental para garantir a qualidade e a precisão dos resultados. Esta análise inicial permitirá uma definição clara das variáveis e um entendimento profundo das particularidades da amostra investigada.

4.1 Caracterização da Base de Dados: Utilização de CTI/UTI Neonatal

O primeiro estudo de caso foca na investigação da probabilidade de recém-nascidos, pertencentes a uma operadora de plano de saúde da região de Belo Horizonte, utilizarem os serviços de CTI/UTI Neonatal durante o primeiro ano de vida. A problemática central é identificar precocemente os fatores de risco associados à internação neonatal intensiva. Para viabilizar essa análise, a base de dados da Operadora compreende dados de 2018 a 2022, seguindo o esquema da Figura 2:

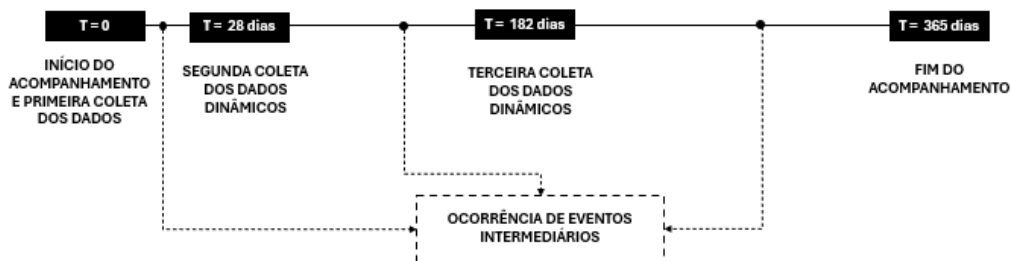


Figura 2: Acompanhamento clínico do estudo de caso de CTI/UTI neonatal.

A análise será conduzida em três momentos distintos, com base em extrações de dados que refletem diferentes fases do desenvolvimento da criança. Serão feitas previsões para os primeiros 28 dias de vida, em seguida será predito para o 6^o mês (182 dias) de vida das crianças que permaneceram no estudo após 28 dias e ainda não utilizaram o CTI/UTI e, por fim, para o primeiro ano de vida para aquelas que permaneceram no estudo até o sexto mês de vida e ainda não utilizaram o CTI/UTI. Essa abordagem permite acompanhar a evolução dos marcadores clínicos ao longo do tempo, ajustando o modelo preditivo conforme as variáveis se comportam em cada estágio crucial da saúde neonatal.

Os dados coletados são em grande parte referentes às mães, partindo da premissa de que a saúde da criança no primeiro ano de vida está intrinsecamente associada à saúde materna durante o período gestacional e pós-gestacional. Essa correlação entre a saúde materna e a utilização de serviços intensivos pelos recém-nascidos destaca a importância de intervenções preventivas direcionadas à saúde da mulher como forma de reduzir a necessidade de internações neonatais. A base de dados contém informações de 15.202 gestações, cujos partos ocorreram entre 2018 e 2022. Em 6,51% dos casos analisados, os recém-nascidos necessitaram de cuidados em CTI/UTI Neonatal ao longo do primeiro ano de vida.

As variáveis utilizadas foram divididas em quatro grandes grupos: 1- a jornada assistencial da mãe (consultas, exames, etc); 2 - ao estado de saúde da mãe (comorbidades, etc), 3 - às características pessoais e da sua gestação (idade, idade gestacional, se é primípara, etc), e 4 - características do primeiro ano de vida das crianças. Sendo as variáveis utilizadas no estudo descritas a seguir:

1. Jornada assistencial da mãe (1 Variável).

- Indicador de Realização do Pré-natal de Acordo com o Protocolo¹ da Cooperativa Médica: Variável binária (1- Segue Protocolo; 0- Não segue o Protocolo);

2. Estado de saúde da mãe (4 Variáveis).

- Indicador de Comorbidade Hipertensiva Aguda a Gestação: Variável binária (1- Possui Comorbidade; 0- Não possui comorbidade);
- Indicador de Comorbidade Metabólica Aguda a Gestação: Variável binária (1- Possui Comorbidade; 0- Não possui comorbidade);
- Indicador de Comorbidade Hipertensiva Crônica: Variável binária (1- Possui Comorbidade; 0- Não possui comorbidade);
- Indicador de Comorbidade Metabólica Crônica: Variável binária (1- Possui Comorbidade; 0- Não possui comorbidade);

3. Características pessoais e da gestação (5 Variáveis).

- Número de Semanas até o Parto (idade gestacional): variável binária segundo a quantidade de semanas até o parto, as mães são classificadas

¹O protocolo de pré-natal da cooperativa médica envolve um conjunto padronizado de exames, vacinas e consultas, visando o acompanhamento sistemático do cuidado das gestantes durante a gestação.

em: 1 - partos prematuro (37 semanas ou menos) ou 0 - a termo (entre 37 e 42 semanas);

- Idade no Momento do Parto: variável categórica de acordo com a idade das mães no momento do parto, os indivíduos são classificados em três faixas de idade de acordo com o grau de risco que a idade pode provocar, sendo: 1 - até 20 anos, 2 - de 21 a 35 anos e 3 - acima de 36 anos;
- Indicador Primípara²: Variável binária (1 - Primeira gravidez; 0 - Teve uma ou mais gestações anteriores);
- Tipo de Parto: Variável binária (1 - Parto cesárea; 0 - Parto vaginal);
- Tempo Entre a Internação e o Parto: variável binária da classificação do tempo entre a internação e o parto: 0 - tempo dentro do esperado (abaixo de 9 horas para mães primíparas e abaixo de 6 horas para mães múltiparas) e 1 - fora do tempo esperado (acima de 9 horas para mães primíparas e acima de 6 horas para mães múltiparas);

4. Características do primeiro ano de vida da criança (5 Variáveis).

- Percentil do Peso (variável tempo-dependente): variável numérica medida entre 0 a 100%, construída com base na referência da Organização Mundial da Saúde (tabela NCHS 77/8 para o gênero masculino);
- Indicador Apgar³ do Primeiro Minuto: Variável numérica medida numa escala de 0 a 10;
- Indicador Apgar do Quinto Minuto: Variável numérica medida numa escala de 0 a 10;
- Indicador de Comorbidade: Variável binária (1- Possui Comorbidade; 0- Não possui comorbidade);
- Utilizou CTI/UTI Neonatal: Variável resposta (1- Utilizou CTI/UTI Neonatal; 0- Não Utilizou CTI/UTI Neonatal).

A base de dados segue o princípio de que todas as observações são centradas na mãe, funcionando como a unidade principal de análise. As variáveis relacionadas aos filhos foram lidas e vinculadas aos respectivos códigos das mães, permitindo a criação de um sistema de relacionamento entre mãe e filho. Dessa forma, toda a

²Mulher que teve apenas uma gestação.

³O índice de Apgar é uma avaliação rápida da condição do recém-nascido, realizada no primeiro e no quinto minuto após o parto, com base em cinco critérios: frequência cardíaca, esforço respiratório, tônus muscular, irritabilidade reflexa e cor da pele.

estrutura da base foi construída a fim de garantir que as características maternas e neonatais estivessem adequadamente conectadas para a realização das análises preditivas.

4.1.1 Breve Descrição dos Dados

Na Tabela 1 são sumarizadas informações sobre as características pessoais e da gestação, além das informações sobre a jornada assistencial das grávidas.

Tabela 1: Distribuição de Frequências (Absoluta e Proporcional) das Variáveis relacionadas as mães.

Variável	f (%)
Primiparidade	
Não é primípara	6.667 (43,9%)
Primípara	8.535 (56,1%)
Tipo de Parto	
Cesárea	9.939 (65,4%)
Vaginal	5.263 (34,6%)
Pré-natal Completo	
Não completo	6.547 (43,1%)
Completo	8.655 (56,9%)

Através da Tabela 1 é possível constatar que mães primíparas representam 56,1% da amostra, essas podem enfrentar mais incertezas e complicações durante a gestação e o parto, o que pode aumentar o risco de internações neonatais. Já as mães múltiplas, mais experientes, podem ter um impacto inverso, reduzindo essa necessidade. Além disso, partos cesáreos, que constituem 65,4% dos casos, tendem a estar associados a complicações neonatais, como dificuldades respiratórias, aumentando o risco de internação em CTI/UTI.

Já sobre a jornada assistencial das mães, destaca-se que 56,9% das mulheres realizaram o acompanhamento pré-natal completo, sendo esse um fator essencial na redução de complicações neonatais. Esta variável deve ser analisada com cautela, uma vez que não ter realizado o pré-natal completo na Operadora de Planos de Saúde (OPS) em questão, não significa que a mãe não realizou todas as etapas do pré-natal, uma vez que exames e consultas podem ter sido feitas de forma apartada do plano e por isso não foram consideradas. Logo, o principal objetivo desta variável é avaliar o impacto na probabilidade da criança utilizar ou não o CTI/UTI dado que a mãe realizou todo o pré natal na OPS.

Quanto à distribuição do número de semanas da gestação apresentado na Tabela 2, observa-se uma alta concentração de partos ocorrendo entre 37 e 42 semanas. Essa faixa de semanas é considerada a termo, ou seja, o período ideal para o nascimento, o que sugere menor risco de complicações neonatais e, conseqüentemente, uma menor probabilidade de utilização de CTI/UTI Neonatal. No entanto, nascimentos prematuros (abaixo de 37 semanas) aparecem em menor frequência, mas podem ter um impacto substancial no risco de internação, visto que bebês prematuros frequentemente apresentam problemas de saúde que exigem cuidados intensivos logo após o nascimento.

Tabela 2: Distribuição de Frequências (Absoluta e Proporcional) do número de semanas da gestação.

Variável	f (%)
Semanas até gestação	
Prematuro	1.762 (11,6%)
Termo	13.340 (88,4%)

A análise da idade materna no momento do parto apresentada na Tabela 3 revela que a maioria das mulheres tem entre 21 e 35 anos. Idades entre 21 e 35 anos são, geralmente, consideradas de menor risco em termos de complicações gestacionais e neonatais, o que pode estar associado a uma menor necessidade de CTI/UTI Neonatal. No entanto, é possível observar um pequeno número de partos em faixas etárias mais jovens (até 20 anos) e mais avançadas (acima de 36 anos), que podem ser consideradas fatores de risco para internação neonatal. Mães muito jovens ou mais velhas estão mais suscetíveis a complicações durante a gestação, o que pode aumentar a probabilidade de os bebês necessitarem de cuidados intensivos no período neonatal. Reforçando a importância de considerar a idade materna como um fator preditivo para o evento de interesse.

Tabela 3: Distribuição de Frequências (Absoluta e Proporcional) da Idade no momento do parto

Variável	f (%)
Classificação da Idade	
Até 20 anos	306 (2,0%)
21 a 35 anos	10.127 (66,6%)
Acima de 36 anos	4.769 (31,4%)

A análise do tempo de internação até o parto apresentado na Tabela 4 revelou uma variação significativa entre os grupos avaliados. A variável foi classificada em

dois grupos: 'tempo dentro do esperado', predominando com 11.953 casos (78,6%), e 'tempo fora do esperado' com cerca de 21%. Dentro da classificação do tempo esperado entre a internação e o parto foi observado em alguns casos tempo igual a zero, evidenciando que o parto ocorreu imediatamente após a internação, possivelmente relacionado a situações emergenciais ou procedimentos previamente agendados.

Tabela 4: Distribuição de Frequências (Absoluta e Proporcional) do Tempo de internação até o parto.

Variável	f (%)
Classificação do Tempo	
Dentro do esperado	11.953 (78,6%)
Fora do esperado	3.249 (21,4%)

A análise da classificação das comorbidades maternas apresentada na Tabela 5 revela que 80,3% das mulheres na amostra não apresentam comorbidades, o que indica um grupo relativamente saudável. No entanto, das mulheres que possuem comorbidades um percentual significativo de 29,87% são agudas à gestação e 70,13% têm condições crônicas, o que pode ter implicações diretas na saúde dos recém-nascidos e na probabilidade de internação em CTI/UTI Neonatal. As comorbidades agudas, como hipertensões gestacionais, podem exigir intervenções imediatas durante a gravidez e no parto, aumentando o risco de partos prematuros ou complicações neonatais. Isso, por sua vez, pode influenciar a necessidade de cuidados intensivos no período pós-natal.

Além disso, as comorbidades crônicas, como doenças metabólicas ou hipertensivas persistentes, também podem afetar negativamente a gestação ao longo do tempo, levando a desfechos adversos para os bebês, como baixo peso ao nascer e problemas respiratórios. Esses fatores de risco crônicos exigem acompanhamento cuidadoso durante a gravidez e podem aumentar a necessidade de internação dos bebês, especialmente em CTI/UTI Neonatal, dependendo da gravidade e do controle dessas condições. Portanto, as comorbidades maternas, tanto agudas quanto crônicas, desempenham um papel crucial na saúde neonatal e na probabilidade de complicações que exijam cuidados intensivos.

Tabela 5: Distribuição de Frequências (Absoluta e Proporcional) das Variáveis relacionadas ao estado de saúde das mães.

Comorbidade	f (%)
Indicador de Comorbidade Hipertensiva Aguda a Gestação	
Possui Comorbidade	702 (4,6%)
Não Possui Comorbidade	14.500 (95,4%)
Indicador de Comorbidade Metabólica Aguda a Gestação	
Possui Comorbidade	385 (2,5%)
Não Possui Comorbidade	14.817 (97,5%)
Indicador de Comorbidade Hipertensiva Crônica	
Possui Comorbidade	948 (6,2%)
Não Possui Comorbidade	14.254 (93,8%)
Indicador de Comorbidade Metabólica Crônica	
Possui Comorbidade	1.774 (11,7%)
Não Possui Comorbidade	13.428 (88,3%)

Com relação às variáveis do estado de saúde das crianças, um dos indicadores relevantes é o Apgar do primeiro minuto, que avalia a condição clínica do recém-nascido logo após o parto. Este escore, que varia de 0 a 10, considera cinco critérios: frequência cardíaca, esforço respiratório, tônus muscular, resposta ao estímulo e coloração da pele. O Apgar é fundamental para identificar rapidamente a necessidade de intervenções médicas, refletindo a saúde inicial da criança e sua adaptação ao ambiente extrauterino.

Devido à importância do indicador Apgar na avaliação das condições de saúde do recém-nascido nos primeiros minutos de vida e sua associação com o estado clínico inicial da criança, este estudo analisará essa variável em relação à necessidade de cuidados intensivos.

Conforme destaca a Tabela 6, a maioria dos indivíduos que não vivenciaram o evento de interesse apresentou uma distribuição de escores de Apgar predominantemente alta, especialmente nos escores 8 e 9, que somam 89,5% da frequência relativa. Isso indica que a maioria das crianças que não precisou de cuidados intensivos neonatais teve um desempenho favorável no primeiro minuto de vida, refletindo a boa adaptação ao ambiente extrauterino. Por outro lado, aqueles que vivenciaram o evento de interesse mostraram uma distribuição mais dispersa nos escores de Apgar, com frequências relativas mais baixas para os escores mais altos. Isso sugere que a pontuação do Apgar pode ser um indicativo útil para prever a necessidade de intervenções em ambientes de cuidados intensivos, ressaltando a importância de

monitorar esse indicador no momento do parto.

Tabela 6: Frequência relativa de Apgar do Primeiro Minuto por indivíduos que utilizaram/não utilizaram CTI/UTI neonatal.

Indicador	Apgar	Freq. Relativa (%)
Não Utilizou		
CTI/UTI Neonatal	0 a 7	7,9
	8 a 9	89,5
	10	2,5
Utilizou		
CTI/UTI Neonatal	0 a 7	25,2
	8 a 9	73,8
	10	0,9

A porcentagem de crianças nascidas com comorbidades⁴ é relativamente alta, atingindo 27,85%, conforme evidencia a Tabela 7. Ressalta-se que o indicador é uma variável chave na previsão de utilização de internações em CTI/UTI Neonatal, uma vez que dado que uma criança tenha doenças pré existente em seu primeiro ano de vida, a chance de desenvolver complicações severas que exigem cuidados especiais é alta.

Tabela 7: Distribuição de Frequências (Absoluta e Proporcional) do estado de saúde das crianças.

Comorbidade das Crianças	f (%)
Não Possui Comorbidade	10.967 (72,1%)
Possui Comorbidade	4.235 (27,9%)

Por fim, são apresentados na Figura 3 a distribuição do percentil do peso para a linha de base entre os indivíduos que utilizaram e não utilizaram o CTI/UTI Neonatal. Destaca-se que pesos abaixo ou acima dos percentis adequados podem ser indicativos de possíveis problemas de saúde, como desnutrição ou obesidade, que podem impactar negativamente o desenvolvimento da criança. Conforme evidencia a Figura 3, dentre os indivíduos que não utilizaram, a média do percentil do peso fica próximo de 50%, dentro da normalidade, contudo, aqueles que utilizaram possuem peso em média no percentil 40%, revelando um baixo peso ao nascer e ao longo do seu primeiro ano de vida.

⁴Foram consideradas qualquer tipo de comorbidade, sem realizar uma seleção específica

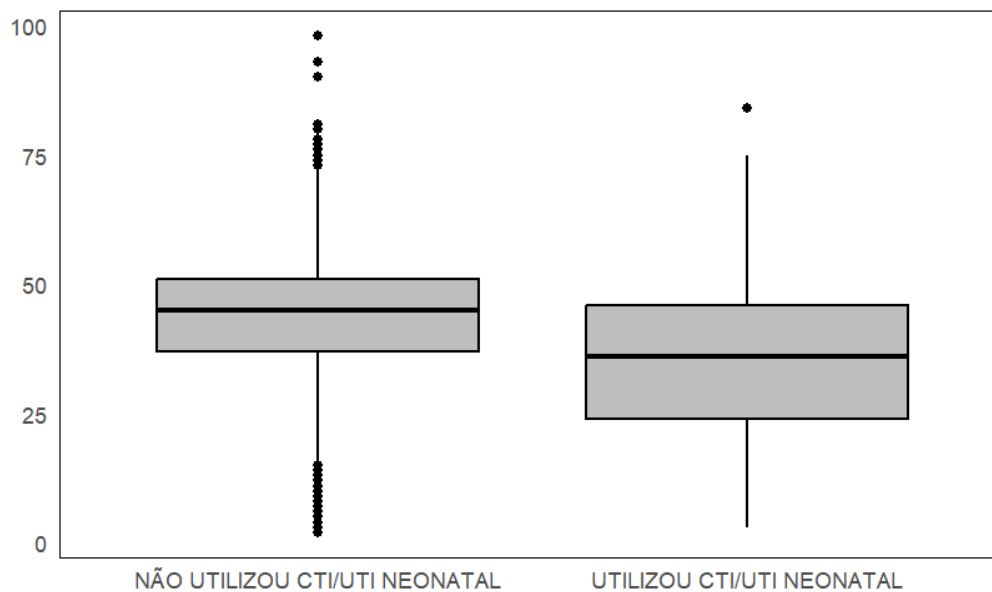


Figura 3: Box-Plot da variável percentil do peso.

4.2 Caracterização da Base de Dados: Coorte de Cardiopatas Chagásicos - SamiTrop

Para a realização do segundo estudo de caso, foi utilizada a coorte prospectiva SaMi-Trop. O estudo conta com 1544 pacientes que foram diagnosticados com cardiopatia chagássica. O estudo foi aplicado em 21 cidades do norte de Minas Gerais, sendo que os pacientes foram acompanhados no período de 2013 até 2018.

O objetivo é estimar a probabilidade de morte nos dois primeiros anos de estudo (estático), e para os indivíduos que sobreviveram aos primeiros dois anos e permanecem no estudo estimou-se a probabilidade de morte para mais dois anos. Para cumprir com o objetivo a base de dados seguindo o esquema mostrado na Figura 4:

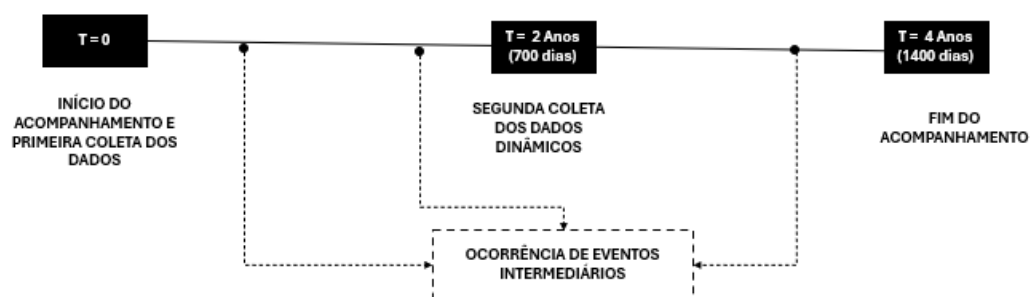


Figura 4: Acompanhamento clínico do estudo de caso de cardiopatas chagásicos.

As variáveis utilizadas no estudo, foram as seguintes:

- Idade do indivíduo: Variável medida em anos;
- Classe Funcional: Variável binária, que segue a classificação New York Heart Association (NYHA). Sendo quatro classes que dizem respeito à capacidade do indivíduo desenvolver atividades cotidianas. As classes foram agrupadas em duas categorias: 0 - classe funcional I e 1 - classe funcional II, III e IV.
- Frequência Cardíaca: variável tempo-dependente, medida em batimentos por minuto;
- ECG: Variável binária que indica se o indivíduo teve alteração em seu exame de eletrocardiograma, por meio do quantitativo de QRS(ms)⁵;
- IC: Variável binária que indica a presença de insuficiência cardíaca nos indivíduos, o diagnóstico é feito por meio do marcador NT-ProIBNP⁶.

Na segunda visita aos indivíduos participantes dos estudos, foram coletados os dados dos pacientes que sobreviveram aos dois primeiros anos e que não foram censurados, sendo que o marcador longitudinal atualizado foi a frequência cardíaca. Foram encontradas informações na segunda visita somente para 283 pacientes.

4.2.1 Breve Descrição dos Dados

A Tabela 8⁸ sumariza as informações referentes à idade dos indivíduos na coorte SaMi-Trop (2013-2014). Os indivíduos que compõem a coorte apresentam uma ampla variação de idade, com valores entre 18 e 96 anos. A mediana, que representa o ponto central da distribuição, é de 59 anos, indicando que metade dos participantes têm menos de 59 anos e a outra metade tem mais. A média é ligeiramente maior, em 59,4 anos, sugerindo que a distribuição é relativamente simétrica. O desvio padrão de 12,7 anos reflete uma dispersão moderada em torno da média, mostrando que há variabilidade considerável nas idades dos indivíduos, com alguns participantes significativamente mais velhos ou mais jovens.

Tabela 8: Descrição da Idade (2013-2014).

Mínimo	Máximo	Mediana	Média	Desvio Padrão
18	96	59	59,4	12,7

⁵Componente do eletrocardiograma que representa a despolarização ventricular, medido em milissegundos.

⁶Biomarcador liberado pelo coração em resposta à sobrecarga de volume ou pressão, utilizado na avaliação da insuficiência cardíaca.

A classe funcional foi transformada em uma variável dicotômica, agrupando as classes II, III e IV em uma única categoria, representando os indivíduos com alguma dificuldade na execução de suas atividades básicas. Observamos que 830 indivíduos (53%) estavam na classe I, enquanto 714 (47%) estavam no agrupamento das classes II, III e IV, ou seja, aproximadamente metade dos indivíduos estudados possuíam alguma dificuldade na execução de suas atividades diárias.

Para a variável ECG, que indica alterações no exame de eletrocardiograma, observamos que 1.277 indivíduos (82%) não apresentaram alterações e 267 (18%) apresentaram alguma alteração. Os resultado sugere que, apesar de todos os pacientes da coorte terem sido diagnosticados com cardiopatia de Chagas, uma proporção significativa não apresenta sinais eletrocardiográficos de complicações cardíacas evidentes. Contudo, os 18% que exibem alterações no ECG podem indicar um grupo de pacientes com maior risco de progressão para insuficiência cardíaca ou morte.

Da mesma forma, para a variável IC, que identifica a presença de insuficiência cardíaca, observamos que 1.330 indivíduos (86%) não apresentaram insuficiência e 214 (14%) apresentaram. Evidenciando a existência de um grupo pequeno dentre os indivíduos estudados que apresentam sinais mais latentes de complicações cardíacas e que possivelmente necessitem de um acompanhamento médico mais rigoroso e intervenções terapêuticas precoces para prevenir a progressão da doença e evitar o óbito.

A Tabela 9 sumariza as informações referentes à frequência cardíaca medida em batimentos por minuto (bpm) na coorte SaMi-Trop na linha de base (2013-2014). Os valores variam entre 21 bpm a 174 bpm, tem-se uma mediana de 64,5 bpm e uma média de 66,2 bpm, indicando que a maioria dos pacientes apresenta uma frequência cardíaca dentro de uma faixa considerada normal para adultos. O desvio padrão de 13,9 bpm reflete uma dispersão significativa em torno da média, o que pode sugerir a existência de subgrupos de pacientes com alterações importantes na frequência cardíaca e que potencialmente irão requer uma atenção médica mais próxima e intervenções mais agressivas.

Tabela 9: Descrição da Frequência cardíaca (2013-2014).

Mínimo	Máximo	Mediana	Média	Desvio Padrão
21	174	64,5	66,2	13,9

Por fim, com o objetivo de avaliar a alteração dos marcadores ao longo dos 4 anos de estudo, investigou-se a centralidade e dispersão dos dados nos dois momentos de coleta de informações. Vale destacar que, na segunda visita (após 2 anos da linha

de base), foram coletados dados de apenas 283 pacientes sobreviventes. Esta análise é crucial, dado o intervalo de dois anos entre as coletas, um período suficientemente longo para que os marcadores estáticos: idade, ECG, IC e Classe Funcional, possam apresentar variações significativas. Dessa forma, a análise foi realizada para garantir que a pesquisa não apresente viés decorrentes do uso de dados discrepantes. A Tabela 10 foi elaborada para ilustrar essas mudanças potenciais e resguardar a integridade da análise.

Tabela 10: Média e mediana das covariáveis quantitativas e frequência relativa das covariáveis categóricas ao longo do tempo.

Covariáveis	1ª Extração dos Dados			2ª Extração dos Dados		
	Média	Mediana	Desvio Padrão	Média	Mediana	Desvio Padrão
Idade	59,4	59	12,7	62,9	62	13,1
Frequência Cardíaca (bpm)	66,2	64,5	13,9	64,3	65,2	16,7
	Frequência Relativa (%)			Frequência Relativa (%)		
Classe Funcional		47%			85%	
ECG		18%			38%	
IC		14%			37%	

Com base nos resultados apresentados nas duas extrações de dados, observa-se que não houve mudanças significativas nas variáveis estudadas ao longo do tempo. A média da idade aumentou de 59,4 para 62,9 anos, o que era esperado devido ao envelhecimento natural da coorte ao longo dos dois anos. No entanto, a dispersão da idade, representada pelo desvio padrão, manteve-se estável, variando de 12,7 para 13,1, sugerindo pouca alteração na distribuição etária dos participantes.

A frequência cardíaca também apresentou uma variação mínima, com a média passando de 66,2 bpm na primeira extração para 64,3 bpm na segunda, e uma leve alteração no desvio padrão, de 13,9 para 16,7 bpm. Essa pequena variação na frequência cardíaca ao longo do tempo sugere estabilidade dessa variável, sem mudanças abruptas entre os dois momentos.

As variáveis dicotômicas, como a classe funcional, ECG e IC, apresentaram aumento nas frequências relativas, sendo essa mudança mais latente na classe funcional.

5 Métodos Computacionais

O modelo de riscos proporcionais de Cox (1972), segue sendo explorado em diversos softwares estatísticos, como R, Stata e Python, devido à sua ampla aplicabilidade em estudos de sobrevivência. Implementações nesses programas oferecem ferramentas para ajuste do modelo, diagnóstico da proporcionalidade dos riscos e visualização das curvas de sobrevivência, tornando-o acessível para pesquisadores de diferentes áreas. No software R, o modelo de Cox é implementado de maneira eficiente, com destaque para o pacote `survival`, que oferece funções como `coxph` para o ajuste de modelos e `survfit` para a análise de curvas de sobrevivência. Além de estimar coeficientes associados às covariáveis, a implementação permite verificar a suposição de proporcionalidade dos riscos, uma premissa fundamental do modelo, por meio de gráficos diagnósticos e testes estatísticos.

Entre as funcionalidades adicionais oferecidas no R, destaca-se a possibilidade de manipular dados censurados à direita, comum em estudos de sobrevivência, além de realizar análises multivariadas, estratificadas e ajustadas. Extensões como o cálculo de razões de risco (hazard ratios) ajustadas e intervalos de confiança ampliam as aplicações práticas do modelo de Cox, integrando-o com outros pacotes especializados.

Dada a limitação do modelo de Cox para lidar adequadamente com covariáveis endógenas e sua não capacidade de modelar explicitamente a interdependência entre os processos longitudinais e de tempo até o evento, surgem alternativas que permitem integrar essas informações de forma mais robusta. Nesse contexto, os modelos conjuntos (joint models) emergem como uma abordagem poderosa, pois modelam simultaneamente os dois processos e capturam as associações dinâmicas entre eles, oferecendo estimativas mais precisas e interpretações mais completas (Rizopoulos, 2014).

A crescente necessidade de modelagens mais abrangentes para superar as limitações dos métodos tradicionais, como o modelo de Cox, incentivou o desenvolvimento de ferramentas computacionais voltadas para a aplicação de modelos conjuntos. Nesse cenário, destaca-se o pacote JM no R, amplamente utilizado para ajustar modelos conjuntos que relacionam um desfecho longitudinal contínuo a um processo de tempo até o evento, sob o paradigma da máxima verossimilhança (Rizopoulos, 2014, 2012, 2010). O JM oferece suporte a diversas estruturas de associação entre os processos e também permite análises que envolvem dados de sobrevivência com riscos concorrentes. Além disso, o pacote proporciona uma funcionalidade robusta para calcular previsões dinâmicas, possibilitando insights preditivos detalhados tanto para os desfechos longitudinais quanto para os de sobrevivência.

Um dos aspectos fundamentais na especificação do modelo é a escolha da distribuição da linha de base para o tempo de falha, que pode influenciar significativamente os resultados da análise. O pacote permite a escolha entre três distribuições principais: Weibull, Spline e Piecewise. A escolha entre essas distribuições deve considerar as características dos dados e a complexidade desejada no ajuste do modelo. O pacote JM foi escolhido neste trabalho para realizar as previsões nos dois estudos de caso explorados.

Além do JM, outras ferramentas computacionais expandem as possibilidades na aplicação de modelos conjuntos. O pacote `joiner` (Philipson et al., 2012), por exemplo, segue uma abordagem semelhante ao ajustar modelos conjuntos para um desfecho longitudinal contínuo e um tempo até o evento, fundamentando-se na formulação de Henderson, Diggle e Dobson (2002). Já no software STATA, o pacote `stjm` (Crowther, 2013) implementa modelagens conjuntas que associam uma resposta longitudinal normal a um tempo até o evento, com ênfase em modelos paramétricos para a análise do tempo de sobrevivência. Alternativas adicionais incluem a implementação de modelos conjuntos no SAS e no WinBUGS, conforme discutido por Guo e Carlin (2004), ampliando o acesso a técnicas avançadas de análise conjunta em diferentes plataformas.

Outro exemplo de avanço significativo é o pacote `lcmm` no R, que incorpora a função `Jointlcmm()` para ajustar modelos conjuntos baseados em classes latentes. Essa abordagem, descrita por Proust-Lima et al. (2013), permite capturar a associação entre os processos longitudinais e de sobrevivência por meio de efeitos aleatórios categóricos, também conhecidos como classes latentes. Essa inovação amplia as possibilidades de modelagem para situações onde os efeitos não podem ser devidamente explicados por variáveis contínuas tradicionais, fornecendo uma estrutura flexível para lidar com a heterogeneidade entre os indivíduos analisados.

O pacote `JMbayes` para o R surge como uma evolução do JM, ao incorporar uma abordagem Bayesiana na implementação de modelos conjuntos, ampliando significativamente as possibilidades de análise. Diferentemente do JM, que utiliza máxima verossimilhança, o `JMbayes` adota métodos Bayesianos para ajustar modelos conjuntos que integram respostas longitudinais contínuas e categóricas, além de fornecer flexibilidade para modelar estruturas de associação complexas entre os desfechos. Entre as funcionalidades adicionais, destaca-se a capacidade de incluir diferentes termos do submodelo longitudinal no preditor linear do submodelo de sobrevivência, como o valor atual da variável longitudinal e sua taxa de variação (derivada), permitindo transformações gerais desses termos (Rizopoulos, 2014).

O pacote `JMbayes` tem como característica central sua robustez em prever dina-

micamente desfechos longitudinais e de tempo até o evento. Essas previsões podem ser combinadas utilizando técnicas Bayesianas, o que aumenta a precisão e a aplicabilidade prática das análises (Rizopoulos, 2014). Além disso, o pacote oferece uma interface web que facilita o uso das previsões no contexto clínico, permitindo que os usuários integrem facilmente os resultados às suas práticas. Adicionalmente, o JMBayes disponibiliza ferramentas avançadas para avaliar a qualidade das previsões, utilizando métricas como discriminação e calibração, aspectos fundamentais para a validação e interpretação dos modelos.

Ademais, tem-se que o pacote JMBayes apresenta uma abrangência metodológica que suporta uma ampla gama de estruturas de modelos conjuntos, incluindo dados de sobrevivência com riscos concorrentes e submodelos longitudinais de diferentes naturezas. Essas características tornam o pacote adequado para uma variedade de aplicações, desde análises biomédicas envolvendo biomarcadores até estudos epidemiológicos complexos. Sua flexibilidade também se reflete na possibilidade de ajuste de modelos paramétricos e não paramétricos, oferecendo maior liberdade ao pesquisador na escolha da abordagem mais adequada.

Em sua última divulgação, Rizopoulos et al. (2024) apresenta a implementação em R do pacote JMBayes2 que ampliou as capacidades de modelagem conjunta no R, tornando-se ainda mais versátil e abrangente. Diferentemente de seu predecessor, o JMBayes2 introduziu suporte para múltiplos desfechos longitudinais de tipos variados, como contínuos, dicotômicos, ordinais e contagens, permitindo a modelagem sob diferentes distribuições, incluindo Gaussiana, t de Student, Gamma, Beta, unit Lindley, Normal censurada, Binomial, Poisson, Binomial Negativa e Beta-Binomial. Além disso, no contexto do processo de tempo até o evento, o pacote passou a lidar com dados censurados à direita, à esquerda e por intervalo, além de incorporar a modelagem de riscos concorrentes e processos multiestado.

A utilização de modelos computacionais adequados não apenas enriquece a análise, mas também fornece insights valiosos para a prática clínica, promovendo uma melhor compreensão dos fatores que influenciam a sobrevida dos pacientes. Neste trabalho, foi adotado o pacote JM, que implementa o ajuste de modelos conjuntos por máxima verossimilhança, enquanto o pacote JMBayes será empregado em investigações futuras para integrar abordagens bayesianas. As metodologias discutidas servem como base para novas aplicações na área de saúde, contribuindo para a evolução das técnicas de análise de sobrevivência em estudos de coorte.

6 Resultados

Nesta seção, são apresentados os resultados obtidos a partir dos dois estudos de caso explorados: a coorte de cardiopatas chagásicos e a utilização do CTI/UTI neonatal. O foco da análise está na avaliação do desempenho das quatro abordagens dinâmicas estudadas em termos das medidas de validação do modelo predito, com o objetivo de identificar qual metodologia apresenta os melhores resultados em cada contexto avaliado. A comparação entre as abordagens será conduzida considerando as especificidades de cada base de dados, fornecendo evidências sobre a aplicabilidade e a eficácia das técnicas no suporte à tomada de decisão clínica.

6.1 Modelos de Predição - Utilização CTI/UTI Neonatal

São detalhados os resultados obtidos a partir da aplicação das abordagens estática e dinâmicas ao conjunto de dados da Cooperativa Médica. O foco da análise é a predição do risco de recém-nascidos necessitarem de internação em CTI/UTI neonatal ao longo do primeiro ano de vida, proporcionando uma avaliação comparativa entre as diferentes metodologias.

6.1.1 Modelo de Cox na Linha de Base

Na Tabela 11, são apresentados os resultados do modelo de Cox, sendo eles: as estimativas, a razão das taxas de falha (RTF), o erro padrão e o p-valor associado aos testes de significância de cada um dos marcadores.

Tabela 11: Estimativas, risco relativo, EP e p-valor - Modelo de Cox.

	Estimativa	\widehat{RTF} (IC 95%)	Erro Padrão	P-Valor
Indicador de Realização do pré-natal de acordo com o Protocolo - Cooperativa Médica	-0,259	0,772 (0,657 - 0,907)	0,082	0,002
Indicador de Comorbidade Hipertensiva Aguda à Gestação	-0,006	0,994 (0,771 - 1,283)	0,130	0,965
Indicador de Comorbidade Metabólica Aguda à Gestação	0,371	1,448 (1,037 - 2,022)	0,170	0,030
Indicador de Comorbidade Hipertensiva Crônica	0,172	1,188 (0,913 - 1,546)	0,135	0,200
Indicador de Comorbidade Metabólica Crônica	-0,119	0,888 (0,703 - 1,122)	0,119	0,320
Número de Semanas do Parto	-1,104	0,332 (0,276 - 0,398)	0,093	<0,001
Idade no momento do parto	-0,145	0,865 (0,747 - 1,002)	0,075	0,054
Indicador Primípara	-0,189	0,828 (0,712 - 0,962)	0,077	0,014
Tipo de Parto	-0,114	0,892 (0,758 - 1,051)	0,083	0,171
Tempo entre a internação e o parto	-0,214	0,807 (0,683 - 0,954)	0,085	0,012
Indicador Apgar do primeiro minuto	-0,119	0,888 (0,846 - 0,932)	0,025	<0,001
Indicador Apgar do quinto minuto	-0,130	0,878 (0,822 - 0,939)	0,034	<0,001
Indicador de Comorbidade Fetal	1,768	5,859 (4,797 - 7,156)	0,102	<0,001
Percentil do Peso ao Nascer	-0,036	0,965 (0,958 - 0,971)	0,004	<0,001

No modelo de Cox ajustado, três variáveis apresentaram efeito positivo, sendo duas delas significativas, indicando que o aumento desses fatores está associado a um maior risco de utilização do CTI/UTI Neonatal pela criança. O Indicador de Comorbidade Metabólica Aguda à Gestação reflete condições metabólicas que surgem

durante a gestação e podem comprometer tanto a saúde materna quanto o desenvolvimento fetal. Já o Indicador de Comorbidade Hipertensiva Crônica abrange hipertensões pré-existentes na mãe, as quais podem dificultar o curso natural da gestação e aumentar as complicações perinatais. Por fim, o Indicador de Comorbidade Fetal está relacionado a condições de saúde adversas do recém-nascido, como malformações ou outras doenças detectadas ao nascimento, que frequentemente demandam cuidados mais intensivos.

As variáveis Indicador de Comorbidade Hipertensiva Aguda à Gestação e Indicador de Comorbidade Metabólica Crônica não apresentaram efeito estatisticamente significativo sobre o desfecho, mas serão mantidas no modelo por seu reconhecido poder preditivo. De forma empírica, sabe-se que essas condições estão associadas a complicações no primeiro ano de vida da criança, conforme também apontado por profissionais médicos envolvidos no estudo.

Adicionalmente, destaca-se o papel da variável dinâmica percentil do peso da criança, que exerce um efeito protetor: quanto maior o percentil de peso ao nascimento, menor o risco de necessidade de cuidados intensivos neonatais. Esse achado reforça a importância de monitorar o crescimento fetal adequado como estratégia para reduzir as complicações neonatais e a necessidade de intervenções em CTI/UTI Neonatal.

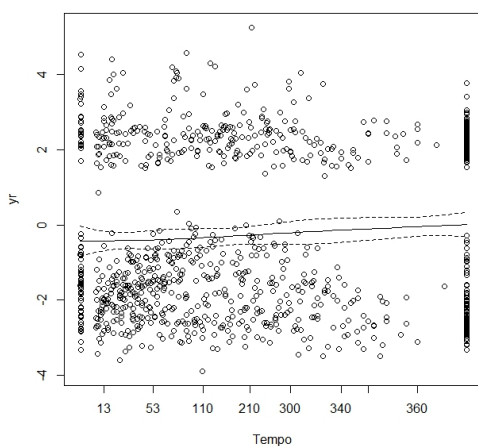
No modelo ajustado, duas variáveis se destacaram, por seu impacto significativo, no risco de internação em CTI/UTI Neonatal: Número de Semanas do Parto e o indicador de comorbidade fetal. Número de Semanas do Parto apresentou uma associação inversa, com cada unidade adicional de semanas que a mãe passa reduz o risco de admissão em CTI/UTI Neonatal em 67%. Esse resultado reforça a hipótese de que as complicações graves no parto estão altamente associadas a prematuridade da criança. Por outro lado, o indicador de comorbidade fetal mostrou o maior efeito positivo no modelo, sendo que o risco de admissão em CTI/UTI Neonatal para os indivíduos com comorbidade fetal é cerca de 6 vezes maior em comparação com aqueles que não possuem comorbidade. Isso evidencia que a presença de condições fetais adversas é um fator determinante para a necessidade de cuidados intensivos neonatais.

Por fim, conclui-se que, embora nem todas as variáveis apresentem significância estatística direta em relação ao desfecho, todas serão mantidas no modelo devido à sua capacidade de contribuir para a previsão global do desfecho. Cada variável, mesmo aquelas com efeitos aparentemente menos expressivos, agrega valor ao modelo ao capturar nuances importantes que, em conjunto, aprimoram a acurácia preditiva. Essa abordagem garante que o modelo considere múltiplas dimensões do risco, per-

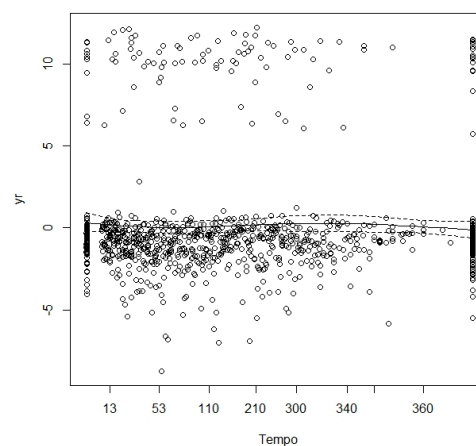
mitindo uma análise mais abrangente e confiável sobre a utilização do CTI/UTI Neonatal.

As estimativas de sobrevivência da linha de base para o tempo de 28 dias e 6 meses são $\hat{S}_0(28 \text{ dias}) = 0,99$ e $\hat{S}_0(182 \text{ dias}) = 0,97$.

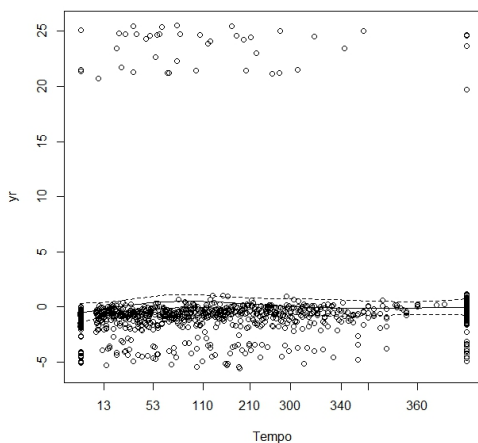
Foram analisados os gráficos dos resíduos de Schoenfeld em função do tempo para os marcadores, a fim de identificar violação do princípio de proporcionalidade dos riscos.



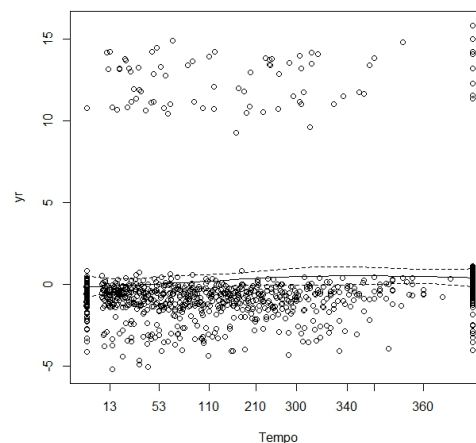
Pré-natal segundo Protocolo – Cooperativa Médica.



Comorbidade Hipertensiva Aguda à Gestação.

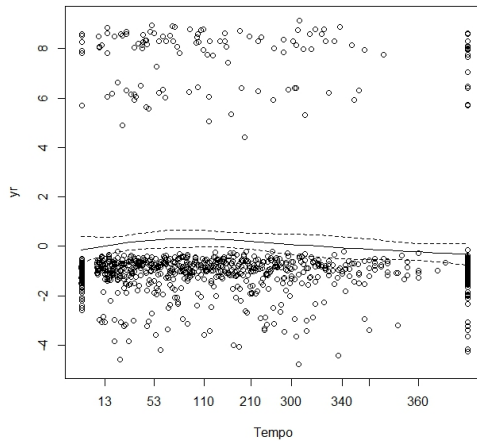


Comorbidade Metabólica Aguda à Gestação.

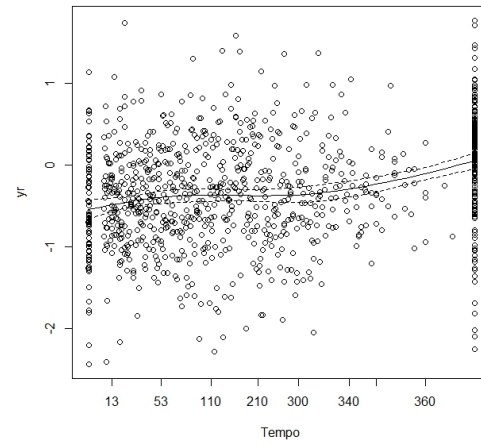


Comorbidade Hipertensiva Crônica.

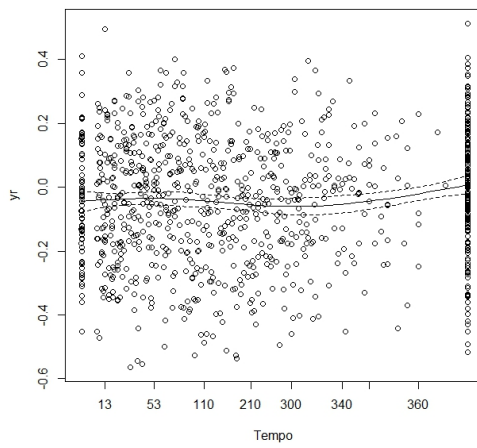
Figura 5: Resíduos de Schoenfeld versus tempo para os marcadores.



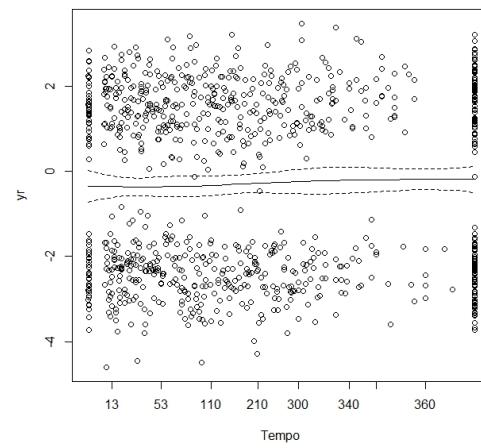
Comorbidade Metabólica Crônica.



Número de Semanas do Parto.

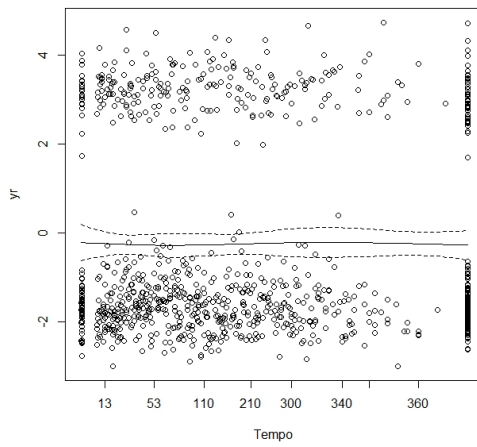


Idade no momento do parto.

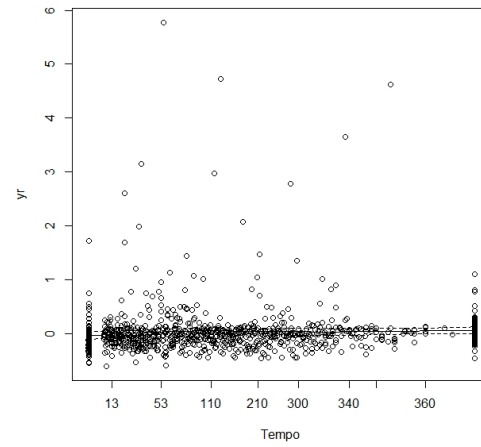


Indicador Primípara.

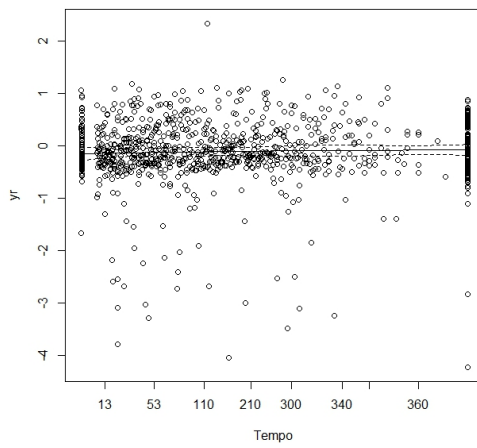
Figura 6: Resíduos de Schoenfeld versus tempo para os marcadores.



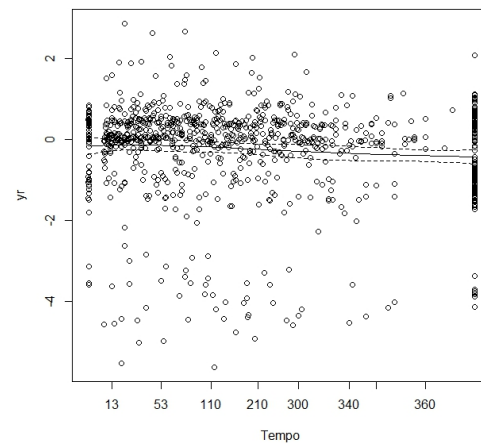
Tipo de Parto.



Tempo entre a internação e o parto.

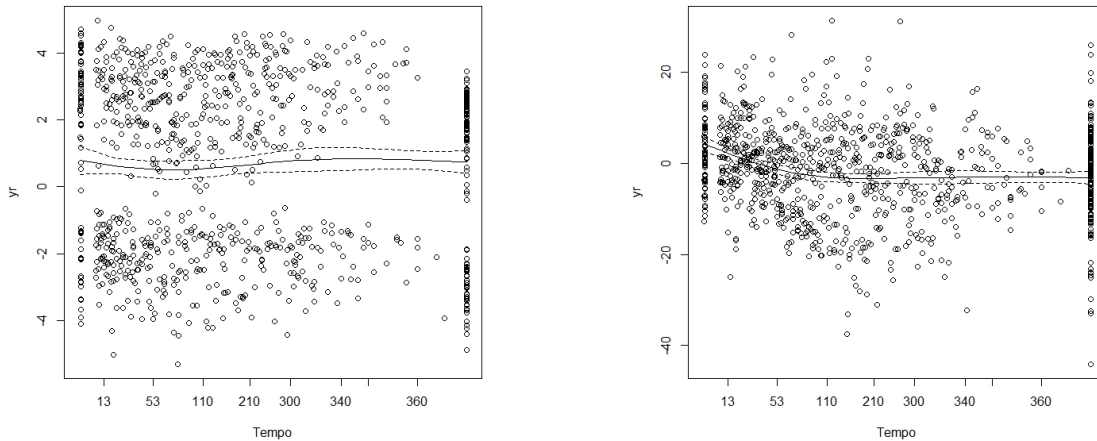


Indicador Apgar do primeiro minuto.



Indicador Apgar do quinto minuto.

Figura 7: Resíduos de Schoenfeld versus tempo para os marcadores.



Comorbidade Fetal.

Percentil do Peso ao Nascer.

Figura 8: Resíduos de Schoenfeld versus tempo para os marcadores.

De maneira geral, a partir dos gráficos apresentados nas Figuras 5, 6, 7, 8 e 9, é possível constatar que todas as variáveis do modelo satisfaz o pressuposto de proporcionalidade dos riscos, com resíduos distribuídos de forma aleatória em torno da linha de referência.

6.1.2 Predição Estática em 28 dias

A Tabela 12 evidencia o alto poder de discriminação do modelo de linha de base ($AUC \geq 0,8$), ou seja, o modelo consegue distinguir os indivíduos que irão utilizar o CTI/UTI Neonatal em 28 dias após o nascimento dos que não irão utilizar de forma excepcional. Além disso, por meio dos testes de Hosmer e Lemenshow, verifica-se que os modelos estão bem calibrados, ou seja, as probabilidades preditas são próximas das observadas.

	AUC (95%)	Teste HL _{ND-VAR}	Teste HL	\widehat{PE}
K-fold	84,26% (79,13% - 88,40%)	0,2607	0,7452	0,0097
Boots	84,39% (83,79% - 85,03%)	0,2885	0,4528	0,0097

Tabela 12: Performance do modelo (discriminação (AUC) e calibração (Teste de Hosmer Lemenshow e Erro de Predição)) de linha de base para prever a utilização de CTI/UTI neonatal em 28 dias.

Os resultados apresentados na Tabela 12 são fruto de 100 simulações por meio dos métodos dos k-folds com 5 folds e 20 repetições do método bootstrap com 200 reamostras. Esse processo de repetição proporciona uma avaliação mais confiável e

estável do desempenho dos modelos, confirmando que os achados são consistentes e não resultam de variações aleatórias nos dados.

Com base nos resultados obtidos, tem-se evidências para afirmar que o modelo construído para prever a utilização de CTI/UTI neonatal é eficaz em discriminar pacientes com alto risco daqueles com baixo risco. Além disso, os testes dão evidência de uma boa calibração.

A Figura 9 apresenta a curva ROC para o modelo de predição em até 28 dias, evidenciando uma área sob a curva (AUC) de 0,846. Esse valor indica uma excelente capacidade discriminativa do modelo, ou seja, sua habilidade em diferenciar corretamente entre os indivíduos que irão ou não experimentar o desfecho em estudo. Já a Figura 10 mostra o gráfico de calibração, comparando os riscos esperados com os riscos observados em diferentes faixas de probabilidade. Os pontos estão distribuídos próximos à linha de 45 graus, sugerindo boa calibração do modelo — ou seja, as probabilidades previstas estão em consonância com as proporções reais de eventos observados. Esses dois resultados, em conjunto, indicam que o modelo apresenta bom desempenho tanto em termos de discriminação quanto de calibração.

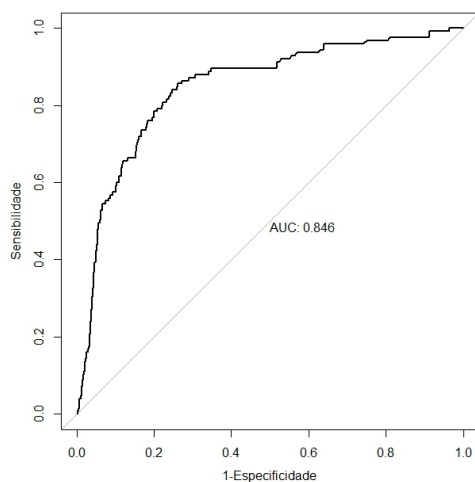


Figura 9: Curva ROC para predição de 28 dias.

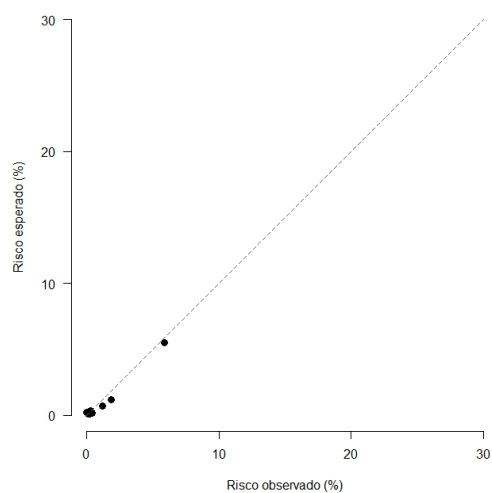


Figura 10: Risco esperados vs. risco observado para predição de 28 dias.

6.1.3 Modelo de Predição Dinâmica - 6 Meses

As quatro metodologias de predição dinâmicas apresentadas no Capítulo 3 serão utilizadas para prever a utilização de CTI/UTI Neonatal em seis meses após o nascimento (182 dias) para as crianças que até o vigésimo oitavo dia ainda não haviam utilizado da internação em CTI/UTI. Para os modelos que utilizam de uma nova linha de base, foram observados 9.448 indivíduos.

A Tabela 13 apresenta os estimadores do modelo de Cox para a linha de base em 28 dias. Os resultados incluem as estimativas dos parâmetros do modelo, acompanhadas de seus respectivas razão da taxa de falha, erros padrão e p-valores.

Tabela 13: Estimativas, razão de taxa de falha, erro padrão e p-valor do modelo LandMark para linha de base em 28 dias para o risco de internação em CTI.

	Estimativa	Razão das Taxas de Falha (IC 95%)	Erro Padrão	P-Valor
Indicador de Realização do pré-natal de acordo com o Protocolo – Cooperativa Médica	-0,350	0,705 (0,521 - 0,954)	0,154	0,024 *
Indicador de Comorbidade Hipertensiva Aguda à Gestação	0,159	1,172 (0,775 - 1,774)	0,211	0,452
Indicador de Comorbidade Metabólica Aguda à Gestação	0,162	1,176 (0,668 - 2,070)	0,289	0,575
Indicador de Comorbidade Hipertensiva Crônica	-0,144	0,865 (0,529 - 1,417)	0,251	0,566
Indicador de Comorbidade Metabólica Crônica	-0,077	0,926 (0,660 - 1,298)	0,173	0,655
Número de Semanas do Parto	-1,284	0,277 (0,202 - 0,380)	0,162	<0,001
Idade no momento do parto	-0,009	0,991 (0,761 - 1,291)	0,135	0,947
Indicador Primípara	-0,100	0,905 (0,697 - 1,176)	0,133	0,454
Tipo de Parto	-0,186	0,830 (0,613 - 1,125)	0,155	0,230
Hora do Parto	-0,147	0,863 (0,653 - 1,141)	0,142	0,302
Apgar no Primeiro Minuto	-0,085	0,919 (0,843 - 1,001)	0,044	0,053
Apgar no Quinto Minuto	-0,140	0,870 (0,735 - 1,029)	0,086	0,103
Indicador de Comorbidade Fetal	1,831	6,243 (4,257 - 9,155)	0,195	<0,001
Percentil do Peso ao Nascer	-0,037	0,964 (0,953 - 0,975)	0,006	<0,001

Os resultados do modelo Landmark indicaram que as variáveis percentil do peso ao nascer e o indicador de comorbidade fetal exerceram os maiores impactos sobre a taxa de falha. O percentil do peso ao nascer apresentou um efeito significativo e negativo, evidenciando que aumentos nessa variável estão fortemente associados a uma redução na ocorrência de internação intensiva neonatal. Por outro lado, o indicador de comorbidade fetal demonstrou alta relevância, mantendo o sentido observado no modelo de linha de base e reforçando o impacto das condições fetais adversas nos desfechos. Além disso, o número de semanas do parto e o indicador de pré-natal completo destacaram-se por seus efeitos estatisticamente significativos, ambos associados à redução do risco.

Para o modelo de sobrevivência, foram utilizadas três formas para a taxa de falha na linha de base: Weibull, Piecewise com cinco nós e Spline com sete nós. Ressalta-se que a escolha do número de nós para a função spline foi realizada por tentativa, visando garantir a convergência da matriz Hessiana, com variações testadas até o limite de sete nós.

Através dos resultados apresentados nas Tabelas 14, 15 e 16, podemos constatar que para as três linhas de base a maior parte das variáveis medidas são significativas

para explicar o tempo até a ocorrência da internação em CTI/UTI Neonatal.

Tabela 14: Estimativas, $\hat{H}R$ e p-valor do submodelo de sobrevivência com especificação da distribuição Weibull na linha de base.

	Estimativa	$\hat{H}R$	P-Valor
Indicador de Realização do Pré-Natal Completo – Cooperativa Médica	-0,2221	0,0729	0,0023
Indicador de Comorbidade Hipertensiva Aguda à Gestação	0,0543	0,1123	0,6288
Indicador de Comorbidade Hipertensiva Crônica	-0,0251	0,1185	0,8321
Indicador de Comorbidade Metabólica Crônica	-0,1092	0,0970	0,2600
Indicador de Comorbidade Metabólica Aguda à Gestação	0,2286	0,1497	0,1268
Número de Semanas do Parto	-0,7605	0,1406	<0,0001
Idade no Momento do Parto	-0,1672	0,0668	0,0123
Indicador Primípara	-0,2817	0,0689	<0,0001
Tipo de Parto	-0,1662	0,0734	0,0235
Hora do Parto	-0,1154	0,0734	0,1159
Apgar no Primeiro Minuto	-0,0931	0,0224	<0,0001
Apgar no Quinto Minuto	-0,1562	0,0326	<0,0001
Indicador de Comorbidade Fetal	1,7201	0,0919	<0,0001
Percentil do Peso ao Nascer	-0,0831	0,0137	<0,0001

Com base nos resultados do modelo conjunto com distribuição de linha de base Weibull, observou-se que o percentil do peso ao nascer apresentou associação estatisticamente significativa com o risco de utilização do CTI/UTI Neonatal, conforme observado nos resultados obtidos para linha de 28 dias, onde esta variável também foi um preditor relevante. O número de semanas do parto manteve-se como um fator fortemente associado à predição da utilização do CTI/UTI Neonatal, reforçando a relevância da prematuridade como preditor de risco em diferentes horizontes de tempo.

Além disso, a presença de comorbidade fetal continuou sendo um dos principais preditores de risco, apresentando valores de significância consistentes com os obtidos no modelo para 28 dias. As variáveis relacionadas às comorbidades agudas maternas apresentaram associação positiva, mas com menor magnitude em comparação ao modelo de linha de 28 dias, sugerindo que o impacto dessas condições pode ser mais expressivo no curto prazo. Em síntese, embora os preditores principais sejam similares em ambos os períodos, a redução da influência das comorbidades agudas no modelo aponta para uma possível modificação no perfil de risco ao longo do tempo, destacando a importância de ajustes temporais para modelos preditivos.

Tabela 15: Estimativas, $\hat{H}R$ e p-valor do submodelo de sobrevivência com especificação da distribuição Piecewise na linha de base.

	Estimativa	$\hat{H}R$	P-Valor
Indicador de Realização do pré-natal de acordo com o Protocolo – Cooperativa Médica	-0,2508	0,0778	0,0013
Indicador de Comorbidade Hipertensiva Aguda à Gestação	0,0355	0,1295	0,7839
Indicador de Comorbidade Hipertensiva Crônica	0,0362	0,1311	0,7827
Indicador de Comorbidade Metabólica Crônica	-0,1176	0,1065	0,2694
Indicador de Comorbidade Metabólica Aguda à Gestação	0,3036	0,1696	0,0734
Número de Semanas do Parto	-0,3061	0,1520	0,0440
Idade no momento do parto	-0,1666	0,0732	0,0229
Indicador Primípara	-0,3649	0,0768	<0,0001
Tipo de Parto	-0,2225	0,0799	0,0054
Hora do Parto	-0,1432	0,0815	0,0788
Apgar no Primeiro Minuto	-0,0877	0,0257	0,0006
Apgar no Quinto Minuto	-0,1712	0,0422	<0,0001
Indicador de Comorbidade Fetal	1,6221	0,0945	<0,0001
Percentil do Peso ao Nascer	-0,1657	0,0163	<0,0001

Os resultados do modelo ajustado com especificação da distribuição piecewise na linha de base confirmam a relevância dos preditores. O indicador de primiparidade manteve-se como o preditor mais fortemente associado ao risco de utilização do CTI/UTI Neonatal. O modelo sugere que mães que não são primíparas apresentam uma redução significativa no risco de suas crianças utilizarem o CTI/UTI Neonatal.

O percentil do peso ao nascer continua sendo um preditor significativo no modelo, destacando a importância do peso como um marcador de risco neonatal. O indicador de comorbidade metabólica aguda à gestação foi significativo e apresentou uma baixa associação em relação ao modelo de 28 dias, indicando que o impacto dessas condições pode diminuir com o aumento do horizonte temporal. Por outro lado, a idade no momento do parto, o indicador de realização de pré-natal completo, o tipo de parto, e os escores de Apgar no primeiro e quinto minuto tiveram significância estatística no modelo, sugerindo que essas variáveis têm um impacto consistente e prolongado no risco de utilização do CTI/UTI Neonatal.

Notavelmente, a comorbidade fetal apresentou uma associação ainda mais forte no modelo, reforçando sua relevância como um dos principais preditores de risco ao longo do tempo. O indicador de comorbidade hipertensiva, embora não significativo em nenhum dos modelos, mostrou-se ligeiramente mais próximo da significância, sugerindo a necessidade de estudos futuros para explorar sua potencial influência em cenários de longo prazo.

Em resumo, enquanto o indicador de primiparidade e o percentil do peso ao nascer permanecem como preditores consistentes de risco, algumas variáveis, como as comorbidades metabólicas, apresentam diminuição de impacto no modelo de longo prazo, destacando a relevância de ajustar os modelos conforme o horizonte temporal

considerado.

Tabela 16: Estimativas, $\hat{H}R$ e p-valor do submodelo de sobrevivência com especificação da distribuição Spline na linha de base.

	Estimativa	$\hat{H}R$	P-Valor
Indicador de Realização do Pré-Natal Completo	-0,2425	0,0759	0,0014
Indicador de Comorbidade Hipertensiva Aguda à Gestação	0,0776	0,1236	0,5303
Indicador de Comorbidade Hipertensiva Crônica	0,0470	0,1269	0,7109
Indicador de Comorbidade Metabólica Crônica	-0,1176	0,1032	0,2546
Indicador de Comorbidade Metabólica Aguda à Gestação	0,3093	0,1634	0,0584
Número de Semanas do Parto	-0,4257	0,1414	0,0026
Idade no momento do parto	-0,0672	0,0706	0,3411
Indicador de Primiparidade	-0,2708	0,0739	0,0002
Tipo de Parto	-0,1513	0,0773	0,0503
Hora do Parto	-0,1112	0,0794	0,1615
Apgar no Primeiro Minuto	-0,0901	0,0249	0,0003
Apgar no Quinto Minuto	-0,0937	0,0417	0,0246
Indicador de Comorbidade Fetal	1,7928	0,0945	<0,0001
Percentil do Peso ao Nascer	-0,1448	0,0147	<0,0001

O submodelo de sobrevivência com especificação spline apresentou resultados em geral consistentes com os observados nos modelos para as demais especificações da linhas de base. As variáveis relacionadas à comorbidade e à idade novamente não demonstraram significância estatística robusta. No entanto, destaca-se que o impacto do número de semanas do parto permaneceu proeminente, com uma estimativa significativa indicando que um aumento no número de semanas está associado à redução do risco de utilização do CTI/UTI Neonatal.

Adicionalmente, o indicador de primiparidade e as variáveis associadas ao estado fetal, como o percentil do peso ao nascer e o Apgar, mantiveram-se significativos, reforçando sua relevância clínica.

A Tabela 17 apresentada os resultados das quatro abordagens em termos de calibração e discriminação dos modelos para a predição dinâmica até os seis meses de vida dos recém nascidos, dado que eles não utilizaram dos serviços de CTI/UTI Neonatal até os 28 primeiros dias de vida. As validações partiram de duas metodologias distintas, a fim de estimar a performance dos modelos e reduzir viés e variância: método dos k-folds considerando 5 folds, e bootstrap, considerando 20 reamostras. Os modelos foram testados repetidas vezes, a fim de verificar a consistência dos resultados em diferentes simulações. Desta forma, os resultados apresentados são a média de 100 simulações para o método dos k-folds, e a média de 20 simulações para

o método de bootstrap.

Tabela 17: Resultados de validação pelas métricas de discriminação (AUC) e calibração (Teste de Hosmer Lemenshow e Erro de Predição) por modelo e método.

Modelo	Método	AUC	IC 2,5% - 97,5%	HL_var	HL	PE
Naive 1	K-Folds	0,8464	(0,7804 - 0,9165)	0,0089	0,8148	0,0167
	Bootstrap	0,8417	(0,7964 - 0,8848)	0,0280	0,5394	0,0167
Naive 2	K-Folds	0,9631	(0,8891 - 0,9918)	0,2841	0,6058	0,0043
	Bootstrap	0,9658	(0,9164 - 0,9890)	0,0016	0,0000	0,0042
LM	K-Folds	0,8430	(0,7826 - 0,8946)	0,0026	0,5916	0,0138
	Bootstrap	0,8429	(0,7965 - 0,8785)	0,1106	0,1280	0,0139
JM (Weibull)	K-Folds	0,9342	(0,8624 - 1,0000)	0,6711	0,0006	0,0065
	Bootstrap	0,9468	(0,8822 - 0,9965)	0,0000	0,5056	0,0066
JM (Spline)	K-Folds	0,9410	(0,8761 - 1,0000)	0,4829	0,0073	0,0032
	Bootstrap	0,9431	(0,8771 - 0,9938)	0,9658	0,0146	0,0032
JM (Piecewise)	K-Folds	0,9344	(0,8662 - 1,0000)	0,4792	0,0228	0,0033
	Bootstrap	0,9179	(0,8727 - 1,0000)	0,9584	0,0457	0,0033

Ao analisar os resultados pelos métodos k-folds e bootstrap para cada modelo, observa-se que os resultados obtidos pelas duas abordagens são consistentes, com diferenças marginais nos valores de AUC e no erro de predição (PE). No entanto, nota-se que o Bootstrap produz intervalos de confiança mais estreitos, indicando maior precisão na estimativa da variabilidade do modelo, sem comprometer a equivalência nos valores centrais entre os métodos.

Por meio da Tabela 17 é possível verificar que os testes de calibração dos modelos, avaliados pelas estatísticas de Hosmer Lemenshow, permitem identificar aqueles que apresentam aderência adequada entre as predições e os valores observados. Com base no nível de significância de 5%, nota-se que todos os modelos em pelo menos um dos dois testes de Hosmer Lemenshow, sendo k-folds ou bootstrap apresentam evidência a favor da hipótese nula de calibração dos modelos. Ademais, destaca-se que sob a ótica do teste que considera a variância do estimador de Kaplan Maier no denominador, apenas o modelo Naive 1 (p-valor 0,008 e 0,02 para k-folds e bootstrap respectivamente) não possui evidências a favor da hipótese nula, isto é, frente ao teste que possui maior robustez na análise de calibração, apenas um dos seis métodos avaliados não satisfaz a condição.

Na análise conjunta de discriminação e calibração, AUC e erro de predição (PE), os modelos JM (Spline, Bootstrap) e Naive 2 (Bootstrap) destacam-se como os mais

robustos. O modelo JM (Spline, Bootstrap) apresenta AUC elevado (0,9431), intervalo de confiança reduzido e o menor PE (0,0032), indicando alta acurácia e baixa variabilidade na predição. Da mesma forma, o Naive 2 (Bootstrap) exibe AUC ainda maior (0,9658) e um PE extremamente baixo (0,0042), reforçando sua excelente capacidade preditiva. O modelo JM (Piecewise, K-folds) também se mostra competitivo, com AUC de 0,9344, e um desempenho similar as abordagens Spline e Naive 2 em termos do erro de predição. Dessa forma, todas as abordagens discutidas neste parágrafo surgem como boas opções, combinando alta discriminação, calibração satisfatória e baixa incerteza preditiva.

6.1.4 Estudo dos Limiares de Risco - 6 Meses

Como a construção dos estimadores de cada um dos modelos explorados na seção anterior foi realizado com tamanhos de amostra distintos, para equiparar as quatro abordagens, foram geradas curvas de Kaplan-Meier baseadas em três categorias de risco: baixo risco, risco intermediário e alto risco. A definição dos limiares para essas categorias foi construída com o apoio de uma equipe técnica composta por médicos e enfermeiros, que analisaram os intervalos de probabilidade e determinaram os níveis de risco correspondentes a cada um deles.

Foram estabelecidos limiares distintos para dois grandes grupos de crianças: aquelas que apresentavam comorbidades ao nascer e aquelas sem condições pré-existentes. Essa diferenciação considerou que crianças com malformações ou outras doenças ao nascimento tendem a apresentar maior probabilidade de utilizar CTI/UTI Neonatal, enquanto aquelas sem comorbidades possuem probabilidades mais baixas. Como o risco pode se modificar ao longo do tempo, foi introduzido o limiar 2, que organiza as categorias com base nos maiores valores dos riscos estimados. Assim, espera-se que a sobrevida seja progressivamente menor entre as categorias de baixo, intermediário e alto risco, respectivamente. A Tabela 18 apresenta a definição dos limiares de risco.

Tabela 18: Limiares para definir as categorias do risco predito.

Comorbidade Fetal	Categoria de Risco	Limiar 1	Limiar 2
Possui Comorbidade	Risco Baixo	$< 2\%$	$< 5\%$
	Risco Intermediário	$\geq 2\%$ e $< 5\%$	$\geq 5\%$ e $< 10\%$
	Risco Alto	$\geq 5\%$	$\geq 10\%$
Não Possui Comorbidade	Risco Baixo	$< 0,3\%$	$< 1\%$
	Risco Intermediário	$\geq 0,3\%$ e $< 0,5\%$	$\geq 1\%$ e $< 2\%$
	Risco Alto	$\geq 0,5\%$	$\geq 2\%$

A Tabela 19 apresenta o número de eventos em CTI/UTI Neonatal e a probabilidade média das crianças utilizarem CTI/UTI Neonatal para cada limite estabelecido, considerando um total de 9.448 indivíduos com registro do marcador longitudinal na segunda onda, entre os 15.202 participantes do estudo.

Tabela 19: Classificação dos limiares, evento e risco para os diferentes modelos. Sendo o evento a internação em CTI/UTI neonatal.

Limiar	Modelo	Classificação	Nº Pacientes	Nº Eventos	Risco Médio
LIMIAR 1	Naive 1	Risco alto	2.983	197	0,232
		Risco intermediário	4.225	48	0,023
		Risco baixo	2.240	10	0,007
	Naive 2	Risco alto	1.858	177	0,191
		Risco intermediário	2.616	48	0,017
		Risco baixo	4.974	30	0,005
	LM	Risco alto	1.547	173	0,172
		Risco intermediário	1.978	48	0,016
		Risco baixo	5.923	34	0,004
	JM Weibull	Risco alto	7.406	220	0,245
		Risco intermediário	2.042	35	0,038
		Risco baixo	0	0	0,017
	JM Piecewise	Risco alto	955	166	0,174
		Risco intermediário	316	13	0,015
		Risco baixo	8.177	76	0,004
	JM Spline	Risco alto	959	168	0,201
		Risco intermediário	243	7	0,014
		Risco baixo	8.246	80	0,004
LIMIAR 2	Naive 1	Risco alto	478	143	0,302
		Risco intermediário	978	41	0,030
		Risco baixo	7.992	71	0,009
	Naive 2	Risco alto	300	119	0,248
		Risco intermediário	552	47	0,022
		Risco baixo	8.596	89	0,006
	LM	Risco alto	329	125	0,223
		Risco intermediário	453	38	0,021
		Risco baixo	8.666	92	0,006
	JM Weibull	Risco alto	974	169	0,319
		Risco intermediário	958	27	0,050
		Risco baixo	7.516	59	0,023
	JM Piecewise	Risco alto	216	94	0,226
		Risco intermediário	531	71	0,019
		Risco baixo	8.701	90	0,005
	JM Spline	Risco alto	198	92	0,262
		Risco intermediário	520	75	0,019
		Risco baixo	8.730	88	0,005

A alocação dos beneficiários no primeiro Limiar de risco variou substancialmente entre os modelos analisados. Dentre os modelos Naive 1, Naive 2 e LandMark, o modelo Naive 1 foi o que mais alocou beneficiários no maior risco (2.983 indivíduos) e apresentou o maior risco médio de utilizar o CTI/UTI Neonatal.

Entre os modelos conjuntos com diferentes especificações de linha de base, o modelo JM Piecewise e o JM Spline destacaram-se por alocar proporcionalmente menos indivíduos no maior risco (955 e 959, respectivamente), mas com menores risco médio (0,174 e 0,201). Esses resultados sugerem diferenças significativas nas abordagens metodológicas, com alguns modelos priorizando a sensibilidade (maior probabilidade

para o risco alto) e outros, a especificidade (maior número de alocação).

No Limiar 2, o modelo Naive 1 classificou 478 indivíduos no grupo de risco alto, atribuindo uma probabilidade de 30% para essa categoria de vivenciar o evento de interesse. O modelo LM, por sua vez, distribuiu 329 indivíduos no risco alto, com uma probabilidade próxima a 22%, demonstrando uma abordagem mais equilibrada entre a quantidade de beneficiários e a precisão na estimativa de risco. Entre os modelos JM, o JM Weibull alocou um número expressivo de 974 indivíduos no risco alto, mas com uma probabilidade de 31,9%, enquanto os modelos JM Piecewise e JM Spline classificaram proporcionalmente menos beneficiários no maior risco (216 e 198, respectivamente), mas atribuíram probabilidades mais baixas (22,6% e 26,2%).

As variações entre os modelos nos dois limiares construídos reforçam a importância de ajustar a escolha do modelo aos objetivos específicos do estudo, como priorizar a identificação de casos de alto risco ou otimizar recursos clínicos.

As Figuras 11 e 12 mostra as curvas de Kaplan-Meier de acordo com os grupos de risco: baixo, intermediário e alto de acordo com o Limiar 1.

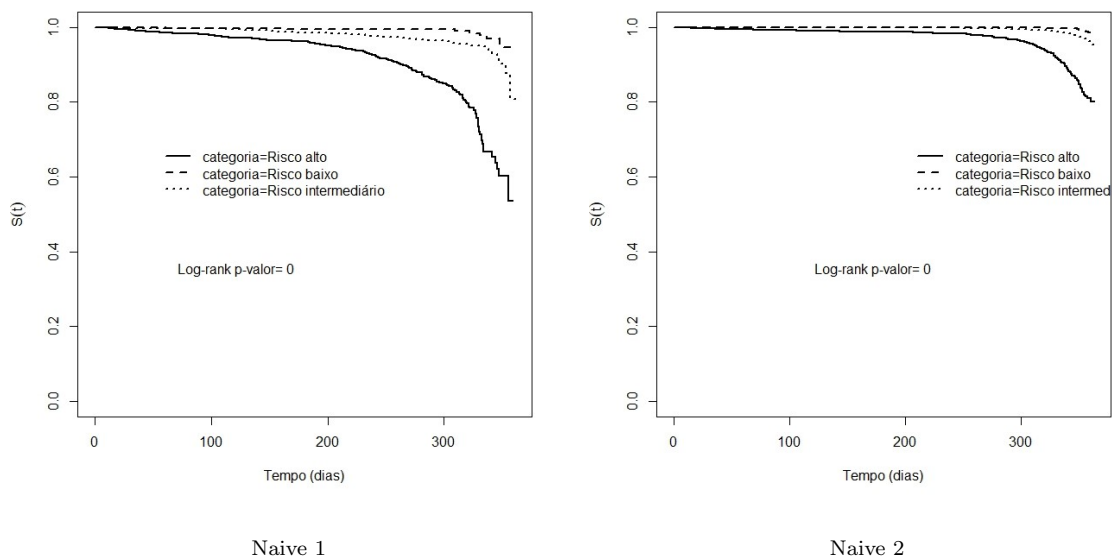


Figura 11: Curvas de Kaplan-Meier para os modelos Naive 1 e Naive 2 com o limiar 1.

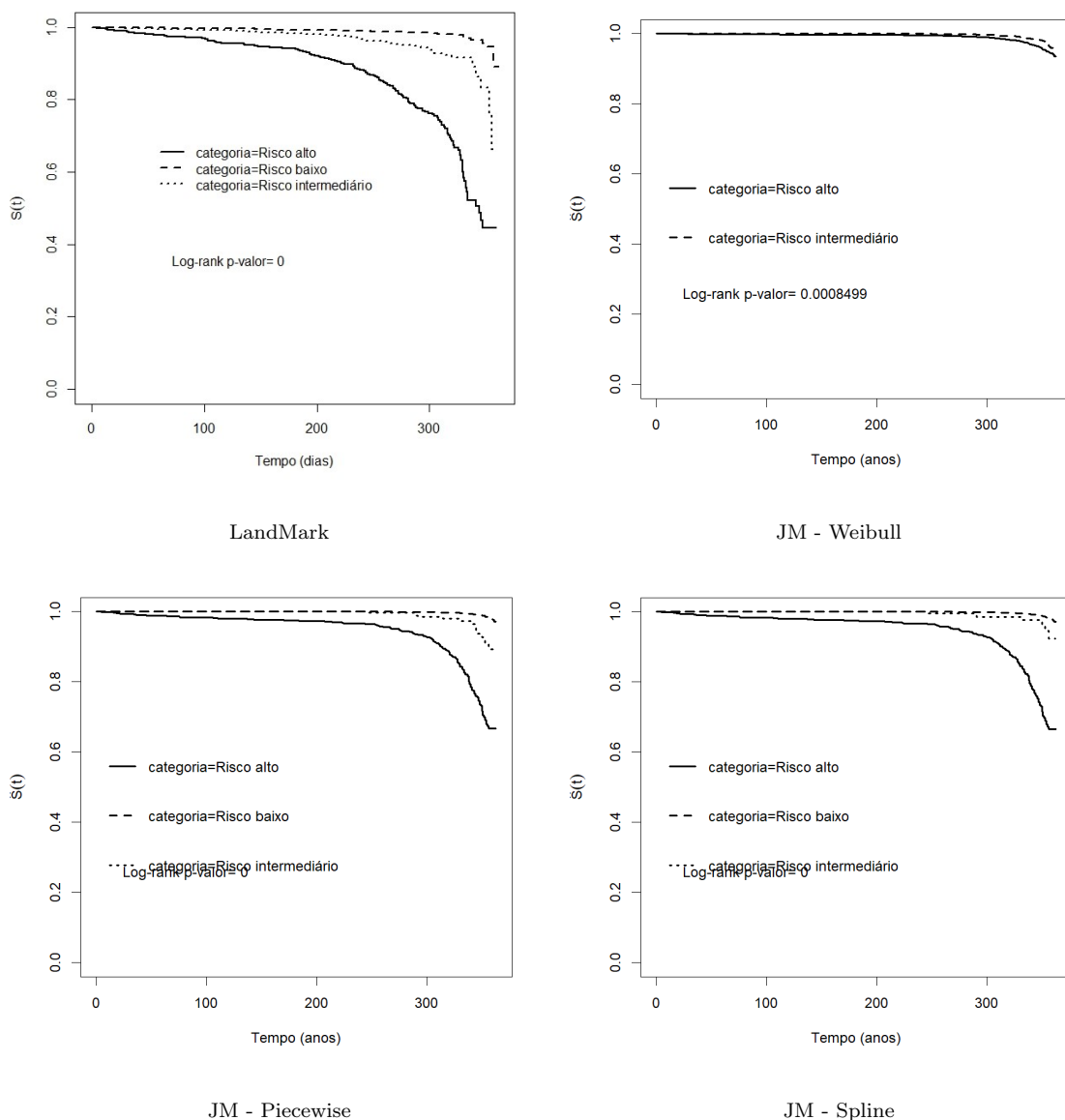


Figura 12: Curvas de Kaplan-Meier para os modelos LandMark e Modelos Conjuntos com o limiar 1.

Para as categorias definidas pelo Limiar 1, o método JM com distribuição Weibull como linha de base não classifica nenhum dos 9.448 pacientes em risco baixo. Conseqüentemente, não há curva de sobrevivência associada a essa categoria. Em todos os modelos pelo menos um indivíduo vivenciou o evento de interesse em cada uma das categorias em que houve alocação de beneficiários. Ademais, observa-se um comportamento bastante similar entre as abordagens Naive 1 e LM, de acordo com esses métodos, as curvas das categorias de risco baixo e intermediário são próximas e se distanciam da curva de risco alto (diferença significativa: Tabela 20). O mesmo é observado para as curvas dos modelos JM Piecewise e JM Spline, curvas de risco baixo e intermediários com comportamento próximos e distantes da curva de risco

alto. No método JM Weibull, verifica-se que não existe uma separação clara entre as curvas de risco alta e intermediária.

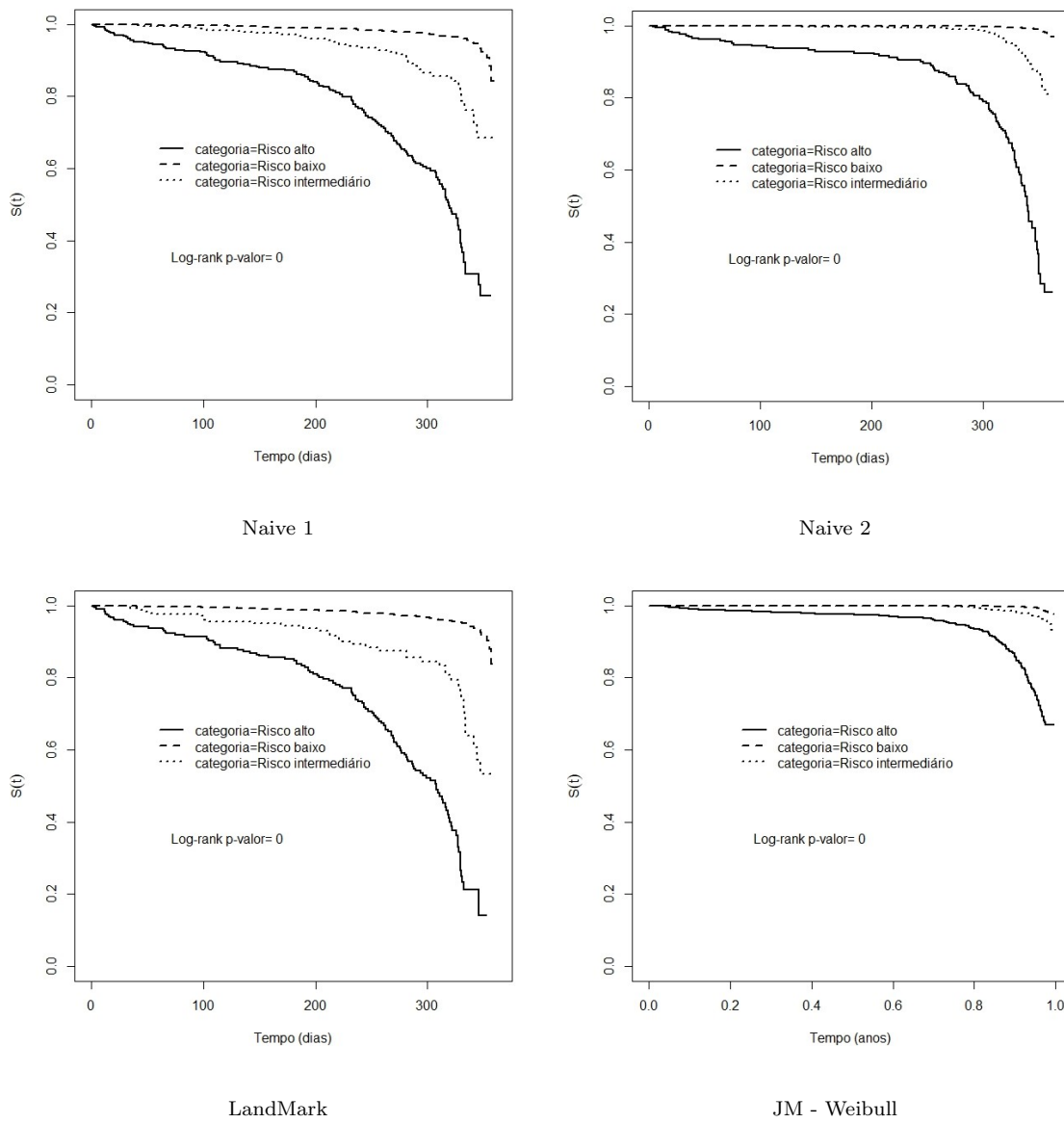


Figura 13: Curvas de Kaplan-Meier para os modelos Naive 1, Naive 2, LandMark e JM - Weibull com o limiar 2.

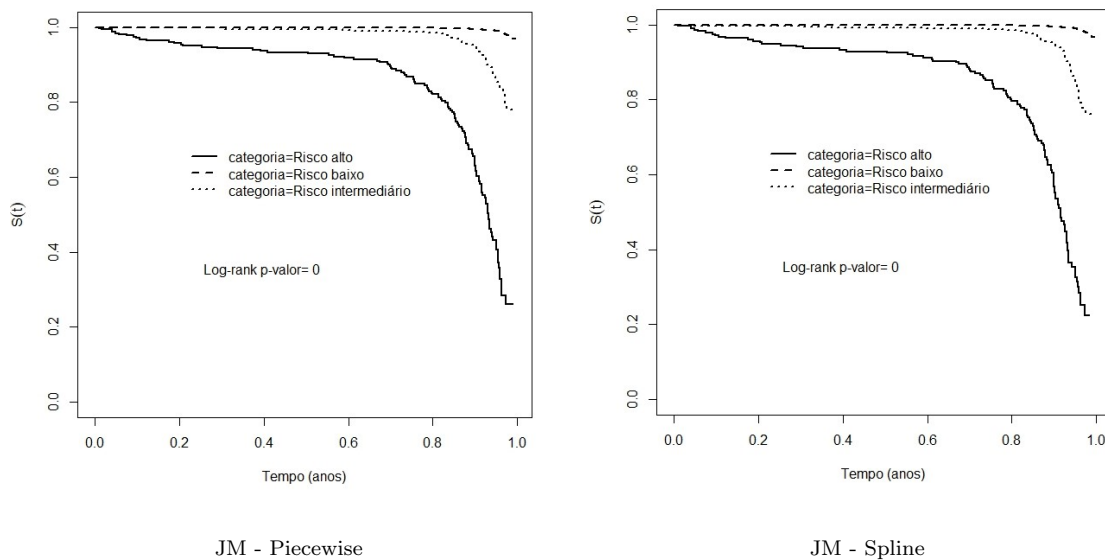


Figura 14: Curvas de Kaplan-Meier para os modelos JM - Piecewise e JM - Spline com o limiar 2.

As Figuras 13 e 14 apresentam as curvas de Kaplan-Meier para os grupos definidos pelo limiar 2. Os grupos de risco gerados, têm um padrão de sobrevida que difere do anterior, sendo que em todos os modelos inclusive o JM Weibull houve a utilização do CTI/UTI Neonatal de pelo menos um paciente em cada grupo de risco. Para os modelos Naive 1 e LandMark que apresentam curvas similares, foi observado uma separação visual das três curvas por categoria de risco. Para o Naive 2, diferente do limiar 1, a queda da curva para o risco alto é mais acentuada ao longo do tempo, porém, como observado no limiar 1 as curvas intermediárias e alta são próximas entre si. O modelo JM Weibull apresenta comportamento próximo ao modelo Naive 2, com uma curva para o risco alto menos acentuada. Os modelos JM Piecewise e Spline possuem comportamentos análogos, com diferença visual da curva de risco alto para as curvas de risco baixo e intermediário.

Para avaliar diferenças significativas nas taxas de sobrevivência entre as classificações de risco, foi aplicado o teste de Log-rank, comparando os grupos de risco definidos em cada limiar construído. Além disso, realizou-se uma análise dois a dois das curvas de sobrevivência, considerando as categorias de risco 'baixo' versus 'intermediário' e 'intermediário' versus 'alto', utilizando o método de Bonferroni para ajustar os resultados. Os testes de Bonferroni foram realizados a um nível de significância corrigido de $\alpha^* = \frac{0.05}{2} = 0,025$.

Tabela 20: Comparação dois-a-dois das curvas de sobrevivência pelo teste de log-rank, Limiar 1.

	Baixo vs. Intermediário		Intermediário vs. Alto	
	χ^2	p -valor	χ^2	p -valor
Naive 1	8,5	0,004	96,4	<0,001
Naive 2	26,6	<0,001	164	<0,001
LM	28,8	<0,001	93,7	<0,001
JM - Weibull	-	-	11,1	0,001
JM - Piecewise	39,4	<0,001	36,3	<0,001
JM - Spline	11,2	0,001	34,7	<0,001

Os resultados do teste de Bonferroni para as comparações dois a dois das curvas de sobrevivência indicaram diferenças significativas entre os grupos de risco, considerando o Limiar 1. Para a comparação entre “baixo” e “intermediário”, todos os modelos apresentaram valores de χ^2 elevados, com destaque para o *JM - Piecewise* ($\chi^2 = 39.4$) e o *JM - Spline* ($\chi^2 = 11.2$), ambos com p -valor menor que o nível corrigido ($\alpha^* = 0,025$), evidenciando uma clara distinção entre essas categorias de risco.

Nas comparações entre “intermediário” e “alto”, os resultados também foram altamente significativos para todos os modelos, com χ^2 particularmente elevados para o *Naive 2* ($\chi^2 = 164$) e o *JM - Spline* ($\chi^2 = 34.7$). Destaca-se ainda o modelo *JM - Weibull*, que apresentou um $\chi^2 = 11.1$ e $p = 0.001$, superando o nível de significância ajustado. Esses achados reforçam a capacidade dos modelos em discriminar adequadamente os níveis de risco definidos, com variações entre as metodologias quanto à magnitude das diferenças identificadas.

Tabela 21: Comparação dois-a-dois das curvas de sobrevivência pelo teste de log-rank, Limiar 2.

	Baixo vs. Intermediário		Intermediário vs. Alto	
	χ^2	p -valor	χ^2	p -valor
Naive 1	63,8	<0,001	79,7	<0,001
Naive 2	249	<0,001	179	<0,001
LM	106	<0,001	52,3	<0,001
JM - Weibull	38,8	<0,001	129	<0,001
JM - Piecewise	608	<0,001	121	<0,001
JM - Spline	693	<0,001	131	<0,001

Para as comparações das curvas de sobrevivência no Limiar 2 como mostrado na Tabela 21, considerando as categorias de risco “baixo” versus “intermediário” e “intermediário” versus “alto”, observou-se diferenças estatisticamente significativas em todas as comparações, indicando que as classificações de risco se distinguem claramente quanto às taxas de sobrevivência.

Entre os modelos avaliados, o modelo JM Spline apresentou o maior valor de χ^2 na comparação “baixo” versus “intermediário” ($\chi^2 = 693$), seguido pelos modelos JM Piecewise ($\chi^2 = 608$) e Naive 2 ($\chi^2 = 249$). Essas métricas refletem uma maior discriminação entre as categorias de risco nesses modelos.

Na comparação entre “intermediário” e “alto”, o modelo Naive 2 destacou-se com $\chi^2 = 179$, enquanto os modelos JM Spline ($\chi^2 = 131$), JM Weibull ($\chi^2 = 129$) e JM Piecewise ($\chi^2 = 121$) também demonstraram elevada capacidade de discriminação. Esses resultados reforçam que o Limiar 2 possibilita uma distinção robusta entre os diferentes grupos de risco em todos os modelos analisados, sendo particularmente evidente nos modelos JM Spline e JM Piecewise, que apresentaram os maiores valores de χ^2 para as comparações realizadas.

6.1.5 Modelo de Predição Dinâmica - 1 Ano

Nesta seção, serão exploradas as quatro abordagens dinâmicas para prever a utilização de CTI/UTI Neonatal dentro de 1 ano, dado que as crianças ainda não haviam utilizado o CTI/UTI Neonatal até o sexto mês de vida. Para as abordagens que utilizam uma nova linha de base, a análise incluiu um total de 7.237 indivíduos.

A Tabela 22 apresenta os resultados obtidos a partir da aplicação do modelo LandMark utilizando a linha de base em 182 dias. Os resultados incluem as estimativas dos parâmetros do modelo, acompanhadas de seus respectivos erros padrão,

valores z e p-valores.

Tabela 22: Estimativas, razão de taxa de falha, erro padrão e p-valor do modelo LandMark – 1 Ano (365 dias).

	Estimativa	Razão das Taxas de Falha (IC 95%)	Erro Padrão	P-Valor
Indicador de Realização do Pré-natal Completo	-0,265	0,767 (0,428 - 1,373)	0,297	0,372
Indicador de Comorbidade Hipertensiva Aguda à Gestação	0,491	1,635 (0,744 - 3,593)	0,402	0,221
Indicador de Comorbidade Metabólica Aguda à Gestação	0,032	1,032 (0,301 - 3,542)	0,629	0,960
Indicador de Comorbidade Hipertensiva Crônica	0,248	1,282 (0,527 - 3,116)	0,453	0,584
Indicador de Comorbidade Metabólica Crônica	-0,070	0,932 (0,458 - 1,899)	0,363	0,846
Número de Semanas do Parto	-0,865	0,421 (0,227 - 0,782)	0,316	0,006
Idade no Momento do Parto	-0,321	0,726 (0,409 - 1,286)	0,292	0,272
Indicador Primípara	-0,492	0,611 (0,353 - 1,058)	0,280	0,079
Tipo de Parto	-0,308	0,735 (0,401 - 1,348)	0,309	0,320
Hora do Parto	-0,277	0,758 (0,418 - 1,377)	0,304	0,363
Apgar no Primeiro Minuto	0,048	1,049 (0,846 - 1,302)	0,110	0,663
Apgar no Quinto Minuto	-0,456	0,634 (0,508 - 0,790)	0,113	<0,001
Indicador de Comorbidade Fetal	2,007	7,444 (3,479 - 15,928)	0,388	<0,001
Percentil do Peso ao Nascer	-0,042	0,959 (0,941 - 0,977)	0,010	<0,001

Os resultados do modelo LandMark para prever desfechos aos 365 dias, considerando indivíduos que sobreviveram até 182 dias, apresentam algumas semelhanças e diferenças em relação ao modelo utilizando linha de base em 28 dias. Novamente, o percentil do peso ao nascer e o indicador de comorbidade fetal continuam exercendo impactos expressivos sobre a taxa de falha. O percentil do peso ao nascer manteve o efeito inverso, com maiores percentis associados a menores taxas de falha, confirmando sua relevância protetiva ao longo do tempo. O indicador de comorbidade fetal também reforçou sua alta relevância, com um aumento ainda mais expressivo na razão de taxas de falha ($HR = 7,44$), evidenciando o impacto significativo das condições fetais adversas mesmo em um horizonte temporal mais longo.

O número de semanas do parto manteve-se como uma variável fortemente associada à redução do risco, com um efeito estatisticamente significativo consistente nos dois modelos. No entanto, o indicador de pré-natal completo, que foi significativo no modelo de 28 dias, não apresentou associação estatisticamente relevante para os 365 dias ($p = 0,37$), sugerindo que seu impacto pode ser mais importante nos primeiros meses após o parto. Essas diferenças indicam que, à medida que o horizonte de tempo se expande, o efeito de algumas variáveis pode se atenuar, enquanto outras, como o impacto de condições fetais adversas, permanecem robustas e até se tornam mais marcantes.

Para a predição do risco de utilização do CTI/UTI neonatal no primeiro ano de vida, serão empregados os estimadores apresentados na seção anterior, conforme detalhado nas Tabelas 14, 15 e 16. Esses estimadores fornecerão a base para a análise preditiva, permitindo uma avaliação precisa dos fatores associados à necessidade de suporte intensivo neonatal.

São apresentados na Tabela 23 os resultados das quatro abordagens em termos de calibração e discriminação dos modelos para a predição dinâmica até o primeiro ano de vida dos recém nascidos, dado que eles não utilizaram dos serviços de CTI/UTI Neonatal até os primeiros seis meses de vida. Os modelos foram testados repetidas vezes, a fim de verificar a consistência dos resultados em diferentes simulações, sendo a média de 100 simulações para o método dos k-folds (5 folds), e a média de 20 simulações para o método de bootstrap (200 reamostras).

Tabela 23: Resultados de validação pelas métricas de discriminação (AUC) e calibração (Teste de Hosmer Lemeshow e Erro de Predição) por modelo e método em 1 ano.

Modelo	Método	AUC	IC 2,5% - 97,5%	HL_var	HL	PE
Naive 1	K-Folds	0,8243	(0,7155 - 0,9161)	0,0538	0,8118	0,0407
	Bootstrap	0,8343	(0,7674 - 0,8969)	0,0002	0,3513	0,0407
Naive 2	K-Folds	0,8781	(0,7856 - 0,9505)	0,1713	0,9729	0,0661
	Bootstrap	0,8829	(0,8238 - 0,9354)	0,0000	0,3708	0,0661
LM	K-Folds	0,8165	(0,7303 - 0,9021)	0,0060	0,2470	0,0213
	Bootstrap	0,8390	(0,7694 - 0,8985)	0,0000	0,0150	0,0209
JM (Weibull)	K-Folds	0,8869	(0,7995 - 0,9598)	0,0000	0,1503	0,0367
	Bootstrap	0,8832	(0,8075 - 0,9589)	0,0000	0,0752	0,0371
JM (Spline)	K-Folds	0,8815	(0,7796 - 0,9603)	0,0015	0,5565	0,0629
	Bootstrap	0,8644	(0,7875 - 0,9414)	0,0000	0,9171	0,0615
JM (Piecewise)	K-Folds	0,8756	(0,7315 - 0,9592)	0,1582	0,4070	0,0658
	Bootstrap	0,8716	(0,7359 - 0,9273)	0,0000	0,9335	0,0659

Os resultados das validações das previsões dinâmicas para 365 dias revelam diferenças importantes entre os modelos em termos de calibração, discriminação e erro preditivo (PE). Dentre os modelos avaliados, o JM (Weibull) destaca-se como o mais robusto nesse horizonte temporal. Ele apresenta os maiores valores de AUC (0,8869 no K-Folds e 0,8832 no Bootstrap), indicando excelente capacidade discriminativa. Além disso, seus intervalos de confiança são relativamente estreitos, e os testes de calibração (Hosmer-Lemeshow) não mostram evidências contra a hipótese nula de aderência às predições, reforçando sua confiabilidade. Embora o modelo Naive 2 também apresente valores elevados de AUC (0,8781 no K-Folds e 0,8829 no Bootstrap), seu erro preditivo (PE) é maior em comparação ao JM (Weibull), o que reduz sua robustez.

Em comparação aos resultados observados para 182 dias, onde o JM (Spline,

Bootstrap) foi identificado como o melhor modelo, observa-se que, para 365 dias, a discriminação e a calibração permanecem relevantes na escolha do modelo ideal, mas o contexto temporal parece favorecer o desempenho do JM (Weibull). Essa diferença destaca o impacto do horizonte de predição na performance dos modelos, indicando que, enquanto o JM (Spline) é mais apropriado para predições de curto prazo, o JM (Weibull) se adapta melhor a predições de médio prazo, combinando alta discriminação, calibração satisfatória e baixa incerteza preditiva.

6.1.6 Estudo dos Limiares de Risco - 1 Ano

Da mesma forma que para o modelo de 182 dias, a fim de equiparar as quatro abordagens, foram geradas curvas de Kaplan-Meier baseadas nas três categorias de risco: baixo risco, risco intermediário e alto risco. Os limiares foram estabelecidos para dois grandes grupos de crianças: aquelas que apresentavam comorbidades ao nascer e aquelas sem condições pré-existentes. A Tabela 24 apresenta a definição dos limiares de risco.

Tabela 24: Limiares para definir as categorias do risco predito.

Comorbidade Fetal	Categoria de Risco	Limiar 1	Limiar 2
Possui Comorbidade	Risco Baixo	$< 10\%$	$< 20\%$
	Risco Intermediário	$\geq 10\%$ e $< 20\%$	$\geq 20\%$ e $< 40\%$
	Risco Alto	$\geq 20\%$	$\geq 40\%$
Não Possui Comorbidade	Risco Baixo	$< 2\%$	$< 5\%$
	Risco Intermediário	$\geq 2\%$ e $< 3\%$	$\geq 5\%$ e $< 10\%$
	Risco Alto	$\geq 3\%$	$\geq 10\%$

A Tabela 25 apresenta o número de eventos em CTI/UTI Neonatal e a probabilidade média das crianças utilizarem CTI/UTI Neonatal para cada limite estabelecido, considerando um total de 7.237 indivíduos com registro do marcador longitudinal na terceira onda.

Tabela 25: Classificação dos limiares, evento e risco para os diferentes modelos. Sendo o evento a internação em CTI/UTI neonatal.

Limiar	Modelo	Classificação	Nº Pacientes	Nº Eventos	Risco Médio
LIMIAR 1	Naive 1	Risco alto	833	32	0,370
		Risco intermediário	1.027	12	0,060
		Risco baixo	5.377	15	0,024
	Naive 2	Risco alto	3.608	46	0,490
		Risco intermediário	3.049	12	0,130
		Risco baixo	580	1	0,051
	LM	Risco alto	115	18	0,320
		Risco intermediário	246	12	0,030
		Risco baixo	6.876	29	0,009
	JM Weibull	Risco alto	675	28	0,300
		Risco intermediário	186	5	0,050
		Risco baixo	6.376	26	0,029
	JM Piecewise	Risco alto	3.247	45	0,170
		Risco intermediário	3.484	13	0,010
		Risco baixo	506	1	0,073
	JM Spline	Risco alto	2.278	44	0,200
		Risco intermediário	3.918	13	0,140
		Risco baixo	1.041	2	0,072
LIMIAR 2	Naive 1	Risco alto	118	19	0,478
		Risco intermediário	407	11	0,083
		Risco baixo	6.712	29	0,031
	Naive 2	Risco alto	540	27	0,639
		Risco intermediário	1.042	17	0,164
		Risco baixo	5.655	15	0,066
	LM	Risco alto	23	10	0,411
		Risco intermediário	68	8	0,036
		Risco baixo	7.146	41	0,011
	JM Weibull	Risco alto	50	7	0,392
		Risco intermediário	461	19	0,069
		Risco baixo	6.726	33	0,037
	JM Piecewise	Risco alto	682	28	0,226
		Risco intermediário	772	15	0,019
		Risco baixo	5.783	16	0,095
	JM Spline	Risco alto	652	27	0,608
		Risco intermediário	628	16	0,180
		Risco baixo	5.957	16	0,093

No Limiar 1, os modelos apresentaram variações expressivas na alocação de beneficiários em cada categoria de risco. O modelo Naive 2 foi o que mais alocou beneficiários na categoria de risco alto, com 3.608 indivíduos, além de atribuir uma probabilidade de vivenciar a utilização do CTI/UTI Neonatal alta (0,49) para essa classificação. Por outro lado, o modelo LM se destacou por atribuir uma probabilidade moderada ao risco alto (0,32), embora tenha classificado apenas 115 indivíduos nesta categoria. Entre os modelos JM, o JM Piecewise alocou um número elevado de beneficiários no risco alto (3247), com uma probabilidade de apenas 0,17, enquanto o JM Weibull foi mais conservador, classificando 675 indivíduos no maior risco e

atribuindo uma probabilidade superior (0,30). Esses resultados refletem abordagens metodológicas distintas, com modelos como o LM priorizando a precisão na identificação de casos graves e outros, como o Naive 1, favorecendo a sensibilidade.

O comportamento do Limiar 2 foi análogo ao observado no limiar 1, com aumentos expressivos nas probabilidades de vivenciar o evento para o modelo LM (0,41), acompanhado por uma menor alocação de beneficiários no risco alto (23 indivíduos). Já os modelos JM, como o JM Weibull, apresentaram um comportamento mais conservador, com apenas 50 indivíduos alocados no risco alto e uma probabilidade de 0,39, enquanto os modelos JM Piecewise e JM Spline mantiveram uma abordagem mais equilibrada, com 682 e 652 indivíduos alocados no risco mais alto.

Abaixo são apresentadas as curvas de Kaplan-Meier de acordo com os grupos de risco: baixo, intermediário e alto de acordo com o Limiar 1.

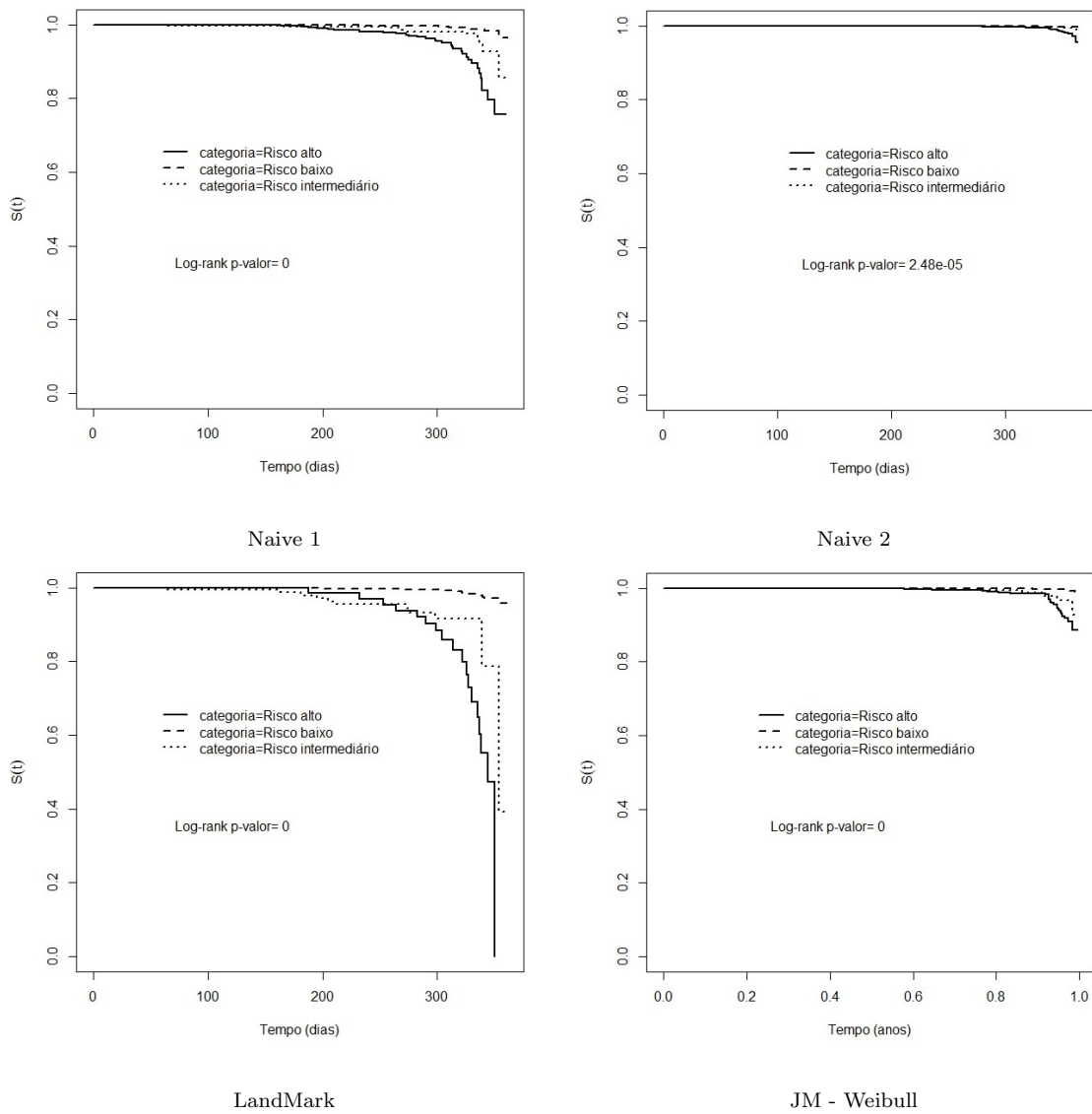


Figura 15: Curvas de Kaplan-Meier para os quatro primeiros modelos.

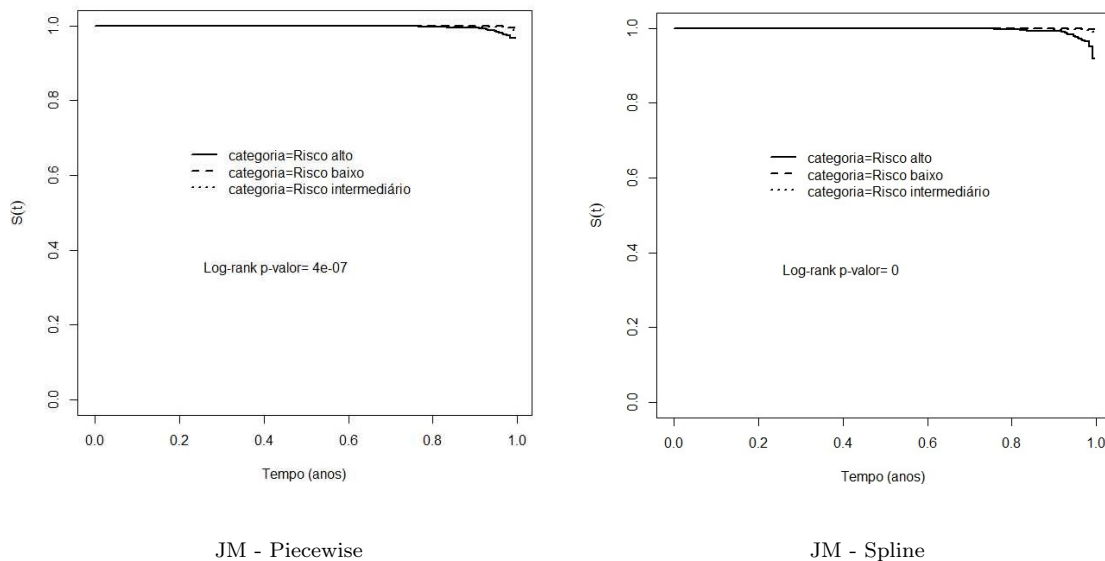


Figura 16: Curvas de Kaplan-Meier para os dois modelos restantes.

Em todos os modelos pelo menos um indivíduo viveu o evento de interesse em cada uma das categorias em que houve alocação de beneficiários. Os modelos Naive 2, JM Piecewise e Spline tiveram um comportamento bastante similar, em que de acordo com esses métodos, a curva das categorias de risco baixo e intermediário e alto são visualmente próximas, com um ligeiro distanciamento das cruvas baixo e intermediário da curva de risco alto. As curvas dos modelos Naive 1 e LandMark possuem certa similaridade entre elas, com o Landmark com maior queda na curva de sobrevida para os riscos intermediários e alto.

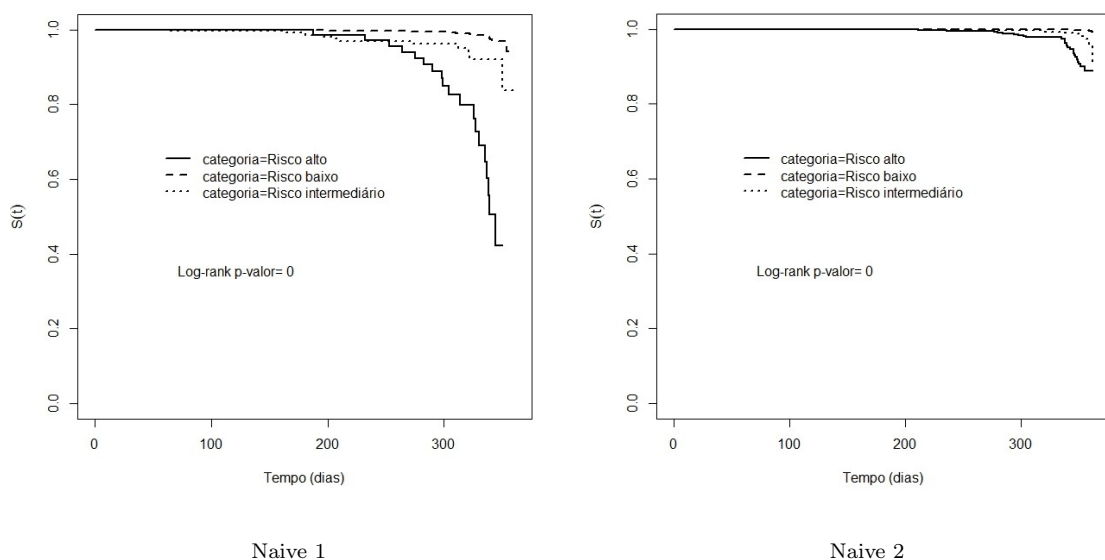


Figura 17: Curvas de Kaplan-Meier para os dois primeiros modelos.

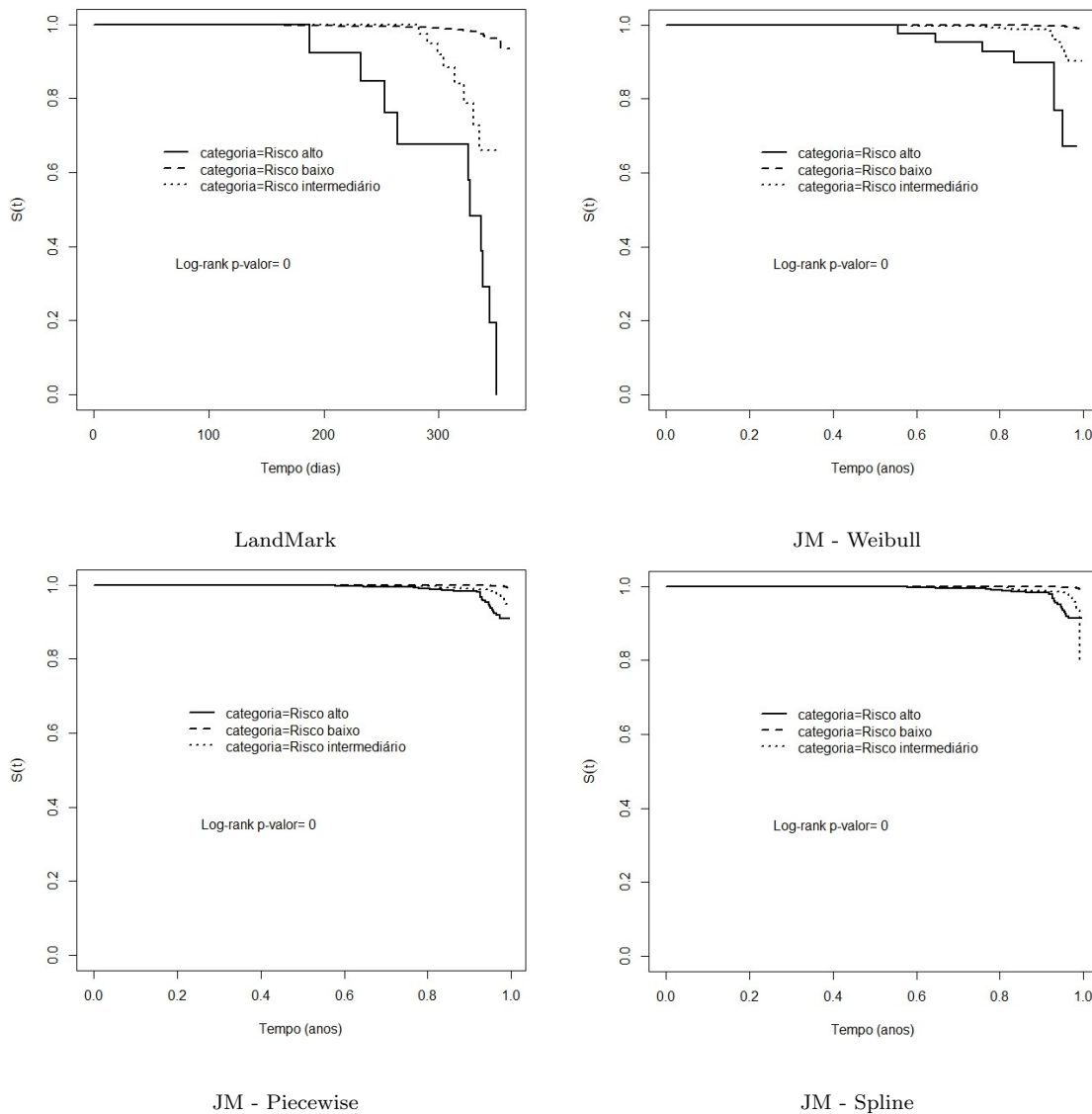


Figura 18: Curvas de Kaplan-Meier para os quatro modelos restantes.

Os modelos Naive 1 e LandMark que apresentam curvas similares, sendo possível indentificar uma separação visual das três curvas por categoria de risco. Para o Naive 2, diferente do limiar 1, a queda da curva para o risco alto é levemente mais acentuada ao longo do tempo, porém, como observado no limiar 1 as curvas intermediária e alta são próximas entre si. O modelo JM Weibull apresenta comportamento próximo ao modelo Naive 2, com uma curva para o risco alto mais acentuada. Os modelos JM Piecewise e Spline possuem comportamentos similares, com diferença visual da curva de risco alto para as curvas de risco baixo e intermediário.

Abaixo são apresentadas as análise par a par das curvas de sobrevivência, considerando as categorias de risco 'baixo' versus 'intermediário' e 'intermediário' versus

'alto', utilizando o método de Bonferroni para ajustar os resultados ($\alpha^* = \frac{0.05}{2} = 0,025$).

Tabela 26: Comparação dois-a-dois das curvas de sobrevivência pelo teste de log-rank, Limiar 1.

	Baixo vs. Intermediário		Intermediário vs. Alto	
	χ^2	p -valor	χ^2	p -valor
Naive 1	10,2	0,001	9,9	0,002
Naive 2	0,6	0,400	16,8	<0,001
LM	55,8	<0,001	6,4	0,010
JM - Weibull	22,4	<0,001	1,9	0,200
JM - Piecewise	0,4	0,500	24,8	<0,001
JM - Spline	0,6	0,400	49,5	<0,001

Com base nos resultados apresentados na Tabela 26, observa-se que o teste de Log-rank, ajustado pelo método de Bonferroni para comparações múltiplas, revelou diferenças estatisticamente significativas entre algumas categorias de risco no Limiar 1. No modelo Naïve 1, foram identificadas diferenças significativas tanto entre os grupos de risco baixo e intermediário ($p = 0.001$) quanto entre os grupos intermediário e alto ($p = 0.002$). No modelo Naïve 2, a diferença entre os grupos intermediário e alto foi altamente significativa ($p = 0.000$), mas não houve evidência de diferença entre os grupos baixo e intermediário ($p = 0.400$). O modelo LandMark apresentou diferenças significativas entre as categorias baixo e intermediário ($p = 0.000$) e intermediário e alto ($p = 0.010$).

Para os modelos conjuntos (JM), os resultados variaram. No JM Weibull, foi encontrada diferença significativa apenas entre os grupos baixo e intermediário ($p = 0.000$). Já nos modelos JM Piecewise e JM Spline, as diferenças significativas foram observadas apenas entre os grupos intermediário e alto ($p = 0.000$). Esses resultados apontam para padrões de separação entre as categorias de risco que variam conforme o modelo adotado, evidenciando uma discriminação mais consistente para o modelo LandMark.

Tabela 27: Comparação dois-a-dois das curvas de sobrevivência pelo teste de log-rank, Limiar 2.

	Baixo vs. Intermediário		Intermediário vs. Alto	
	χ^2	p -valor	χ^2	p -valor
Naive 1	26,6	<0,001	18,0	<0,001
Naive 2	34,9	<0,001	24,3	0,000
LM	52,9	<0,001	6,7	0,010
JM - Weibull	113	<0,001	16,5	<0,001
JM - Piecewise	44,4	<0,001	9,1	0,003
JM - Spline	66,9	<0,001	5	0,030

No Limiar 2, os resultados do teste de Log-rank ajustado pelo método de Bonferroni, conforme mostra a Tabela 27, mostram diferenças estatisticamente significativas entre todas as comparações realizadas, exceto para o intermediário vs. alto para o JM - Spline. Em comparação com o Limiar 1, o Limiar 2 apresenta uma discriminação geral mais consistente e acentuada entre todas as categorias de risco, sugerindo que este limiar pode oferecer maior sensibilidade na separação de grupos de risco, especialmente nos modelos conjuntos.

6.2 Modelos de Predição - Coorte Prospectiva SamiTrop

Neste capítulo a abordagem estática e dinâmicas apresentadas nas Seções 2 e 3 serão aplicadas ao conjunto de dados SaMi-Trop. O objetivo é prever o risco de morte dos pacientes com cardiopatia chagásica.

6.2.1 Modelo de Cox na Linha de Base

A Tabela 28, apresenta os resultados do modelo de Cox, sendo eles: as estimativas dos coeficientes, a razão das taxas de falha, o erro padrão e o p-valor associado aos testes de significância de cada um dos marcadores.

Tabela 28: Estimativas, razão das taxas de falha, erro padrão e p-valor do modelo de Cox

	Estimativa	Razão das Taxas de Falha (IC 95%)	Erro Padrão	P-Valor
IC	1,760	5,813 (3,883 - 8,702)	0,205	<0.001
ECG	0,528	1,696 (1,114 - 2,581)	0,214	0,013
CLASSE FUNCIONAL	0,468	1,597 (1,068 - 2,389)	0,205	0,022
IDADE	0,036	1,037 (1,022 - 1,053)	0,007	<0.001
FREQUÊNCIA CARDÍACA	0,016	1,016 (1,005 - 1,028)	0,005	0,004

Como podemos observar através da Tabela 28, todos os marcadores são significantes para estimar o risco de óbito, ou seja, aumentam o risco de óbito em algum grau. A Insuficiência Cardíaca (IC) apresenta uma estimativa de 1,76, com uma razão de taxas de falha (HR) de 5,809 (IC 95%: 3,881 - 8,695), o que indica que pacientes com IC têm um risco aproximadamente 5,8 vezes de vir a óbito em comparação com aqueles sem IC. Este resultado, altamente significativo ($p < 0,001$), demonstra que a insuficiência cardíaca é um preditor crítico no contexto da cardiopatia chagásica, confirmando sua forte relação com a mortalidade.

O ECG, que avalia alterações eletrocardiográficas, também está associado a um risco aumentado para o evento de interesse. Embora a magnitude do efeito seja menor do que a da IC, seu valor p significativo ($p = 0,013$) indica que a alteração no ECG é um preditor relevante de mortalidade. Já a classe funcional, medida pela classificação NYHA, também apresentou uma estimativa significativa. Esse resultado indica que indivíduos com pior classe funcional têm aproximadamente 1,6 vezes o risco de apresentar o evento adverso, óbito.

Por fim, é possível identificar através da Tabela 28 que a idade e a frequência cardíaca dos pacientes mostraram-se preditores importante com razões de taxas de falha de 1,037 e 1,016, respectivamente. Cada aumento na idade e na frequência cardíaca está associado a um risco 3,7% e 1,6% maior da ocorrência do óbito.

As estimativas de sobrevivência da linha de base para o tempo de 2 e 4 anos são $\hat{S}_0(2) = 0,978$ e $\hat{S}_0(4) = 0,973$, respectivamente.

Antes de avaliar o desempenho preditivo dos modelos através da discriminação e calibração, foram plotados os gráficos dos resíduos de Schoenfeld versus tempo para os marcadores. Através destes gráficos é possível constatar que as variáveis apresentam um bom ajuste, não apresentando nenhuma tendência acentuada ao longo do tempo, conforme evidencia as análises gráficas apresnetadas na Figura 19.

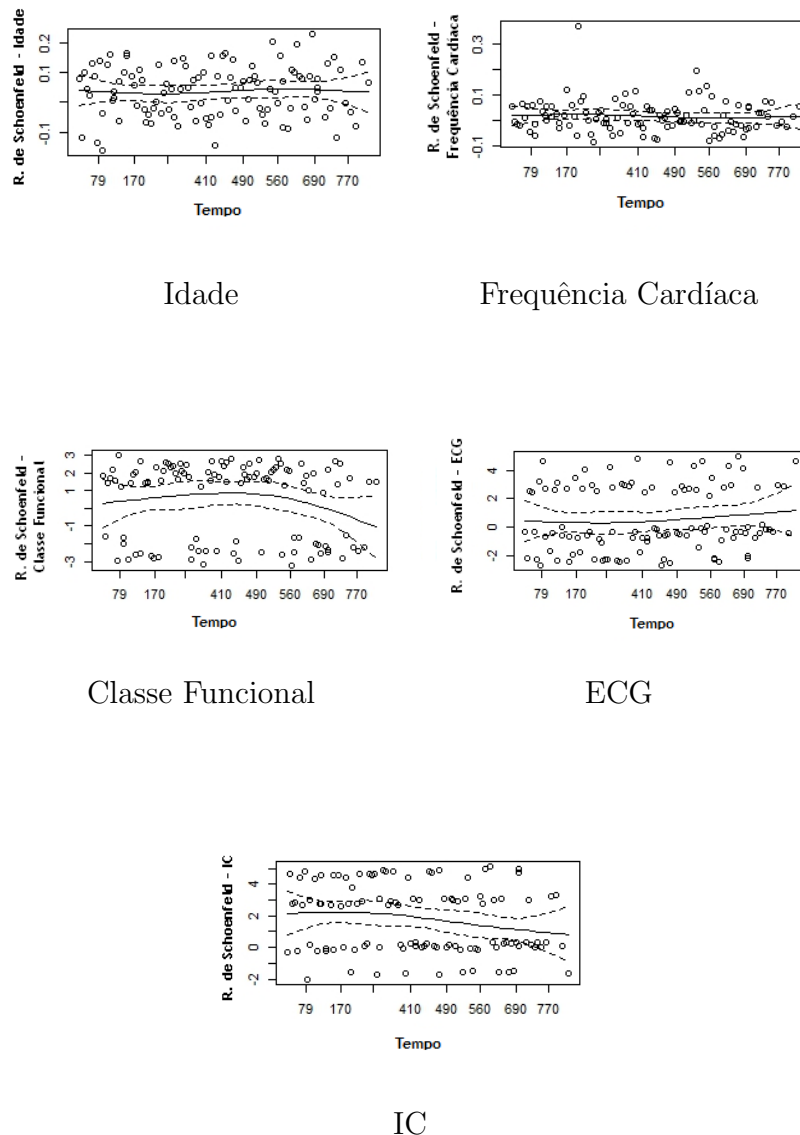


Figura 19: Resíduos de Schoenfeld versus tempo para os marcadores.

A análise dos resíduos de Schoenfeld em relação ao tempo para a variável 'Frequência Cardíaca' demonstrou uma boa aderência ao pressuposto de proporcionalidade dos riscos, conforme evidenciado pela maioria dos pontos próximos à linha de referência.

No entanto, foi identificado um ponto extremo correspondente a um valor de 174 bpm em um paciente masculino de 53 anos. Embora esse ponto se destaque visualmente como um outlier, o comportamento geral dos resíduos sugere que este caso isolado não compromete substancialmente o modelo ajustado. Esse tipo de ponto extremo pode ocorrer devido a variações biológicas individuais ou a possíveis particularidades clínicas do paciente, o que justifica uma investigação mais aprofundada para verificar sua influência no modelo de forma isolada.

Como podemos observar pela Tabela 29, o modelo apresentou poder de discriminação satisfatório ($AUC \geq 0,8$), ou seja, o modelo consegue distinguir os indivíduos que irão a óbito até 2 anos dos que não irão de forma consideravelmente bem. Além disso, os testes de Hosmer-Lemeshow indicam evidências a favor da hipótese nula de calibração dos modelos, sugerindo que o modelo está bem ajustado.

	AUC (95%)	Teste HL_{ND}	Teste HL_{ND-VAR}	Teste HL	\widehat{PE}
K-fold	80,21% (73,17% - 84,23%)	0,914	0,133	0,966	0,054
Boots	80,69% (79,76% - 81,39%)	0,977	0,421	0,874	0,053

Tabela 29: Resultados de validação pelas métricas de discriminação (AUC) e calibração (Teste de Hosmer Lemeshow e Erro de Predição) do modelo de linha de base.

A fim de verificar a consistência dos modelos, tanto o método k-fold (5 folds) e bootstrap (200 reamostras) foram processados por 10 repetições seguidas. O objetivo foi observar se as métricas de validação oscilam ou se permanecem consistentes ao longo dos 10 experimentos. Isso permite comprovar que os modelos são robustos e que os resultados obtidos não são meramente fruto da combinação de folds ou reamostras geradas. Os resultados são apresentados na Tabela 30:

Método	Rep	AUC(95%)	pvalueHL _{DN}	pvalueHL _{DN-VAR}	pvalueHL	\widehat{PE}
K-fold	1	80,14% (71,91% - 85,84%)	0,941	0,195	0,916	0,054
	2	80,57% (78,17% - 83,79%)	0,918	0,313	0,932	0,054
	3	80,21% (76,80% - 84,06%)	0,893	0,215	0,811	0,055
	4	80,64% (73,51% - 86,34%)	0,950	0,154	0,875	0,053
	5	80,35% (77,06% - 83,76%)	0,879	0,226	0,893	0,054
	6	80,82% (75,63% - 85,71%)	0,840	0,168	0,967	0,054
	7	80,98% (72,76% - 86,49%)	0,966	0,157	0,909	0,054
	8	80,29% (73,80% - 85,10%)	0,980	0,098	0,939	0,054
	9	80,11% (67,32% - 90,73%)	0,990	0,168	0,776	0,054
	10	80,60% (77,01% - 84,93%)	0,904	0,265	0,815	0,054
Boots	1	80,75% (79,75% - 81,42%)	0,980	0,456	0,895	0,053
	2	80,71% (79,70% - 81,42%)	0,974	0,456	0,864	0,053
	3	80,72% (79,94% - 81,36%)	0,990	0,419	0,881	0,054
	4	80,77% (79,78% - 81,42%)	0,984	0,468	0,887	0,053
	5	80,77% (79,76% - 81,43%)	0,986	0,439	0,885	0,053
	6	80,79% (79,98% - 81,46%)	0,992	0,409	0,880	0,053
	7	80,78% (79,97% - 81,39%)	0,995	0,440	0,877	0,053
	8	80,72% (79,59% - 81,40%)	0,983	0,439	0,888	0,053
	9	80,74% (79,85% - 81,40%)	0,983	0,426	0,880	0,053
	10	80,73% (79,73% - 81,44%)	0,979	0,434	0,860	0,053

Tabela 30: Resultados de validação pelas métricas de discriminação (AUC) e calibração (Teste de Hosmer Lemenshow e Erro de Predição) por modelo e método em 10 repetições, utilizando 5 folds e 200 reamostras bootstrap

Por meio dos resultados, temos evidência para afirmar que o modelo construído para o risco de óbito de pacientes chagásicos em 2 anos é capaz de discriminar evento e não evento. Além disso, o modelo retorna predições com boa acurácia.

A Figura 20 apresenta a curva ROC para o modelo de predição de óbito em até 2 anos, evidenciando uma área sob a curva (AUC) de 0,812. Esse valor indica uma excelente capacidade discriminativa do modelo. Já a Figura 21 mostra o gráfico de calibração, comparando os riscos esperados com os riscos observados em diferentes faixas de probabilidade. Os pontos estão distribuídos próximos à linha de 45 graus, sugerindo boa calibração do modelo.

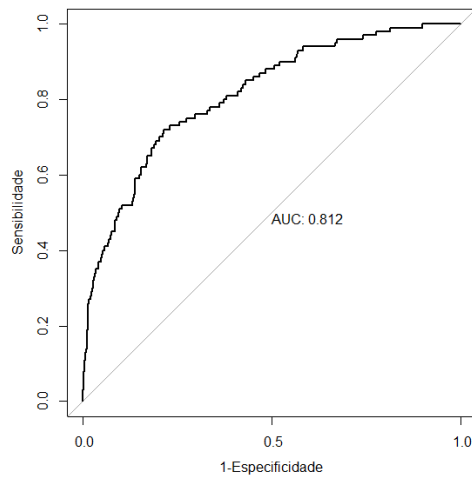


Figura 20: Curva ROC.

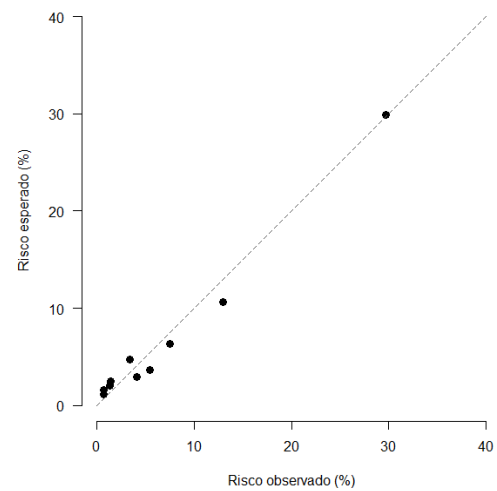


Figura 21: Risiko esperados vs. risco observados.

6.2.2 Modelo de Predição Dinâmica

Os resultados dos quatro métodos de predição dinâmicos são apresentados nesta seção. Os métodos que utilizaram da nova linha de base contaram com 283 indivíduos, enquanto o modelo conjunto contou com 2978 observações.

A Tabela 31 apresenta o ajuste utilizando a linha de base em 2 anos. É possível constatar, que as variáveis Classe Funcional e Frequência Cardíaca não são significativas para explicar o desfecho, diferente do modelo de Cox na linha de base.

Tabela 31: Estimativas, razão das taxas de falha, erro padrão e p-valor do modelo de Cox em uma nova linha de base (Landmark)

	Estimativa	Razão das Taxas de Falha (IC 95%)	Erro Padrão	P-Valor
IC	0,839	2,314 (1,473 - 3,635)	0,233	0,000
ECG	0,695	2,003 (1,272 - 3,152)	0,233	0,002
Classe Funcional	0,264	1,302 (0,643 - 2,638)	0,360	0,464
Idade	0,020	1,020 (1,003 - 1,037)	0,009	0,023
Frequência Cardíaca	0,002	1,002 (0,989 - 1,015)	0,007	0,795

Os resultados do modelo de Cox demonstram que parte das variáveis têm um impacto significativo no aumento do risco de ocorrência do desfecho estudado. A variável IC aumenta o risco em 131% ($R\hat{T}F = 2,314$), indicando que a presença dessa condição mais que duplica o risco de óbito. Similarmente, ECG está associado a um aumento de 100,1% no risco ($R\hat{T}F = 2,003$), sugerindo que esta condição dobra o risco. Por outro lado, a Idade também contribui para um aumento do

risco, embora de forma mais moderada, com um incremento de 0,2% para cada ano adicional ($R\hat{T}F = 1,002$). As variáveis Classe Funcional e Frequência Cardíaca não apresentaram associações estatisticamente significativas, o que implica que, dentro do intervalo de confiança, seu impacto no risco pode ser nulo.

No submodelo de sobrevivência, foram atribuídas a taxa de falha na linha de base três formas, sendo elas: Weibull, Piecewise e Spline. Através dos resultados apresentados nas Tabelas 32, 33 e 34, podemos constatar que para as três linhas de base a maior parte das variáveis medidas na segunda coleta são significativas para explicar o tempo até o óbito, exceto a frequência cardíaca.

Tabela 32: Estimativas, $\hat{H}R$ e p-valor do submodelo de sobrevivência com especificação da distribuição de Weibull na linha de base.

	Estimativa	$\hat{H}R$	P-Valor
IC	1,474	0,159	0,000
ECG	0,460	0,159	0,003
Classe Funcional	0,891	0,171	0,003
Idade	0,029	0,006	0,000
Frequência Cardíaca	0,008	0,007	0,223

Com base nos resultados do submodelo de sobrevivência com especificação Weibull, a insuficiência cardíaca (IC) apresentou forte associação com o risco de morte, indicando risco substancialmente maior para pacientes com IC. Alterações no eletrocardiograma e a classe funcional também foram fatores preditivos significativos de mortalidade.

A idade mostrou-se positivamente associada ao risco de morte, sugerindo que o risco aumenta com o avanço da idade. Já a frequência cardíaca não apresentou associação significativa com a mortalidade neste modelo.

Tabela 33: Estimativas, $\hat{H}R$ e p-valor do submodelo de sobrevivência com especificação da distribuição Piecewise na linha de base (5 nós).

	Estimativa	$\hat{H}R$	P-Valor
IC	1,483	0,160	0,000
ECG	0,460	0,158	0,003
Classe Funcional	0,900	0,171	0,005
Idade	0,029	0,005	0,000
Frequência Cardíaca	0,008	0,007	0,236

Com relação ao submodelo de sobrevivência na forma Piecewise, a insuficiência cardíaca (IC) manteve-se significativa. Além disso, como no modelo anterior, o ECG e a classe funcional também demonstraram associações significativas, reforçando sua relevância como preditores.

A idade permanece como um fator significativo e com o mesmo sentido apresentado no modelo anterior. E a frequência cardíaca permanece não apresentando associação estatisticamente significativa com o desfecho.

Por fim, é apresentado abaixo o submodelo de sobrevivência com especificação spline, que de forma geral apresenta o mesmo comportamento das demais linhas de base em relação a significância das variáveis, com ligeiras diferenças nos resultados dos estimadores.

Tabela 34: Estimativas, \hat{HR} e p-valor do submodelo de sobrevivência com especificação da distribuição Spline na linha de base (5 nós)

	Estimativa	\hat{HR}	P-Valor
IC	1,504	0,160	0,000
ECG	0,456	0,158	0,004
Classe Funcional	0,9149	0,171	0,005
Idade	0,027	0,005	0,000
Frequência Cardíaca (2)	0,007	0,007	0,326

Os resultados relacionados à discriminação e calibração são apresentados na Tabela 35. As validações foram baseadas em dois métodos: pelo método dos k-folds considerando 5 folds, e por meio do bootstrap, considerando 200 reamostras. Os modelos foram testados repetidas vezes, a fim de verificar a consistência dos resultados em repetidas simulações. Desta forma, os resultados apresentados são a média de 100 simulações de cada um dos experimentos considerando o método k-folds, e a média de 20 simulações considerando o método bootstrap.

Tabela 35: Resultados de validação pelas métricas de discriminação (AUC) e calibração (Teste de Hosmer Lemeshow e Erro de Predição) por modelo e método.

Modelo	Método	AUC	IC 2,5% - 97,5%	HL_var	HL	PE
Naive 1	K-Folds	0,6905	(0,5485 - 0,8181)	0,6227	0,8765	0,0952
	Bootstrap	0,6878	(0,5847 - 0,7828)	0,0877	0,3147	0,0952
Naive 2	K-Folds	0,6499	(0,4874 - 0,7999)	0,2527	0,2031	0,0740
	Bootstrap	0,6479	(0,5409 - 0,7436)	0,0411	0,0033	0,0739
LM	K-Folds	0,6808	(0,5354 - 0,8154)	0,5706	0,8642	0,0854
	Bootstrap	0,7062	(0,6051 - 0,8095)	0,3919	0,7098	0,0831
JM (Weibull)	K-Folds	0,6782	(0,4144 - 0,9314)	0,8131	0,8866	0,0865
	Bootstrap	0,6368	(0,4745 - 0,7991)	0,0456	0,5227	0,0814
JM (Spline)	K-Folds	0,6705	(0,3977 - 0,9285)	0,6500	0,7955	0,0738
	Bootstrap	0,6459	(0,4879 - 0,8038)	0,7141	0,8571	0,0754
JM (Piecewise)	K-Folds	0,6704	(0,4092 - 0,9173)	0,5780	0,8307	0,0750
	Bootstrap	0,6441	(0,4845 - 0,8037)	0,5327	0,9069	0,0734

Atrvés dos resultados apresentados tem-se que os valores de AUC foram semelhantes entre os dois métodos de validação (K-folds e Bootstrap), indicando que ambos são consistentes na avaliação da performance preditiva. Contudo, o método Bootstrap apresentou intervalos de confiança mais estreitos, o que sugere uma maior precisão na estimativa dos parâmetros.

Os resultados indicam que as metodologias Naive 1, LandMark e JM Weibull apresentam maior similaridade em termos do erro de predição. Por outro lado, os métodos Naive 2, JM Piecewise e JM Spline também exibem um padrão semelhante de erro predito, sugerindo que esses grupos de modelos compartilham características em sua capacidade de calibração.

Entre os modelos avaliados, os métodos Naive1 e LM apresentam maiores valores de AUC, indicando a melhor discriminação na predição do desfecho. No entanto, estes modelos não apresentaram os menores erros de predição. Em relação aos testes de Hosmer-Lemeshow (HL), apenas o Naive 2 e JM (Weibull) rejeitaram a hipótese de modelo calibrado.

De forma geral, os métodos Naive 1 e LM têm melhor discriminação e os métodos Naive2 e JM melhor calibração.

6.2.3 Estudo dos Limiares de Risco

Como a construção dos estimadores de cada um dos modelos explorados na seção anterior foi realizado com tamanhos de amostra distintos, para equiparar as quatro abordagens, foram geradas curvas de Kaplan-Meier baseadas em três categorias de risco: baixo risco, risco intermediário e alto risco, como definido na Tabela 36.

As definições dos limiares foram construídas com o apoio de uma equipe técnica especializada, que analisou os intervalos de probabilidade e determinaram os níveis de risco correspondentes a cada um deles. Como o comportamento do risco pode ser alterado com o passar do tempo, foi considerado o Limiar 2, que se baseia em definir as categorias de acordo com os valores maiores para as probabilidades estimadas. Dessa forma, espera-se que indivíduos classificados como de alto risco apresentem maiores probabilidades de óbitos, enquanto aqueles com risco baixo apresentem menores probabilidades, refletindo uma tendência decrescente de sobrevivência das categorias de baixo para alto risco.

Tabela 36: Limiares para definir as categorias do risco predito.

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

A Tabela 37 apresenta o número de óbitos, número de pacientes, assim como a probabilidade de morte para cada limiar definido, considerando um total de 283 indivíduos que sobreviveram a primeira onda e que tiveram registro do marcador longitudinal na segunda.

Tabela 37: Classificação de risco, óbitos e probabilidade de morte para os diferentes modelos.

Limiar	Modelo	Classificação	Nº Pacientes	Nº Óbitos	Risco Médio	
LIMIAR 1	Naive 1	Risco alto	121	52	0,270	
		Risco intermediário	150	27	0,050	
		Risco baixo	12	0	0,010	
	Naive 2	Risco alto	16	8	0,140	
		Risco intermediário	102	41	0,050	
		Risco baixo	165	30	0,010	
	LM	Risco alto	105	49	0,160	
		Risco intermediário	178	30	0,060	
		Risco baixo	0	0	-	
	JM Weibull	Risco alto	142	59	0,250	
		Risco intermediário	141	20	0,070	
		Risco baixo	0	0	-	
	JM Piecewise	Risco alto	108	46	0,200	
		Risco intermediário	173	33	0,050	
		Risco baixo	2	0	0,020	
	JM Spline	Risco alto	100	42	0,180	
		Risco intermediário	171	37	0,050	
		Risco baixo	12	0	0,020	
	LIMIAR 2	Naive 1	Risco alto	74	34	0,350
			Risco intermediário	132	34	0,090
			Risco baixo	77	11	0,030
Naive 2		Risco alto	2	1	0,230	
		Risco intermediário	57	24	0,080	
		Risco baixo	224	54	0,020	
LM		Risco alto	20	11	0,240	
		Risco intermediário	198	60	0,090	
		Risco baixo	65	8	0,040	
JM Weibull		Risco alto	82	34	0,320	
		Risco intermediário	171	40	0,090	
		Risco baixo	30	5	0,030	
JM Piecewise		Risco alto	42	18	0,280	
		Risco intermediário	146	47	0,100	
		Risco baixo	95	14	0,030	
JM Spline		Risco alto	31	13	0,260	
		Risco intermediário	128	47	0,100	
		Risco baixo	124	19	0,030	

De acordo com a Tabela 37, é possível observar que no Limiar 1, o modelo 'JM Weibull' aloca o maior número de óbitos na categoria de alto risco, seguido pelo modelo LM, indicando maior capacidade de identificar indivíduos com alta probabilidade de óbito. Já no Limiar 2, o modelo LM concentra 60 óbitos na categoria de risco intermediário, mas apresenta o menor número de beneficiários no risco baixo, sugerindo maior restrição em suas classificações. No geral, os modelos conjunto ('JM Weibull' e 'JM Piecewise') destacam-se por distribuir beneficiários de forma mais equilibrada entre os limiares, mantendo a coerência nas probabilidades atribuídas.

Abaixo são apresentadas as curvas de Kaplan-Meier de acordo com os grupos de

risco baixo, intermediário e alto de acordo com o Limiar 1.

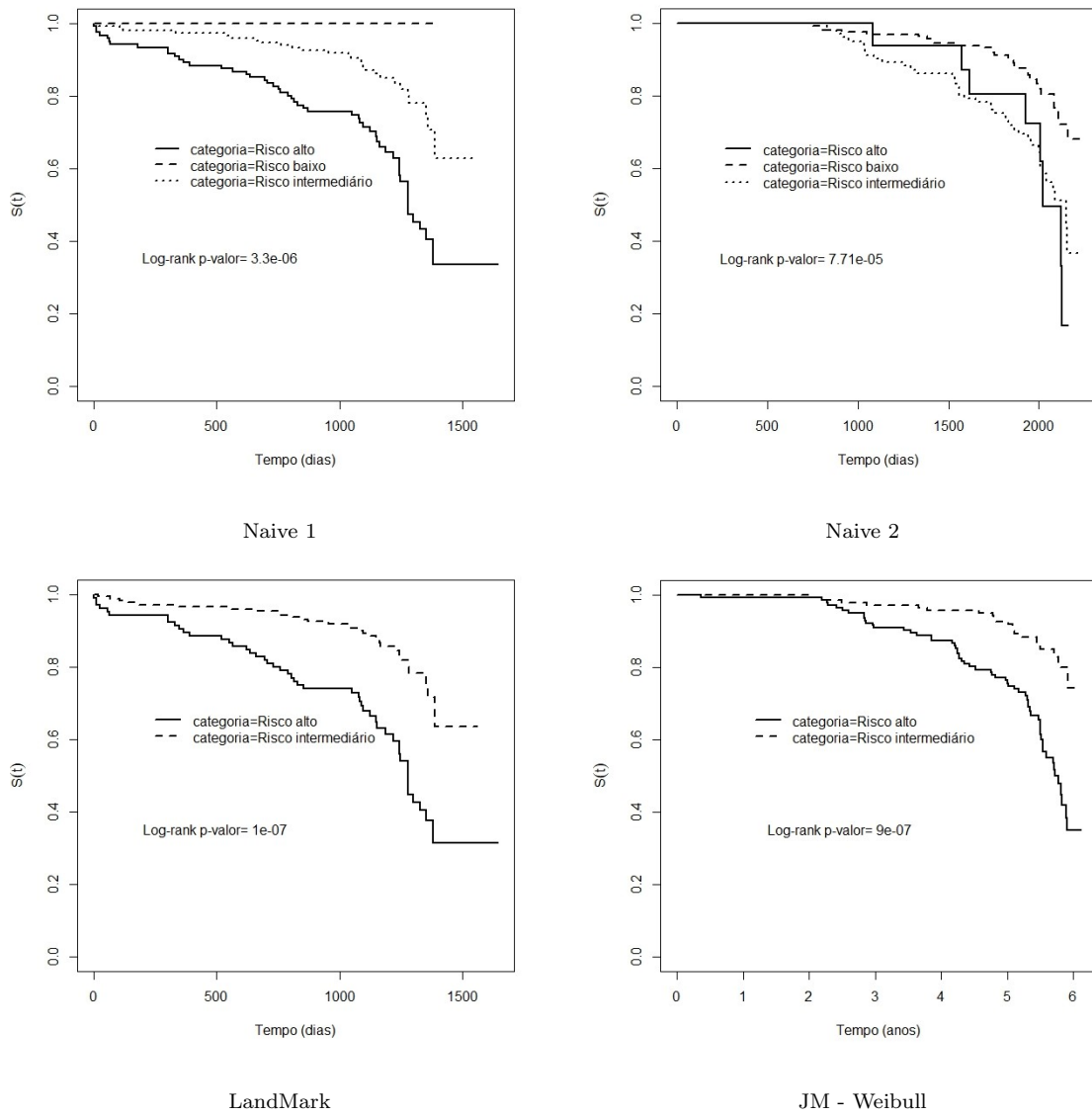


Figura 22: Curvas de Kaplan-Meier segundo os grupos de risco (baixo, intermediário e alto) conforme o limiar 1 – abordagens: Naive 1, Naive 2, LandMark e JM - Weibull.

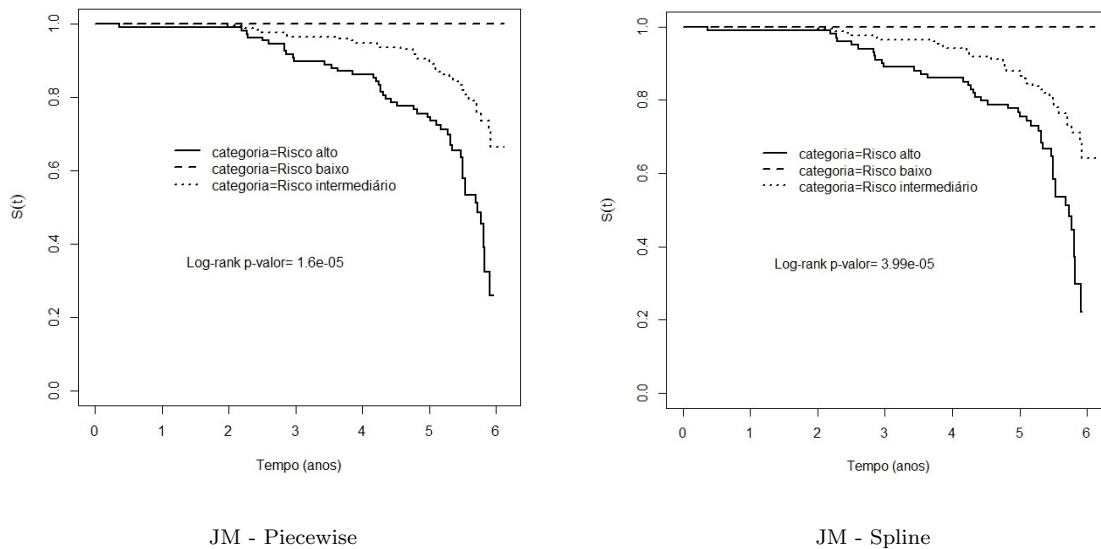
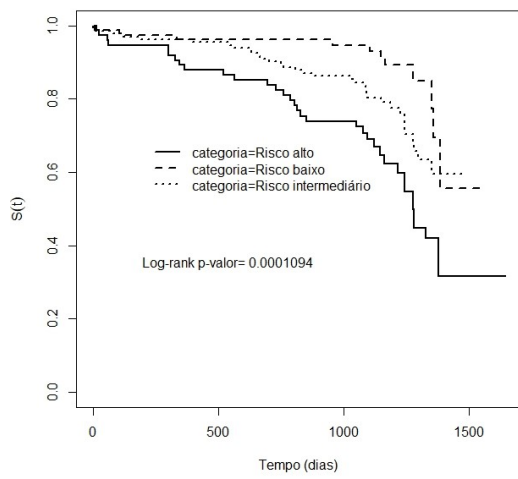


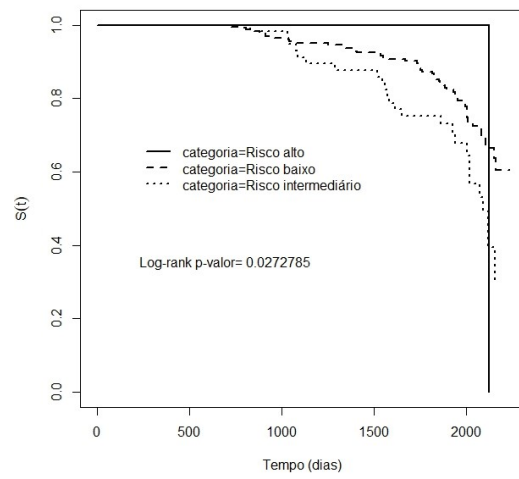
Figura 23: Curvas de Kaplan-Meier segundo os grupos de risco (baixo, intermediário e alto) conforme o limiar 1 – abordagens: JM - Piecewise e JM - Spline.

Para as categorias obtidas com o Limiar 1 (Figuras 22 e 23), o método LM e JM usando como linha de base a Weibull não atribui, para nenhum dos 283 pacientes, estimativas do risco menores que 2%, por isso não há curva de sobrevivência para essa categoria. Observa-se um comportamento bastante similar entre as abordagens Naive 1 e JM (Piecewise e Spline), de acordo com esses métodos, nenhum paciente presente na categoria de risco baixo sofreu o evento, além disso, observa-se que há uma separação das curvas de risco intermediário e risco alto (diferença significativa: Tabela 38). No método Naive 2, verifica-se que não existe uma separação clara entre as curvas de risco alta e intermediária, diferente da curva de risco baixa, que se distancia significativamente das demais (p-valor=0.0003).

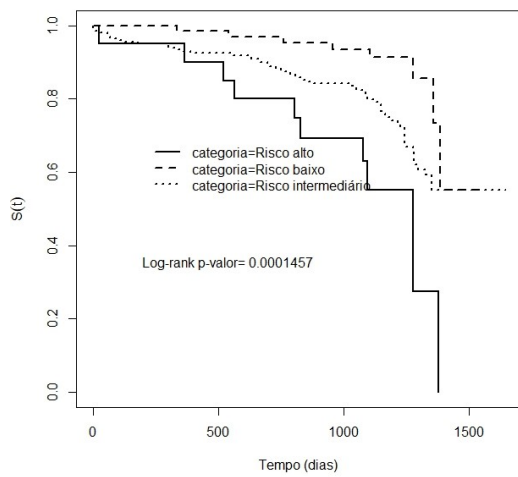
O mesmo foi realizado para o limiar 2, sendo apresentado abaixo as curvas de Kaplan-Meier de acordo com os grupos de risco baixo, intermediário e alto.



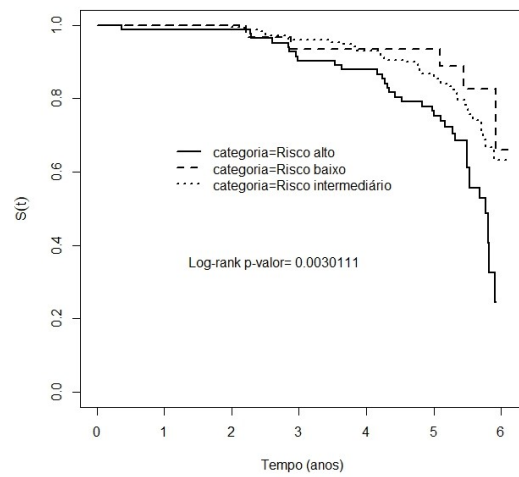
Naive 1



Naive 2



LandMark



JM - Weibull

Figura 24: Curvas de Kaplan-Meier segundo os grupos de risco (baixo, intermediário e alto) conforme o limiar 2 – abordagens: Naive 1, Naive 2, LandMark e JM - Weibull.

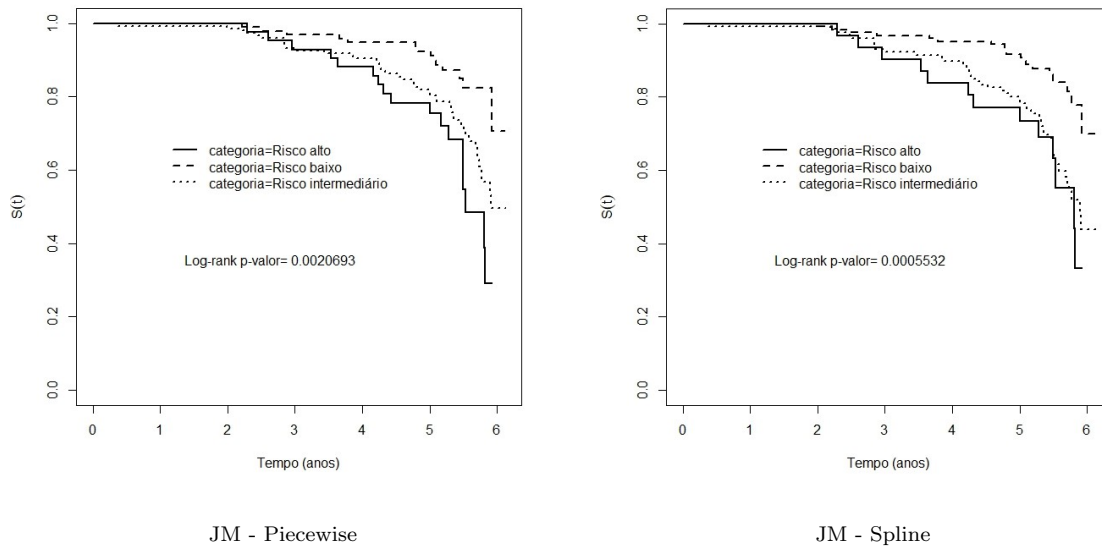


Figura 25: Curvas de Kaplan-Meier segundo os grupos de risco (baixo, intermediário e alto) conforme o limiar 2 – abordagens: JM - Piecewise e JM - Spline.

Os grupos de risco gerados a partir do Limiar 2, têm um padrão de sobrevida que difere do anterior, sendo que em todos os modelos houve a morte de pelo menos um paciente em cada grupo de risco. Para o Naive 2, houve um considerável aumento no número de indivíduos que vieram a óbito no grupo baixo risco (aumento de 24 óbitos). No caso do LM, observa-se que a mudança de limiar provocou uma queda da sobrevida para as três categorias estudadas, sendo que as curvas do LM e Naive 1 são similares, com boa separação entre as classes. Para os modelos JM as curvas de risco alto e intermediário são muito próximas e se diferenciam da curva de baixo risco, contudo para o método Weibull é possível observar uma aproximação entre as curvas de baixo e intermediário risco.

Além das curvas, a fim de testar diferenças significativas entre as sobrevivências por classificação de risco foi realizado teste de log-rank entre os grupos de risco para cada limiar contruído. Foi realizada uma comparação dois-a-dois das curvas para 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 $\alpha^* = \frac{0.05}{2} = 0,025$.

Tabela 38: Comparação dois-a-dois das curvas de sobrevivência pelo teste de log-rank, Limiar 1.

	Baixo vs. Intermediário		Intermediário vs. Alto	
	χ^2	p -valor	χ^2	p -valor
Naive 1	1,700	0,200	21,300	<0,001
Naive 2	16,400	<0,001	0,100	0,800
LM	-	-	28,700	<0,001
JM - Weibull	-	-	24,200	<0,001
JM - Piecewise	0,600	0,500	21,000	<0,001
JM - Spline	2,900	0,009	15,200	<0,001

Os resultados apresentados na Tabela 38 destacam diferenças significativas entre os grupos de risco para o Limiar 1. O modelo Naive 1 apresentou significância estatística na comparação “intermediário vs. alto” ($p = < 0,001$), mas não identificou diferenças significativas entre “baixo vs. intermediário” ($p = 0,200$). O modelo Naive 2 demonstrou desempenho inverso, com significância na comparação “baixo vs. intermediário” ($p = < 0,001$), mas não para “intermediário vs. alto” ($p = 0,800$). Entre os modelos conjuntos (JM), tanto o Weibull quanto o Piecewise apresentaram significância na comparação “intermediário vs. alto” ($p = < 0,001$), mas não detectaram diferenças significativas entre os grupos “baixo vs. intermediário”. O modelo Landmark (LM) foi significativo apenas para “intermediário vs. alto” ($p = < 0,001$). Esses resultados reforçam que, no contexto do Limiar 1, a capacidade dos modelos em discriminar entre os grupos varia amplamente, destacando o potencial dos modelos conjuntos para diferenciar grupos com maior risco de eventos.

Tabela 39: Comparação dois-a-dois das curvas de sobrevivência pelo teste de log-rank, Limiar 2.

	Baixo vs. Intermediário		Intermediário vs. Alto	
	χ^2	p -valor	χ^2	p -valor
Naive 1	4,500	0,030	6,900	0,009
Naive 2	6,800	0,090	0,000	0,900
LM	6,900	0,090	7,400	0,070
JM - Weibull	0,700	0,400	8,900	0,003
JM - Piecewise	7,100	0,008	2,100	0,100
JM - Spline	12,700	0,000	0,300	0,600

Os resultados apresentados na Tabela 39 evidenciam diferenças no desempenho dos modelos na identificação de significância estatística entre os grupos de risco para o limiar 2. O modelo Naive 1 apresentou diferenças significativas na comparação “intermediário vs. alto” ($p < 0.009$), mas os grupos “baixo” e “intermediário” ($p = 0.030$) tem curvas estatisticamente iguais. Já o modelo Naive 2 não identificou diferenças significativas entre os grupos em nenhuma das comparações. Dentre os modelos conjuntos (JM), o Weibull destacou-se ao apresentar significância para a comparação “intermediário vs. alto” ($p = 0.003$), enquanto o Piecewise não atingiu o nível crítico em nenhuma das análises. O JM Spline foi o único a demonstrar significância para a comparação “baixo vs. intermediário” ($p = 0.000$), mas não para “intermediário vs. alto”. Esses resultados indicam que a sensibilidade na identificação de diferenças entre os grupos varia consideravelmente entre os modelos e suas parametrizações, reforçando a importância da escolha do modelo no contexto da análise de sobrevivência.

Em resumo, as predições baseadas nos modelos Naive 1 e LM parecem ter melhor comportamento para separar os 3 grupos do que do Naive2 e JM. Estes resultados de discriminação estão de acordo com aqueles apresentados na Tabela 38.

7 Conclusão e Próximos Passos

Nesta dissertação foram explorados quatro métodos distintos para predição dinâmica em análises de sobrevivência: Naive 1, Naive 2, Landmark e Modelagem Conjunta (Joint Modeling). Enquanto os métodos Naive 1 e Naive 2 oferecem abordagens mais simples e diretas, o método Landmark introduz uma perspectiva dinâmica com a atualização periódica dos estimadores ao longo do tempo. Já a Modelagem Conjunta, o método mais sofisticado dentre os avaliados, integra de forma simultânea os métodos longitudinais e o de sobrevivência, proporcionando uma análise mais robusta e adaptada às mudanças temporais.

No contexto da modelagem conjunta foram avaliados três diferentes modelos de linha de base: Weibull, Piecewise e Spline. O modelo Weibull assume uma função de risco proporcional ao tempo, com uma estrutura paramétrica mais rígida. O modelo Piecewise, por sua vez, divide o tempo em intervalos, permitindo que o risco varie de forma mais flexível entre esses segmentos. Já o modelo Spline adota uma abordagem ainda mais flexível, utilizando funções suavizadas para capturar variações mais complexas no tempo. Essas diferenças metodológicas influenciam diretamente na calibração e discriminação dos modelos, destacando a importância da escolha adequada do método conforme as características do conjunto de dados e do objetivo do estudo.

O principal objetivo desta dissertação foi avaliar a adequação de diferentes modelos de predição em análises de sobrevivência aplicados a dados dinâmicos, utilizando dois estudos de caso distintos: (i) a predição da utilização de CTI/UTI neonatal e (ii) ocorrência de óbito em pacientes com cardiopatia chagásica. Além disso, buscou-se explorar a influência do envelhecimento das variáveis longitudinais na performance dos modelos, analisando seu impacto na calibração e discriminação. Essa investigação permitiu compreender como os modelos se adaptam a diferentes contextos clínicos e temporais, destacando suas vantagens e limitações em cenários reais.

Para o primeiro estudo de caso que avaliou a utilização do CTI/UTI Neonatal no primeiro ano de vida das crianças, o modelo de regressão de Cox na linha de base apresentou boa aderência, oferecendo um ponto de partida robusto para avaliações posteriores. Ao analisar os tempos dinâmicos de 182 e 365 dias, o melhor desempenho foi alcançado pela abordagem de Modelagem Conjunta, considerando discriminação, calibração e posteriormente as análises utilizando as curvas de Kaplan-Meier. Notou-se que, com o envelhecimento da variável longitudinal, houve uma ligeira redução na acurácia, destacando a importância de incorporar atualizações contínuas nos modelos. Essa aderência sugere que a modelagem conjunta oferece vantagens consideráveis para a predição dinâmica nesse cenário.

Já para o estudo da coorte prospectiva dos pacientes chagásicos, a regressão de Cox também demonstrou boa aderência para a linha de base, identificando fatores de risco relevantes. No tempo dinâmico, a modelagem conjunta destacou-se com bom desempenho em termos de calibração, porém teve resultado inferior na discriminação em relação aos métodos Naive 1 e LM. A análise dinâmica também revelou que, à medida que as variáveis longitudinais envelhecem, a efetividade da predição é parcialmente reduzida mas sem comprometer a capacidade de discriminação e calibração dos modelos.

O trabalho destacou que a utilização de diferentes técnicas de predição dinâmica, com ênfase na Modelagem Conjunta (JM), oferece vantagens significativas em relação a predição estática. Enquanto a regressão de Cox é amplamente utilizada devido à sua simplicidade e capacidade de identificar fatores de risco em um único ponto no tempo, ela não consegue captar de forma adequada as alterações que ocorrem em variáveis longitudinais ao longo do tempo. Por outro lado, a Modelagem Conjunta demonstrou ser uma ferramenta poderosa para integrar essas variáveis em tempo real, permitindo que a predição se adapte à evolução das condições clínicas. Essa abordagem foi especialmente eficaz nos dois casos analisados, onde medições contínuas, como o percentil de peso ao nascer e a frequência cardíaca, desempenharam um papel crítico na predição dos desfechos.

No primeiro estudo de caso, voltado à predição do uso de CTI/UTI Neonatal, o modelo conjunto com distribuição Spline apresentou o melhor desempenho geral, destacando-se nas métricas de discriminação e calibração para prever, aos 6 meses, o risco de utilização entre os indivíduos que não haviam utilizado o CTI/UTI nos primeiros 28 dias de vida. Além disso, o modelo conjunto com distribuição Weibull obteve o melhor desempenho na predição do risco de utilização do CTI/UTI em até um ano, considerando os indivíduos que permaneceram sem uso até os 6 meses. Já no segundo estudo de caso, embora o modelo conjunto com distribuição Spline não tenha atingido o maior valor de discriminação, seu desempenho foi muito próximo dos melhores resultados, sendo o modelo que apresentou o melhor ajuste em termos de calibração, o que reforça sua robustez na previsão do desfecho de interesse.

A escolha da distribuição da linha de base dentro da Modelagem Conjunta revelou-se um fator decisivo para o desempenho dos modelos. O modelo Weibull mostrou-se adequado em cenários com padrões de risco proporcionais e bem definidos ao longo do tempo, oferecendo simplicidade e estabilidade nos ajustes. Já o modelo Piecewise permitiu maior flexibilidade, dividindo o tempo em segmentos para capturar variações específicas de risco em diferentes períodos. Essa diversidade metodológica ressalta que a seleção da linha de base não é trivial e deve ser

guiada pelas características dos dados e pelos objetivos da análise, como priorizar calibração, discriminação ou sensibilidade às alterações temporais.

Os achados desta dissertação destacam que as técnicas de predição dinâmica, especialmente a Modelagem Conjunta, são ferramentas fundamentais para análises de sobrevivência em contextos clínicos complexos. Ao combinar informações longitudinais e tempos de evento, essas técnicas não apenas aumentam a precisão das predições, mas também fornecem insights valiosos sobre o curso temporal das condições analisadas. Isso reforça a necessidade de estudos futuros que explorem a aplicabilidade dessas abordagens em outros contextos, contribuindo para a personalização do cuidado em saúde.

Portanto, esta dissertação conclui que a aplicação de predições dinâmicas em análises de sobrevivência é uma abordagem indispensável para maximizar a precisão e a utilidade clínica em cenários complexos e dinâmicos. Ao explorar técnicas como Landmark e Modelagem Conjunta (JM), evidenciou-se que essas abordagens não apenas superam as limitações do modelo de linha de base, mas também oferecem ferramentas robustas para integrar variáveis longitudinais e tempos de eventos de forma simultânea.

Esses achados reforçam a relevância de estratégias preditivas que considerem não apenas a dinâmica das variáveis longitudinais, mas também as nuances dos desfechos em cada contexto clínico. A aplicação dessas técnicas não só aprimora o entendimento sobre o comportamento dos eventos de interesse, mas também contribui para uma prática clínica mais personalizada e orientada por evidências.

Para trabalhos futuros, sugere-se expandir a aplicação de modelos de predição dinâmica para outros cenários clínicos, incorporando variáveis longitudinais que reflitam diferentes aspectos das condições de saúde em análise. Além disso, é recomendável explorar os avanços computacionais mais recentes, como os pacotes JMBayes e JMBayes2, que adotam uma abordagem Bayesiana para a Modelagem Conjunta. Esses pacotes oferecem maior flexibilidade na especificação dos modelos, permitindo uma integração mais robusta entre variáveis longitudinais e desfechos de sobrevivência. Outra direção promissora é a investigação de modelos que considerem múltiplas variáveis dinâmicas simultaneamente, permitindo capturar interações complexas entre diferentes marcadores clínicos ao longo do tempo. Esses avanços podem contribuir significativamente para melhorar a precisão e a aplicabilidade dos modelos em cenários clínicos cada vez mais desafiadores.

8 Referências

1. 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.
2. 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.
3. 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.
4. COLOSIMO, E. A.; GIOLO, S. R. Análise de sobrevivência aplicada. São Paulo: Edgard Blucher, 2024.
5. COX, D. R. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, v. 34, n. 2, p. 187–220, 1972.
6. COX, D. R. Partial likelihood. *Biometrika*, v. 62, n. 2, p. 269–276, 1975.
7. CROWTHER, M. J.; ABRAMS, K. R.; LAMBERT, P. C. Joint modeling of longitudinal and survival data. *The Stata Journal*, v. 13, n. 1, p. 165–184, 2013.
8. 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.*
9. 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, 2015.
10. GRAMBSCH, P. M.; THERNEAU, T. M. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, v. 81, n. 3, p. 515–526, 1994.
11. GOULD, A.; BOYE, M.; CROWTHER M.; IBRAHIM J.; QUARTEY G.; MICALLEF S.; BOIS F.; Joint modeling of survival and longitudinal non-

- survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group. *Stat Med.* 2015 Jun 30;34(14):2181-95.
12. GUO, X.; CARLIN, B. P. Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician*, v. 58, p. 16–24, 2004.
 13. HARRELL, F. E. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. New York: Springer-Verlag, 2001.
 14. 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).
 15. HERNDON, J. E. II; HARRELL, F. E. Jr. The restricted cubic spline as baseline hazard in the proportional hazards model with step function time-dependent covariables. *Statistics in Medicine*, v. 14, p. 2119–2129, 1995.
 16. HENDERSON, R.; DIGGLE, P.; DOBSON, A. Identification and efficacy of longitudinal markers for survival. *Biostatistics*, v. 3, n. 1, p. 33–50, 2002.
 17. 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.
 18. HOUWELINGEN, H. van; PUTTER, H. Dynamic prediction in clinical survival analysis. Boca Raton: CRC Press, 2012.
 19. HOUWELINGEN, H. van; PUTTER, H. Landmarking 2.0: Bridging the gap between joint models and landmarking. Netherlands: Statistics in Medicine, 2022.
 20. 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.
 21. LIMA, E. Predição em análise de sobrevivência: aplicação em estudo envolvendo óbito de pacientes chagásicos cardiopatas. Belo Horizonte: Universidade Federal de Minas Gerais, 2019.

22. NATIONAL CENTER FOR HEALTH STATISTICS (NCHS). Percentile of weight-for-age: NCHS 77/8. In: United States Growth Charts. Washington, D.C.: National Center for Health Statistics, 1977.
23. NELSON, W. A short life test for comparing a sample with previous accelerated test results. *Technometrics*, v. 14, p. 175–185, 1972.
24. PHILIPSON, P.; DIGGLE, P.; SOUSA, I.; KOLAMUNNAGE-DONA, R.; WILLIAMSON, P.; HENDERSON, R. *joiner*: Joint modelling of repeated measurements and time-to-event data. 2012.
25. PROUST-LIMA, C.; AMIEVA, H.; JACQMIN-GADDA, H. Analysis of multivariate mixed longitudinal data: a flexible latent process approach. *The British Journal of Mathematical and Statistical Psychology*, v. 66, n. 3, p. 470–487, 2013.
26. 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.
27. 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.
28. RIZOPOULOS, D. Joint models for longitudinal and time-to-event data: With applications in R. New York: Chapman and Hall/CRC, 2012.
29. RIZOPOULOS, D.; HATFIELD, L. A.; CARLIN, B. P.; TAKKENBERG, J. J. M. Combining dynamic predictions from joint models for longitudinal and time-to-event data using Bayesian model averaging. *Journal of the American Statistical Association*, v. 109, n. 508, p. 1385–1397, 2014.
30. 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.
31. RIZOPOULOS, D.; TAYLOR, J. M. G. Optimizing dynamic predictions from joint models using super learning. *Statistics in Medicine*, v. 43, n. 7, p. 1315–1328, 2024.

32. ROSENBERG, M.; SCHOOLER, C.; SCHOENBACH, C.; ROSENBERG, F. Global self-esteem and specific self-esteem: different concepts, different outcomes. *American Sociological Review*, v. 60, p. 141–156, 1995.
33. SURITA, F. G.; PAULINO, D. S. M.; PINHO-POMPEU, M. Health-related behaviors in pregnancy: A key to achieve better outcomes. *Revista Brasileira de Ginecologia e Obstetrícia*, v. 42, n. 3, p. 121–123, 2020.
34. WEGMAN, E. J.; WRIGHT, I. W. Splines in statistics. *Journal of the American Statistical Association*, v. 78, p. 351–365, 1983.
35. WHITTEMORE, A. S.; KELLER, J. B. Survival estimation using splines. *Biometrics*, v. 42, n. 3, p. 495–506, 1986.
36. 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.