



OSDETE CORREA DE CARVALHO

**AVALIAÇÃO DE SOBREVIVENTES DE LONGO PRAZO USANDO MODELOS DE
FRAÇÃO DE CURA: APLICAÇÃO A DADOS DE INSUFICIÊNCIA CARDÍACA**

MARINGÁ - PR

2025



OSDETE CORREA DE CARVALHO

**AVALIAÇÃO DE SOBREVIVENTES DE LONGO PRAZO USANDO MODELOS DE
FRAÇÃO DE CURA: APLICAÇÃO A DADOS DE INSUFICIÊNCIA CARDÍACA**

Dissertação apresentada ao Programa de Pós-Graduação em Bioestatística da Universidade Estadual de Maringá como requisito para obtenção do título de Mestre em Bioestatística.

Orientador: Prof. Dr. Edson Zangiacomi Martinez.

Universidade Estadual de Maringá - UEM
Departamento de Estatística - DES
Programa de Pós-Graduação em Bioestatística

Maringá - PR

2025

OSDETE CORREA DE CARVALHO

**Avaliação de sobreviventes de longo prazo usando Modelos de Fração de
Cura: Aplicação a dados de Insuficiência Cardíaca**

Dissertação apresentada ao Programa de Pós-Graduação em Bioestatística do Centro de Ciências Exatas da Universidade Estadual de Maringá, como requisito parcial para a obtenção do título de Mestre em Bioestatística.

BANCA EXAMINADORA

Prof. Dr. Edson Zangiacomi Martinez
Universidade de São Paulo – PBE/USP

Prof. Dr. Vanderly Janeiro
Universidade Estadual de Maringá – PBE/UEM

Prof. Dr. Davi Casale Aragon
Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo – FMRP/USP

Maringá, 18 de dezembro de 2025.



À minha mãe e ao meu pai,
guardiões de tudo o que sou e do que ainda sonho ser.
Vocês, que me ensinaram o amor sem medida,
serão para sempre, em mim,
como um eco que atravessa o tempo,
como raízes que sustentam,
e estrelas que guiam, mesmo distantes.

Nas horas mais escuras, encontro o brilho de seus olhos,
e na alegria mais simples, escuto o riso que herdei.
São eternos em cada batida do meu coração,
em cada escolha que carrega a marca de suas mãos,
em cada pedaço do meu mundo onde seus abraços deixaram calor.

Vocês são o sempre,
o lugar onde eu retorno,
mesmo quando os caminhos parecem se perder.



AGRADECIMENTOS

Agradeço à Deus, pela força, saúde, coragem e persistência que me trouxeram até aqui.

Aos meus pais, por todo suporte, mesmo do outro lado do país.

Ao meu orientador, Dr. Edson Zangiacomi Martinez e ao professor Dr. Ricardo Puziol de Oliveira, por me mostrarem o caminho ao longo desta jornada.

Ao meu amigo Tailon Gustavo Küster Azeredo, cuja indicação e incentivo foram determinantes para que eu abraçasse essa oportunidade no mestrado.

À Thaisa Fernanda Cazatti, por sempre acreditar que sou capaz.

À minha família e amigos, por todo amor, compreensão e suporte emocional durante as fases mais desafiadoras desta caminhada. Vocês são minha base e minha força.

E, por fim, a todos que, direta ou indiretamente, contribuíram para este momento. Suas palavras de apoio, gestos de incentivo e presença marcaram profundamente esta etapa da minha vida.

"A estatística será, em breve, tão essencial à cidadania quanto a leitura e a escrita: um instrumento para compreender o mundo em sua incerteza."

— *H.G. Wells*

RESUMO

Introdução: A insuficiência cardíaca (IC) é uma condição de elevada morbimortalidade e apresenta grande variação no prognóstico dos pacientes. As análises tradicionais de sobrevivência, embora amplamente utilizadas, nem sempre conseguem identificar grupos que apresentam menor risco ao longo do tempo. Por isso, os modelos de frações de cura surgem como uma alternativa para explorar melhor essa heterogeneidade. **Objetivo:** Avaliar o desempenho de diferentes modelos de fração de cura aplicados a pacientes hospitalizados com IC e identificar os fatores associados tanto ao tempo de sobrevivência quanto à probabilidade de cura. **Métodos:** Foram utilizados dados clínicos e laboratoriais de um estudo disponível na literatura. Ajustou-se modelos de frações de cura nas versões de mistura e não-mistura. A comparação entre os modelos foi feita por meio dos critérios AIC, BIC e AICc, além da análise gráfica dos resíduos de Cox–Snell modificados. **Resultados:** Entre os modelos testados, o Weibull com regressão apresentou o melhor ajuste aos dados, com resíduos bem distribuídos. Variáveis como idade, fração de ejeção, som, creatinina, plaquetas e CPK demonstraram influência importante no risco e na fração de cura. **Conclusão:** Os resultados indicam que os modelos de fração de cura são úteis na avaliação prognóstica do IC, especialmente o modelo Weibull contínuo com covariáveis, que apresentou o desempenho mais consistente. Esses modelos contribuem para uma estratificação de risco mais precisa e podem ajudar no planejamento do cuidado clínico.

Palavras-chave: Insuficiência cardíaca; Análise de sobrevivência; Modelos de fração de cura; Weibull; estatística aplicada à saúde.

ABSTRACT

Introduction: Heart failure is a condition associated with high morbidity and mortality and presents significant variation in patient prognosis. Traditional survival analyses, although widely used, do not always succeed in identifying groups that exhibit a lower risk over time. Therefore, cure fraction models emerge as an alternative approach to better explore this heterogeneity. **Objective:** To evaluate the performance of different cure fraction models applied to hospitalized HF patients and to identify the factors associated with both the time to survival and the probability of cure. **Methods:** Clinical and laboratory data from a study available in the literature were used. Cure fraction models were fitted in both the mixture and non-mixture formulations. Model comparison was performed using the criteria AIC, BIC, and AICc, in addition to the graphical analysis of modified Cox–Snell residuals. **Results:** Among the tested models, the Weibull regression model showed the best fit to the data, with well-distributed residuals. Variables such as age, ejection fraction, sodium, creatinine, platelets, and CPK demonstrated a significant influence on both the risk and the cure fraction. **Conclusion:** The results indicate that cure fraction models are useful in the prognostic evaluation of HF, especially the continuous Weibull model with covariates, which presented the most consistent performance. These models facilitate a more precise risk stratification and can aid in the planning of clinical care.

Keywords: Heart failure; Survival analysis; Cure fraction models; Weibull; Applied statistics in healthcare.

LISTA DE ABREVIATURAS E SIGLAS

AIC – Akaike Information Criterion
AICc – Akaike Information Criterion Corrigido
BIC – Bayesian Information Criterion
CPK – Creatina Fosfoquinase
Crea. – Creatinina
DP – Desvio Padrão
EP – Erro-Padrão
FE / FEVE – Fração de Ejeção (do Ventrículo Esquerdo)
GAMLSS – *Generalized Additive Models for Location, Scale and Shape*
HAS – Hipertensão Arterial Sistêmica
HR – *Hazard Ratio* (Razão de Risco)
IC – Insuficiência Cardíaca
IC95% – Intervalo de Confiança de 95%
ICFER – Insuficiência Cardíaca com Fração de Ejeção Reduzida
ICFEP – Insuficiência Cardíaca com Fração de Ejeção Preservada
KM – *Kaplan-Meier*
NYHA – *New York Heart Association*
Plaq. – Plaquetas
ROC – *Receiver Operating Characteristic*
SRAA – Sistema renina-angiotensina-aldosterona
TTT – *Total Time on Test*
UTI – Unidade de Terapia Intensiva
WAIC – *Watanabe-Akaike Information Criterion*

LISTA DE TABELAS

Tabela 1 – Variáveis analisadas no estudo	21
Tabela 2 – Estatística descritiva e valores de referência	42
Tabela 3 – Distribuição das variáveis categóricas da amostra	45
Tabela 4 – Medidas de assimetria e curtose das variáveis contínuas	49
Tabela 5 – Resultados do modelo de regressão de Cox para os fatores associados à sobrevivência em pacientes com insuficiência cardíaca	52
Tabela 6 – Estimativas dos parâmetros dos modelos de fração de cura sem regressão	61
Tabela 7 – Estimativas dos parâmetros dos modelos de fração de cura com regressão no parâmetro p (fração de cura)	65
Tabela 8 – Estimativas dos parâmetros dos modelos de fração de cura com regressão nos parâmetros de locação e no parâmetro p	66

LISTA DE FIGURAS

Figura 1 — Boxplots das variáveis clínicas e laboratoriais: (a) idade, (b) FE, (c) sódio, (d) log(Creatinina), (e) log(Plaquetas) e (f) log(CPK).	47
Figura 2 — Curva de sobrevivência Kaplan–Meier para pacientes com insuficiência cardíaca.	50
Figura 3 — Resíduos de Martingale para o modelo de Cox.	54
Figura 4 — Resíduos de Schoenfeld.	55
Figura 5 — Resíduos Deviance em função do tempo.	57
Figura 6 — TTT Plot (total time on test).	58
Figura 7 — Função de risco $h(t)$ estimada por método não paramétrico com suavização B-splines.	59
Figura 8 — Curva de sobrevivência observada (Kaplan–Meier) e ajustada pelos modelos de fração de cura.	62
Figura 9 — Matriz de correlações entre as variáveis independentes	63
Figura 10 — Medidas do Fator de Inflação da Variância (FIV) entre as variáveis independentes.	64
Figura 11 — Envelopes simulados para os resíduos de Cox–Snell modificados dos modelos de fração de cura (a) Weibull; (b) Gama; (c) Lindley; (d) Gumbel; (e) Exponencial; (f) Rayleigh.	68

SUMÁRIO

1 INTRODUÇÃO	12
2 JUSTIFICATIVA	16
3 OBJETIVOS	18
3.1 Objetivo Geral	18
3.2 Objetivos Específicos	18
4 HIPÓTESES	19
5 METODOLOGIA	20
5.1 Delineamento e Fonte dos Dados	20
5.2 Critérios de Inclusão e Exclusão	21
5.3 Variáveis do Estudo	21
5.4 Descrição e Preparação dos Dados	22
5.5 Estudos de Sobrevida	22
5.6 Mecanismos de Censura	26
5.7 Estimador de Kaplan–Meier	28
5.8 Modelo de Riscos Proporcionais de Cox	29
5.9 Modelos de Fração de Cura	29
5.9.1 Modelo de Mistura	30
5.9.2 Modelo de Não-Mistura	31
5.10 Distribuições de Probabilidade	31
(a) Weibull	32
(b) Gama	32
(c) Lindley	33
(d) Gumbel	34
(e) Exponencial	35
(f) Rayleigh	35
5.11 Inferência em Análise de Sobrevida	36
5.12 Diagnóstico do Modelo: Resíduos de Cox–Snell	37
5.13 Estrutura de Regressão	39
5.14 Considerações Éticas	40
6 RESULTADOS E DISCUSSÃO	42
6.1 Contextualização e Descrição dos Dados	42
6.2 Análise de Qualidade e Caracterização dos Dados	43
6.3 Estimador de Kaplan–Meier	50
6.4 Modelo de Riscos Proporcionais de Cox	51
6.5 Estimação da Função de Risco Empírica	57
6.6 Análise por Modelos de Fração de Cura	59
6.6.1 Modelos com Regressão	63
7 CONCLUSÃO	70
REFERÊNCIAS	71

INTRODUÇÃO

A insuficiência cardíaca (IC) é uma síndrome clínica de grande relevância em saúde pública devido a sua alta prevalência, caracterizada pela incapacidade progressiva do coração de suprir adequadamente as demandas hemodinâmicas do organismo. É uma condição complexa, associada a fadiga, dispneia e edema, decorrentes de anormalidades estruturais e funcionais, e isso resulta em débito cardíaco e elevam as pressões intracardíacas. A sua classificação se dá em dois subtipos: insuficiência cardíaca com fração de ejeção reduzida (ICFEr) e preservada (ICFEp). Ambas representam diferentes mecanismos fisiopatológicos e possuem implicações prognósticas e terapêuticas distintas (Heidenreich *et al.*, 2022).

Para compreender a IC é fundamental recordar a função primordial do coração: atuar como uma bomba eficiente, enviando sangue rico em oxigênio e nutrientes para todos os tecidos do corpo e recebendo o sangue venoso, pobre em oxigênio, para reoxigená-lo nos pulmões. O ciclo começa quando o sangue venoso, retornando do corpo, chega ao átrio direito pelas veias cavas e, em seguida, passa para o ventrículo direito, que o bombeia para os pulmões através da artéria pulmonar. Nos pulmões, o sangue libera gás carbônico e recebe oxigênio. Este sangue agora oxigenado retorna ao átrio esquerdo pelas veias pulmonares e desce para o ventrículo esquerdo. O ventrículo esquerdo, que é a câmara mais muscular do coração, se contrai e ejeta o sangue para o resto do corpo através da aorta (Guyton & Hall, 2017).

A fração de ejeção do ventrículo esquerdo (FEVE) é a métrica central para avaliar a eficácia dessa bomba. Ela representa a porcentagem de sangue que é ejetada do ventrículo esquerdo a cada batimento, em relação ao volume total de sangue que estava preenchido nele no final da diástole (fase de relaxamento e enchimento). O ventrículo nunca ejeta todo o seu conteúdo; uma FEVE considerada normal situa-se acima de 55%. Isso significa que, em um indivíduo saudável, mais da metade do volume de sangue no ventrículo esquerdo é bombeada para o corpo a cada contração (Lang *et al.*, 2015).

A IC se instala quando, por algumas condições (como hipertensão arterial, infarto ou doenças nas válvulas), o coração se torna incapaz de desempenhar essa função de bombeamento de forma adequada. A FEVE é obtida por métodos de imagem, principalmente a ecocardiografia, e é o critério que divide a IC em três subgrupos de acordo com a Diretriz Brasileira de IC Crônica e Aguda (Rohde *et al.*, 2018).

1. Na ICFEr, o músculo cardíaco está enfraquecido ou "remodelado," e não consegue se contrair com força suficiente, resultando em uma FEVE reduzida (<40%).

2. Já na ICFEp, o problema principal não é a contração (sístole), mas sim o relaxamento e o enchimento do ventrículo (diástole). O músculo cardíaco fica rígido, impedindo que se encha de sangue adequadamente. A FEVE é $\geq 50\%$, sendo o valor de corte clínico para o subtipo ICFEp, mas é crucial notar que "preservada" ($\geq 50\%$) não é sinônimo de "normal" ($\geq 55\%$), indicando que a função sistólica pode estar ligeiramente comprometida, embora não reduzida.
3. A IC com fração de ejeção levemente reduzida constitui um grupo intermediário, com FEVE entre 40% e 49%. Em todos os casos, o débito cardíaco se torna insuficiente para atender às demandas do organismo, desencadeando os sintomas característicos da síndrome, também nomeados “zona cinzenta” da fração de ejeção.

Essa diferença na FEVE leva a pensar em diferentes mecanismos fisiopatológicos. Enquanto na ICFEr há predomínio do remodelamento e perda da contração ventricular, na ICFEp prevalecem rigidez diastólica, fibrose e alterações microvasculares, frequentemente associadas a comorbidades como obesidade, diabetes e hipertensão arterial. A IC com fração de ejeção levemente reduzida constitui um grupo intermediário, ainda em investigação, mas de relevância clínica. A progressão da doença envolve a ativação contínua de sistemas neuro-hormonais, como o sistema renina–angiotensina–aldosterona (SRAA) e o sistema nervoso simpático, que, quando estimulados de forma persistente, favorecem a vasoconstrição, a retenção de sódio e água, além de promoverem hipertrofia e fibrose do miocárdio, sustentando o ciclo de disfunção ventricular. Como resposta compensatória, há a liberação de peptídeos natriuréticos, que exercem efeito vasodilatador e diurético, mas sua ação vai se tornando insuficiente conforme a doença avança. Outros processos também participam da progressão, como a inflamação crônica, o estresse oxidativo e a disfunção endotelial. Esses fatores tem impacto ainda maior nos pacientes com ICFEp, em que as comorbidades metabólicas atuam como elementos desencadeadores e mantenedores da síndrome (Rohde *et al.*, 2018).

A classificação da *New York Heart Association* (NYHA) é utilizada para fazer a estratificação dos pacientes de acordo com a intensidade dos sintomas e o grau de limitação nas atividades do dia a dia.

1. Na classe I, os pacientes apresentam cardiopatia estrutural, mas permanecem assintomáticos durante esforços diários comuns.
2. Na classe II, observa-se leve limitação, com manifestações causadas por atividades comuns.
3. A classe III corresponde a pacientes com limitação mais evidente, nos quais pequenos esforços já provocam dispneia ou fadiga, embora o repouso seja tolerado.

4. Na classe IV, os sintomas ocorrem mesmo em repouso, comprometendo amplamente a qualidade de vida. Além de prática, essa classificação possui valor prognóstico, pois níveis mais avançados de limitação funcional estão associados a maior mortalidade e risco de reinternações (Heidenreich *et al.*, 2022).

Diversos fatores contribuem para a progressão clínica da IC, entre eles a disfunção ventricular, alterações eletrolíticas, comorbidades (como hipertensão arterial sistêmica e diabetes mellitus) e complicações renais (Heidenreich *et al.*, 2022). A gravidade da doença é acentuada por elevadas taxas de hospitalização e mortalidade precoce, especialmente em pessoas com disfunção sistólica avançada. A identificação de fatores prognósticos e a estratificação de risco são fundamentais para orientar condutas e planejar intervenções médicas. Biomarcadores, exames de imagem e escores clínicos são muito usados na predição de desfechos adversos. Abordagens baseadas no perfil clínico e em diretrizes, contribuem para melhorar a qualidade de vida e a sobrevida desses pacientes (Rohde *et al.*, 2018).

Estima-se que mais de 64 milhões de pessoas no mundo convivam com a doença, o que a coloca entre as principais síndromes cardiovasculares da atualidade. No Brasil, representa uma das maiores causas de internação hospitalar pelo Sistema Único de Saúde (SUS), sobretudo em pessoas acima de 60 anos, gerando custos elevados e sobrecarga aos serviços de saúde (Groenewegen *et al.*, 2020).

Para descrever e analisar dados de insuficiência cardíaca, diversas abordagens estatísticas podem ser consideradas. No entanto, em função das características dessa síndrome e dos desfechos clínicos usualmente investigados (como óbito ou reinternação), a análise de sobrevivência assume um papel importante. Métodos tradicionais, como o estimador de *Kaplan–Meier* e o modelo dos riscos proporcionais de Cox, são amplamente utilizados, possibilitando incorporar informações sobre o tempo até a ocorrência de um evento de interesse, lidar adequadamente com a presença de censura e identificar variáveis associadas à evolução clínica, oferecendo, assim, uma visão mais abrangente da progressão da doença. No entanto, tais métodos assumem de forma implícita que todos os indivíduos permanecem em risco de apresentar o evento, o que nem sempre corresponde à realidade clínica da doença. Em determinadas situações, parte da população pode ser considerada “curada”, afirmação não contemplada por modelos tradicionais (Therneau & Grambsch, 2000).

Os modelos de fração de cura são uma alternativa para lidar com essa situação, permitindo estimar simultaneamente a proporção de indivíduos suscetíveis ao evento e aqueles considerados curados (Oliveira, 2015). Existem duas formulações principais:

1. Modelos de mistura, que pressupõem dois subgrupos distintos na população: os suscetíveis, ainda em risco, e os “curados”, livres do evento de interesse;
2. Modelos de não-mistura, que pressupõe que a função de sobrevivência irá convergir para um platô positivo em longo prazo.

A aplicação de modelos de fração de cura é mais comum em estudos oncológicos, mas ainda pouco explorada em doenças crônicas cardiovasculares. A maior parte da literatura utiliza distribuições contínuas para modelar o tempo até o evento, o que permite representar de forma mais realista a progressão temporal dos desfechos clínicos (Oliveira, 2015), embora modelos baseados em distribuições discretas também sejam possíveis.

Diante desse cenário, o presente trabalho propõe aplicar modelos de fração de cura para analisar dados de 299 pacientes com insuficiência cardíaca, acompanhados entre 4 e 285 dias, dos quais 96 (32%) evoluíram para óbito. No estudo original de Ahmad *et al.*, (2017), de onde estes dados foram obtidos, foram utilizadas as abordagens tradicionais de *Kaplan–Meier* e do modelo de Cox, que identificaram idade, fração de ejeção e creatinina como variáveis associadas ao risco de mortalidade.

Foram aplicados modelos de fração de cura em sua formulação de mistura, utilizando inicialmente as distribuições Weibull, Gama, Lindley, Gumbel, Exponencial e Rayleigh, com o propósito de investigar a possível existência de um subgrupo de pacientes clinicamente estáveis ao longo do seguimento. Ao comparar esses modelos com métodos tradicionais de análise de sobrevivência, como o modelo de Cox, buscou-se avaliar com maior precisão o comportamento dos tempos de vida e a eventual presença de indivíduos com risco praticamente nulo no período analisado. Dessa forma, o objetivo deste estudo é investigar o perfil de sobrevivência de pacientes com IC, avaliando a capacidade dos modelos de fração de cura, de aprimorar a estratificação de risco e oferecer subsídios mais refinados para a tomada de decisão clínica.

2 JUSTIFICATIVA

A IC é uma síndrome clínica grave com alta prevalência, que traz sofrimento para o paciente, sobrecarga para as famílias e desafios constantes para as equipes multiprofissionais, desde o diagnóstico até o tratamento. A experiência da autora, como enfermeira intensivista e atuante na educação permanente em saúde, levou-a a observar cotidianamente o impacto dessa doença no ambiente hospitalar: desde a instabilidade clínica e a vulnerabilidade dos pacientes até a necessidade de decisões rápidas, baseadas em evidências, que podem significar a diferença entre a vida e a morte. Diante disso, foram pensadas estratégias estatísticas para melhorar a predição de mortalidade nesses pacientes e guiar de forma mais refinada as decisões clínicas.

Os métodos tradicionais de análise de sobrevivência, como o estimador de *Kaplan–Meier* e o modelo de Cox, oferecem contribuições válidas, mas partem do pressuposto de que todos os indivíduos permanecem em risco de ocorrência do evento de interesse. A realidade clínica demonstra que, embora a insuficiência cardíaca ainda apresente elevada mortalidade e recorrência de hospitalizações, parte dos pacientes consegue alcançar estabilidade clínica sustentada, especialmente quando submetidos a tratamento otimizado e acompanhamento multidisciplinar. Estudos recentes indicam melhoras graduais na sobrevida em longo prazo, com um subconjunto de pacientes vivendo por vários anos após a hospitalização inicial, com melhor qualidade de vida e controle sintomático (Antohi, Adamo & Chioncel, 2022). Reconhecer esse subgrupo, fração de “cura”, é importante para uma estratificação de risco mais realista, para otimizar recursos assistenciais e, também, para oferecer recursos ancorados em dados reais.

A escolha de modelos de fração de cura, se justifica na natureza dos dados (medidos em dias de acompanhamento com possibilidade de censura e platô na curva de sobrevivência) quanto na necessidade de maior adequação do modelo estatístico aos dados. Essa abordagem permite não apenas avaliar fatores prognósticos, mas também identificar a possibilidade de existência de um grupo de pacientes sem risco de óbito no período analisado. Do ponto de vista clínico trata-se de uma inovação metodológica e informação prática para equipes de saúde, ou seja, reconhecer quais pacientes demandam vigilância intensiva contínua e quais podem ter um acompanhamento ambulatorial, por exemplo.

Cientificamente, este trabalho amplia o portfólio da bioestatística aplicada à cardiologia, incorporando técnicas modernas capazes de responder a questões ainda pouco exploradas na literatura. Do ponto de vista profissional, traz contribuições para o cuidado em saúde, ao possibilitar que pacientes com IC sejam contemplados por condutas mais assertivas.

Ao unir a vivência da autora da dissertação na assistência no meio intensivista mais a atuação em educação permanente e a aplicação de modelos estatísticos avançados, este trabalho busca não apenas produzir conhecimento, mas também gerar impacto real na vida dos pacientes, na qualidade da assistência multiprofissional e disseminar o uso de modelos de fração de cura entre investigadores da área da saúde.

3 OBJETIVOS

3.1 Objetivo Geral

Investigar a sobrevida de pacientes com IC a partir de uma reanálise de dados introduzidos na literatura, aplicando modelos de fração de cura e comparando-os com métodos tradicionais (*Kaplan-Meier* e modelo de Cox), a fim de identificar subgrupos com risco nulo ou reduzido de mortalidade.

3.2 Objetivos Específicos

- I. Estimar a curva de sobrevivência utilizando o estimador de *Kaplan-Meier*;
- II. Identificar fatores associados ao risco de mortalidade pelo modelo de riscos proporcionais de Cox;
- III. Ajustar, via método da máxima verossimilhança, os modelos de fração de cura sob as formulações de mistura, com base nas distribuições Weibull, Gama, Lindley, Gumbel, Exponencial e Rayleigh;
- IV. Incorporar aos modelos de fração de cura estruturas de regressão no parâmetro de cura;
- V. Comparar o desempenho dos modelos usando critérios de informação (AIC, BIC, AICc);
- VI. Quantificar a fração de pacientes considerados curados no horizonte temporal do estudo;
- VII. Avaliar o papel de variáveis clínicas e laboratoriais (idade, anemia, fumo, fração de ejeção, log(creatinina), sódio, log(plaquetas), log(CPK), entre outras) na probabilidade de cura e no risco de óbito entre pacientes suscetíveis.

4 HIPÓTESES

H1: Existe um subgrupo de pacientes com insuficiência cardíaca que apresenta risco nulo de mortalidade no período analisado.

H3: Os modelos de fração de cura apresentam melhor aderência aos dados na avaliação da probabilidade de sobrevivência e identificação dos fatores associados a essa probabilidade do que as formulações tradicionais.

5 METODOLOGIA

Este estudo é uma reanálise de dados previamente coletados em um centro hospitalar especializado no acompanhamento de pacientes com IC sistólica, classificados nas classes funcionais III e IV da NYHA, 1973, uma organização cardiológica de referência que estabeleceu essa classificação. É importante notar que a NYHA é apenas uma ferramenta utilizada para estimar o grau de limitação do paciente, e sua avaliação deve ser complementada por critérios objetivos (como a FEVE e biomarcadores) para o diagnóstico e o estadiamento completo da IC.

Na reanálise foram incluídos 299 indivíduos (100% da amostra original) acompanhados em regime ambulatorial e/ou hospitalar, com tempos de seguimento entre 4 e 285 dias. O desfecho de interesse foi definido como óbito por qualquer causa durante o período de acompanhamento.

O início do seguimento foi considerado a data da admissão hospitalar, e o acompanhamento estendeu-se até a ocorrência do óbito ou o término do período de observação. Os indivíduos que não apresentaram o evento (óbito) até o final do seguimento foram considerados censurados, representando pacientes que permaneceram vivos até a alta hospitalar ou até o encerramento do estudo.

A base de dados original, proveniente de prontuários hospitalares e registros clínicos informatizados, contemplava variáveis sociodemográficas (idade, sexo), clínicas (classe funcional, fração de ejeção (FE), comorbidades como hipertensão e diabetes) e laboratoriais (creatinina sérica, ureia, sódio, potássio, creatina-fosfoquinase (CPK), entre outras).

5.1 Delineamento e Fonte dos Dados

População-alvo: pacientes diagnosticados com IC por disfunção sistólica do ventrículo esquerdo, classes funcionais III e IV.

Amostra: 299 pacientes.

Período de seguimento: mínimo de 4 dias e máximo de 285 dias.

Desfecho primário: óbito por quaisquer causas.

5.2 Critérios de Inclusão e Exclusão

No estudo original, e também nessa reanálise de dados que considerou toda a amostra, foram incluídos pacientes com idade ≥ 18 anos, diagnóstico confirmado de insuficiência cardíaca por disfunção sistólica do ventrículo esquerdo, classificados nas classes funcionais III e IV (NYHA), acompanhados ambulatorialmente ou em regime hospitalar no centro de referência.

Foram excluídos pacientes com registros incompletos para a variável desfecho (tempo ou óbito), prontuários duplicados ou com inconsistências, bem como aqueles com cardiopatias não classificadas como insuficiência cardíaca sistólica.

5.3 Variáveis do estudo

A Tabela 1 apresenta as variáveis consideradas no estudo:

Tabela 1 — Variáveis analisadas no estudo.

Variável	Tipo	Unidade/Definição	Transformação	Papel no modelo
Óbito	Binária	0 = censurado; 1 = evento	—	Componente da Resposta
Tempo até evento	Contínua	Dias (4 a 285)	—	Resposta
Idade	Contínua	Anos	—	Preditor
Sexo	Categórica	Masculino/ Feminino	—	Preditor
Fração de ejeção	Contínua	%	—	Preditor
Creatinina	Contínua	mg/dL	$\log(\text{Creatinina} + 1)$	Preditor
Sódio	Contínua	mEq/L	—	Preditor
Plaquetas	Contínua	mil/mm ³	$\log(\text{Plaquetas} + 1)$	Preditor
CPK	Contínua	U/L	$\log(\text{CPK} + 1)$	Preditor
Comorbidades (HAS, DM, anemia, tabagismo)	Binária	Presente/ Ausente	—	Preditor

Fonte: Elaboração própria com base nos dados do estudo.

5.4 Descrição e Preparação dos Dados

Inicialmente, foi conduzida uma etapa de limpeza e validação dos dados, verificando a consistência de faixas clínicas e a ausência de valores faltantes. A caracterização da amostra foi feita por meio de:

1. Medidas de tendência central e dispersão: média, mediana, desvio-padrão e amplitude interquartil.
2. Distribuição das variáveis: avaliação gráfica (*boxplots*) e medidas de assimetria e curtose.
3. Tratamento de assimetrias: em variáveis com forte assimetria, como CPK, creatinina e plaquetas, foi aplicada transformação logarítmica às variáveis CPK, creatinina e plaquetas a fim de reduzir a influência de valores extremos e aproximando a distribuição da normalidade.
4. Adicionalmente, variáveis contínuas foram categorizadas em alguns momentos para análise exploratória (ex.: faixas etárias), a fim de facilitar a interpretação clínica.

Durante a etapa de verificação da consistência, constatou-se ausência de valores faltantes para as variáveis incluídas na análise. Dessa forma, não houve necessidade de procedimentos de imputação ou exclusão de dados.

5.5 Estudos de Sobrevivência

Dados de sobrevivência, dados de tempo de vida, dados de tempo até a falha ou dados de tempo até o evento são termos usados para descrever dados que medem o tempo até a ocorrência de algum evento e surgem em diversos campos aplicados, como medicina, biologia, saúde pública, epidemiologia, engenharia, economia, demografia, entre muitos outros. Ainda segundo Miot (2017), em todos esses campos, o foco está na modelagem do tempo até que ocorre um evento específico, ou seja, dados de tempo até o evento, bem como na previsão da probabilidade de resposta, sobrevivência ou tempo médio de vida, comparações das distribuições de sobrevivência e a identificação de fatores de risco ou previsões relacionadas às respostas fornecidas pelos tempos de vida.

O evento pode ser a morte, como o termo sugere, mas também pode ser aplicado em diferentes cenários. Por exemplo, na pesquisa médica, o evento pode ser alguma ocorrência que altere o estado de saúde dos indivíduos, como a cura de uma doença, o tempo de remissão para um tipo específico de câncer, o tempo até que indivíduos expostos sejam infectados por uma

doença, o tempo até o nível de estímulos ou o tempo de ocorrência a um tratamento em pares de pulmões, enxárgues, olhos ou ouvidos de humanos; na análise de confiabilidade, o evento pode estar relacionado a quebra, reparo de máquinas ou tempo de vida útil de componentes de um sistema de engenharia (Cox, 1972; Klein e Moeschberger, 1997).

Outras possibilidades também são apresentadas em estudos de toxicologia, cujos tempos de falha podem ser observados para as crias em cada ninhada, ou em estudos relacionados aos efeitos de duas drogas na atividade cerebral; tempo até a liberdade condicional de condenados (criminologia); duração do primeiro casamento (sociologia); tempo de assinatura de jornal ou revista (marketing); e pedidos de indenização trabalhista (seguros) e seus diversos fatores de risco ou previsões; entre muitas outras aplicações (Cox, 1972; Klein e Moeschberger, 1997).

Estudos que analisam o tempo até o evento possuem natureza temporal, pois consideram o período transcorrido até a ocorrência de determinado desfecho. As informações sobre o tempo e demais variáveis podem ser obtidas retrospectivamente, como em estudos de caso-controle, ou prospectivamente, por meio do acompanhamento dos participantes ao longo do tempo (Bordalo, 2006).

Um estudo prospectivo de sobrevivência pode ser um estudo de coorte observacional ou um ensaio clínico randomizado. Um estudo observacional de acompanhamento nos permite caracterizar a história natural da doença em termos quantitativos. Depois de descrevermos a gravidade de uma doença, podem então ser prioridades para serviços clínicos e programas de saúde pública. Por outro lado, geralmente conduzimos um ensaio para modificar a história natural de uma doença, de modo a prevenir ou retardar um desfecho adverso, como morte ou incapacidade, e melhorar a saúde de um paciente individual ou da população em geral. Um ensaio clínico planejado pode avaliar a eficácia e os efeitos colaterais de novas formas de intervenção. Tanto os estudos observacionais quanto os ensaios clínicos estudados são amplamente praticados na medicina, e a escolha do delineamento depende do objetivo do estudo, da doença específica e das limitações práticas (Freire e Patussi, 2018).

Para exemplificar, basta considerar um paciente selecionado aleatoriamente da população, cujo tempo até a ocorrência do evento de interesse seja representado por T . A análise de sobrevivência busca descrever o comportamento dessa variável ao longo do tempo, permitindo responder a questões práticas, como: por quanto tempo um paciente com determinada condição pode permanecer vivo com uma certa probabilidade? ou se um tratamento A proporciona maior sobrevida do que um tratamento B. Embora seja possível ajustar distribuições de probabilidade específicas para modelar T , também existem abordagens

não paramétricas que estimam a função de sobrevivência sem pressupor uma forma paramétrica para os dados (Armitage; Berry; Matthews, 2002).

Para responder, usa-se a função de sobrevivência e a função de risco. A função de sobrevivência pode ser interpretada e definida como a probabilidade de observar um tempo de sobrevivência maior que um valor fixo t , isto é:

$$S(t) = P(T \geq t) = 1 - F(t) \quad (1)$$

Em que $F(t)$, denotado como função de distribuição acumulada, é a probabilidade de um indivíduo selecionado aleatoriamente morrer antes do tempo t . Ainda segundo Lawless (1982), a função de sobrevivência é uma função decrescente e contínua à esquerda, isto é, $S(0) = 1$ e $S(\infty) = 0$.

Outra função importante relacionada à descrição da distribuição de sobrevivência é a *função de risco*. Ela pode ser definida como a probabilidade instantânea de ocorrência do evento (por exemplo, morte) no instante t , dado que o indivíduo tenha sobrevivido até pelo menos o tempo t (Lawless, 1982), ou seja :

$$h(t) = \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} \quad (2)$$

onde Δt é um *incremento infinitesimal de tempo*. Em algumas aplicações práticas, $h(t)$ também é chamada de *taxa instantânea de falha* e pode ser expressa como $h(t) = f(t)/S(t)$, onde $f(t)$ é a *função de densidade de probabilidade* (ou de massa, no caso discreto). O *risco acumulado*, $H(t)$, também pode ser obtido em termos da função de sobrevivência como:

$$S(t) = \exp \{-H(t)\} \quad (3)$$

Em resumo, as funções $f(t)$, $S(t)$, $H(t)$ e $h(t)$, estão inter-relacionadas e, conhecendo-se uma delas, é possível obter as demais por meio de relações integrais e diferenciais apropriadas. Por exemplo, a partir da função de risco $h(t)$ pode-se obter o risco acumulado $H(t) = \int_0^t h(u)du$ e, consequentemente, derivar a função de sobrevivência $S(t) = e^{-H(t)}$ e a densidade de probabilidade $f(t) = h(t)S(t)$ (Lawless, 1982).

Uma estratégia de estimação não-paramétrica da função de risco $h(t)$ com suavização baseada diretamente nos dados é descrita por Rebora, Salim e Reilly (2014), em que funções B-splines são utilizadas para estimar a forma de $h(t)$ em uma abordagem de modelos lineares generalizados de efeitos mistos. Particionando o tempo em pequenos intervalos, o número de eventos em cada intervalo segue aproximadamente uma distribuição de Poisson com média $\mu(t) = h(t)P(t)$. Nesta expressão, $P(t)$ é o número de pessoas-tempo no intervalo correspondente. Se não há censuras, $P(t)$ é o número de pessoas em risco no início do intervalo, multiplicado pela amplitude do intervalo. Rebora, Salim e Reilly (2014) consideram:

$$\ln \mu(t) = g(t) + \ln P(t), \quad (4)$$

em que $g(t) = \ln h(t)$ é o logaritmo da função de risco, estimada por B-splines. B-splines são uma generalização da chamada “curva de Bézier”, que são usadas para modelar curvas suaves e contínuas definidas por conjunto de “pontos de controle” discretos. Assim, B-splines são definidos por uma ordem m e um número de “nós” interiores. A função “bshazard” do pacote R utiliza a sintaxe:

`bshazard(Surv(time, cens) ~ 1, data, nbin, nk, degree)`

em que “time” corresponde ao tempo até o evento, “cens” é variável indicadora que estabelece as censuras, data é a base de dados, nbin é o número de intervalos igualmente espaçados, nk é o número de “nós” (o default utilizado pelo pacote é 31, incluindo o mínimo e o máximo), e degree é a ordem (o default é 1, correspondente a segmentos lineares). O pacote “bshazard” permite gerar um gráfico da função $h(t)$ assim estimada.

Na teoria da probabilidade, os momentos das distribuições de probabilidade são medidas-resumo importantes das características da distribuição. A média e a variação são exemplos familiares e representam o primeiro e o segundo momento, respectivamente. Contudo, para a maioria das distribuições de sobrevivência, é mais relevante resumir a distribuição de sobrevivência pelos momentos da distribuição truncada, uma vez que não é apropriado assumir uma distribuição simétrica. Uma dessas configurações úteis é a *vida residual média*, que é definida como:

$$r(t) = E(T \geq t) = \frac{1}{S(t)} \int_t^{\infty} S(u) du \quad (5)$$

e pode ser interpretada como a expectativa de vida após um indivíduo ter sobrevivido até o tempo t . Outra medida importante na análise de sobrevivência é o *percentil*. O percentil, definido por $F^{-1}(p)$, da distribuição de sobrevivência $F(t)$ é descrito pela equação (Kaplan; Meier, 1958):

$$F^{-1}(p) = \inf \{t : F(t) > p\} \quad (6)$$

O percentil 50° é o tempo médio de sobrevivência e é interpretado como o tempo no qual metade dos indivíduos da população em estudo ainda sobrevivem. Ele oferece duas vantagens em relação à média do tempo de sobrevivência. Primeiro, sua estimativa amostral é menos afetada por valores extremamente grandes ou pequenos, enquanto a média pode ser sensível a um pequeno número de valores atípicos. Em segundo lugar, devemos observar todos os eventos da população em estudo para calcular a média, enquanto para a mediana não é necessário (Kaplan; Meier, 1958).

5.6 Mecanismos de Censura

Segundo Klein e Moeschberger (1997), Giolo e Colosimo (2006) e Collett (2015), a principal característica dos dados de sobrevivência é a presença de observações censuradas ou truncadas. O tempo de sobrevivência de um paciente é censurado quando o ponto de interesse não foi observado para esse paciente. Em geral, isso ocorre porque, por algum motivo, o acompanhamento do paciente foi interrompido ou porque o paciente faleceu por uma causa diferente da pesquisa científica. Na literatura, três tipos comuns de esquemas de censura são considerados :

1. Censura Tipo I: Muitos estudos têm fundos limitados, e os pesquisadores não podem esperar até que todos os indivíduos desenvolvam o evento de interesse. Assim, é aceitável observar por um período fixo, por exemplo, nove meses ou cinco anos. Os tempos de sobrevivência dos indivíduos que desenvolveram o desfecho de interesse durante o período fixo são observações não censuradas. Os tempos de sobrevivência dos indivíduos que ainda estão vivos no final do estudo não são conhecidos exatamente, mas

são registrados como tendo a duração total do estudo. Na censura do tipo I, todos os indivíduos que não apresentam o evento até o final do estudo têm seus tempos de sobrevivência censurados no instante correspondente ao término do acompanhamento, o que padroniza os dados de censura e simplifica a estimação de parâmetros de sobrevivência.

2. **Censura Tipo II:** Aparece geralmente em estudos laboratoriais com animais. O evento de interesse é frequentemente o desfecho fatal. O estudo prossegue até que uma proporção ou número fixo de indivíduos morra, por exemplo, 45% da amostra ou 120 indivíduos. Nesse esquema, o tempo de censura corresponde ao maior tempo de sobrevivência observado entre os indivíduos que não experimentaram o evento.
3. **Censura Tipo III:** Em muitos estudos clínicos e epidemiológicos, o período de acompanhamento é fixo em calendário, e pacientes são incluídos no estudo em momentos diferentes. Alguns indivíduos podem desenvolver o evento antes do término do estudo, outros podem se retirar ou se perder durante o acompanhamento, e outros ainda podem não apresentar o evento até o final do período de acompanhamento. Nesse contexto, os tempos de censura variam entre os indivíduos e, portanto, comportam-se como variáveis aleatórias.

O tipo mais comum de censura apresentado em muitos estudos é uma censura à direita. Esse fato está relacionado ao fato de que muitos estudos têm recursos limitados e os pesquisadores não podem esperar até que todos os sujeitos desenvolvam o evento de interesse. Matematicamente, esse esquema de censura pode ser definido usando duas verificações examinadas, sendo T variável a associada ao tempo até a falha de um paciente e C uma variável pesquisada independente de T , relacionada ao tempo de censura. Nesse caso, o tempo observado e o esquema de censura são dados por:

$$\begin{aligned} t &= \min(T, C) \\ \delta &= \begin{cases} 1, & \text{se } T \leq C \\ 0, & \text{se } T > C \end{cases} \end{aligned} \quad (7)$$

Por outro lado, o esquema de censura à esquerda também pode ser definido usando duas variáveis avaliadas, como no esquema de censura à direita. Nesse caso, t é igual a T se o tempo de vida para coletado e δ indica se o tempo exato de vida T foi coletado ($\delta = 1$) ou não ($\delta = 0$). Portanto, para a censura à esquerda, os dados coletados são dados por:

$$t = \max(T, C) \quad (8)$$

$$\delta = \begin{cases} 1, & \text{se } T \geq C \\ 0, & \text{se } T < C \end{cases}$$

É importante ressaltar que, ao não considerar a presença de observações censuradas, as técnicas clássicas, como a análise de regressão padrão com erros normais e dados geralmente transformados, poderiam ser usadas na análise de dados de sobrevivência. Entretanto, essas técnicas geralmente baseadas na suposição de normalidade ou métodos não paramétricos podem ser inconvenientes na presença de investigações censuradas ou taxas de cura. Assim, novas técnicas têm sido apresentadas na literatura para lidar com observações censuradas ou taxas de cura. A técnica mais popular é o modelo de riscos proporcionais de Cox (Cox, 1972); modelos de longo prazo (também chamados de modelos de taxa de cura) e modelos de fragilidade (Price e Manatunga, 2001).

5.7 Estimador de *Kaplan–Meier*

Em estudos de sobrevida, observa-se com frequência que a curva de sobrevivência relativa tende a atingir um platô após determinado período de acompanhamento. De acordo com Lambert *et al.* (2007), esse efeito de platô ocorre quando a taxa de mortalidade dos indivíduos doentes se iguala à taxa de mortalidade esperada na população geral, ou seja, quando a mortalidade em excesso se torna nula. Tal comportamento sugere a presença de uma fração de cura na população-alvo um subgrupo de pacientes considerados efetivamente livres do risco associado à doença. No contexto de IC, esse fenômeno também pode ocorrer, particularmente entre pacientes com resposta terapêutica prolongada ou estabilização clínica sustentada. A identificação dessa possível fração de pacientes “curados” demanda o uso de abordagens analíticas adequadas, sendo a análise não-paramétrica de *Kaplan–Meier* uma etapa inicial fundamental. Nesse sentido, a função de sobrevivência dos pacientes com IC foi estimada por meio do estimador de *Kaplan–Meier*, utilizado em estudos clínicos com dados censurados. Para definir este estimador, sejam $t_1 < \dots < t_m$ os instantes observados de ocorrência do evento (óbito). Para cada tempo t_j , define-se d_j como número de óbitos observados nesse instante e r_j como o número de pacientes em risco imediatamente antes de t_j . O *estimador de Kaplan–Meier* para a função de sobrevivência, denotado por $\hat{S}(t)$, é dado por:

$$\hat{S}(t) = \prod_{t_j \leq t} \left(1 - \frac{d_j}{r_j}\right) \quad (9)$$

5.8 Modelo de Riscos Proporcionais de Cox

Com o objetivo de avaliar de forma conjunta o efeito de múltiplas covariáveis clínicas e laboratoriais sobre o risco de óbito em pacientes com IC, foi ajustado, como segunda parte da análise, o modelo de riscos proporcionais de Cox. Esse modelo, de natureza semi-paramétrica, permite estimar a função de risco condicional, definida por:

$$h(X) = h_0(t)e^{\beta^T X} \quad (10)$$

em que $h_0(t)$ é a função de risco *baseline*, e β é o vetor de coeficientes associados às covariáveis X . A estimação do vetor de parâmetros β é realizada pela maximização da função verossimilhança parcial definida por:

$$L(\beta) = \prod_{j=1}^m \frac{e^{\beta^T X_j}}{\sum_{i \in R_j} e^{\beta^T X_i}} \quad (11)$$

em que X_j representa o vetor de covariáveis do indivíduo que faleceu no tempo t_j , e R_j é o conjunto de indivíduos em risco imediatamente antes de t_j . Após o ajuste do modelo de riscos proporcionais de Cox, foram examinados os resíduos de Martingale com o objetivo de avaliar a adequação da suposição de linearidade nas covariáveis contínuas. A presença de padrões sistemáticos na relação entre esses resíduos e as covariáveis indicaria possíveis violações desse pressuposto, sugerindo a necessidade de transformações apropriadas, como a inclusão de termos polinomiais ou o uso de funções suaves. Os resíduos de Deviance foram inspecionados com vistas à detecção de observações influentes ou potenciais outliers, cuja presença pode afetar substancialmente a estabilidade das estimativas dos coeficientes.

5.9 Modelos de Fração de Cura

Modelos de fração de cura são uma ferramenta valiosa na análise de sobrevivência, especialmente quando nem todos na população estudada estão sob risco do evento de interesse (Lambert, 2007). Diferente dos modelos tradicionais que esperam que todos eventualmente experienciem o evento, os modelos de fração de cura consideram a possibilidade de que uma parte da população seja "curada", ou seja, não mais suscetível ao desfecho. Quando isso ocorre, as curvas de sobrevivência de *Kaplan-Meier* (Kaplan; Meier, 1958) tendem a exibir platôs, indicando que uma parcela significativa de pacientes, após um certo momento, não está mais em risco. Por exemplo, no contexto clínico da insuficiência cardíaca, essa abordagem é importante para descrever a sobrevida de pacientes que respondem bem ao tratamento e alcançam um estado de estabilidade clínica duradoura.

Formalmente, os modelos de fração de cura assumem que a população estudada é composta por dois subgrupos:

(I) um grupo curado com probabilidade p de ocorrência, que nunca experienciará o evento;

(II) e um grupo suscetível com probabilidade $1 - p$, que permanece em risco e cujo tempo até o evento segue uma distribuição especificada.

Entre outras abordagens, a função de sobrevivência populacional $S(t)$ pode então ser expressa de duas formas distintas, de acordo com o paradigma adotado. Estas duas formas são chamadas de modelos de *mistura* e *não-mistura* (Martinez *et al.*, 2013). Para este estudo, em particular, será considerado o primeiro.

5.9.1 Modelo de Mistura

Nos modelos de mistura, a função de sobrevivência populacional é expressa como uma combinação entre a fração de curados e a função de sobrevivência dos suscetíveis:

$$S(t) = p + (1 - p)S_0(t) \quad (12)$$

em que $S(t)$ é a função de sobrevivência; $0 < p < 1$ é a proporção de curados; $S_0(t)$ representa a função de sobrevivência entre os indivíduos suscetíveis, isto é, a função *baseline*. Nesta formulação, observa que, se $S_0(t)$ converge a zero quando t tende a infinito, então $S(t)$ converge

a p nesta situação. A interpretação estatística e clínica do parâmetro p é direta: trata-se da estimativa da proporção de indivíduos potencialmente livres do evento, mesmo em longo prazo (Martinez *et al.*, 2013).

5.9.2 Modelo de Não-Mistura padrão

Os modelos de não-mistura partem de uma estrutura alternativa, onde todos os indivíduos compartilham um mesmo processo latente de risco, mas a ocorrência do evento depende da ativação de um ou mais fatores internos de progressão. Nessa abordagem, a função de sobrevivência assume a forma:

$$S(t) = \exp\{-\theta F(t)\}, \quad \theta > 0 \quad (13)$$

em que $F(t)$ é uma função de distribuição acumulada arbitrária associada ao tempo até o evento; θ é um parâmetro positivo que controla a intensidade da suscetibilidade populacional (Tsodikov, 2002).

Nesse caso, a fração de cura não aparece explicitamente como um parâmetro p , mas pode ser inferida pela tendência assintótica da função de sobrevivência:

$$p = S(t) = \exp(-\theta) \quad (14)$$

sendo $p = \exp(-\theta)$ a estimativa da fração de indivíduos livres do evento. Essa modelagem é essencial em situações em que se deseja capturar a multiplicidade de mecanismos que podem levar ao evento — por exemplo, múltiplos caminhos fisiopatológicos para descompensação em pacientes com insuficiência cardíaca (Tsodikov, 2002).

Outras abordagens, que não serão incluídas nessa dissertação, são baseadas em distribuições defectivas, ou seja, distribuições de probabilidade cuja área sob a curva densidade não soma 1 e $S(t)$ não converge a zero, quando $t \rightarrow \infty$.

5.10 Distribuições de Probabilidade (*Baselines*)

Para a análise considerada neste estudo, serão adotados, como *funções basais* seis modelos de probabilidade, escolhidos de acordo com sua flexibilidade e capacidade

interpretativa. Tais modelos são: Weibull, Gama, Lindley, Gumbel, Exponencial e Rayleigh. Os detalhes de cada um dos modelos são descritos a seguir.

(a) Distribuição Weibull

A distribuição Weibull foi proposta em 1951 por Waloddi Weibull (Oliveira, 2015). Essa distribuição, em particular, é conhecida por seu vasto campo de aplicações em praticamente todas as áreas da ciência. No contexto de análise de sobrevivência, diz-se que uma variável aleatória T segue uma distribuição Weibull se sua função de sobrevivência for definida por:

$$S_0(t) = \exp \left[- \left(\frac{t}{\gamma} \right)^\alpha \right], \quad t \geq 0 \quad (15)$$

em que $\gamma > 0$ é o parâmetro de escala, e $\alpha > 0$ é o parâmetro de forma. A forma funcional da Weibull permite modelar riscos crescentes ($\alpha > 1$), decrescentes ($\alpha < 1$) ou constantes ($\alpha = 1$, equivalente à distribuição exponencial), oferecendo grande flexibilidade para representar o comportamento do risco de mortalidade ao longo do tempo. Nos modelos de mistura com fração de cura, essa função é incorporada da seguinte maneira:

$$S(t) = p + (1 - p) \exp \left[- \left(\frac{t}{\gamma} \right)^\alpha \right] \quad (16)$$

em que $0 < p < 1$ representa a fração de indivíduos curados (ou seja, que jamais apresentarão o evento – parâmetro de fração de cura). A escolha da distribuição Weibull, neste caso, se justifica pela natureza dos dados observacionais e pela presença de indícios empíricos de heterogeneidade no risco de morte entre os pacientes. A incorporação da fração de cura nesse modelo, em particular, possibilita capturar nuances clínicas importantes, como a identificação de um subgrupo de pacientes com insuficiência cardíaca que pode ter alcançado um estado de recuperação sustentada (Oliveira, 2015).

(b) Distribuição Gama

Diz-se que uma variável aleatória T segue uma distribuição Gama quando sua função de sobrevivência é definida em termos de sua função de distribuição acumulada, a qual depende de dois parâmetros principais: o parâmetro de forma e o parâmetro de escala. Esses parâmetros conferem flexibilidade à distribuição, permitindo representar diferentes padrões de risco ao longo do tempo. Isto é,

$$S_0(t) = 1 - \frac{1}{\Gamma(k)} \gamma\left(k, \frac{t}{\theta}\right), \quad t \geq 0 \quad (17)$$

em que $\gamma(\cdot)$ é a função gama incompleta superior, $k > 0$ é o parâmetro de forma, e $\theta > 0$ é o parâmetro de escala. No modelo de fração de cura, tem-se que:

$$S(t) = 1 - \frac{(1-p)}{\Gamma(k)} \gamma\left(k; \frac{t}{\theta}\right) \quad (18)$$

em que $0 < p < 1$ representa a fração de indivíduos curados (Lawless, 1982). A escolha da distribuição Gama se justifica pela sua capacidade de modelar tempos até o evento com alta variabilidade, o que é relevante em populações heterogêneas de pacientes com insuficiência cardíaca, em que a evolução clínica pode variar.

(c) Distribuição Lindley

A distribuição Lindley foi proposta por D. V. Lindley em 1958, inicialmente no âmbito da estatística bayesiana. Desde então, passou a ser explorada em diferentes contextos devido à sua simplicidade matemática e capacidade de modelar tempos de vida e dados positivos com assimetria. Na análise de sobrevivência, considera-se que uma variável aleatória T segue uma distribuição Lindley quando sua função de sobrevivência é expressa a partir de uma combinação entre componentes exponenciais e gama, o que a torna uma alternativa interessante a modelos clássicos. Sua função de sobrevivência é descrita por:

$$S_0(t) = \frac{\theta + 1 + \theta t}{\theta + 1} e^{-\theta t}, \quad t \geq 0 \quad (19)$$

em que o parâmetro $\theta > 0$ está associado à escala (ou taxa). Para a construção do modelo de fração de cura, tem-se que:

$$S(t) = p + (1 - p) \frac{\theta + 1 + \theta t}{\theta + 1} e^{-\theta t} \quad (20)$$

em que $0 < p < 1$ representa a fração de indivíduos curados (Lindley, 1958). O uso da Lindley se justifica pela sua capacidade de captar padrões crescente de risco; por exemplo, pacientes com evolução rápida para óbito, algo compatível com a heterogeneidade observada em insuficiência cardíaca.

(d) Distribuição Gumbel

A distribuição Gumbel foi introduzida por Emil Julius Gumbel em 1935, no contexto da teoria dos valores extremos. Essa distribuição é utilizada, em particular, na modelagem do comportamento de máximos ou mínimos de variáveis aleatórias, sendo aplicada em áreas como hidrologia, engenharia, meteorologia e análise de riscos. No âmbito da análise de sobrevivência, considera-se que uma variável aleatória T segue uma distribuição Gumbel quando sua função de sobrevivência é definida a partir de uma forma exponencial, capaz de capturar caudas longas e assimetria. Assim,

$$S_0(t) = 1 - \exp \left\{ - \exp \left(- \frac{t - \mu}{\beta} \right) \right\}, \quad t \geq 0 \quad (21)$$

em que μ é o parâmetro de locação e $\beta > 0$ o parâmetro de escala. Essa forma permite modelar eventos ligados a máximos ou extremos, em que o risco é governado por situações críticas. No contexto de fração de cura, tem-se que:

$$S(t) = 1 - (1 - p) \exp \left\{ - \exp \left(- \frac{t - \mu}{\beta} \right) \right\} \quad (22)$$

em que $0 < p < 1$ representa a fração de indivíduos curados (Gumbel, 1960). Sua utilização é justificada pela possibilidade de representar padrões de risco associados a episódios extremos na evolução clínica, como descompensações súbitas e agudas na insuficiência cardíaca.

(e) Distribuição Exponencial

A distribuição Exponencial foi introduzida no início do século XX e consolidou-se como um dos modelos mais fundamentais e essenciais na teoria da confiabilidade e da análise de sobrevivência. Sua simplicidade matemática e interpretação direta conferem-lhe grande popularidade em aplicações práticas. No contexto da análise de sobrevivência, em particular, diz-se que uma variável aleatória T segue uma distribuição Exponencial quando sua função de sobrevivência assume forma decrescente exponencial, caracterizada por um parâmetro de taxa constante. Isto é,

$$S_0(t) = e^{-\lambda t}, \quad t \geq 0 \quad (23)$$

sendo λ o parâmetro de taxa. Essa forma funcional implica que o risco instantâneo (*hazard rate*) é constante ao longo do tempo, o que torna a distribuição Exponencial adequada para modelar situações em que a probabilidade de falha ou ocorrência de um evento não depende da duração já observada. No contexto de fração de cura, tem-se que:

$$S(t) = p + (1 - p) e^{-\lambda t} \quad (24)$$

em que $0 < p < 1$ representa a fração de indivíduos curados (Kalbfleisch; Prentice, 2002). Uma nota importante é que a distribuição Exponencial é um caso particular da distribuição Weibull quando $\alpha = 1$, servindo como *modelo de referência*. Embora mais restritiva, sua

simplicidade facilita a interpretação clínica, quando adequada aos dados, sobretudo em situações em que o risco não apresenta variação significativa ao longo do tempo.

(f) Distribuição Rayleigh

A distribuição Rayleigh foi introduzida por Lord Rayleigh em 1888, no estudo de fenômenos relacionados à propagação de ondas acústicas. Desde então, passou a ser amplamente empregada em diversas áreas, incluindo engenharia, comunicações, física e análise de confiabilidade. No contexto da análise de sobrevivência, considera-se que uma variável aleatória T segue uma distribuição Rayleigh quando sua função de sobrevivência é expressa em termos de uma forma exponencial quadrática, dependente de um parâmetro de escala. Isto é,

$$S_0(t) = \exp\left(-\frac{t^2}{2\sigma^2}\right), \quad t \geq 0 \quad (25)$$

em que σ é o parâmetro de escala. Nesse caso, o risco instantâneo aumenta linearmente com o tempo, o que torna a distribuição Rayleigh especialmente apropriada para modelar tempos de vida em situações nas quais a taxa de falha cresce de maneira progressiva. No modelo de fração de cura, tem-se:

$$S(t) = p + (1 - p) \exp\left(-\frac{t^2}{2\sigma^2}\right) \quad (26)$$

em que $0 < p < 1$ representa a fração de indivíduos curados (Roy, 2004). Sua escolha se justifica para modelar cenários em que a deterioração clínica se intensifica com o tempo, o que pode ocorrer em pacientes com insuficiência cardíaca em estágio avançado.

5.11 Inferência em análise de sobrevivência

A inferência estatística em análise de sobrevivência tem como propósito estimar os parâmetros dos modelos adotados, testar hipóteses e quantificar a incerteza associada a essas estimativas. Esse processo, em geral, envolve duas etapas principais: estimação de parâmetros e avaliação de ajuste. Entre os métodos disponíveis, o método da Máxima Verossimilhança

(MV) é amplamente empregado devido à sua flexibilidade e propriedades assintóticas desejáveis, como consistência e normalidade assintótica (Giolo & Colosimo, 2016; Lawless, 1982). O princípio da MV consiste em determinar o conjunto de parâmetros que maximiza a probabilidade de observar os dados obtidos, levando em consideração a presença de censura.

A função de verossimilhança é válida para mecanismos de censura dos tipos I e II, desde que se assuma que a censura seja não-informativa, ou seja, que o processo de censura não forneça informações adicionais sobre os parâmetros do modelo. Em outras palavras, a probabilidade de um indivíduo ser censurado deve ser independente do tempo real até a ocorrência do evento. Essa validade também se estende a mecanismos de censura aleatória, sob a mesma suposição de não-informatividade.

Para dados de sobrevivência com censura à direita, a função de verossimilhança pode ser escrita como:

$$L(\theta) = \prod_{i=1}^n [f(t_i | \theta)]^{\delta_i} [S(\theta)]^{1-\delta_i} \quad (27)$$

em que t_i é o tempo observado (evento ou censura) para o indivíduo i ; δ_i é o indicador de falha ($\delta_i = 1$ se o evento ocorreu e $\delta_i = 0$ se censurado); $f(\cdot)$ é a função densidade de probabilidade; $S(\cdot)$ é a função de sobrevivência; θ é o vetor de parâmetros a serem estimados.

A incerteza das estimativas pode ser quantificada pela matriz de informação de Fisher, permitindo construir intervalos de confiança assintóticos e aplicar testes de hipóteses. Para os modelos de fração de cura, a estimação via MV é direta, incorporando a formulação específica do modelo (de mistura) e a distribuição *baseline* assumida para os tempos de falha na Equação (26). Quando necessário, métodos numéricos iterativos, como o algoritmo de Newton-Raphson, são utilizados para maximizar a função de verossimilhança (Lawless, 1982).

Neste estudo, em particular, todo o processo de estimação foi conduzido por meio do *software* R, versão 4.5.1 (R Core Team, 2025), executado em ambiente *Windows* 11 (64 bits). As configurações adotadas consistiram na implementação de verossimilhança (Equação 26) para cada um dos modelos de fração de propostos, no uso do método de otimização de *Nelder-Mead* para determinar os estimadores dos parâmetros dos modelos via máxima verossimilhança e, também, a informação de *Fisher*. E, para finalizar o processo, foi considerado como critério de parada do algoritmo um número de iteração máximo de 1000 iterações, sendo a última iteração adotada como estimativa final dos parâmetros.

5.12 Diagnóstico do Modelo: Resíduos de *Cox-Snell*

Após a estimação dos parâmetros em modelos de sobrevivência, como o modelo de riscos proporcionais de Cox ou modelos paramétricos, a avaliação da adequação do ajuste (*goodness-of-fit*) é um passo fundamental. Os resíduos de *Cox-Snell* são uma ferramenta diagnóstica particularmente útil para este fim, pois permitem verificar se o modelo proposto captura de forma satisfatória a estrutura de risco dos dados (Cox & Snell, 1968).

A construção desses resíduos baseia-se na teoria de que, se o modelo for corretamente especificado, o tempo de sobrevida de cada indivíduo, quando transformado pela função de risco cumulativo estimada do modelo, deverá seguir uma distribuição exponencial padrão com taxa de risco igual a 1 (Cox & Snell, 1968). Em outras palavras, o resíduo de *Cox-Snell* para o i -ésimo indivíduo, r_{C_i} , é definido pela função de risco cumulativo (hazard cumulativo) estimada do modelo $\hat{H}(t_i)$:

$$r_{C_i} = \hat{H}(t_i) = H_0(t_i) \exp(x_i^T \hat{\beta}) \quad (28)$$

em que t_i é o tempo de sobrevida observado, x_i é o vetor de covariáveis e $\hat{\beta}$ são os coeficientes estimados. A avaliação do ajuste é realizada inspecionando a distribuição desses resíduos. Para um modelo bem ajustado, os r_{C_i} devem se comportar como uma amostra aleatória de uma distribuição exponencial. Isso é verificado por meio de um gráfico de probabilidade cumulativa (*cumulative hazard plot*). Neste gráfico, a função de risco cumulativo dos resíduos, estimada por meio de um método não-paramétrico (como a estimativa de Nelson-Aalen), é plotada em função dos próprios resíduos. Se o modelo estiver bem ajustado, a curva resultante deverá se aproximar de uma linha reta com inclinação de 45 graus (eixo y igual ao eixo x), passando pela origem. Desvios significativos dessa linha reta indicam que o modelo não está adequadamente ajustado, sugerindo que a forma funcional da função de risco base ou a relação com as covariáveis pode estar incorreta.

Quando se considera a presença de uma fração de cura no modelo de sobrevivência, a definição e a interpretação dos resíduos de Cox-Snell precisam ser modificadas, pois parte da população é considerada imune ao evento de interesse, e, portanto, o pressuposto de que todos os indivíduos falham em algum instante não é válido. Nesse cenário, o tempo de sobrevivência observado t_i é modelado condicionalmente à probabilidade de suscetibilidade, e o resíduo de Cox-Snell deve incorporar a estrutura do modelo de cura. Para um indivíduo i , o resíduo passa

a ser definido em termos da função de risco cumulativo condicional para os suscetíveis, ajustada pela probabilidade de cura, isto é,

$$r_{C_i} = -\log \left(\frac{\hat{S}(t_i)}{\hat{p}_i + (1 - \hat{p}_i)\hat{S}_0(t)\exp(x_i^T \hat{\beta})} \right) \quad (29)$$

onde $\hat{S}(t_i)$ é a função de sobrevivência estimada, $\hat{p}_i$ representa a probabilidade estimada de cura para o indivíduo i , $\hat{S}_0(t)$ é a função de sobrevivência *baseline*, e $\hat{\beta}$ são os coeficientes estimados do modelo. A interpretação gráfica, neste caso, também sofre alteração. Como a presença da fração de cura implica que a função de sobrevivência não tende a zero, mas sim a um platô correspondente à proporção de curados, a função de risco cumulativo dos resíduos não seguirá uma distribuição exponencial padrão para toda a amostra. A avaliação do ajuste deve, portanto, ser restrita ao subconjunto de indivíduos suscetíveis. Para estes, a análise gráfica segue de maneira análoga ao caso clássico: a curva estimada do risco cumulativo dos resíduos deve se aproximar da reta com inclinação de 45 graus.

5.13 Estrutura de Regressão

Uma característica importante dos modelos de fração de cura é a de que eles podem ser estendidos para a inclusão de covariáveis tanto no parâmetro de cura p quanto nos parâmetros da função de sobrevivência dos indivíduos suscetíveis. Essa incorporação, em particular, permite avaliar como características clínicas e laboratoriais dos pacientes influenciam a probabilidade de pertencimento ao grupo curado e, simultaneamente, o risco de morte entre os não curados. Essa flexibilidade é relevante em contextos de elevada heterogeneidade clínica, como a insuficiência cardíaca, em que fatores como idade, função renal e fração de ejeção podem modificar substancialmente a evolução dos pacientes (Colosimo; Giolo, 2006).

A modelagem da fração de cura é geralmente realizada por meio de uma estrutura de regressão logística. Seja p_i a probabilidade de cura para o i -ésimo indivíduo, condicionado ao vetor de covariáveis $x = (x_1, \dots, x_{ki})'$. Define-se:

$$\ln \left(\frac{p_i}{1 - p_i} \right) = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} \quad (30)$$

em que $\beta_0, \beta_1, \dots, \beta_k$ são os parâmetros de regressão a serem estimados. Essa formulação assegura que $0 < p_i < 1$, garantindo interpretação probabilística adequada (Colosimo; Giolo, 2006). Além dessa inclusão, é possível incluir covariáveis nos parâmetros da função de sobrevivência dos suscetíveis. Por exemplo, em um modelo de fração de cura Weibull, a função de sobrevivência condicional pode ser expressa como:

$$S(t | x_i, z_i) = p_i + (1 - p_i) \exp\{-(\lambda t)^{\alpha \exp(z_i^T \gamma)}\} \quad (31)$$

em que p_i : proporção de curados estimada via regressão logística; z_i : covariáveis associadas ao risco dos suscetíveis; λ, α : parâmetros de escala e forma da Weibull; γ : vetor de parâmetros do efeito das covariáveis. Dessa forma, a estrutura de regressão permite capturar dois aspectos distintos da evolução clínica:

- (i) a chance de estar livre do risco de óbito dentro do período de acompanhamento, e
- (ii) a velocidade/probabilidade com que o evento ocorre entre os suscetíveis.

Tal abordagem é fundamental em insuficiência cardíaca, pois possibilita identificar subgrupos de pacientes que alcançam estabilização sustentada, ao mesmo tempo em que se avalia o impacto de covariáveis prognósticas sobre aqueles ainda em risco (Colosimo; Giolo, 2006).

5.14 Considerações Éticas

O presente estudo constitui uma reanálise de dados secundários provenientes do artigo “*Survival analysis of heart failure patients: A case study*” publicado no periódico *PLOS ONE* (Ahmad et al., 2017). O banco de dados original foi composto por 299 pacientes acompanhados em um hospital universitário no Paquistão, todos com insuficiência cardíaca sistólica em classes funcionais III e IV da NYHA.

No estudo original, os autores obtiveram aprovação ética institucional junto ao *Institutional Review Board of Government College University, Faisalabad – Pakistan*, e declararam seguir os princípios da Declaração de Helsinque. Além disso, o consentimento informado foi obtido diretamente dos pacientes, garantindo respeito à autonomia e confidencialidade das informações clínicas coletadas.

Para a presente dissertação, a utilização dos dados se caracteriza como pesquisa de base secundária, não envolvendo contato direto com seres humanos nem coleta de novas

informações. Os dados encontram-se disponibilizados em acesso aberto, conforme política editorial da revista *PLOS ONE*, de modo que a reanálise aqui conduzida respeita a legislação nacional e internacional aplicável à ética em pesquisa.

Este trabalho está em conformidade com a Resolução nº 466/2012 do Conselho Nacional de Saúde do Brasil, que disciplina pesquisas em saúde envolvendo seres humanos. Como se trata de reuso de dados anonimizados e previamente publicados, não se configura necessidade de novo consentimento informado, tampouco de submissão ao Comitê de Ética em Pesquisa da instituição, segundo as normativas do Sistema CEP/CONEP para pesquisas exclusivamente documentais.

Reforçamos que todas as etapas de análise estatística e apresentação dos resultados foram conduzidas de forma a preservar a integridade, privacidade e confidencialidade dos pacientes incluídos no estudo original.

6 RESULTADOS E DISCUSSÃO

6.1 Contextualização e descrição dos dados

A IC é uma síndrome clínica de elevada prevalência e complexidade. Seu desfecho clínico pode ser agravado por idade avançada, alterações renais, desequilíbrios eletrolíticos, diabetes mellitus e hipertensão arterial, por exemplo. A redução da FEVE constitui um dos principais marcadores de gravidade. Nesse contexto, identificar fatores associados ao risco de mortalidade é fundamental para o direcionamento de intervenções clínicas e para a melhoria dos prognósticos (Heidenreich *et al.*, 2022).

O tempo de seguimento dos 299 pacientes acompanhados variou de 4 a 285 dias. Ao longo do acompanhamento, 96 pacientes (32%) evoluíram a óbito e 68% dos pacientes permaneceram vivos, caracterizando a proporção censurada. As análises incluíram variáveis demográficas (idade, sexo), clínicas (diabetes, hipertensão, anemia e tabagismo) e laboratoriais (creatinina sérica, sódio, contagem plaquetária e CPK, além da FEVE).

6.2 Análise de qualidade e caracterização dos dados

Foi realizada uma avaliação da completude do banco de dados, com atenção à presença de valores ausentes. Nenhuma das variáveis apresentou valores ausentes, evidenciando a completude e consistência do banco de dados (Little & Rubin, 2019).

Na sequência, as variáveis foram exploradas por meio de estatística descritiva, permitindo uma caracterização detalhada do perfil clínico e laboratorial dos pacientes. Os resultados para as 299 amostras mais os valores de referência estão sintetizados na Tabela 2.

Tabela 2 — Estatística descritiva e valores de referência.

Variável (Unidade)	Média	DP	Mediana	Q1	Q3	Mín	Máx	Valores de referência
Idade (anos)	60,83	11,89	60,00	51,00	70,00	40,00	95,00	-
Fração de Ejeção	38,08	11,83	38,00	30,00	45,00	14,00	80,00	≥ 55% (Lang <i>et al.</i> , 2015).
Sódio (mEq/L)	136,63	4,41	137,00	134,00	140,00	113,00	148,00	136 – 145 (Braun <i>et al.</i> , 2015).
log(Creatinina) (mg/dL)	0,82	0,29	0,74	0,64	0,88	0,41	2,34	≤ 1.2 (KDIGO, 2013).
Log(CPK) (U/L)	5,67	1,13	5,53	4,77	6,37	3,18	8,97	H: 39-308; M: 26-192 (Aujla <i>et al.</i> , 2025).
log(Plaquetas) (mil/mm ³)	12,41	0,40	12,48	12,27	12,62	10,13	13,65	150 - 450 x 10 ³ /mm ³ (Hoffman <i>et al.</i> , 2018).

Legenda: DP: desvio padrão; Q1: primeiro quartil; Q3: terceiro quartil; mín: mínimo; máx: máximo.

Fonte: Elaboração própria com base nos dados do estudo.

Os resultados apresentados na Tabela 2 sugerem variabilidade entre as características clínicas e laboratoriais, o que elenca a diferença entre casos clínicos, frequentemente encontrada em pessoas com IC. A idade variou de 40 a 95 anos, com mediana de 60 anos, apontando um perfil predominantemente idoso. Esse achado condiz com a literatura, que aponta a maior prevalência da doença em idosos, em função do acúmulo de fatores de risco cardiovasculares ao longo da vida (Freitas; Batlouni; Gamarsky, 2012).

A fração de ejeção apresentou mediana de 38%, com grande dispersão, evidenciando a presença de pacientes com ICFer (<40%) e também uma parcela com ICFEp (≥50%). Essa diversidade reforça a importância de estratificação fenotípica em pesquisas clínicas, levando a equipe assistencial a pensar que cada paciente exige cuidados individualizados.

O sódio sérico tem valor de referência aproximado para adultos 136 – 145 mEq/L (Braun; Barstow; Pyzocha, 2015). Embora a mediana tenha permanecido dentro da normalidade, observou-se valor mínimo de 113 mEq/L, indicando que parte dos pacientes

apresentava hiponatremia. Hiponatremia é definida como concentração sérica de sódio inferior a 135 mEq/L, conforme descrito por Braun, Barstow e Pyzocha (2015). Essa alteração é frequente em pacientes com IC em virtude da retenção hídrica mediada pela ativação neuro-hormonal e está associado a pior prognóstico, em alguns casos (Spinetti, 2019).

A creatinina sérica possui valor de referência aproximado de $\leq 1,2$ mg/dL para adultos (KDIGO, 2013). No presente estudo, a variável foi transformada para a escala logarítmica com o objetivo de reduzir assimetria e melhorar o ajuste dos modelos paramétricos. Após a transformação, os valores apresentaram média de 0,82 e mediana de 0,74 na escala logaritmada, com amplitude variando de 0,41 a 2,34. Embora esses valores não correspondam diretamente à unidade mg/dL, a dispersão observada sugere a presença de pacientes com níveis elevados de creatinina na escala original, o que pode indicar disfunção renal. A função renal comprometida está associada a pior prognóstico, maior tempo de hospitalização e maior risco de mortalidade em pacientes com IC (Kang *et al.*, 2022).

A CPK foi analisada após transformação logarítmica. A mediana foi de 5,53, com valores variando entre 3,18 e 8,97, indicando marcada variabilidade entre os pacientes. Embora esses valores não possam ser diretamente comparados às faixas de referência clínicas (39–308 U/L para homens e 26–192 U/L para mulheres), a amplitude observada sugere a presença de indivíduos com CPK elevada. Níveis aumentados estão associados a injúria muscular, disfunção miocárdica e até complicações renais agudas, sendo um marcador importante do metabolismo energético e cardíaco (Aujla; Zubair; Patel, 2025).

O número de plaquetas também foi transformado para a escala logarítmica. Os valores de referência para adultos situam-se entre 150 e $450 \times 10^3/\text{mm}^3$ (Hoffman *et al.*, 2017). Na escala logaritmada, os valores apresentaram mediana de 12,48, com amplitude entre 10,13 e 13,65. A dispersão observada sugere que a maior parte dos indivíduos provavelmente apresentava contagens plaquetárias dentro da faixa de normalidade clínica. Alterações mais extremas podem se justificar por processos inflamatórios ou distúrbios hematológicos, principalmente nos estágios avançados da doença (Cardoso *et al.*, 2010).

Na tabela 3 está descrita a distribuição das variáveis categóricas da amostra, bem como os valores de referência.

Tabela 3 – Distribuição das variáveis categóricas da amostra (n = 299).

Variável	Categoria	n	%	Critério de Definição / Valores de Referência	Fonte
Gênero	0 = Feminino	105	35,1%	0 = Feminino; 1 = Masculino	Ahmad <i>et al.</i> , 2017
	1 = Masculino	194	64,9%		
Fumante	0 = Não	203	67,9%	Classificação baseada em anotações médicas (1 = fumante atual).	Ahmad <i>et al.</i> , 2017
	1 = Sim	96	32,1%		
Diabetes	0 = Não	174	58,2%	Diagnóstico prévio de Diabetes Mellitus e/ou uso de hipoglicemiantes orais/insulina.	American Diabetes Association, 2015
	1 = Sim	125	41,8%		
Hipertensão	0 = Não	194	64,9%	Diagnóstico prévio de Hipertensão Arterial Sistêmica (HAS) e/ou uso de anti-hipertensivos.	Williams <i>et al.</i> , 2018
	1 = Sim	105	35,1%		
Anemia	0 = Não	170	56,9%	Hematócrito < 36%.	McClellan <i>et al.</i> , 2002
	1 = Sim	129	43,1%		

Fonte: Elaboração própria com base nos dados do estudo.

A amostra descreve um perfil clínico frequentemente observado em pacientes com IC, sendo majoritariamente composta por indivíduos do sexo masculino (64,9%). Este predomínio está alinhado com a literatura, que reporta uma maior prevalência da síndrome em homens, possivelmente refletindo diferenças fisiopatológicas, de exposição a fatores de risco ou de acesso ao diagnóstico (Ahmad *et al.*, 2017).

A análise das comorbidades destaca a diabetes mellitus, estando presente em 41,8% da coorte, um achado consistente com seu papel estabelecido como fator de risco independente e agravante do prognóstico na IC (American Diabetes Association, 2015). A hipertensão arterial sistêmica (HAS) foi identificada em 35,1% dos pacientes (Williams *et al.*, 2018). Vale ressaltar

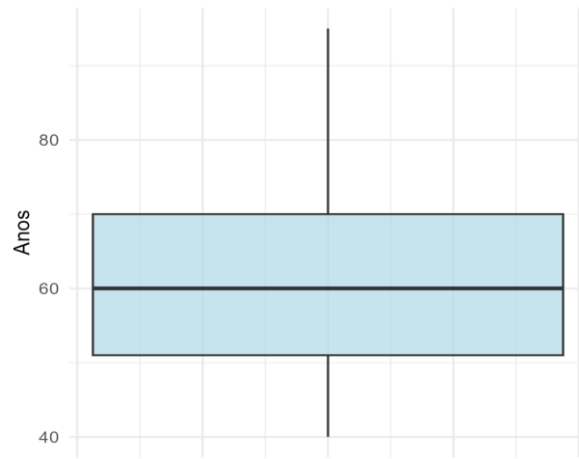
que a HAS não foi classificada em estádios, foi registrada presença ou ausência da comorbidade. A amostra estudada não representa um ambulatório típico e sim pacientes internados com IC aguda, onde a etiologia pode não ser exclusivamente hipertensiva.

Observou-se prevalência de anemia (43,1%), definida por um hematócrito inferior a 36%, parâmetro associado à síndrome cardiorrenal e a processos inflamatórios crônicos (McClellan *et al.*, 2002). A prevalência de tabagismo foi de 32,1%, fator de impacto na mortalidade em estágios mais avançados da doença.

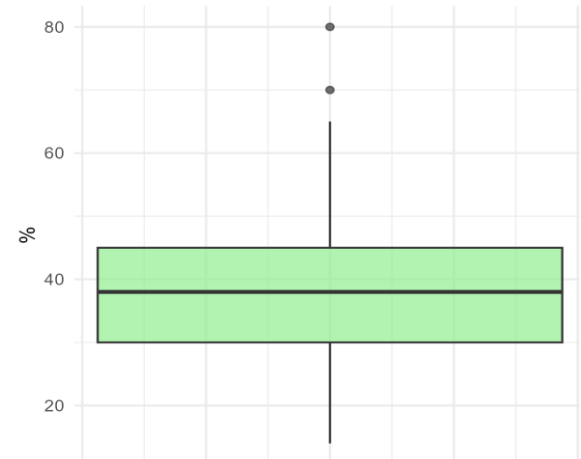
A Figura 1 apresenta, por meio de *boxplots*, a distribuição das variáveis contínuas da amostra. Os painéis mostram: (a) idade, (b) FE, (c) sódio, (d) log(Creatinina), (e) log(Plaquetas) e (f) log(CPK). As transformações logarítmicas foram aplicadas às variáveis com forte assimetria e presença de valores extremos (creatinina, plaquetas e CPK), com o objetivo de reduzir a influência de outliers e melhorar a adequação dos dados aos procedimentos estatísticos subsequentes. Visualmente, observa-se a concentração da idade entre 50 a 70 anos, dispersão importante nos marcadores laboratoriais e a ocorrência de valores atípicos em diferentes magnitudes, compatíveis com a variabilidade clínica esperada em insuficiência cardíaca.

Figura 1 — *Boxplots* das variáveis clínicas e laboratoriais: (a) idade, (b) FE, (c) sódio, (d) log(Creatinina), (e) log(Plaquetas) e (f) log(CPK).

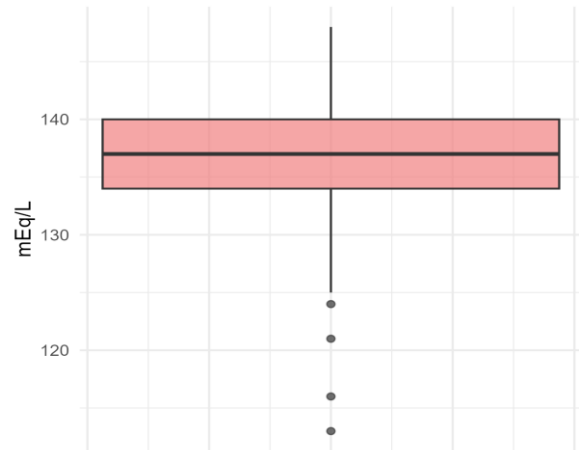
(a) Idade



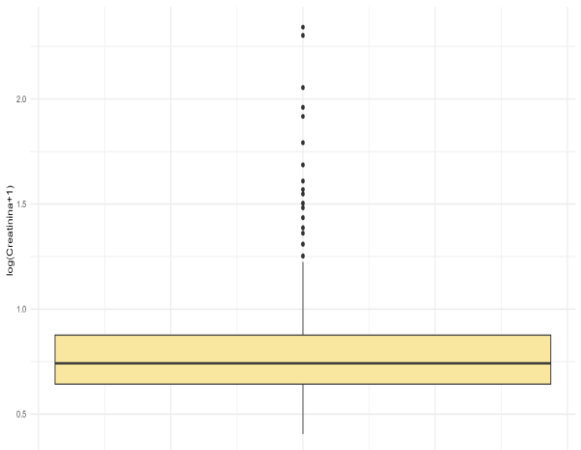
(b) FE



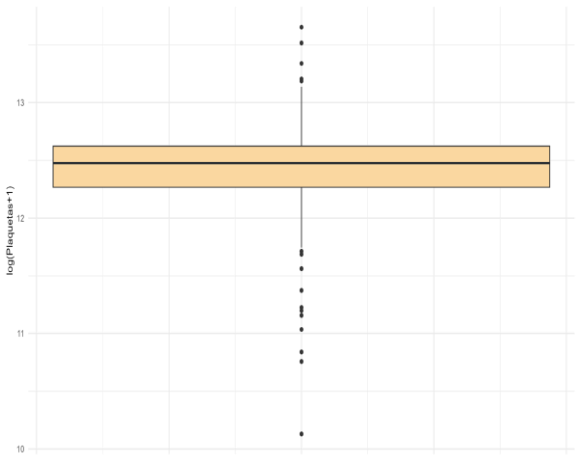
(c) Sódio



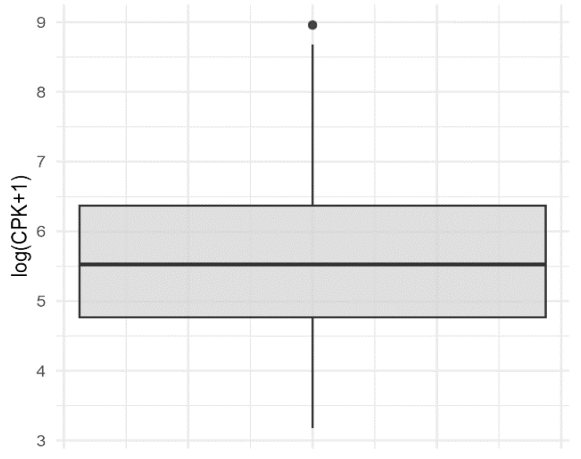
(d) log(Creatinina)



(e) log(Plaquetas)



(f) log(CPK)



Fonte: Elaboração própria com base nos dados do estudo.

A FE apresentou mediana em torno de 38–40%, indicando disfunção ventricular esquerda. A variação entre os valores, com valores bem abaixo da normalidade clínica ($\geq 55\%$, Lang *et al.*, 2015), sugere a presença de pacientes com IC de FE reduzida e também casos limítrofes.

Quanto ao sódio, os valores centrais se situam entre 135-140 mEq/L, próximo do normal clinicamente falando (136-145 mEq/L). Alterações no sódio são marcadores prognósticos independentes nos casos agudos e crônicos (Spinetti, 2019).

A creatinina, mesmo após aplicação de log (creatinina+1), mostra assimetria positiva e cauda longa à direita nos *boxplots*, representando pacientes com disfunção renal importante. Achados clínicos comprovam que níveis elevados de creatinina ou declínio da função renal (por exemplo, taxa de filtração glomerular reduzida ou aumento da creatinina) estão associados a pior prognóstico, incluindo morte e readmissão hospitalar (Kang *et al.*, 2022).

As plaquetas, transformadas em log, apresentam dispersão relativamente moderada; no entanto, há outliers extremos que sugerem alterações hematológicas ou inflamatórias em alguns pacientes. Embora não haja estudos tão robustos quanto para creatinina ou sódio no contexto de IC que usem plaquetas como variável prognóstica primária, trabalhos sobre marcadores inflamatórios e parâmetros hematológicos confirmam que alterações nessas variáveis influenciam o estado clínico e a gravidade da doença.

A CPK, após transformação log(CPK+1), apresenta uma distribuição mais simétrica do que na escala bruta, reduzindo o impacto visual de outliers nos gráficos. Ainda assim, existem valores extremos que podem representar injúria miocárdica ou muscular, ou complicações secundárias, indicando que mesmo transformada, a variável mantém potencial informativo. Estudos mostram que fluxos de CK (*creatine kinase*) diminuídos correlacionam-se com desfechos adversos em pacientes com cardiomiopatia ou IC (Aujla; Zubair; Patel, 2025).

Os *boxplots* descrevem variabilidade clínica expressiva na amostra distribuições assimétricas, presença de outliers e diferenças entre pacientes mais estáveis e outros em situação crítica. Demonstram claramente que pacientes com IC não formam um grupo homogêneo, tendo perfis distintos de gravidade, com diferentes graus de disfunção renal, injúria cardíaca e descompensações.

A Tabela 4 apresenta as medidas de assimetria e curtose das variáveis contínuas. Os resultados evidenciam a ocorrência de distribuições com diferentes graus de distorção em relação à normalidade. Para a variável idade, observou-se *skewness* de 0,42 e curtose de -0,22, indicando distribuição aproximadamente simétrica, com discreta cauda à direita e leve achatamento, sem grandes desvios de normalidade. A FE apresentou *skewness* de 0,55 e curtose

de 0,00, caracterizando uma distribuição simétrica e mesocúrtica, compatível com um perfil relativamente estável em relação à normalidade.

Tabela 4 — Medidas de assimetria e curtose das variáveis contínuas.

Variável (Unidade)	Assimetria	Curtose
Idade (anos)	0.42	-0.22
Fração de Ejeção (%)	0.55	0.00
Sódio (mEq/L)	-1.04	3.98
log(Creatinina) (mg/dL)	2.39	7.00
log(Plaquetas) (mil/mm ³)	-1.27	5.27
log(CPK) (U/L)	0.42	-0.36

Fonte: Elaboração própria com base nos dados do estudo.

No caso do sódio sérico, o valor de *skewness* foi de -1,04 e a curtose de 3,98, sugerindo assimetria negativa influenciada por pacientes com valores mais baixos. A creatinina, mesmo após transformação logarítmica, apresentou *skewness* de 2,39 e curtose de 7,00, evidenciando forte assimetria positiva e distribuição leptocúrtica, com concentração de valores próximos à mediana, mas com casos extremos compatíveis com disfunção renal significativa. Esse achado é consistente com a literatura, que aponta a deterioração da função renal como fator prognóstico desfavorável (Kang *et al.*, 2022).

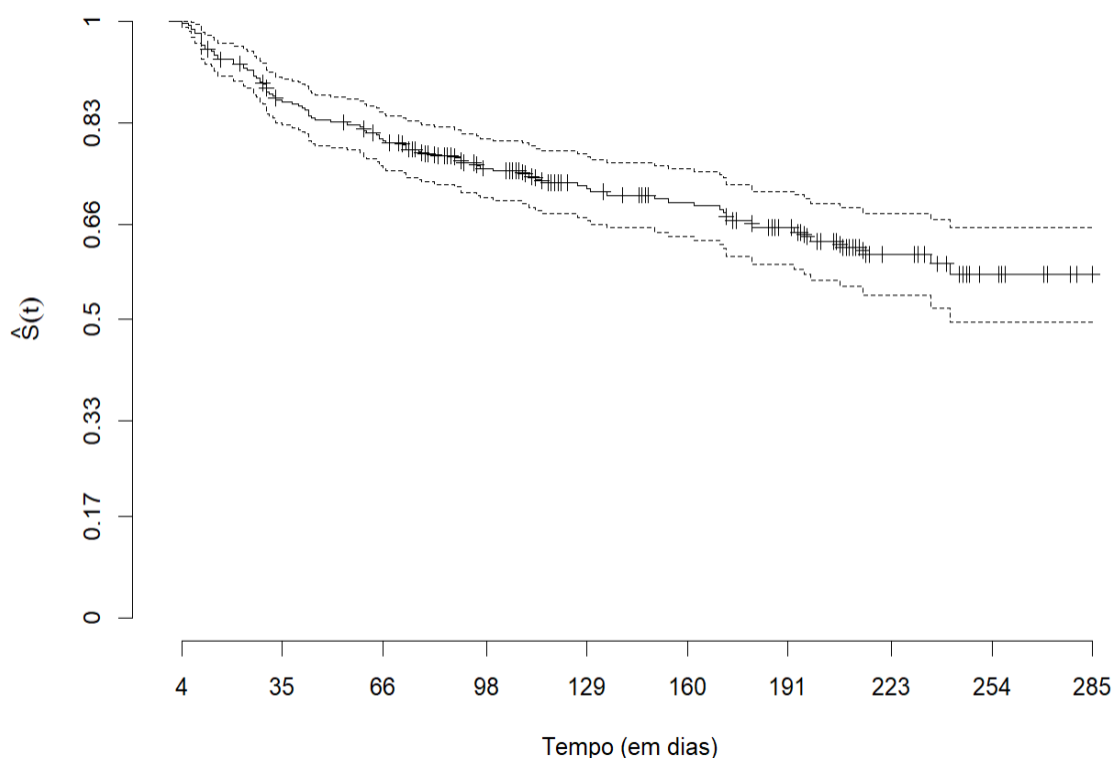
As plaquetas, transformadas em log, exibiram *skewness* de -1,27 e curtose de 5,27, caracterizando assimetria negativa e presença de valores extremos em direção a contagens muito baixas. Essas alterações podem refletir estados inflamatórios ou infecciosos em determinados pacientes, ainda que não haja consenso quanto ao seu papel prognóstico direto na insuficiência cardíaca. Por fim, a variável CPK, transformada em log, apresentou assimetria (*skewness*) de 0,42 e curtose de -0,36, indicando uma distribuição aproximadamente simétrica e platicúrtica, com comportamento estatístico mais estável em comparação às demais variáveis laboratoriais.

De modo geral, a análise conjunta das medidas de assimetria e curtose, associada à inspeção gráfica, mostra que idade, FE, sódio e CPK apresentam distribuições mais próximas da normalidade, enquanto creatinina e plaquetas evidenciam maior distorção e presença de valores extremos. Esse padrão reforça a necessidade de cautela na interpretação clínica e estatística dessas variáveis, uma vez que podem impactar tanto o ajuste dos modelos quanto a compreensão da diversidade dos pacientes com IC.

6.3 Estimador de *Kaplan-Meier*

A Figura 2 apresenta a curva de sobrevivência estimada pelo método de *Kaplan-Meier*, refletindo a probabilidade de sobrevida dos pacientes com IC ao longo do tempo de seguimento. A linha sólida representa a estimativa pontual da função de sobrevivência, enquanto a faixa sombreada corresponde ao intervalo de confiança de 95%, fornecendo uma medida da incerteza associada às estimativas em cada ponto temporal. Os resultados evidenciam um declínio na probabilidade de sobrevivência nos primeiros 100 dias de seguimento. Esse padrão reflete a gravidade clínica dos pacientes analisados, muitos dos quais apresentavam disfunção ventricular significativa e múltiplas comorbidades. Estudos prévios destacam que a IC descompensada está associada a taxas elevadas de mortalidade precoce, especialmente no período inicial após o diagnóstico ou hospitalização (Arruda *et al.*, 2022).

Figura 2 – Curva de sobrevivência *Kaplan-Meier* para pacientes com IC.



Fonte: Elaboração própria com base nos dados do estudo.

Após 200 dias de acompanhamento, observa-se uma estabilização relativa da curva de sobrevivência, formando um platô em torno de 60%, o que pode sugerir a presença de um subgrupo de pacientes que não experimenta o evento nesse período. Esse comportamento indica

que indivíduos que superam a fase inicial mais vulnerável podem apresentar prognóstico relativamente mais favorável a médio e longo prazo.

Entretanto, a literatura destaca que a simples observação de um platô não é suficiente, por si só, para afirmar a existência de pacientes “curados”. A estabilização da curva representa, em geral, uma evidência indireta dessa possibilidade, devendo sempre ser interpretada à luz de informações clínicas e de acompanhamento prolongado. O platô pode refletir a consequência da presença de pacientes clinicamente estáveis e não a causa da cura em si.

Estudos recentes têm demonstrado que a identificação de um “platô” na curva de *Kaplan–Meier* envolve grande subjetividade, uma vez que diferentes especialistas podem divergir amplamente sobre o ponto em que a curva realmente se estabiliza ou se ainda existe risco residual de evento. Jackson *et al.* (2025) observaram fraca concordância entre cinco especialistas ao avaliarem dez curvas de sobrevivência oriundas de ensaios oncológicos, evidenciando que a noção de platô carece de um consenso formal e de critérios objetivos. Essa constatação reforça que a interpretação visual deve ser complementada por modelagens matemáticas robustas, capazes de estimar a fração de cura com base em pressupostos explícitos sobre o comportamento do risco ao longo do tempo.

Além disso, os autores demonstram que o surgimento de uma fração de cura positiva depende fortemente do tipo de modelo adotado (exponencial, Weibull ou flexível paramétrico), e que tal resultado não garante, necessariamente, a existência biológica de pacientes curados, mas apenas uma estabilização estatística da função de risco. Assim, a adoção de modelos de cura deve estar sempre ancorada na plausibilidade clínica, na maturidade dos dados e no acompanhamento suficiente para excluir eventos tardios, evitando interpretações precipitadas sobre a presença de cura verdadeira (Jackson *et al.*, 2025).

A interpretação de platôs deve ser feita com cautela. Conforme reforçam Price e Manatunga (2001), a aplicabilidade dos modelos de fração de cura deve estar ancorada em fundamentos clínicos robustos, de modo a evitar inferências equivocadas sobre a existência de cura estatística.

6.4 Modelo de Riscos Proporcionais de Cox

Como segunda etapa da análise, foi considerado o modelo de regressão de Cox, utilizado para identificar os fatores associados ao risco de mortalidade entre os pacientes com IC. Os resultados, coeficientes, razões de risco (*hazard ratios*), intervalos de confiança de 95% e valores de *p*, serão exibidos na Tabela 5. As estatísticas globais do modelo (teste da razão de

verossimilhança, teste de Wald e teste de escore, todos com $p < 0,0001$) confirmam o ajuste significativo, com concordância de 0,723, indicando boa capacidade discriminativa.

Tabela 5 — Resultados do modelo de regressão de Cox para os fatores associados à sobrevivência em pacientes com insuficiência cardíaca.

Variável	β	EP	HR	IC95% Inferior	IC95% Superior	z	p
Gênero	-0.2495	0.2504	0.7792	0.4769	1.2730	-0.9962	0.3191
Fumante	0.1960	0.2510	1.2166	0.7438	1.9899	0.7809	0.4349
Diabetes	0.1471	0.2228	1.1585	0.7486	1.7928	0.6602	0.5092
Hipertensão	0.4915	0.2148	1.6348	1.0731	2.4906	2.2886	0.0221
Anemia	0.4505	0.2128	1.5691	1.0339	2.3812	2.1167	0.0343
Idade (anos)	0.0442	0.0095	1.0452	1.0259	1.0648	4.6576	<0.0001
Fração de Ejeção (%)	-0.0438	0.0103	0.9571	0.9381	0.9766	-4.2710	<0.0001
Sódio (mEq/L)	-0.0357	0.0235	0.9649	0.9215	1.0104	-1.5202	0.1285
log(Creatinina+1)	1.3835	0.3031	3.9888	2.2021	7.2253	4.5642	<0.0001
log(Plaquetas)	-0.1097	0.2434	0.8961	0.5561	1.4440	-0.4506	0.6523
log(CPK+1)	0.0855	0.0987	1.0893	0.8978	1.3217	0.8669	0.3860

Fonte: Elaboração própria com base nos dados do estudo.

A idade mostrou-se um dos principais preditores de mortalidade, com aumento aproximado de 4% no risco de morte a cada ano adicional (HR = 1,043; IC95%: 1,025–1,062; $p < 0,001$). Esse achado está em consonância com a literatura, que demonstra o papel do envelhecimento como determinante de pior prognóstico em insuficiência cardíaca, em função do acúmulo de comorbidades e da redução progressiva da reserva funcional cardíaca (Arruda *et al.*, 2022).

A fração de ejeção apresentou efeito aparentemente protetor, de forma que cada aumento de 1% reduziu em cerca de 4% o risco de mortalidade (HR = 0,959; IC95%: 0,940–0,979; $p < 0,001$). Esses resultados reforçam o papel da função ventricular esquerda como um dos principais marcadores prognósticos da síndrome, em consonância com estudos que associam valores mais altos de fração de ejeção à sobrevida prolongada (Danzmann *et al.*, 2023).

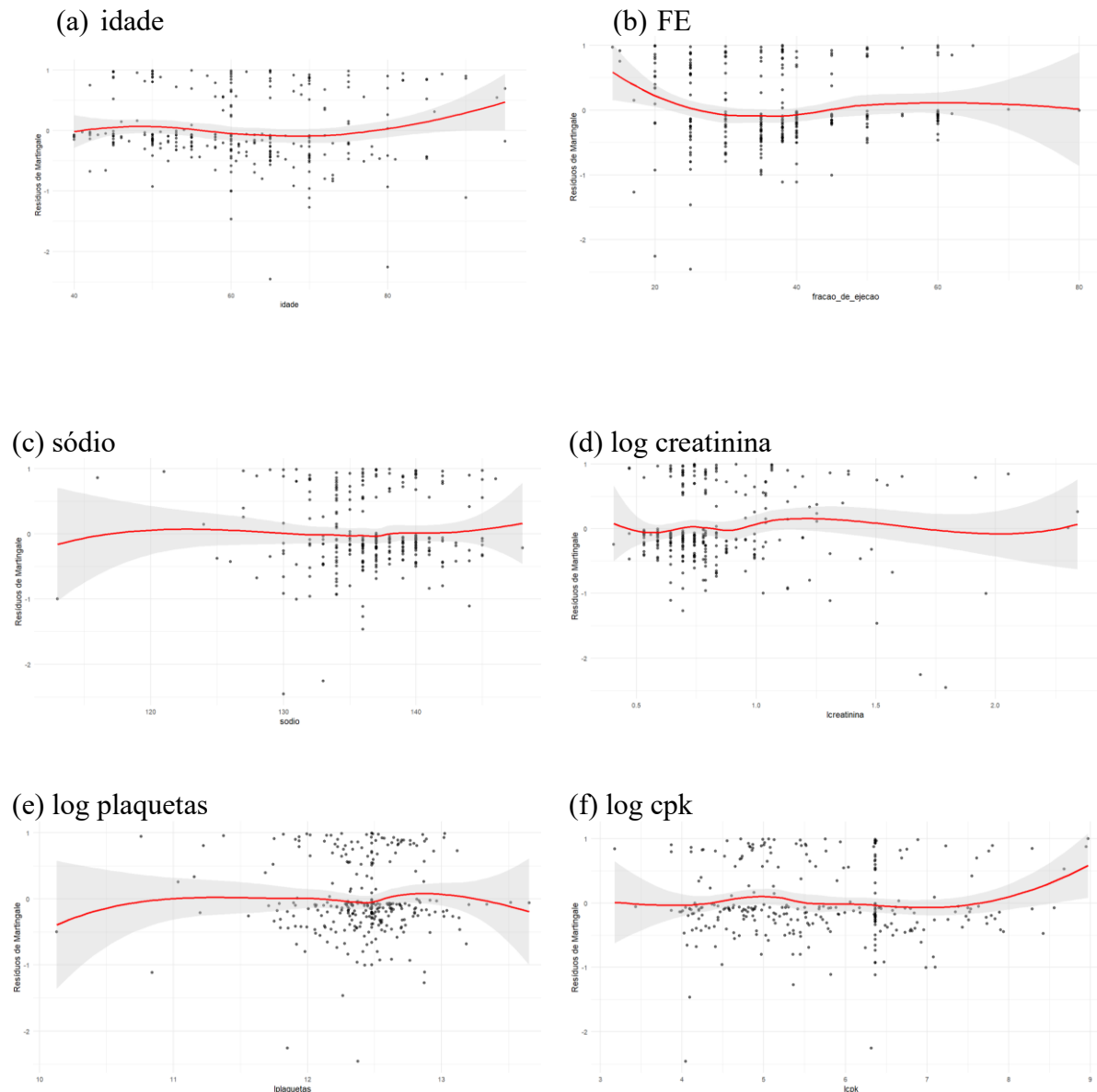
A creatinina sérica log-transformada também se destacou como variável significativa, com risco de mortalidade quase quatro vezes maior nos pacientes com níveis mais elevados (HR = 3,87; IC95%: 2,10–7,14; $p < 0,001$). Esse resultado ilustra a importância da síndrome cardiorrenal, frequentemente descrita como um dos principais determinantes de desfechos adversos em insuficiência cardíaca, exigindo monitoramento e intervenções específicas para retardar a progressão da disfunção renal (Damman *et al.*, 2023).

Por outro lado, as variáveis sódio, plaquetas e log(CPK) não apresentaram associação estatisticamente significativa com a mortalidade no modelo múltiplo. Ainda assim, os sinais de efeito observados possuem relevância clínica. O sódio apresentou tendência negativa ($HR < 1$), o que pode refletir a ocorrência de hiponatremia, condição reconhecida como marcador prognóstico em insuficiência cardíaca, embora seu impacto isolado seja limitado nesta amostra (Rodriguez *et al.*, 2019).

Já o CPK, mesmo sem resultados marcantes, apresentou coeficiente positivo, sugerindo possível influência de injúria muscular ou miocárdica em episódios agudos de descompensação, aspecto descrito em estudos sobre biomarcadores na insuficiência cardíaca (Lee *et al.*, 2011).

Em síntese, os resultados do modelo de Cox confirmam o papel da idade, da fração de ejeção e da creatinina como determinantes prognósticos, em consonância com evidências consolidadas na literatura. Outras variáveis laboratoriais, embora não tenham alcançado significância estatística nesta amostra, permanecem relevantes clinicamente e foram incluídas na modelagem, contribuindo para o ajuste adequado do modelo e refletindo fatores conhecidos na literatura.

Foi realizada a análise dos resíduos de Martingale, com o objetivo de verificar a adequação do ajuste e a linearidade das covariáveis contínuas. Essa avaliação é fundamental, pois desvios sistemáticos nos resíduos podem indicar a necessidade de transformações adicionais ou a inadequação da forma funcional escolhida para as variáveis contínuas (Oliveira *et al.*, 2023).

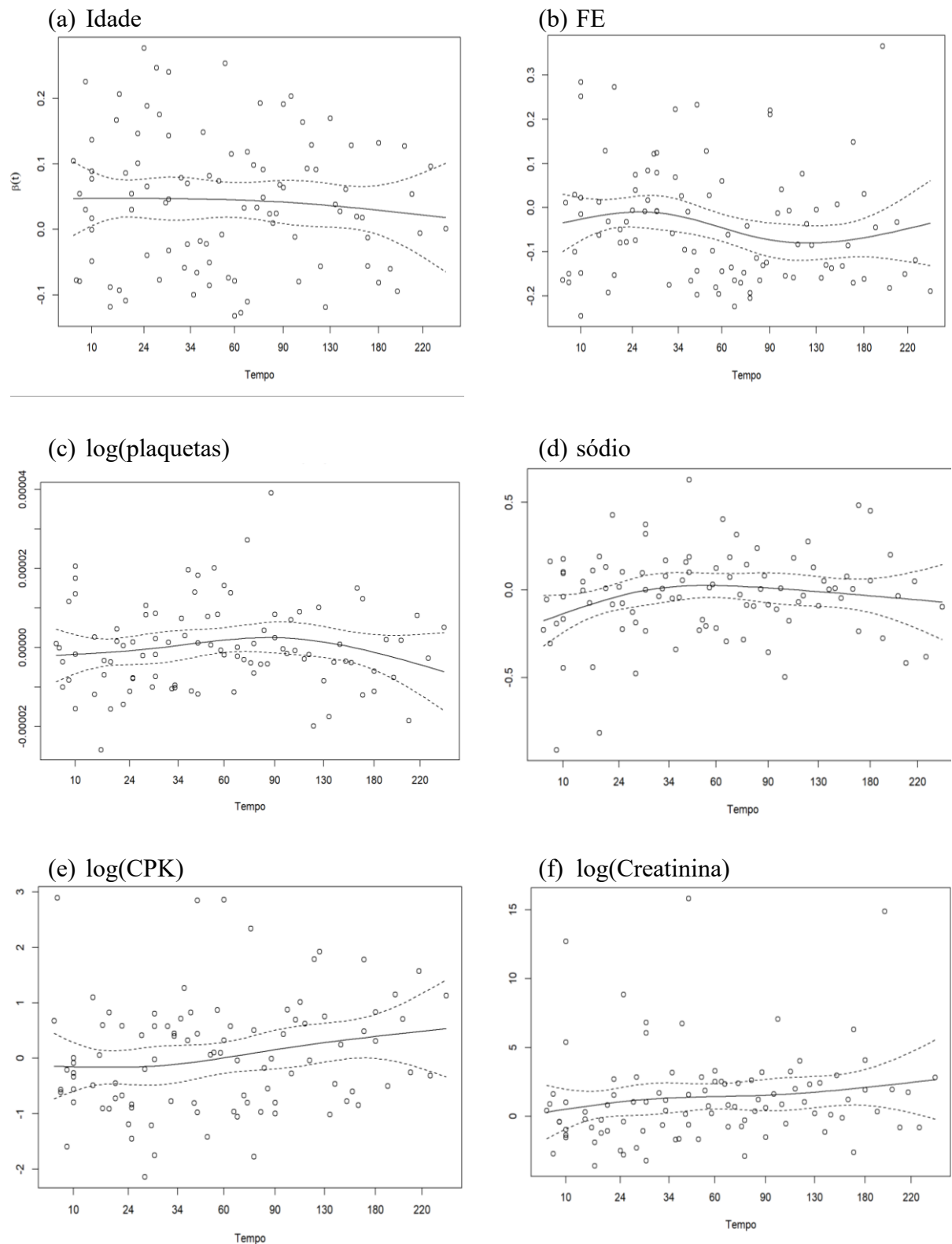
Figura 3 – Resíduos de Martingale para o modelo de Cox.

Fonte: Elaboração própria com base nos dados do estudo

Conforme ilustrado na Figura 3, os resíduos de Martingale não apresentaram padrão sistemático. A linha de suavização se manteve próxima ao zero, com leve oscilação dentro da faixa de confiança, indicando que a relação entre idade e risco de mortalidade pode ser considerada aproximadamente linear no contexto do modelo (Oliveira *et al.*, 2023).

Avaliou-se o pressuposto de proporcionalidade dos riscos por meio dos resíduos de Schoenfeld. O objetivo foi verificar se os efeitos das covariáveis se mantêm constantes ao longo do tempo. Na Figura 4, o eixo X representa o tempo de acompanhamento e o eixo Y mostra os resíduos estimados $\beta(t)$, onde a linha sólida a tendência média dos resíduos e as linhas tracejadas os intervalos de confiança.

Figura 4 – Resíduos Schoenfeld.



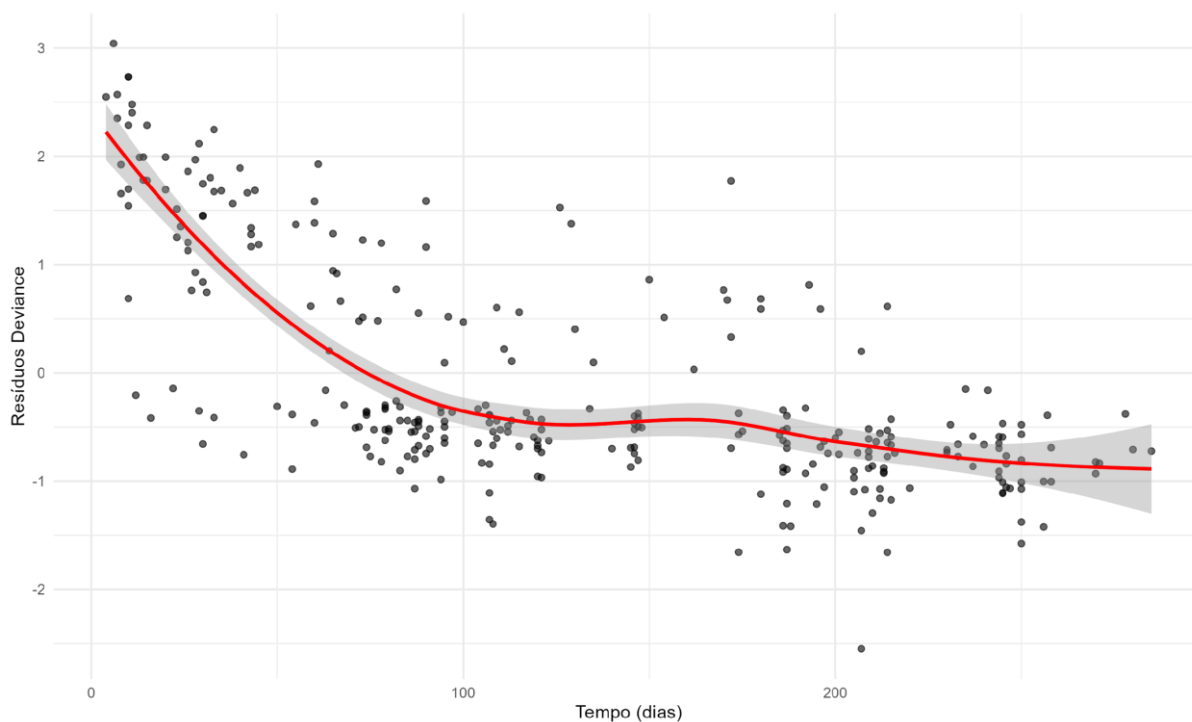
Fonte: Elaboração própria com base nos dados do estudo

Os resultados demonstram que, para a maioria das variáveis incluídas, os resíduos se distribuem de forma aleatória em torno de zero, sem apresentar padrões sistemáticos ou tendências claras ao longo do tempo. Esse comportamento sugere que o efeito das covariáveis sobre o risco de mortalidade pode ser considerado estável durante o período analisado, reforçando a adequação do pressuposto de proporcionalidade. Ainda que em períodos mais avançados de seguimento (acima de 150 dias) se observe discreta dispersão, os valores permanecem dentro de limites aceitáveis e não configuram violação relevante do pressuposto.

A ausência de variações sistemáticas na relação entre covariáveis e risco ao longo do tempo reforça a consistência do modelo, evitando vieses que poderiam comprometer a interpretação dos resultados. Desde o trabalho de Cox (1972), a verificação do pressuposto de proporcionalidade é reconhecida como etapa indispensável na modelagem de sobrevivência. Estudos mais recentes, como os de In e Lee (2019), reforçam que a análise dos resíduos de Schoenfeld representa uma ferramenta fundamental para garantir a validade e a robustez de modelos aplicados em contextos clínicos.

Também foram analisados os resíduos Deviance, apresentados na Figura 5, com o objetivo de avaliar graficamente a qualidade do ajuste. A inspeção dos resíduos mostra uma dispersão predominantemente homogênea em torno de zero, sem padrões evidentes ou tendências sistemáticas ao longo do tempo que indiquem inadequação do modelo. A linha suavizada acompanha esse comportamento, sugerindo estabilidade do ajuste, sobretudo no período intermediário do segmento, onde há maior concentração de observações.

Figura 5 – Resíduos Deviance em função do tempo.



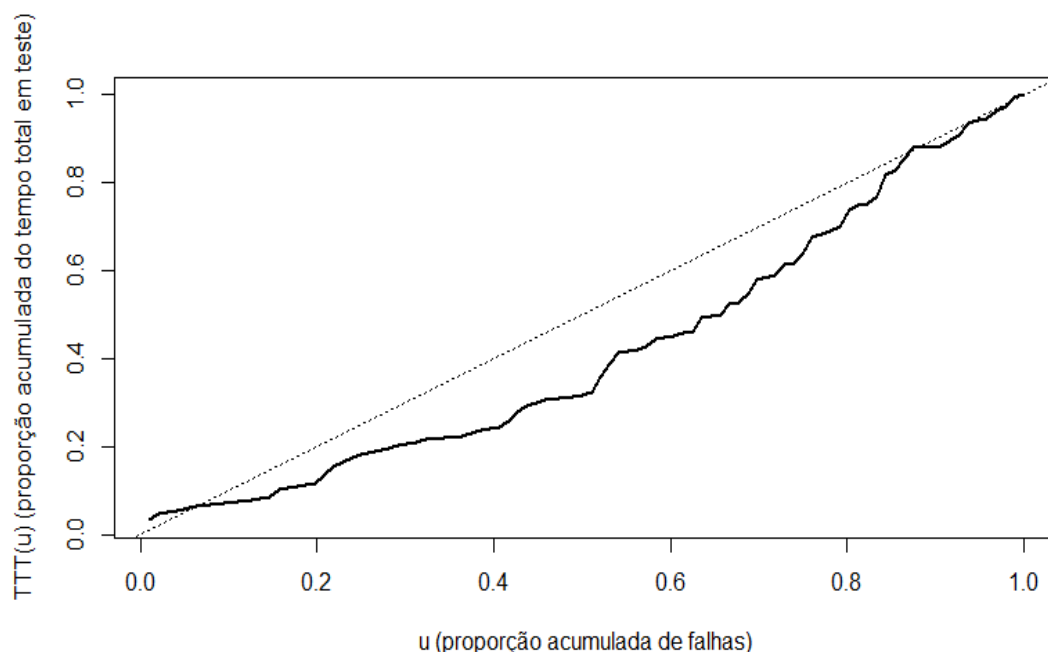
Fonte: Elaboração própria com base nos dados do estudo

Em contrapartida, nota a presença de alguns valores extremos nos períodos iniciais e finais do acompanhamento. Esses outliers podem estar relacionados tanto à menor quantidade de indivíduos em risco nesses intervalos quanto a características clínicas particulares, como maior gravidade da doença ou comorbidades não contempladas no modelo. Apesar disso, os resíduos não evidenciam um padrão sistemático, o que indica que a adequação global do modelo permanece preservada.

6.5. Estimação da função risco empírica

A figura 6 apresenta o gráfico TTT (*Total Time on Test*), que permite analisar o risco acumulado de mortalidade ao longo do tempo de acompanhamento dos pacientes. O comportamento observado sugere uma curva de risco decrescente, informação relevante para a seleção de modelos probabilísticos futuros, os quais devem ser capazes de representar adequadamente uma função de risco decrescente. Na figura observa-se a elevação mais acentuada nos estágios iniciais, seguida por estabilização ao longo do seguimento. Esse padrão indica que a mortalidade foi mais frequente no período inicial, comportamento esperado em pacientes com IC descompensada, nos quais a vulnerabilidade clínica é maior nos primeiros meses após o diagnóstico ou internação.

Figura 6 – TTT Plot (total time on test).



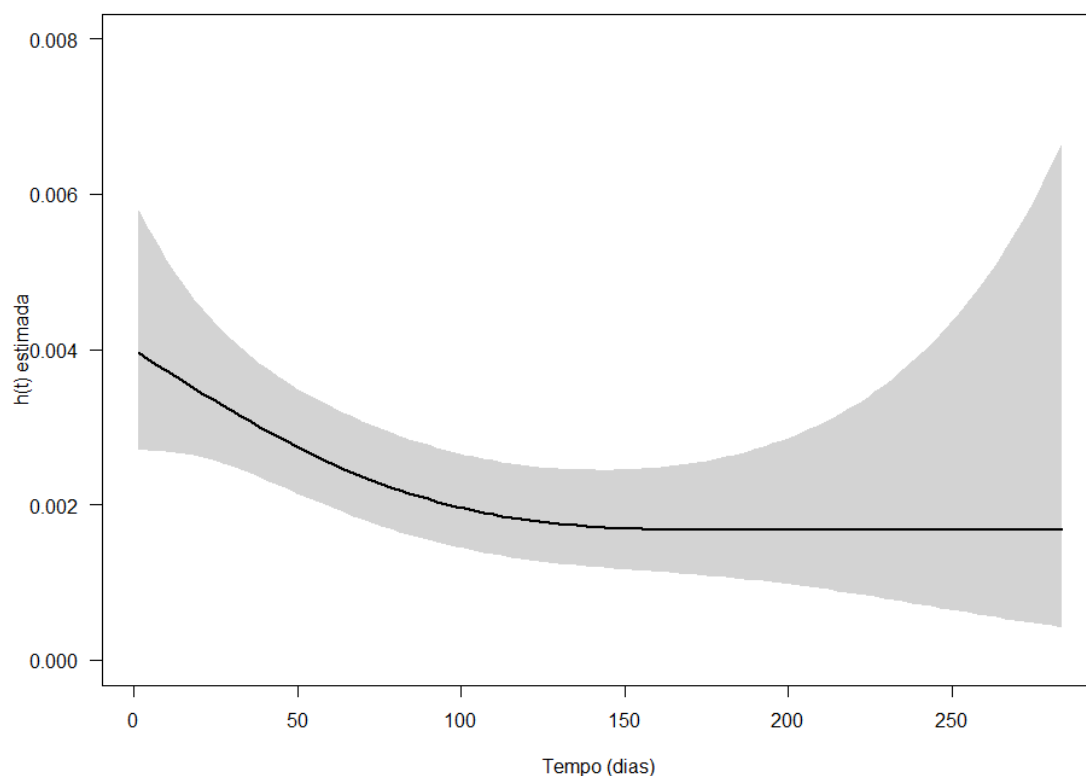
Fonte: Elaboração própria com base nos dados do estudo

A Figura 7 apresenta a função de risco $h(t)$ estimada pelo método não paramétrico com suavização baseada em B-splines, conforme Rebora *et al.* (2004), utilizando 100 intervalos igualmente espaçados. A área sombreada representa a banda de confiança de 95%. Observa-se um risco inicial elevado, seguido por uma queda acentuada nos primeiros 100 dias de seguimento.

Após aproximadamente 150 dias, a função de risco se estabiliza, formando um platô. Esse comportamento descarta a hipótese de risco constante, característica de modelos exponenciais, e também não é compatível com distribuições unimodais. Em vez disso, sugere um comportamento compatível com modelos flexíveis, como Weibull ou Gama, que permitem risco monotônico decrescente.

Do ponto de vista clínico, a curva indica maior vulnerabilidade dos pacientes nos primeiros meses após a internação, seguida por um período de relativa estabilidade entre aqueles que sobrevivem à fase inicial crítica. A estabilização do risco no final do seguimento é coerente com o platô observado na curva de *Kaplan–Meier* e sustenta a hipótese estatística da presença de uma fração de cura dentro da população analisada.

Figura 7 – Função de risco $h(t)$ estimada pelo método não-paramétrico com suavização baseada em B-splines.



Fonte: Rebora *et al.* (2004).

6.6 Análise por Modelos de Fração de Cura

Nesta etapa, procedeu-se à análise de sobrevivência utilizando modelos de fração de cura, com o objetivo de investigar a existência de uma proporção de pacientes que, após determinado tempo de acompanhamento, não apresentam mais risco associado ao evento de interesse. A utilização desses modelos é particularmente adequada em contextos clínicos como a insuficiência cardíaca, nos quais avanços terapêuticos e estratégias de manejo podem conferir estabilidade prolongada a subgrupos de pacientes (Lambert, 2007).

Inicialmente, foi realizada uma comparação entre os modelos propostos para a análise com a fração de cura, sob a formulação de mistura e sem inclusão de covariáveis. Para essa comparação, foram empregados o Critério de Informação de Akaike (AIC), Critério de Informação de Akaike Corrigido (AIC), e Critério de Informação de Bayesiano (BIC), como medida de qualidade relativa do ajuste, na qual o menor valor obtido nos critérios indica melhor desempenho do modelo (Colosimo; Giolo, 2006; Klein; Moeschberger, 1997).

A Tabela 6 apresenta as estimativas dos parâmetros obtidos pelos modelos de fração de cura sem regressão. As estimativas apresentadas foram obtidas por máxima verossimilhança,

utilizando rotinas próprias implementadas no R. A otimização foi realizada por meio da função `optim()`, com o método de Nelder-Mead, e a matriz de informação de Fisher foi aproximada pela inversa da Hessiana numérica, permitindo a obtenção dos erros-padrão e intervalos de confiança de 95%. Essa abordagem foi aplicada aos modelos de mistura sem regressão (Weibull, Gama, Lindley, Gumbel, Exponencial e Rayleigh), conforme o código de inferência desenvolvido para este estudo. O ajuste foi realizado considerando convergência numérica pelo método de Newton-Raphson e seleção de modelos com base nos critérios AIC e BIC. Observa-se que todos os modelos forneceram estimativas plausíveis para os parâmetros de escala e forma, bem como para a fração de cura. O modelo exponencial de mistura apresentou o menor valor de AIC (1342.13), sugerindo desempenho ligeiramente superior entre os modelos paramétricos considerados. A Figura 9 apresenta a curva de Kaplan–Meier sobreposta às curvas ajustadas pelos modelos de fração de cura Weibull, Gama, Lindley, Gumbel, Exponencial e Rayleigh.

As estimativas da fração de cura variaram de aproximadamente 49,7% no Weibull a 60,2% no Lindley, sugerindo a presença de uma proporção significativa de indivíduos que podem ser considerados *long-term survivors*. Essa variabilidade entre distribuições é esperada, dado que cada modelo impõe diferentes hipóteses sobre a forma da função de risco ao longo do tempo.

Outro aspecto a ser considerado é a largura dos intervalos de confiança, especialmente nos modelos Gama e Gumbel, onde se verificou maior imprecisão em função da dispersão dos tempos de sobrevivência. Esse resultado é coerente com a literatura, que ressalta a dificuldade em obter estimativas estáveis de parâmetros em distribuições mais complexas quando o tamanho amostral é limitado (Colosimo; Giolo, 2006).

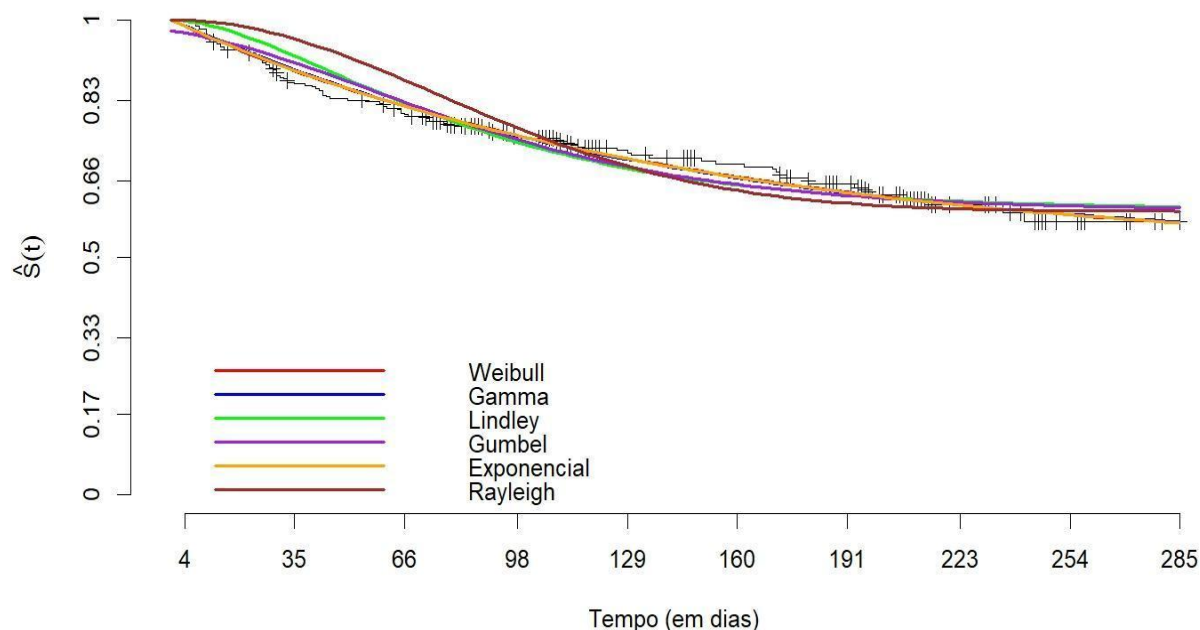
Observa-se que todos os modelos capturam de forma satisfatória o comportamento geral da curva de sobrevivência ao longo do tempo. Algumas diferenças pequenas são notadas quando se compara o ajuste em diferentes regiões. O modelo Exponencial, por exemplo, mostrou-se relativamente mais rígido, apresentando menor flexibilidade na cauda da distribuição, o que pode comprometer a descrição adequada dos tempos mais longos de seguimento.

Tabela 6 — Estimativas dos parâmetros sem regressão.

Modelo	Parâmetro	Estimativa	Erro-Padrão	IC95%		AIC	BIC	AICc
				LI	LS			
Weibull	α	0,987	0,116	0,758	1,215	1344,12	1355,22	1344,20
	γ	149,03	53,84	43,51	254,55			
	p	0,497	0,102	0,298	0,696			
Gama	α	1,029	0,159	0,716	1,341	1344,10	1355,20	1344,18
	γ	135,73	59,90	18,32	253,14			
	p	0,512	0,085	0,345	0,679			
Lindley	θ	0,022	0,002	0,018	0,027	1363,63	1371,03	1363,67
	p	0,602	0,036	0,531	0,672			
Gumbel	μ	56,71	6,87	43,24	70,18	1380,47	1391,57	1380,55
	β	54,01	6,40	41,47	66,55			
	p	0,598	0,037	0,525	0,671			
Exponencial	λ	0,0069	0,0016	0,0038	0,0100	1342,13	1349,53	1342,17
	p	0,503	0,070	0,366	0,640			
Rayleigh	σ	75,94	5,01	66,13	85,76	1408,85	1416,25	1408,89
	p	0,597	0,035	0,529	0,665			

Fonte: Elaboração própria com base nos dados do estudo.

Figura 8 – Curva de sobrevivência observada (Kaplan–Meier) e ajustadas pelos modelos de fração de cura.



Fonte: Elaboração própria com base nos dados do estudo.

O modelo de Rayleigh, embora apresente curva próxima à empírica nos tempos intermediários, mostrou maior divergência no início do segmento, superestimando a probabilidade de sobrevivência. Esse comportamento indica que o modelo tende a escapar da curva observada em períodos de risco inicial elevado, o que compromete seu desempenho global, refletido também pelos maiores valores de AIC e BIC.

Os modelos Weibull e Gama apresentaram curvas bastante próximas à observada, sugerindo melhor aderência, em consonância com os critérios de informação (Tabela 6 e Figura 8).

O modelo Gumbel, embora tenha descrito bem a porção intermediária da curva, demonstrou maior divergência na cauda, indicando possível limitação para capturar o comportamento assintótico dos sobreviventes de longo prazo.

O modelo baseado na distribuição Lindley produziu uma curva de ajuste relativamente satisfatória, mas com maior variabilidade nos extremos, refletindo o ajuste estatístico intermediário observado.

Assim, a análise gráfica reforça os resultados obtidos por meio dos critérios de informação, em que os modelos Weibull e Gama se destacaram como alternativas mais adequadas para descrever os dados de insuficiência cardíaca estudados. Esse achado está em concordância com estudos prévios que utilizaram modelos de fração de cura em contextos clínicos semelhantes, os quais destacam a importância da escolha criteriosa do modelo para

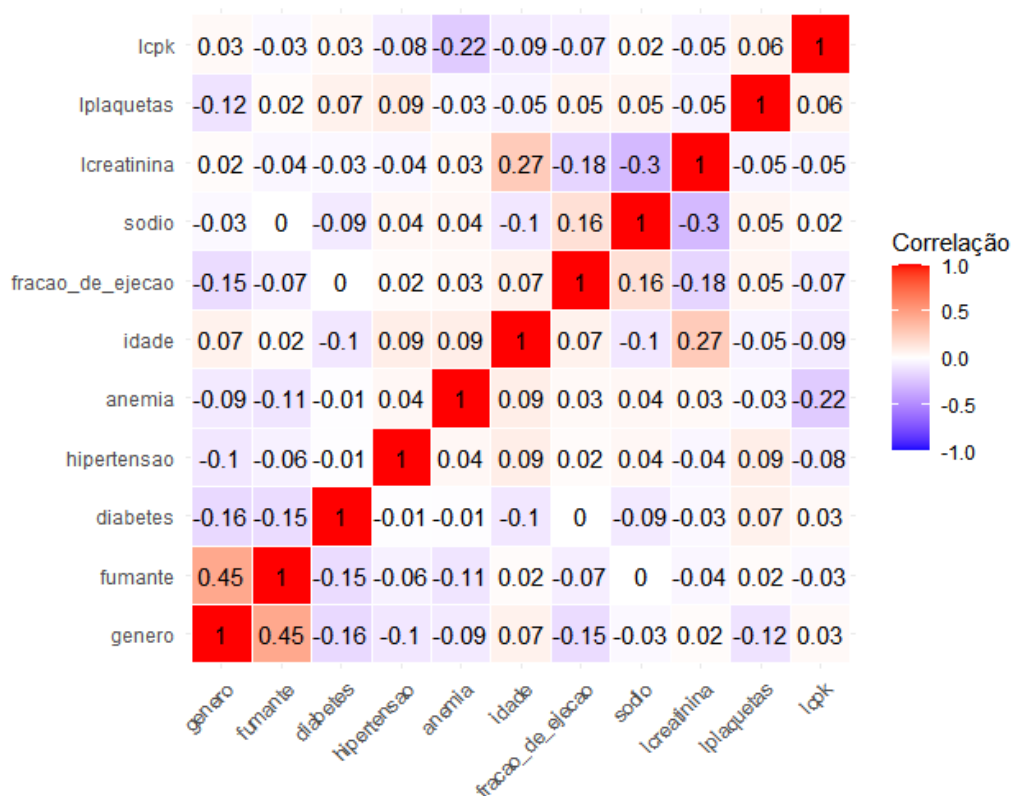
garantir estimativas confiáveis da proporção de curados e do risco residual na população (Martinez *et al.*, 2013).

Esses resultados reforçam que, embora o modelo Exponencial tenha apresentado o menor AIC entre os modelos sem regressão, suas suposições restritivas limitam sua interpretação clínica, enquanto Weibull e Gama demonstram maior capacidade de representar adequadamente a dinâmica da sobrevivência em pacientes com insuficiência cardíaca.

6.6.1. Modelos de regressão

A Figura 9 mostra as correlações entre as variáveis independentes. Utilizou-se a correlação monotônica de Spearman para a mensurar a correlação entre duas variáveis contínuas, a correlação ponto-bisserial nos casos de uma variável binária e outra contínua (por exemplo, entre gênero e sódio), e a correlação de Matthews nos casos de duas variáveis binárias. Os maiores coeficientes de correlação foram encontrados entre tabagismo e gênero (0,45) e entre log-creatinina e idade (0,27).

Figura 9 – Matriz de correlações entre as variáveis independentes. Utilizou-se a correlação de Spearman para a correlação entre duas variáveis contínuas, a correlação ponto-bisserial nos casos de uma variável binária e outra contínua, e a correlação de Matthews nos casos de duas variáveis binárias.

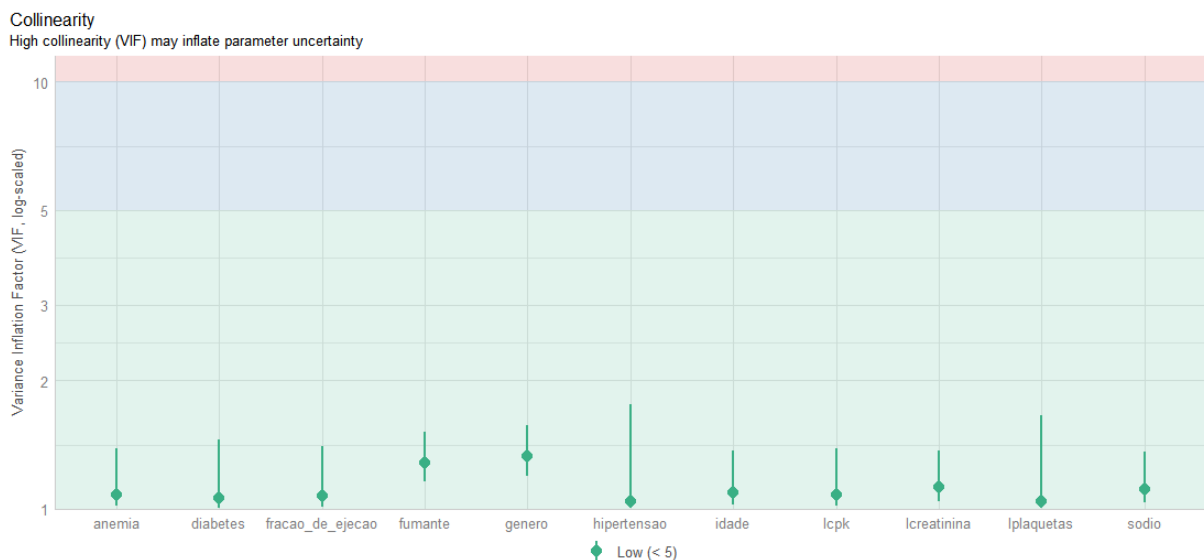


Fonte: Elaboração própria com base nos dados do estudo.

As correlações apresentadas na Figura 9 mostram, de modo geral, valores baixos entre as variáveis independentes, o que sugere ausência de multicolinearidade relevante no conjunto de dados. Os maiores coeficientes foram entre tabagismo e gênero (0,45) e entre log-creatinina e idade (0,27), valores considerados moderados, mas ainda insuficientes para comprometer a análise.

A Figura 10 apresenta as medidas do fator de inflação da variância (FIV) entre as variáveis independentes, obtidas por meio da função *check_collinearity()* do pacote *performance* no programa R. Valores de FIV acima de 0,5 poderiam sugerir multicolinearidade; entretanto, observou-se que todas as variáveis exibiram valores abaixo desse limite. Aliado às correlações relativamente baixas identificadas na Figura 9, esses achados sustentam a plausibilidade de ajustar um modelo de regressão múltiplo incluindo todas as covariáveis.

Figura 10 - Medidas do fator de inflação da variância (FIV) entre as variáveis independentes.



Fonte: Elaboração própria com base nos dados do estudo.

Entre todas as distribuições avaliadas, apenas os modelos Weibull e Gama apresentaram convergência estável, erros-padrão definidos e parâmetros clinicamente interpretáveis. Por essa razão, somente esses modelos foram considerados para a etapa de regressão com fração de cura, conforme apresentado na Tabela 7.

Tabela 7 — Estimativas dos parâmetros dos modelos de fração de cura com regressão em relação ao parâmetro p .

Modelo	Variável	Estimativa	EP	IC95%		AIC	BIC	AICc
				LI	LS			
Weibull	α	0.2034	0.0201	0.1640	0.2428	1646.86	1698.67	1648.34
	γ	1.5316	0.3515	0.8428	2.2204			
	Intercepto	-0.2089	6.6763	-13.2942	12.876			
	Gênero	-0.0160	0.3673	-0.7358	0.7039			
	Fumante	0.4772	0.3710	-0.2499	1.2043			
	Diabetes	0.2103	0.3152	-0.4074	0.8280			
	Hipertensão	0.4179	0.3360	-0.2407	1.0765			
	Anemia	-1.0745	0.3338	-1.7287	-0.4202			
	Idade	-0.0717	0.0154	-0.1019	-0.0416			
	FE	0.0658	0.0151	0.0363	0.0954			
	Sódio	0.0162	0.0358	-0.0539	0.0863			
	log(Crea.)	-0.0222	0.5518	-1.1038	1.0593			
	log(Plaq.)	0.2226	0.3771	-0.5165	0.9617			
	log(CPK)	-0.3545	0.1467	-0.6421	-0.0669			
Gama	α	4.0982	0.2512	3.6058	4.5906	1797.18	1848.98	1798.65
	γ	14.9717	0.7288	13.5434	16.400			
	Intercepto	6.2655	12.459	-18.1555	30.686			
	Gênero	-2.7769	0.6806	-4.1109	-1.4430			
	Fumante	1.2525	0.7466	-0.2107	2.7157			
	Diabetes	-3.8492	0.4032	-4.6395	-3.0588			
	Hipertensão	1.3241	0.5628	0.2211	2.4271			
	Anemia	-2.4311	0.6429	-3.6911	-1.1710			
	Idade	-0.2298	0.0252	-0.2791	-0.1805			
	FE	0.1143	0.0272	0.0610	0.1676			
	Sódio	0.6822	0.0827	0.5201	0.8443			
	log(Crea.)	-10.4786	0.6753	-11.8021	-9.1550			
	log(Plaq.)	-4.7631	0.5760	-5.8921	-3.6341			
	log(CPK)	-3.0033	0.1206	-3.2397	-2.7669			

Fonte: Elaboração própria com base nos dados do estudo.

Após a análise exploratória dos modelos sem regressão (Tabela 6), buscou-se avaliar o impacto de covariáveis clínicas e laboratoriais na fração de cura. A literatura destaca que modelos de fração de cura com regressão permitem captar a heterogeneidade clínica da população, fornecendo estimativas mais precisas e úteis à prática em saúde (Klein & Moeschberger, 1997). A Tabela 8 apresenta as estimativas dos parâmetros dos modelos de fração de cura com regressão, contemplando diferentes distribuições, Weibull, Gama, Lindley, Gumbel, Exponencial e Rayleigh, incluindo coeficientes, erros-padrão, intervalos de confiança de 95% e critérios de informação (AIC, BIC e AICc).

Tabela 8 — Estimativas dos parâmetros dos modelos de fração de cura com regressão em relação aos parâmetros de locação e ao parâmetro p de fração de cura

	Variável	Estimativa	EP	IC95% LI	IC95% LS	AIC	BIC	AICc
Weibull	α (Intercepto)	-0.0126	2.7461	-5.3948	5.3696	1714.84	1807.35	1719.60
	α (Gênero)	0.0188	0.1922	-0.3579	0.3956			
	α (Fumante)	0.0905	0.1765	-0.2554	0.4363			
	α (Diabetes)	0.0044	0.1636	-0.3162	0.3251			
	α (HAS)	0.1412	0.1593	-0.1711	0.4534			
	α (Anemia)	0.0670	0.1588	-0.2442	0.3782			
	α (Idade)	0.0113	0.0062	-0.0009	0.0235			
	α (FE)	-0.0066	0.0073	-0.0210	0.0077			
	α (Sódio)	-0.0063	0.0146	-0.0350	0.0225			
	α (log(Crea.))	0.1374	0.2322	-0.3177	0.5925			
	α (log(Plaq.))	-0.1465	0.1771	-0.4935	0.2006			
	α (log(CPK))	0.0058	0.0748	-0.1408	0.1524			
	γ	0.5525	0.1374	0.2832	0.8219			
	p (Intercepto)	0.0314	6.1318	-11.9867	12.0496			
	p (Gênero)	0.1416	0.3427	-0.5300	0.8133			
	p (Fumante)	0.0571	0.3337	-0.5971	0.7112			
	p (Diabetes)	0.0754	0.2936	-0.5000	0.6509			
	p (HAS)	-0.0353	0.2982	-0.6197	0.5492			
	p (Anemia)	-0.6163	0.2957	-1.1958	-0.0367			
	p (Idade)	-0.0204	0.0121	-0.0441	0.0032			
	p (FE)	0.0419	0.0129	0.0166	0.0672			
	p (Sódio)	0.0007	0.0328	-0.0635	0.0649			
	p (log(Crea.))	-0.0155	0.5043	-1.0040	0.9730			
	p (log(Plaq.))	0.0124	0.3424	-0.6588	0.6835			
	p (log(CPK))	0.0015	0.1304	-0.2540	0.2569			
Gama	$\log(\alpha)$	1.0286	0.1594	0.7161	1.3410	1344.10	1355.20	1344.18
	$\log(\beta)$	135.7311	59.9037	18.3198	253.1424			
	$\logit(p)$	0.5119	0.0851	0.3450	0.6788			
Exponencial	$\log(\lambda)$	-4.6052	—	—	—	4.00	11.40	4.04
	$\logit(p)$	0.0000	—	—	—			

Fonte: Elaboração própria com base nos dados do estudo.

Os resultados obtidos na Tabela 8 permitem avaliar o impacto das covariáveis sobre a probabilidade de cura. O modelo Weibull de mistura apresentou covariáveis clinicamente coerentes como estatisticamente significativas, especialmente idade, fração de ejeção e anemia. A idade mostrou associação negativa com a probabilidade de cura, enquanto a fração de ejeção exibiu relação positiva, evidenciando melhor prognóstico entre indivíduos com função cardíaca mais preservada. A presença de anemia se associou a pior prognóstico, em consonância com achados previamente descritos na literatura de insuficiência cardíaca. Resultados semelhantes foram observados nos modelos Gama e Rayleigh, que também identificaram efeitos consistentes das covariáveis mencionadas.

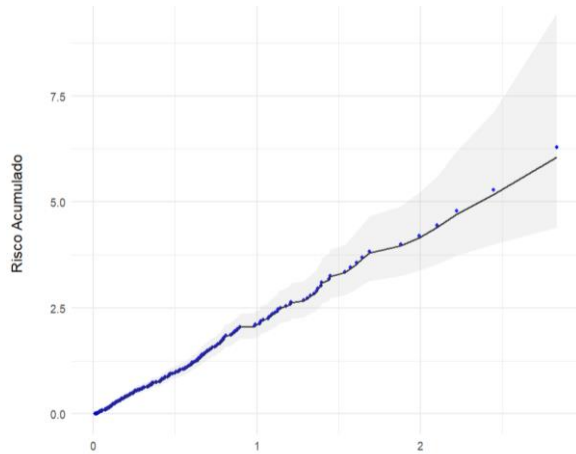
Por outro lado, o modelo Gumbel apresentou estimativas mais extremas, sugerindo instabilidade numérica ou maior sensibilidade à forma da distribuição. Ainda assim, covariáveis como hipertensão e diabetes mantiveram associações relevantes, reforçando a importância de fatores clínicos clássicos. O modelo Lindley, porém, demonstrou estimativas imprecisas, refletidas por erros-padrão elevados e intervalos de confiança amplos, o que limita sua aplicabilidade. Entre todos os modelos avaliados, o modelo Exponencial de mistura apresentou o menor AIC (1326.81), sugerindo melhor ajuste relativo. No entanto, como essa distribuição implica risco constante ao longo do tempo, sua utilização deve ser interpretada com cautela, pois tal suposição dificilmente reflete a trajetória clínica da insuficiência cardíaca. Mesmo assim, os resultados reforçam a utilidade de abordagens paramétricas com regressão para identificar variáveis prognósticas e estimar a fração de indivíduos considerados curados (Kim, 2023).

De modo geral, a análise confirma que os modelos de fração de cura com regressão fornecem estimativas coerentes com a plausibilidade clínica, contribuindo para a compreensão da heterogeneidade residual do risco entre pacientes. Essas abordagens são especialmente relevantes em condições como a insuficiência cardíaca, na qual coexistem indivíduos de alto risco para eventos precoces e outros com sobrevida prolongada quando adequadamente tratados.

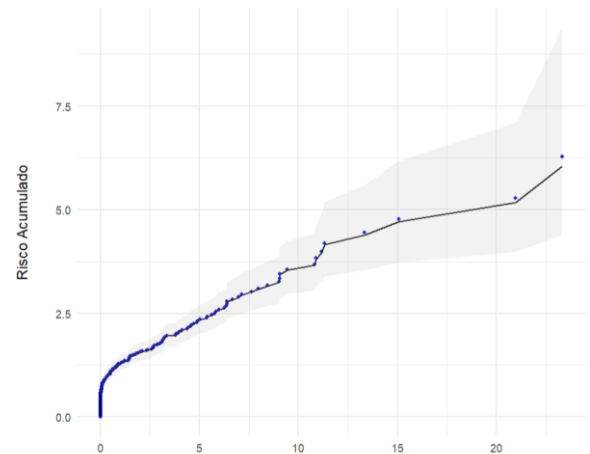
A avaliação da qualidade de ajuste dos modelos foi realizada por meio dos resíduos de Cox–Snell modificados com envelopes simulados, técnica amplamente utilizada em modelos paramétricos de sobrevivência (Colosimo & Giolo, 2006; Klein & Moeschberger, 1997). A Figura 11 apresenta os envelopes simulados correspondentes aos modelos Weibull, Gama, Lindley, Gumbel, Exponencial e Rayleigh. Idealmente, os pontos devem alinhar-se à reta de 45° e permanecer dentro da faixa de confiança simulada, indicando bom ajuste.

Figura 11: Envelopes simulados para os resíduos de Cox–Snell modificados dos modelos de fração de cura (a) Weibull, (b) Gama, (c) Lindley, (d) Gumbel, (e) Exponencial e (f) Rayleigh.

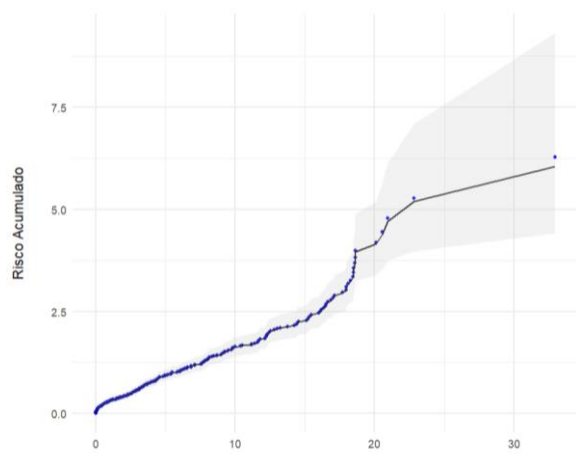
(a) Weibull



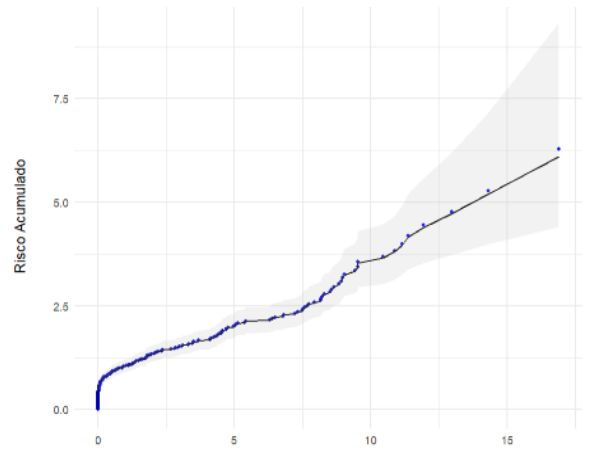
(b) Gama



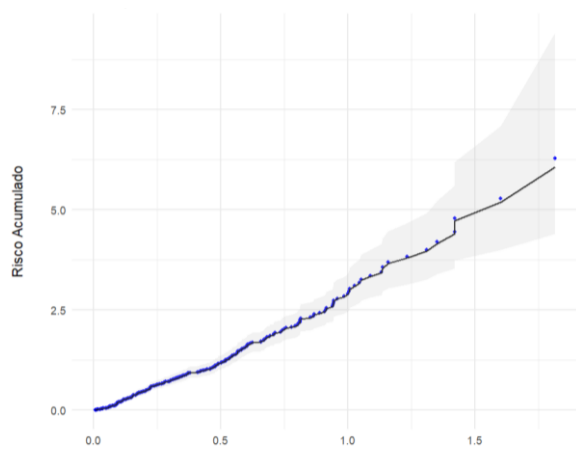
(c) Lindley



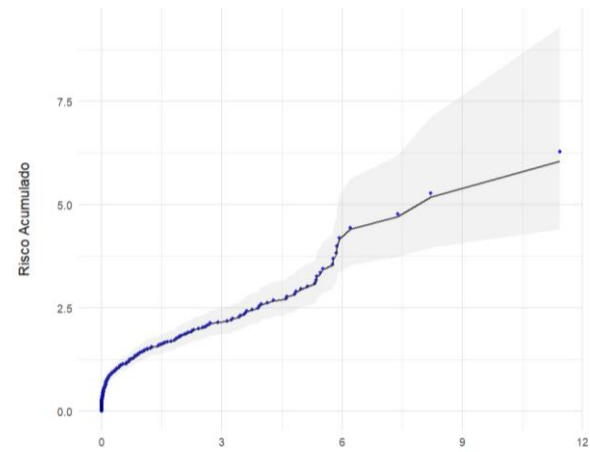
(d) Gumbel



(e) Exponencial



(f) Rayleigh



Fonte: Elaboração própria com base nos dados do estudo.

No presente estudo, o modelo Weibull foi o que apresentou maior adequação visual: os resíduos se mantiveram próximos à reta de referência, com pequenos desvios nas extremidades. O modelo Exponencial também exibiu ajuste aceitável, embora um pouco mais disperso, o que já era esperado devido à restrição do risco constante. O modelo Gama mostrou desempenho intermediário, com desvios mais evidentes na cauda. Já o modelo Rayleigh apresentou afastamentos mais pronunciados, sugerindo que sua suposição de risco estritamente crescente não se ajusta bem aos dados. Os modelos Lindley e Gumbel exibiram instabilidade, com envelopes amplos e resíduos pouco alinhados, reforçando limitações observadas anteriormente.

Esses achados demonstram a importância de combinar métricas quantitativas (AIC, BIC, AICc) e avaliação gráfica para selecionar o modelo mais adequado. Embora o modelo Exponencial tenha apresentado o menor AIC, o conjunto das evidências sugere que o modelo Weibull oferece o melhor equilíbrio entre consistência clínica, estabilidade numérica e qualidade de ajuste. Essa conclusão é coerente com a literatura, que reconhece a flexibilidade do modelo Weibull para representar riscos monotônicos típicos de doenças crônicas (Klein & Moeschberger, 1997; Colosimo & Giolo, 2006).

Por fim, foram incluídos apenas os modelos Weibull e Gama na etapa final da análise, pois foram os únicos a apresentar convergência estável e estimativas interpretáveis quando incorporados os efeitos de regressão nos parâmetros de locação. As demais distribuições apresentaram falhas de convergência ou parâmetros indefinidos, inviabilizando sua interpretação.

Os resultados indicam que a idade, função renal e fração de ejeção são determinantes prognósticos para a mortalidade em insuficiência cardíaca. Os modelos de fração de cura, especialmente o modelo Weibull, sugerem a existência de um subgrupo de pacientes com risco residual reduzido após a fase inicial de maior vulnerabilidade. Esses achados reforçam a utilidade de abordagens paramétricas flexíveis para a compreensão da heterogeneidade clínica e podem orientar estratégias de estratificação de risco mais precisas na prática assistencial.

CONCLUSÃO

O presente estudo investigou a sobrevivência de pacientes com IC utilizando tanto métodos tradicionais quanto modelos de fração de cura. As análises mostraram que idade, FE e creatinina se mantêm como importantes preditores independentes de mortalidade, resultado coerente com o que já é amplamente descrito na literatura. O modelo de riscos proporcionais de Cox apresentou o desempenho mais consistente, refletindo boa capacidade de explicar a variação do risco em uma população clinicamente heterogênea.

A inclusão dos modelos de fração de cura permitiu explorar um aspecto adicional da sobrevivência: a possibilidade de existência de um grupo de pacientes que, após determinado período, apresenta risco residual muito reduzido. Embora esses modelos dependam mais fortemente das suposições distribucionais e nem sempre sejam tão parcimoniosos quanto o modelo de Cox, eles ofereceram informações complementares importantes sobre o comportamento da sobrevivência em longo prazo.

Entre os modelos avaliados, o Weibull foi o que apresentou desempenho mais adequado no contexto da modelagem com fração de cura. Ele mostrou convergência estável, parâmetros interpretáveis e qualidade de ajuste superior tanto pelos critérios de informação quanto pela avaliação gráfica dos resíduos. Esses achados reforçam a utilidade dos modelos de fração de cura quando o objetivo é compreender padrões de risco ao longo do tempo e identificar subgrupos de sobreviventes prolongados.

De modo geral, os resultados indicam que a insuficiência cardíaca, pela própria heterogeneidade clínica, pode ser analisada de forma mais completa quando se combinam abordagens semi-paramétricas e paramétricas. Essa integração oferece uma visão mais ampla do risco, contribuindo para a estratificação de pacientes e auxiliando na tomada de decisões clínicas.

REFERÊNCIAS

- AHMAD, Tanvir; MUNIR, Assia; BHATTI, Sajjad Haider; AFTAB, Muhammad; RAZA, Muhammad Ali. Survival analysis of heart failure patients: A case study. **PLOS ONE**, v. 12, n. 7, e0181001, 2017.
- AMERICAN DIABETES ASSOCIATION. Classification and diagnosis of diabetes. **Diabetes Care**, v. 38, suppl. 1, p. S8–S16, 2015.
- ARMITAGE, Peter; BERRY, Geoffrey; MATTHEWS, J. N. S. **Statistical Methods in Medical Research**. 4. ed. Oxford: Blackwell Science, 2002.
- ARRUDA, V. L.; MACHADO, L. M. G.; LIMA, J. C.; SILVA, P. R. S. Tendência da mortalidade por insuficiência cardíaca no Brasil: 1998 a 2019. **Revista Brasileira de Epidemiologia**, v. 25, p. 1-15, 2022.
- AUJLA, R. S.; ZUBAIR, M.; PATEL, R. **Creatine Phosphokinase**. In: StatPearls. Treasure Island (FL): StatPearls Publishing, 2025. Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK546624/>
- BORDALO, A. A. Estudo transversal e/ou longitudinal. **Revista Paraense de Medicina**, v. 20, n. 4, 2006.
- BRAUN, M. M.; BARSTOW, C. H.; PYZOCHA, N. J. Diagnosis and management of sodium disorders: hyponatremia and hypernatremia. **American Family Physician**, v. 91, n. 5, p. 299-307, 2015.
- CARDOSO, Juliano; BRITO, Michel Ibrahim; OCHIAI, Marcelo Eidi; NOVAES, Milena; BERGANIN, Fabrício; THICON, Tatiana; FERREIRA, Elaine C.; REGINA, Kelly; REIS, Cristina Martins dos; BARRETTO, Antonio Carlos Pereira. Anemia nos pacientes com insuficiência cardíaca avançada. **Arquivos Brasileiros de Cardiologia**, v. 95, n. 4, p. 1-9, 2010.
- COLETT, D. **Modelling Survival Data in Medical Research**. CRC Press, 2015.
- COX, D. R. Regression models and life tables (with discussion). **Journal of the Royal Statistical Society – Series B**, v. 34, p. 187–220, 1972.
- COX, D. R.; SNELL, E. J. A general definition of residuals. **Journal of the Royal Statistical Society – Series B**, v. 30, n. 2, p. 248-265, 1968.
- DANZMANN, L. C.; BODANESE, L. C.; PETZOLD, A. P.; TSCHIEKA, A. P.; MAGEDANZ, E. H.; PETERSEN, L. C.; BELYAVSKIY, E.; CHIEZA, F. L. Sobrevida em 10 anos de pacientes com insuficiência cardíaca com FEVE 40–59%. **Arquivos Brasileiros de Cardiologia**, v. 120, n. 1, p. 1-10, 2023.
- ESTRELA, C. (Org.). **Metodologia Científica**. Porto Alegre: Artes Médicas, 2018. (Obs.: Mantive capítulo citado abaixo.)

FREIRE, M. C. M.; PATTUSSI, M. P. **Tipos de estudos**. In: ESTRELA, C. (Org.). Metodologia científica. Porto Alegre: Artes Médicas, 2018. p. 109–127.

FREITAS, E. V.; BATLOUNI, M.; GAMARSKY, R. Heart failure in the elderly. **Journal of Geriatric Cardiology**, v. 9, n. 2, p. 101-107, 2012.

GIOLO, S. R.; COLOSIMO, E. A. **Análise de Sobrevivência Aplicada**. São Paulo: Blucher, 2006.

GROENEWEGEN, A.; RUTTEN, F. H.; MOSTERD, A.; HOES, A. W. Epidemiology of heart failure. **European Journal of Heart Failure**, v. 22, n. 8, p. 1342–1356, 2020.

GUMBEL, E. J. Bivariate exponential distributions. **Journal of the American Statistical Association**, v. 55, n. 292, p. 698–707, 1960.

GUYTON, Arthur C.; HALL, John E. **Tratado de Fisiologia Médica**. 13. ed. Rio de Janeiro: Elsevier, 2017.

HEIDENREICH, P. A.; BOZKURT, B.; AGUILAR, D.; ALLEN, L. A.; BYUN, J. J.; COLVIN, M. M.; DESWAL, A.; DRAZNER, M. H.; DUNLAY, S. M.; EVERS, L. R.; FANG, J. C.; FEDSON, S. E.; FONAROW, G. C.; HAYEK, S. S.; HERNANDEZ, A. F.; KHAZANIE, P.; KITTLESON, M. M.; LEE, C. S.; LINK, M. S.; MILANO, C. A.; NNACHETA, L. C.; SANDHU, A. T.; STEVENSON, L. W.; VARDENY, O.; VEST, A. R.; YANCY, C. W. 2022 AHA/ACC/HFSA guideline for the management of heart failure. **Journal of the American College of Cardiology**, v. 79, n. 17, p. e263–e421, 2022.

HOFFMAN, Ronald; BENZ, Edward J.; SILBERSTEIN, Leslie E.; HESLOP, Helen E.; WEITZ, Jeffrey I.; ANASTASI, John; SALAMA, Mohammed E.; ABUTALIB, Syed. **Hematology: Basic Principles and Practice**. 7. ed. Elsevier, 2017.

IN, J.; LEE, D. K. Survival analysis: Applied clinical data analysis. **Korean Journal of Anesthesiology**, v. 72, n. 5, p. 441–457, 2019.

JACKSON, D. et al. **Cure models: what is meant by a survival plateau?** Pharmacoeconomics, 2025.

KALBFLEISCH, J. D.; PRENTICE, R. L. **The Statistical Analysis of Failure Time Data**. 2. ed. New York: Wiley, 2002.

KANG, Y. et al. Relação entre a razão nitrogênio ureico/creatinina e prognóstico de insuficiência cardíaca. **Arquivos Brasileiros de Cardiologia**, 2022.

KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. **Journal of the American Statistical Association**, v. 53, n. 282, p. 457–481, 1958.

KDIGO WORK GROUP. **KDIGO guideline for CKD**. Kidney International Supplements, v. 3, n. 1, p. 1–150, 2013.

KIM, M. **Appropriate use of parametric and nonparametric methods**. Stat, v. 12, n. 1, e606, 2023.

- KLEIN, J. P.; MOESCHBERGER, M. L. **Survival Analysis: Techniques for Censored and Truncated Data**. Springer, 1997.
- LAMBERT, P. C. Modeling of the cure fraction in survival studies. **The Stata Journal**, v. 7, n. 3, p. 351–375, 2007.
- LAWLESS, J. F. **Statistical Models and Methods for Lifetime Data**. New York: Wiley, 1982.
- LEE, Christopher S. et al. Biomarkers of myocardial stress and systemic inflammation. **Journal of Cardiovascular Nursing**, v. 26, n. 4, p. 321–328, 2011.
- LINDLEY, D. V. Fiducial distributions and Bayes' theorem. **Journal of the Royal Statistical Society – Series B**, v. 20, p. 102–107, 1958.
- LITTLE, R. J. A.; RUBIN, D. B. **Statistical Analysis with Missing Data**. Hoboken: Wiley, 2019.
- MARTINEZ, E. Z.; ACHCAR, J. A.; JACOME, A. A.; SANTOS, J. S. Mixture and non-mixture cure fraction models... **Computer Methods and Programs in Biomedicine**, v. 112, n. 3, p. 343–355, 2013.
- MCCLELLAN, W. M. et al. Anemia and renal insufficiency... **Journal of the American Society of Nephrology**, v. 13, n. 7, p. 1928–1936, 2002.
- MIOT, H. A. Análise de sobrevivência em estudos clínicos e experimentais. **Jornal Vascular Brasileiro**, v. 16, n. 4, p. 269–275, 2017.
- NEW YORK HEART ASSOCIATION. **Diseases of the Heart and Blood Vessels: Nomenclature and Criteria for Diagnosis**. 7. ed. Boston: Little, Brown, 1973.
- OLIVEIRA, E. C. de. **Modelos de fração de cura...** Dissertação – UNESP, 2015.
- OLIVEIRA, L. E. F.; SANTOS, L. S.; FABIO, L. C.; FERREIRA, P. H.; CARRASCO, J. M. F. Análise de resíduos para o modelo logístico generalizado dependente do tempo. **Trends in Computational and Applied Mathematics**, v. 24, n. 4, p. 635–670, 2023.
- PRICE, D. L.; MANATUNGA, A. K. Modelling survival data with a cured fraction. **Statistics in Medicine**, v. 20, p. 1515–1527, 2001.
- R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. Vienna: R Foundation, 2012.
- REBORA, P.; SALIM, A.; REILLY, M. bshazard: A flexible tool for smoothing. **The R Journal**, v. 6, n. 2, p. 114–122, 2014.
- RODRIGUEZ, Mario et al. Hyponatremia in heart failure. **Current Cardiology Reviews**, v. 15, n. 4, p. 252–261, 2019.

ROHDE, L. E. et al. Diretriz Brasileira de Insuficiência Cardíaca Crônica e Aguda. **Arquivos Brasileiros de Cardiologia**, v. 111, n. 3, p. 436–539, 2018.

ROY, D. Discrete Rayleigh distribution. **IEEE Transactions on Reliability**, v. 53, n. 2, p. 255–260, 2004.

SPINETI, P. P. M. Avaliando a adesão à restrição de sódio. **Arquivos Brasileiros de Cardiologia**, v. 112, n. 2, p. 171–173, 2019.

THERNEAU, T. M. **A package for Survival Analysis in R**. Version 3.5-0, 2024.

THERNEAU, T. M.; GRAMBSCH, P. M. **Modeling Survival Data: Extending the Cox Model**. New York: Springer, 2000.

TSODIKOV, A. Semi-parametric models of long- and short-term survival. **Statistics in Medicine**, v. 21, n. 6, p. 895–920, 2002.

WILLIAMS, B. et al. ESC/ESH Guidelines for the management of arterial hypertension. **European Heart Journal**, v. 39, n. 33, p. 3021–3104, 2018.