

## ESTIMATIVAS DE RISCO DE LESÃO EM MARATONISTA PELA REGRESSÃO DE COX

*CARVALHO JUNIOR, Sergio<sup>1</sup>; FAIAL, Vanessa Esteves<sup>2</sup>; EVERTON, Adriana Nunes da Fonseca<sup>3</sup>; BRITO, Diogo de Freitas<sup>4;5;6</sup>; BARRETO, Ana Cristina Lopes y Glória<sup>5</sup>; BRASIL, Roxana Macedo<sup>7</sup>; JUNIOR, Homero da Silva Nahum<sup>5;8</sup>*

152

### Resumo

Pelo emprego da regressão de Cox objetivou-se modelar o risco de lesão em grupo de maratonista, contendo 40 homens e 16 mulheres, utilizando as variáveis idade, tempo de prática, ocorrência de lesão e número de dias antes do evento lesivo, considerando o macrociclo de treinamento de 2023. Em Sexo, a categoria de referência foi Feminino, e para tempo de prática imposto foi o intervalo 5-7 anos. A codificação foi realizada em linguagem R 4.3.3. Os resultados revelaram que a cada ano vivido o risco aumentaria 2,60%. Enquanto ser Sexo = Masculino e Prática de 8 a 10, 11 a 13 e 14 a 16 anos detinham HR, respectivamente de 1,43, 1,37, 1,69 e 0,50. Então, possível foi concluir que havia indícios de padrões, mas sem sustentação estatística.

**Palavras-chave:** Planejamento. Modelagem. Administração. Economia. Estatística.

### Abstract

Using Cox regression, this study aimed to model the risk of injury in a group of marathon runners, comprising 40 men and 16 women, using the variables age, training experience, injury occurrence, and number of days before the injury event, considering the 2023 training macrocycle. The reference category for sex was Female, and the imposed training experience interval was 5-7 years. Coding was performed using R 4.3.3. The results revealed that the risk increased by 2.60% for each year of training. However, being Male and having 8-10, 11-13, and 14-16 years of training had HRs of 1.43, 1.37, 1.69, and 0.50, respectively. Therefore, it was possible to conclude that there were indications of patterns, but without statistical support.

**Keywords:** Planning. Modelling. Administration. Economics. Statistics.

<sup>1</sup> Pesquisador convidado Biodesa;

<sup>2</sup> Graduanda do Curso de Educação Física do Centro Universitário Celso Lisboa;

<sup>3</sup> Profissional de Educação Física da Adriana Nunes Consultoria de Corrida;

<sup>4</sup> Docente do Curso de Gestão Desportiva e do Lazer do Centro Universitário Celso Lisboa;

<sup>5</sup> Docentes do Curso de Educação Física do Centro Universitário Celso Lisboa;

<sup>6</sup> Consultor Iceberg Business Academy;

<sup>7</sup> Docente Ph.D. em Educação Física;

<sup>8</sup> Docente da Escola de Saúde da Universidade Cândido Mendes.

## Introdução

A Regressão de Cox ou Modelo de Riscos Proporcionais de Cox seria método semiparamétrico de análise de sobrevivência para investigar a associação entre tempo de sobrevivência de indivíduos e covariáveis (variáveis preditoras). O efeito dessas tomado parametricamente, enquanto ao tempo de vida não se imporia distribuição de probabilidade específica. Dessa forma estimado seria o risco de determinado evento ocorrer em certo momento, não o tempo de vida, propriamente dito (Cox, 1972).

A ideia norteadora seria a constância no domínio do tempo do risco relativo entre dois grupos (Pressuposto dos Riscos Proporcionais), o que permitiria a comparação de grupos em razão do risco, originando a razão de riscos. Essa característica seria fundamental na investigação de fenômenos cuja idiosincrasia seria o desconhecimento sobre tempo até a existência do evento, o qual poderia não se manifestar (Colosino e Giolo, 2006). Peculiaridade não raramente presente em ciências Humanas Aplicadas (Tomazela, 2007; Alves, 2009; Nasu *et al.*, 2022; Lins, 2023; Souza, Golgher e Silva, 2024), Humanas (Gonçalves, 2008; Lima Junior, Silveira e Ostermann, 2012; Campos, 2016; Saccaro, França e Jacinto, 2019; Errandonea, 2024) e Saúde (Botelho, Silva e Cruz, 2009; Cachapuz, 2010; Ferraz e Moreira-Filho, 2017; Miot, 2017; Arias *et al.*, 2018; Galvão e Roncalli, 2020; Freitas *et al.*, 2021; Schneider *et al.*, 2022; Collet, 2023; Ferreira *et al.*, 2024; Guimarães, 2024; Shehadeh *et al.*, 2024; Vieira, 2024).

A probabilidade de certo evento ocorrer em determinado tempo  $t$ , sabendo que não houve lesão anterior desde o início do estudo seria dada pela função de risco, *hazard function* ( $h$ ), cujo coeficiente (*hazard ratio* – HR) associado à covariável  $X_i$  seria denotado por  $\beta_i$ , permitindo interpretações (Quadro I) em razão do valor estimado (Cox, 1972).

$$h(t|X) = h_0(t) \cdot \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n) \quad (I)$$

onde:  $h(t|X)$ : função de risco no tempo  $t$  para um indivíduo com definido conjunto de covariáveis  $X$ ;  $h_0(t)$ : função de risco basal, dependente somente do tempo, não guardando especificidade, forneceria o risco quando todas as covariáveis fossem nulas.

A grande inovação de Cox foi contornar a necessidade de conhecer  $h_0(t)$ , usando uma função de verossimilhança parcial (*partial likelihood*). A ideia seria comparar, a cada momento em que um evento ocorre, o risco do indivíduo que sofreu o evento com o risco de todos os outros indivíduos ainda submetidos ao estímulo (por exemplo: corrida,

competição ou treino) naquele mesmo momento. Então, os coeficientes  $\beta$  seriam estimados maximizando aquela verossimilhança parcial, que não dependeria da forma da função de risco basal (Kleibaum e Klein, 2011).

**Quadro I:** Relações entre  $\beta_i$  e HR, Interpretações e Exemplos Simplificados (mantidas constantes as demais variáveis).

$\beta_i$	HR	Interpretação	Exemplos
< 0	< 1	Aumento de $X_i$ estaria associado ao menor risco do evento (sentidos inversos).	HR = 0,70, reduziria o risco em 0,30 a cada unidade de $X_i$ .
= 0	= 1	$X_i$ não influenciaria o risco.	
> 0	> 1	Aumento de $X_i$ estaria associado ao aumento do risco do evento (mesmo sentido).	HR = 1,40, aumentaria o risco em 0,40 a cada unidade de $X_i$ .

**Fonte:** Os Autores (2026).

Como consequência do exposto, eliminadas foram as suposições sobre a distribuição dos tempos de sobrevivência, mitigando erros e favorecendo a robustez do modelo. Adicionalmente, a interpretação de HR se tornaria intuitiva, favorecendo a compreensão do efeito de covariáveis, mesmo em número elevado (Ma, Webb e Hudson, 2024). Essas variáveis poderiam ser controladas por fatores de confusão (variável que influência a exposição e o resultado, distorcendo essa relação), ajustando o efeito principal por aquele de outra variável (Kleibaum e Klein, 2011; Carvalho *et al.*, 2021). O modelo de Cox lidaria adequadamente com censura temporal, ou seja, quando os indivíduos fossem expostos ao estímulo por prazo determinado, por exemplo, a duração de maratona ou macrociclo de treinamento. Por conseguinte, para cada atleta haveria as seguintes possibilidades: 1) sofrer lesão em determinado momento do período de observação; 2) abandonar a exposição (treinamento ou disputa) sem lesão; e 3) findar o período sem acometimento lesivo. As duas últimas ocorrências caracterizariam dados censurados (Cox, 1972; Colosino e Giolo, 2006; Carvalho *et al.*, 2021).

Na residência das desvantagens estariam 1) a impossibilidade de estimar o risco absoluto para novo indivíduo, como consequência da não especificação da função de risco basal; 2) a influência de valores extremos (*outliers*) sobre as estimativas de coeficientes; e 3) o pressuposto de independência temporal das covariáveis, ou melhor, assumida seria a constância das covariáveis no domínio do tempo, elevando a complexidade da elaboração do modelo para recepcionar variabilidade da frequência cardíaca (Jandre *et al.*, 2024), pressão arterial (Barreto *et al.*, 2023), frequência respiratória (Balula, 2024), nível de lactato

(Reis *et al.*, 2013) e consumo máximo de oxigênio (Braga e Stein, 2025) dentre outras variáveis. As estimativas poderiam sofrer influência de vieses, caso ocorresse a violação do pressuposto de risco proporcionais, o que tornaria a realização de testes sobre os resíduos fundamental (Cox, 1972; Kleibaum e Klein, 2011; Carvalho *et al.*, 2021; Ma, Webb e Hudson, 2024).

Tais aspectos não foram suficientes para comprometer a investigação epidemiológica do câncer de pulmão tendo por fatores de risco tabagismo, *cannabis sativa* e sistemas eletrônicos de administração de nicotina (Groot *et al.*, 2018). A regressão de Cox foi empregada para estimar a mortalidade em 251 pacientes com insuficiência cardíaca. Durante 12,30 anos de acompanhamento, aproximadamente 19,12% fizeram óbito, tendo diferença estatisticamente significativa (valor-p < 0,05) em relação aos indivíduos não acometidos, HR = 4,46 (IC<sub>95,00%</sub> = [3,30; 5,90]). As taxas de sobrevivência nos instantes 2,00, 6,00, 10,00 e 12,30 anos foram 96,02%, 88,84%, 82,07% e 80,08%, respectivamente (Lédo *et al.*, 2025).

Com relação à corrida recreativa, 226 homens e 306 mulheres foram acompanhados durante 13 semanas de treinamento para disputa de 6,70 km. Possível foi identificar que, aproximadamente, 20,00% do grupo apresentou, pelo menos, um evento lesivo durante o período. Todavia, o modelo revelou que os fatores preditivos para homens eram IMC, lesão prévia nos últimos 12 meses e inexistência de carga axial, enquanto dentre as mulheres, somente a pronação excessiva do pé (“queda do navicular”) mostrou significância estatística. A associação entre risco de lesões e erros de treino não foi conclusiva, pois os resultados foram conflitantes na consideração de infecção respiratória aguda e variáveis de treinamento como volume, intensidade, frequência e duração (Nielsen *et al.*, 2012). Então, objetivou-se modelar por regressão de Cox o risco de lesão em maratonista.

## Metodologia

O banco de dados utilizado foi desenvolvido por Carvalho Junior *et al.* (2025), contendo 40 homens e 16 mulheres. Essas tinham Idade = 41,13 ± 4,70 anos (Mediana = 42,00 anos e Coeficiente de Variação = 11,44%), Prática = 8,88 ± 3,26 anos (Mediana = 8,00 anos e Coeficiente de Variação = 36,77%) e quatro ocorrências lesivas. O grupo Masculino detinha Idade = 37,23 ± 4,95 anos (Mediana = 36,50 anos e Coeficiente de Variação = 13,29%), Prática = 9,98 ± 3,68 anos (Mediana = 10,00 anos e Coeficiente de Variação = 36,92%) e 11 acometidos por lesão. Além dessas variáveis, o banco de dados

possuía o campo Status, que indicava a ocorrência da lesão (0: inexistente; 1: existente). A Data Inicial = 1/1/2023 era constante. A variável Data Final recebia o dia da ocorrência de lesão, possibilitando determinar o Tempo sem lesão para cada atleta. Esses dados foram registrados ao longo do macrociclo de 2023.

A modelagem desconsiderou a utilização de riscos competitivos (Prentice et al., 1978), mais de uma causa à lesão, por ausência de convergência à abrangência pretendida. A Prática foi categorizada em 5-7, 8-10, 11-13 e 14-16 anos, tomando a primeira como referência. Em Sexo, ao Feminino foi atribuída essa característica. Essas determinações foram estabelecidas deliberadamente em razão das ordens numérica e alfabética, respectivamente. A codificação foi realizada em linguagem R 4.3.3. com o emprego dos pacotes readxl 1.4.5, splines 4.4.3, survival 3.8-3 e survminer 0.5.1.

## Resultados e Discussão

Cada a variável tenderia a aumentar o risco de lesão,  $HR > 1,00$  (Tabela 1), mantidas as demais constantes (*ceteris paribus*), ou seja, tratar-se-ia de efeito líquido ou ajustado. Assim, em indivíduos do Sexo Masculino a chance seria 42,51% ( $\approx 43,00\%$ ) maior do que entre as mulheres. Interessante seria analisar  $HR_{Idade} = 2,60\%$  ( $\approx 3,00\%$ ), pois significaria que a cada ano vivido, o risco aumentaria no percentual estimado, dado que o fenômeno não seria linear (aditivo), então para  $n$  anos, o  $HR = 1,026^n$ , conseqüentemente, após 10 anos,  $HR = 1,026^{10} \approx 1,2925 \approx 29,25\%$  (Figura 1).

Comparativamente ao tempo de Prática 5-7 (Tabela 1), os indivíduos com 8 a 10 anos e 11 a 13 anos de experiência estariam sob maiores riscos, nessa ordem, 36,69% ( $\approx 37,00\%$ ) e 68,83% ( $\approx 69,00\%$ ), portanto esse guardaria maior vulnerabilidade. Em oposição, os indivíduos com 14 a 16 anos de Prática detinham 49,90% ( $\approx 100,00\% - 50,10\%$ ) menor risco de acometimento por lesão, talvez pelo desenvolvimento de táticas de prevenção ou adaptações favorecedores da preservação. Em suma, essa variável teria comportamento parabólico com concavidade negativa ( $\cap$ ), porque o risco se elevaria, sofreria inflexão e diminuiria.

As variáveis tempo de conclusão de maratona e Idade descreveriam uma parábola de concavidade positiva ( $\cup$ ), independentemente da dicotomia sexual (Lara, Salinero e Del Coso, 2014). E, em geral, maratonistas de elite teriam pico de desempenho entre 37 e 32 anos, contra o período entre 30 e 40 anos dentre amadores (Lara, Salinero e Del Coso,

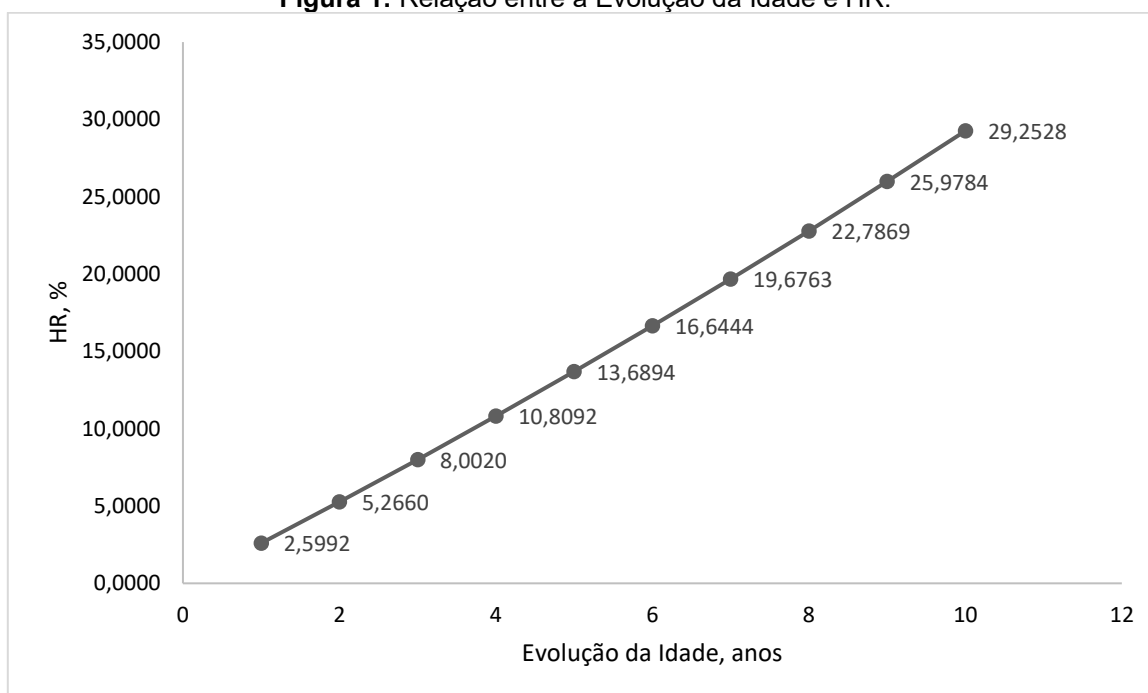
2014; Soloaga, 2021; Carvalho e Rebutini, 2023). Esses fatos poderiam explicar parcialmente os resultados, em que pese a ausência de significância estatística de todas as estimativas, valor-p > 0,05 (Tabela 1). Ou seja, as estimativas identificaram tendências, as quais poderia ser efeitos do acaso, dado que não houve evidência estatística suficiente para ratificá-las.

**Tabela 1:** Estimativas de Coeficiente ( $\beta_i$ ), HR( $\exp(\beta_i)$ ) e Erro Padrão ( $\varepsilon$ ),  $\alpha = 0,05$ .

Variável	$\beta_i$	$\varepsilon(\beta_i)$	HR	IC <sub>95,00%</sub>	z	Pr(> z ) - valor-p
Sexo Masculino	0,35	0,64	1,43	[0,40; 5,01]	0,55	0,58
Idade	0,03	0,06	1,03	[0,92; 1,15]	0,45	0,66
Prática 8-10	0,31	0,70	1,37	[0,34; 5,40]	0,45	0,66
Prática 11-13	0,52	0,65	1,69	[0,47; 6,06]	0,80	0,42
Prática 14-16	-0,69	1,10	0,50	[0,06; 4,34]	-0,63	0,53

Fonte: Os Autores (2026).

**Figura 1:** Relação entre a Evolução da Idade e HR.



Fonte: Os Autores (2026).

Os testes globais, *Likelihood ratio*, Wald e Score (Tabela 2), revelaram que as covariáveis como conjunto não explicariam significativamente o risco de lesão, valor-p > 0,05. Estatisticamente, o modelo não rejeitaria a hipótese de efeito nulo das covariáveis. A Concordância estimaria a capacidade de identificar o atleta que sofreria o evento antes, tendo existência no intervalo: 0,50 – discriminação ao acaso e 1,00 – discriminação perfeita. O modelo conquistou C-índice = 0,61, indicando capacidade modesta (≈61,00%)

e com ampla incerteza, porque o intervalo de confiança (IC<sub>95,00%</sub>) estaria definido no intervalo [0,45; 0,77] ([0,61 - 1,96 . 0,08; 0,61 + 1,96 . 0,08]).

Isso pode ter ocorrido em virtude do 1) tamanho do grupo, especialmente da quantidade de lesionados (15 pessoas – 26,78%), 2) as covariáveis não tinham capacidade explicativa adequada aqueles indivíduos, ou 3) poderia haver relação não linear entre risco e Idade e Prática. As consequências seriam incerteza sobre o desempenho do modelo e não confiabilidade à tomada de decisão, porque as estimativas tenderiam a ser imprecisas (Carvalho *et al.*, 2021).

**Tabela 2:** Resultados Inferenciais com Graus de Liberdade = 5,00 e Erro Padrão\*,  $\alpha = 0,05$ .

Teste	Estimativa	valor-p
Concordância	0,61	0,08*
<i>Likelihood ratio</i>	1,86	0,90
Wald	1,65	0,90
Score (logrank)	1,73	0,90

Fonte: Os Autores (2026).

O teste de correlação dos resíduos de Shoenfeld com o tempo (Tabela 3) informaria a existência de tendência sistemática longitudinal dos erros (Ma, Webb e Hudson, 2024), assim, quando valor-p < 0,05 significaria a existência de violação do pressuposto de riscos proporcionais pelo modelo (Global) ou por alguma covariável. Como em todos os casos valor-p > 0,05, o modelo foi confirmado como válido (consistente), mesmo estatisticamente sendo pouco informativo (baixa robustez).

**Tabela 3:** Resultados do Teste de Wald sobre as Covariáveis,  $\alpha = 0,05$ .

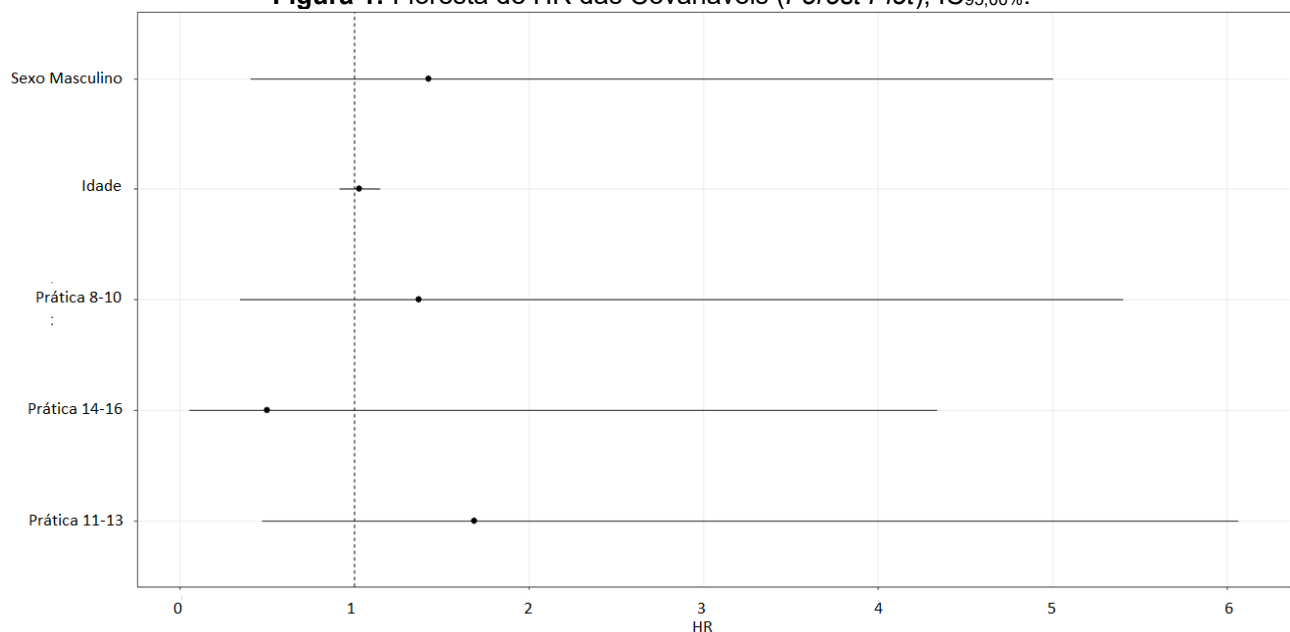
Variáveis	$\chi^2$	GL	Valor-p
Global	4,40	5	0,49
Sexo	1,61	1	0,20
Idade	1,64	1	0,20
Experiência	2,62	3	0,45

Fonte: Os Autores (2026).

A representação de HR das Covariáveis (Figura 1) apresentaria as estimativas (pontos) e respectivos Intervalos de Confiança para  $\alpha = 0,05$  (linhas horizontais). A linha vertical pontilhada seria a de não-efeito (HR = 1,00), conseqüentemente, toda estimativa à direita (HR > 1,00) indicaria variável de risco, mas quando à esquerda (HR < 1,00) identificaria fator de proteção. Em outros termos, 1) à direita: o aumento na variável se

associaria à elevação do risco; e 2) à esquerda: o aumento na variável se associaria à redução do risco. Caso, a estimativa residisse sobre a linha de não-efeito, então a variável não teria associação com o risco. A ausência de interseção entre as linhas vertical e horizontal (não inclusão de HR = 1,00) indicaria a significância da estimativa (Collett, 2023), o que não ocorreu no modelo desenvolvido, ratificando a Tabela 1.

**Figura 1:** Floresta de HR das Covariáveis (*Forest Plot*), IC<sub>95,00%</sub>.



**Fonte:** Os Autores (2026).

A análise de resíduos de Martingale e Deviance objetivaria investigar a forma funcional das variáveis contínuas, assim como a existência de valores extremos (*outliers*). Mais claramente, a regressão de Cox assumiria que o  $\log(\text{HR})$  seria influenciado pelo comportamento linear de cada covariável, portanto as representações gráficas dos resíduos permitiriam averiguar tal hipótese. O Resíduo de Martingale (II) se concentraria no número de eventos de cada indivíduo. Um atleta poderia ter uma lesão ou nenhuma, logo haveria tendência de assimetria, dado que se o maratonista poderia apresentar valores: 1) negativos elevados ( $< -0,50$ ), indicando indivíduo censurado com alta predição de ocorrência de lesão; ou 2) positivos elevados ( $> 0,50$ ) ou próximos de +1, a predição do modelo foi de baixa probabilidade, mas a lesão ocorreu. Isso permitiria ajustar o modelo e se a covariável deveria ser transformada (Schneckenberg Filho, 2020).

$$M_i = \delta_i - \hat{H}_i(t) \quad (II)$$

onde:  $\delta_i$ : evento observado, 0: não lesionado e 1: lesionado;  $\hat{H}_i(t)$ : valor esperado acumulado do risco para o atleta  $i$ .

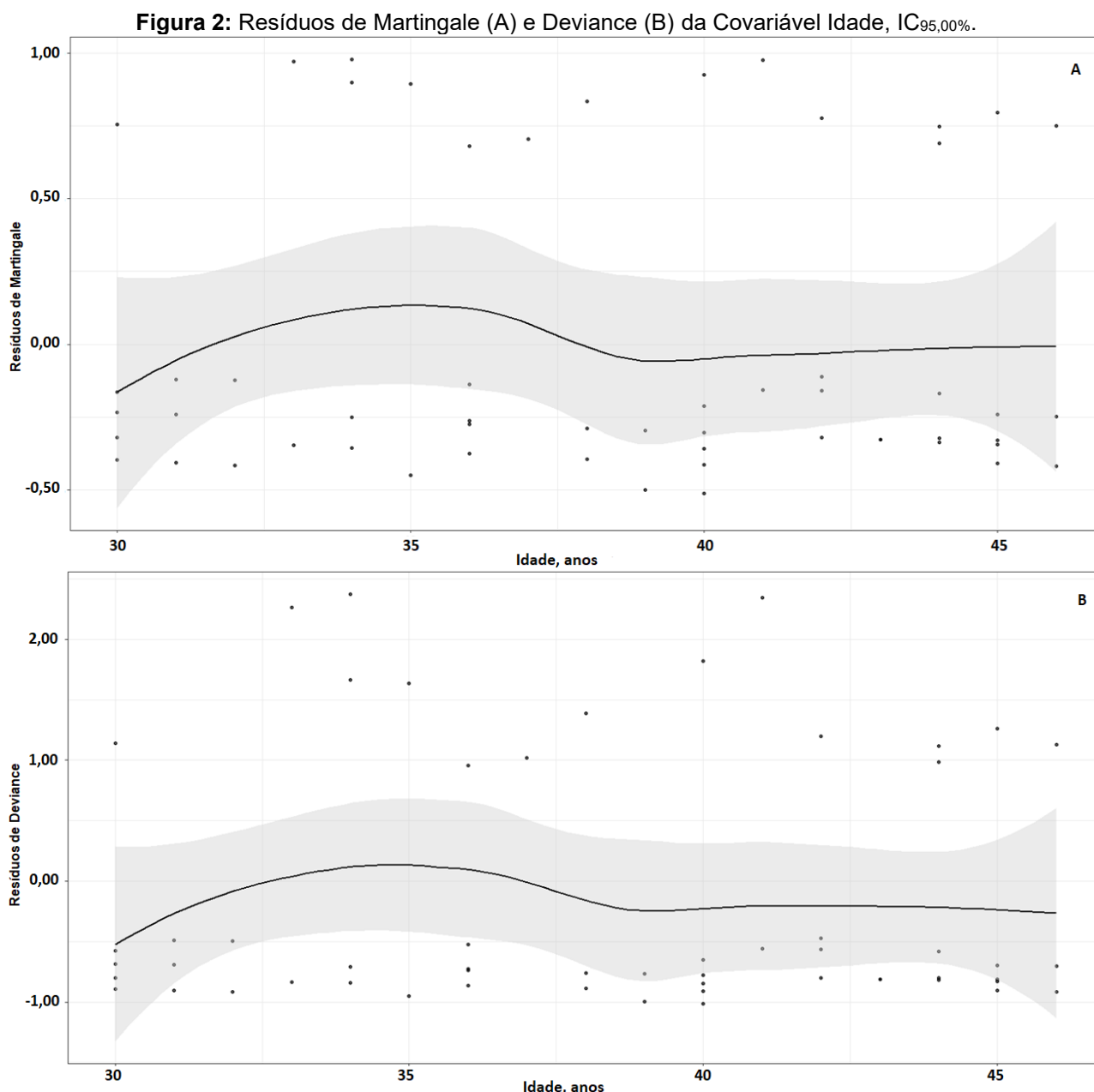
A linha horizontal em  $Y = 0,00$  indicaria o ajuste perfeito, portanto, resíduo próximo do valor nulo seria a adequada predição, considerando a área em cinza (Figura 2A), determinada Intervalo de Confiança (IC<sub>95,00%</sub>). Residentes nessa situação eram 18 esportistas ( $\approx 32,14\%$ ). A curva que se apresentou teria formato perceptível, indicando não linearidade entre Idade e  $\log(\text{HR})$ , o que poderia ser resolvido com transformação,  $\log(\text{Idade})$  ou  $\text{Idade}^2$ , por exemplo, ou adicionando termo polinomial,  $\text{Idade} + \text{Idade}^2$ . Valores muito positivos ( $0,50 \leq \text{Resíduos} \leq 1,00$ ) totalizaram 15 ocorrências ( $\approx 26,78\%$ ) de lesão, porém o modelo as identificou como de baixo risco, mas elegíveis a valores extremos seriam. Somente dois atletas ( $\approx 3,57\%$ ) conquistaram Resíduo =  $-0,50$ . Entre esse valor e a nulidade, porém externamente ao Intervalo de Confiança, houve 20 maratonistas ( $\approx 35,71\%$ ), informando moderada predição de lesão, porém sem acometimento do evento.

Os resíduos de Deviance (III) seriam a suavização dos anteriores, apresentando maior nível de simetria, o que tenderia a favorecer a análise do ajuste e identificação de valores extremos. Porém, preservando as ideias da predição perfeita e intervalo de confiança. Usualmente, valores extremos apresentariam Resíduo  $\approx \pm 2,00$ , considerados influentes seriam aqueles com Resíduo  $\approx \pm 3,00$  (López Fidalgo, 2000). Situação essa que não ocorreu em Idade (Figura 2B), todavia como discrepantes haveria três observações ( $\approx 5,35\%$ ). Interessante atentar para o fato de que a suavização reduziu a quantidade de atletas durante do Intervalo de Confiança, uma vez que em Deviance 16 pessoas (28,57%) se encontravam nessa situação. Valeria destacar que não, necessariamente, seriam as mesmas, porque Martingale se basearia na diferença absoluta entre observado e predito, enquanto Deviance consideraria a magnitude ajustada e simetrizada (Schneckenberg Filho, 2020).

$$D_i = \text{sin}(\alpha) \sqrt{-2(M_i + \delta_i \cdot \ln(\delta_i - M_i))} \quad (III)$$

onde:  $\delta_i$ : evento observado, 0: não lesionado e 1: lesionado;  $M_i$ : resíduo de Martingale;  

$$\text{sin}al(M_i) = \begin{cases} -1, & \text{se } x < 0 \\ 0, & \text{se } x = 0 \\ 1, & \text{se } x > 0 \end{cases}$$
 : para preservar o sinal do resíduo de Martingale.

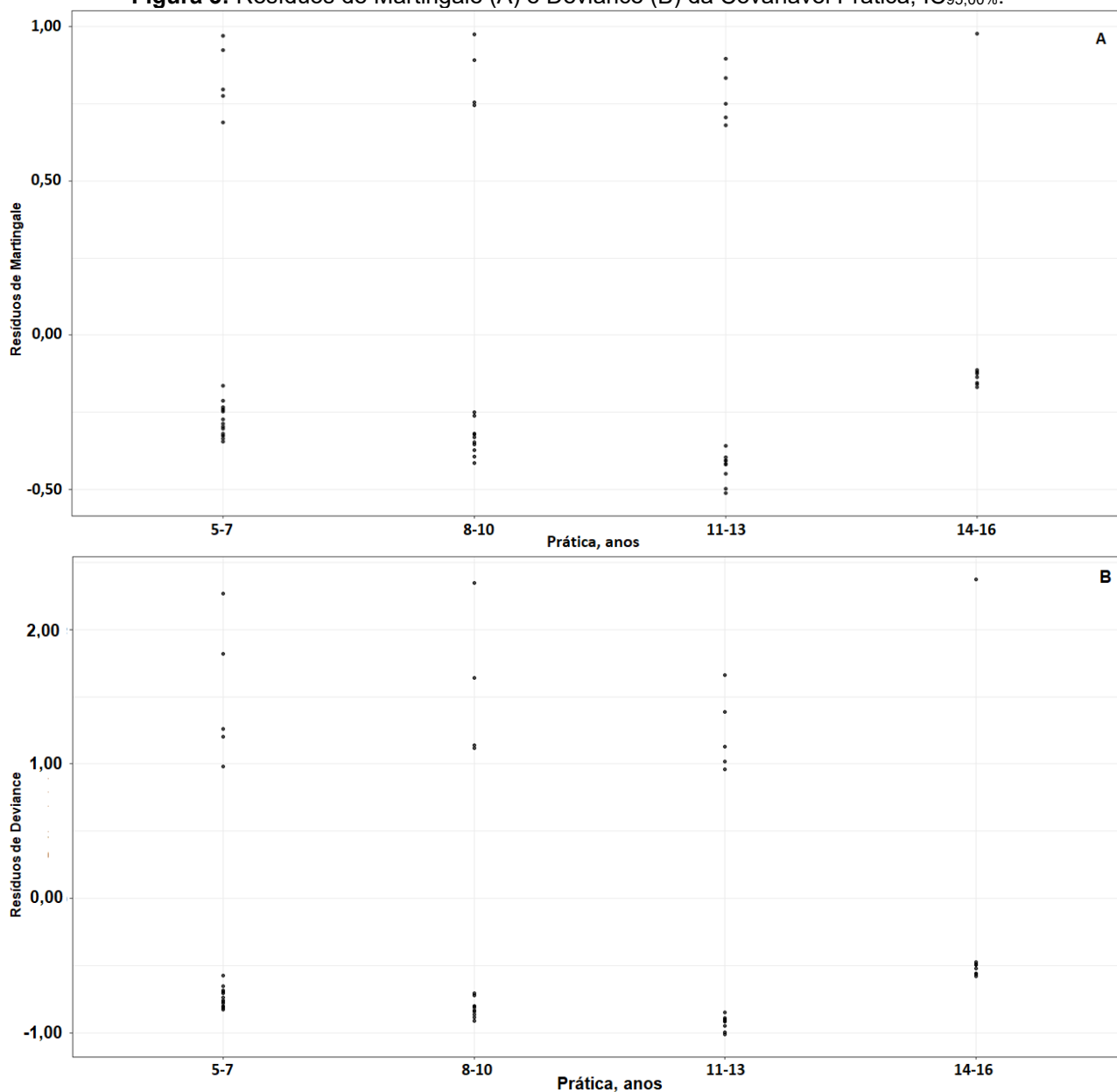


Fonte: Os Autores (2026).

No referente aos resíduos de Martingale de Prática (Figura 3A), os valores muito positivos ( $0,50 \leq \text{Resíduos} \leq 1,00$ ) reuniram 15 eventos lesivos ( $\approx 26,78\%$ ) e dois atletas ( $\approx 3,57\%$ ) conquistaram Resíduo  $\leq -0,50$ , similarmente ao ocorrido com a Idade, também poderiam ser 17 observações extremas. Nos resíduos de Deviance (Figura 3B), três

observações ( $\approx 5,35\%$ ) foram identificadas como impactantes. As análises realizadas corroboraram a validade do modelo, tanto quanto a carência de robustez.

**Figura 3:** Resíduos de Martingale (A) e Deviance (B) da Covariável Prática, IC<sub>95,00%</sub>.



**Fonte:** Os Autores (2026).

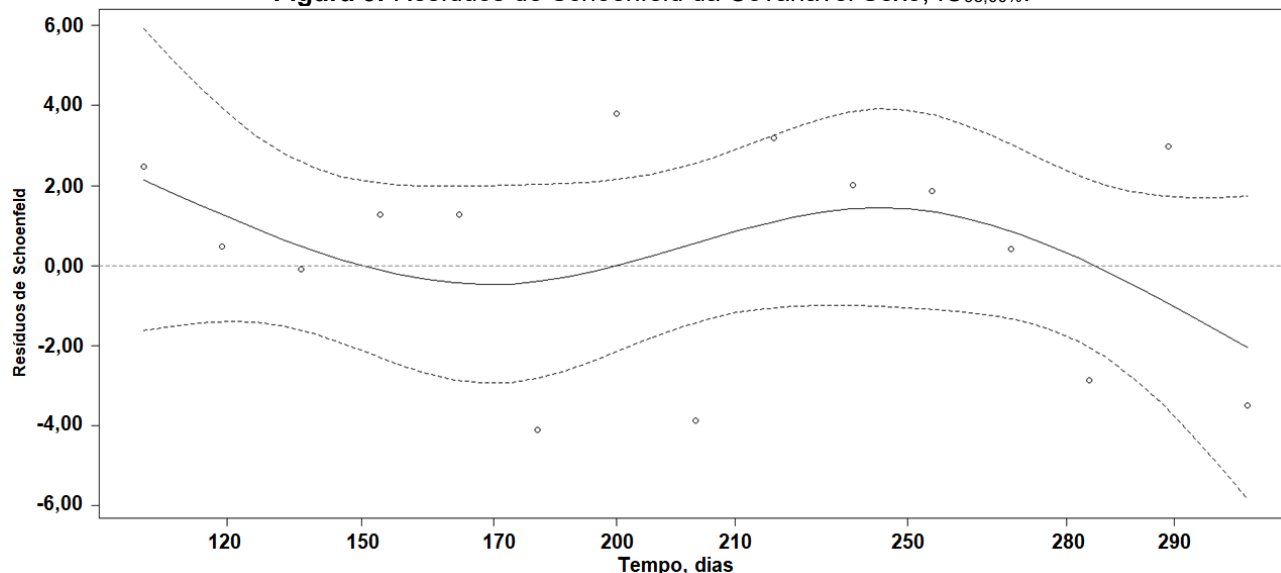
Finalmente, a análise dos resíduos de Schoenfeld (IV) permitiria diagnosticar ofensas ao pressuposto de riscos proporcionais, aspecto fundamental da regressão de Cox, pois consideraria que o efeito de determinada covariável seria constante no domínio do tempo. A violação implicaria em estimativas de RH enviesadas. Esses resíduos estimariam a diferença entre os valores esperado e observado da covariável no momento do evento,

dessarte, somente calculado nos instantes de ocorrência da lesão. As informações do gráfico dependeriam das linhas de 1) Ajuste, sobreposta aos pontos, auxiliando a visualização da tendência; e 2) Referência ( $Y = 0,00$ ), representando o HR constante no tempo. Logo, o distanciamento dela indicaria a violação do pressuposto. Pragmaticamente, se as linhas coincidissem, o modelo estaria perfeitamente ajustado. A Linha de Ajuste ascendente indicaria que o efeito da covariável se elevaria no tempo, logo, o risco aumentaria. Esse diminuiria, caso aquela linha se apresentasse descendente (Schumacher e Schulgen, 2008). O modelo foi constatado como válido, pois em todas as covariáveis, a maioria dos resíduos se manteve no interior do Intervalo de Confiança, haja vista que em Sexo (Figura 5), Idade (Figura 6) e Prática (Figura 7) extrapolaram esse limite, respectivamente, cinco ( $\approx 8,93\%$ ), quatro ( $\approx 7,14\%$ ) e três ( $\approx 5,36\%$ ) estimativas.

$$r_i = x_i - E[X|R(t_i)] \tag{IV}$$

onde:  $x_i$ : valor da covariável para o atleta lesionado;  $E[X|R(t_i)]$ : valor esperado da covariável, dado o conjunto de maratonistas sob risco naquele instante;  $t_i$ : tempo em que ocorreu a lesão.

Figura 5: Resíduos de Schoenfeld da Covariável Sexo, IC<sub>95,00%</sub>.

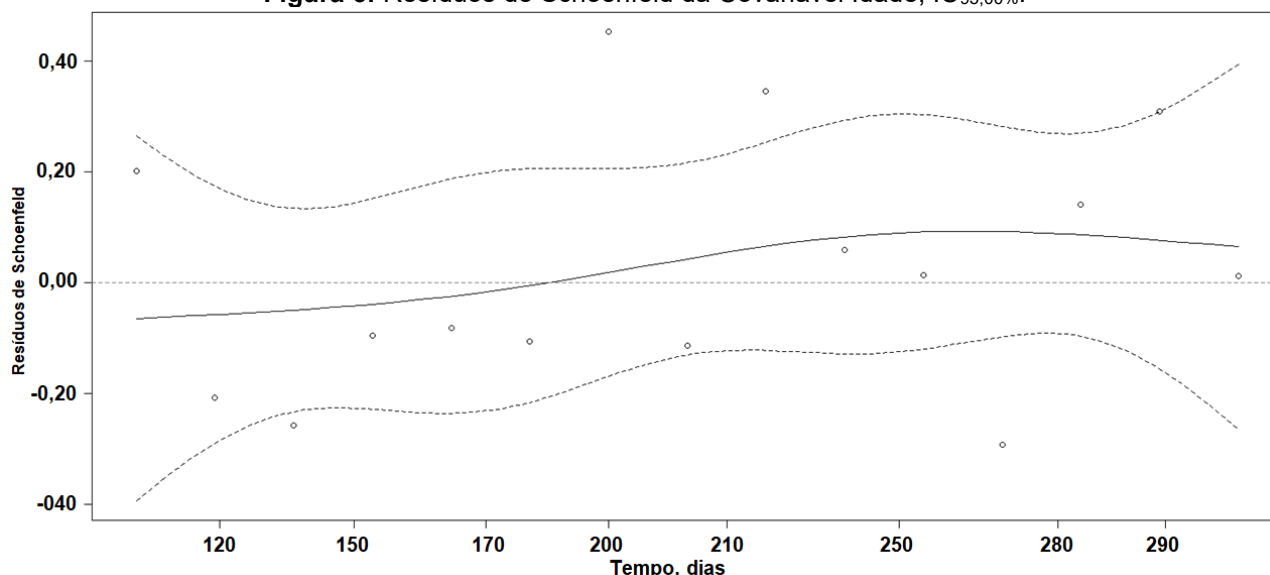


Fonte: Os Autores (2026).

A ausência de significância possibilitou conjecturar que a relação entre as covariáveis contínuas e o risco de lesão não fosse linear, então, desenvolvidas foram funções matemáticas flexíveis (*splines*, curvas suaves e contínuas que se aproximariam ou

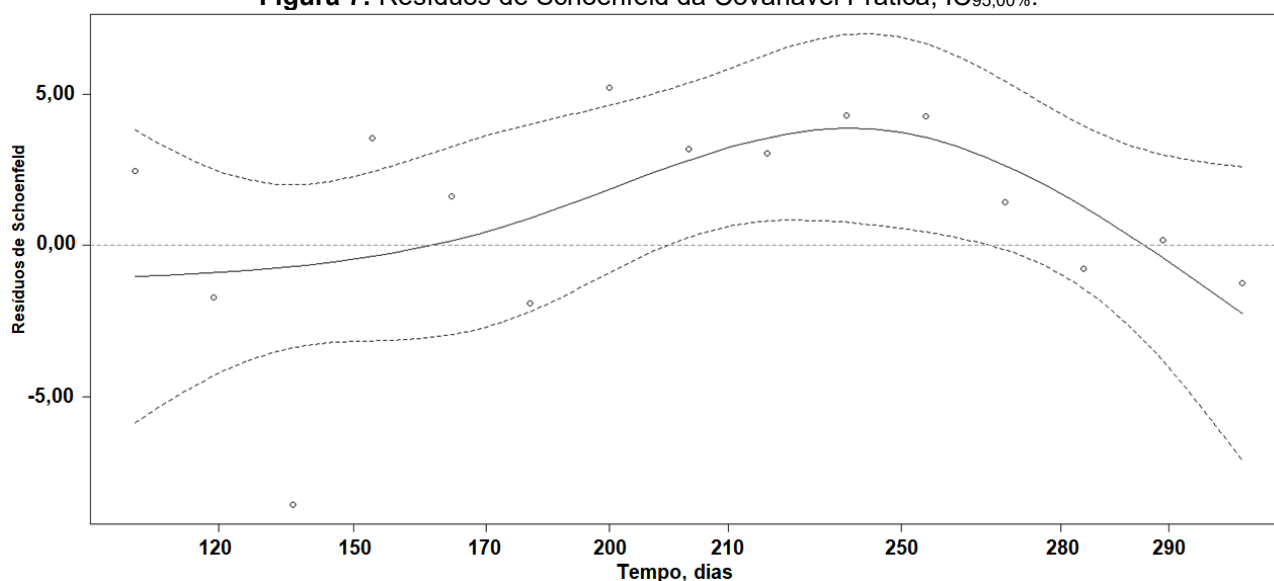
passariam pelos pontos observados) para modelar as relações postas. O objetivo foi capturar possíveis sutilezas ou mudanças em distintas faixas de valores das covariáveis, para tanto, pontos de dados foram conectados a polinômios cúbicos (Therneau e Grambsch, 2000). A expectativa era de que até certa Idade houvesse relação estável com o risco, o qual se tornaria crescente com o avanço cronológico. Enquanto indivíduos com experiência intermediária detivessem maior probabilidade de ocorrência de lesão, essa diminuiria com o acúmulo de anos de Prática. Em ambos os casos, desde que as estimativas fossem significativas (valor-p < 0,05).

**Figura 6:** Resíduos de Schoenfeld da Covariável Idade, IC<sub>95,00%</sub>.



Fonte: Os Autores (2026).

**Figura 7:** Resíduos de Schoenfeld da Covariável Prática, IC<sub>95,00%</sub>.



Fonte: Os Autores (2026).

Na percepção *lato sensu*, o modelo conquistou melhores estimativas (Tabela 4) e ajuste global (Tabela 5), contudo os efeitos careceram de significância estatística, por conseguinte, a funcionalidade linear não era adequada, porém a flexibilização não foi suficiente para obter associações relevantes, ratificaria a suposição de influência do baixo quantitativo de lesões. Talvez, as covariáveis utilizadas não fossem as determinantes de evento lesivo no grupo estudado, o que justificaria a baixa capacidade preditiva, C-índice = 0,64.

**Tabela 4:** Estimativas de Coeficiente ( $\beta_i$ ), HR( $\exp(\beta_i)$ ) e Erro Padrão ( $\varepsilon$ ) com Flexibilização Polinomial Cúbica (ns),  $\alpha = 0,05$ .

Variável	$\beta_i$	$\varepsilon(\beta_i)$	HR	IC <sub>95,00%</sub>	z	Pr(> z ) - valor-p
Sexo Masculino	0,41	0,65	1,51	[0,42; 5,43]	0,63	0,53
ns(Idade)1	-0,26	11,81	0,77	[0,08; 7,79]	-0,22	0,82
ns(Idade)2	2,26	21,33	9,56	[0,15; 624,80]	1,06	0,29
ns(Idade)3	0,16	0,84	1,17	[0,22; 6,10]	0,85	0,85
ns(Prática)1	1,82	12,47	6,16	[0,53; 70,89]	0,15	0,15
ns(Prática)2	-0,17	20,12	0,84	[0,02; 43,37]	0,93	0,93
ns(Prática)3	-1,24	13,41	0,29	[0,02; 4,01]	0,36	0,36

Fonte: Os Autores (2026).

**Tabela 5:** Resultados Inferenciais com Graus de Liberdade = 5,00 e Erro Padrão\* com Flexibilização Polinomial Cúbica,  $\alpha = 0,05$ .

Teste	Estimativa	valor-p
Concordância	0,64	0,07*
<i>Likelihood ratio</i>	5,06	0,70
Wald	4,27	0,70
Score (logrank)	4,56	0,70

Fonte: Os Autores (2026).

Possivelmente, reveladores seriam os aspectos mecânicos dos atletas como comprimento de membros inferiores, massa corporal e IMC, dadas as influências musculares e articulares (Silva e Moreira, 2011; Rossi e Tirapegui, 2011). A ocorrência de lesão seria influenciada pelo tipo de superfície (Abiko *et al.*, 2017; Vedovelli *et al.*, 2017) e prescrição de treinamento (Nielsen *et al.*, 2012; Melo *et al.*, 2018; Sousa, 2023), aspectos não considerados pela ausência de dados. Essas variáveis permitiram a segmentação dos atletas, o que impactaria na variabilidade e, conseqüentemente nas estimativas conquistadas (López Fidalgo, 2000; Therneau e Grambsch, 2000; Schumacher e Schulgen, 2008; Kleinbaum e Klein, 2011; Schneckenberg Filho, 2020; Carvalho *et al.*, 2021; Collett, 2023; Ma, Webb e Hudson, 2024).

## Considerações Finais

Objetivando modelar o risco de lesão em 56 maratonistas, a regressão de Cox foi desenvolvida. O modelo se mostrou válido, porém carente de robustez, logo caracterizado como tendo pouca capacidade informativa sobre a probabilidade de ocorrência de eventos lesivos. Então, possível foi concluir que o objetivo foi alcançado, demonstrando que Idade, Sexo e Prática na modalidade não seriam suficientes para entender o fenômeno da lesão.

Investigações futuras devem empregar a regressão de Cox com efeitos aleatórios, o qual incorpora a heterogeneidade. Investigar a distribuição de tempo por modelos paramétricos como log-logístico ou Weibull pode conquistar resultados efetivamente explicativos. Empregar como covariáveis condições atmosféricas e nutricionais, características psíquicas, fisiológicas e biomecânicas e o histórico de lesões refinariam o modelo, tendendo a torná-lo informativo. Esse ganho poderia ser potencializado com a segmentação da prova por variáveis como a distância na qual a lesão ocorreu, cadência por trecho e os eventos de variação na aceleração.

## Referências

- ABIKO, RH *et al.* Prevalência de lesões musculoesqueléticas e fatores associados em corredores de rua. **Ciência & Saúde**, v. 10, n. 2, p. 109-113, 2017.
- ALVES, KLF. **Análise de sobrevivência de bancos privados no Brasil**. Dissertação (Mestrado em Engenharia de Produção) – Departamento de Engenharia de Produção. Escola de Engenharia. Universidade de São Paulo. São Carlos (SP), 2009.
- ARIAS, DA *et al.* Cáncer de cabeza y cuello en Uruguay. Análisis de sobrevida en dos centros de referencia. **Revista Médica del Uruguay**, v. 34, n. 1, p. 21-28, 2018.
- BALULA, VL. **Treinamento muscular respiratório em atletas: uma revisão sistemática**. Trabalho de Conclusão de Curso (Bacharelado em Fisioterapia) – Centro Universitário Lusíada. Santos (SP), 2023.
- BARRETO, AC *et al.* Heart rate variability and blood pressure during and after three CrossFit® sessions. **Retos: Nuevas Tendencias en Educación Física, Deporte y Recreación**, n. 47, p. 311-316, 2023.
- BIUST, I *et al.* Predictors of running-related injuries in novice runners enrolled in a systematic training program: a prospective cohort study. **The American Journal of Sports Medicine**, v. 38, n. 2, p. 273-280, 2010.
- BOTELHO, F; SILVA, C; CRUZ, F. Epidemiologia explicada – análise de sobrevivência. **Acta Urológica**, v. 26, n. 4, p. 33-38, 2009.
- BRAGA, F; STEIN, R. Além dos Limites: O que 366 Maratonas Consecutivas Revelam sobre o Coração Humano. **Arquivos Brasileiros de Cardiologia**, v. 122, n. 5, e20250231, 2025.
- CACHAPUZ, DR. **Todo mundo se quebra de vez em quando: um estudo sobre os processos de subjetivação em idosas hospitalizadas por fratura de fêmur**. Dissertação (Mestrado em Psicologia

Social e Institucional) – Programa de Pós-graduação em Psicologia Social e Institucional. Instituto de Psicologia. Universidade Federal do Rio Grande do Sul. Porto Alegre (RJ), 2010.

CAMPOS, JDS. **Fatores explicativos para a evasão no ensino superior através da análise de sobrevivência**: o caso da UFPE. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-graduação em Engenharia de Produção. Universidade Federal de Pernambuco. Recife (PE), 2016.

CARVALHO JUNIOR, S *et al.* Análise de sobrevivência e regime de lesões em maratonistas. **Revista Presença**, v. 11, n. 26, p. 169-183, 2025.

CARVALHO, LS; REBUSTINI, F. Corrida e envelhecimento ativo: um estudo misto sobre os motivos de prática, estilo de vida e desempenho. **Concilium**, v. 23, n. 18, p. 701-725, 2023.

CARVALHO, MS *et al.* **Análise de sobrevivência**: teoria e aplicações em saúde. Rio de Janeiro: Fiocruz, 2021.

COLLETT, D. **Modelling survival data in medical research**. London (UK): Chapman & Hall, 2023.

COLOSINO, EA; GIOLO, SR. **Análise de sobrevivência aplicada**. São Paulo: Blucher, 2006.

COX, DR. Regression models and life-tables. **Journal of the Royal Statistical Society. Series B (Methodological)**, v. 34, n.2, p. 187–220, 1972.

ERRANDONEA, G. Método de Kaplan-Meier. Análisis de supervivencia por modelado parcial: Una alternativa metodológica para el estudio de las trayectorias estudiantiles en educación superior. **Páginas de Educación**, v. 17, n. 2, e4115, 2024.

FERRAZ, RO; MOREIRA-FILHO, DC. Análise de sobrevivência de mulheres com câncer de mama: modelos de riscos competitivos. **Ciência & Saúde Coletiva**, v. 22, n. 11, p. 3743-3753, 2017.

FERREIRA, AS *et al.* Análise de sobrevivência de pacientes com COVID-19. **Hematology, Transfusion and Cell Therapy**, v. 46, n. S4, p. S1236, 2024.

FREITAS, NCC *et al.* Dezesseis anos de transplante cardíaco em coorte aberta no Brasil: análise de sobrevivência de pacientes em uso de imunossupressores. **Arquivos Brasileiros de Cardiologia**, v. 116, n. 4, p. 744-753, 2021.

GALVÃO, MHR; RONCALLI, AG. Fatores associados a maior risco de ocorrência de óbito por COVID-19: análise de sobrevivência com base em casos confirmados. **Revista Brasileira de Epidemiologia**, v. 23, e200106, 2020.

GONÇALVES, ME. **Análise de sobrevivência e modelos hierárquicos logísticos longitudinais**: uma aplicação à análise da trajetória escolar (4<sup>a</sup> a 8<sup>a</sup> série – ensino fundamental). Tese (Doutorado em Demografia) – Centro de Desenvolvimento e Planejamento Regional. Faculdade de Ciências Econômicas. Universidade Federal de Minas Gerais. Belo Horizonte (MG), 2008.

GROOT, PM *et al.* The epidemiology of lung cancer. **Translational Lung Cancer Research**, v. 7, n. 3, p. 220–233, 2018.

GUIMARÃES, ER. **Análise de sobrevivência e aprendizado de máquina na previsão do tempo de espera para transplante renal**: uma abordagem otimizada. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Escola Politécnica e de Artes. Pontifícia Universidade Católica de Goiás. Goiânia (GO), 2024.

JANDRE, GRS *et al.* Modelagem estocástica da variabilidade da frequência cardíaca. **Revista Presença**, v. 10, n. 22, p. 237-252, 2024.

KLEINBAUM, DG; KLEIN, M. **Survival analysis**: a self-learning text. New York (USA): Springer, 2011.

LARA, B; SALINERO, JJ; DEL COSO, J. The relationship between age and running time in elite marathoners is U-shaped. **AGE**, v. 36, p. 1003–1008, 2014.

LÉDO, APO *et al.* Análise de sobrevida e fatores associados a mortalidade em portadores de insuficiência cardíaca na coorte ELSA-Brasil. **Arquivos Brasileiros de Cardiologia**, v. 122, n. 6, p. e20240705, 2025.

LIMA JUNIOR, P; SILVEIRA, FL; OSTERMANN, F. Análise de sobrevivência aplicada ao estudo do fluxo escolar nos cursos de graduação em física: um exemplo de uma universidade brasileira. **Revista Brasileira de Ensino de Física**, v. 34, n. 1, a. 1403, 2012.

LINS, PBS. **Difusão consórcios públicos intermunicipais no Brasil**: uma análise de sobrevivência. Dissertação (Mestrado em Ciências Políticas) – Centro de Filosofia e Ciências Humanas. Universidade Federal de Pernambuco. Recife (PE), 2023.

LÓPEZ, MJR; FIDALGO, JL. **Análisis de supervivencia**. Madrid (España): Arco Libros - La Muralla, S.L, 2000.

MA, J; WEBB, A; HUDSON, HM. **Likelihood methods in survival analysis**: with R examples. London (UK): Chapman & Hall/CRC Biostatistics Series, 2024.

MELO, OUM *et al.* Tempo de provas curtas e volume de treinamento como preditores do desempenho de maratona. **Revista Brasileira de Ciências do Esporte**, v. 40, n. 2, p. 117-122, 2018.

MIOT, HA. Análise de sobrevivência em estudos clínicos e experimentais. **Jornal Vascular Brasileiro**, v. 16, n. 4, p. 267-269, 2017.

NASU, VH *et al.* Da origem ao desfecho da solicitação: análise de sobrevivência aplicada ao contexto de acesso à informação pública. **Race**, v. 21, n. 3, p. 361-392, 2022.

NIELSEN, RO *et al.* Training errors and running related injuries: a systematic review. **International Journal of Sports Physical Therapy**, v. 7, n. 1, p. 58-75, 2012.

REIS, VM *et al.* Inclusão do equivalente do lactato sanguíneo O<sub>2</sub> na regressão de intensidade de exercício VO<sub>2</sub> aumenta o gasto energético de corrida e diminui sua precisão. **Revista Brasileira de Medicina do Esporte**, v. 19, n. 6, p. 427-430, 2013.

ROSSI, L; TIRAPGUI, J. Avaliação antropométrica segmentar comparativa de triatletas e maratonistas. **O Mundo da Saúde**, v. 35, n. 4, p. 422-426, 2011.

SACCARO, A; FRANÇA, MTA; JACINTO, PA. Fatores associados à evasão no ensino superior brasileiro: um estudo de análise de sobrevivência para os cursos das áreas de Ciência, Matemática e Computação e de Engenharia de Produção e Construção em instituições públicas e privadas. **Estudos Econômicos**, v. 49, n. 2, p. 337-373, 2019.

SCHNECKENBERG FILHO, DL. **Modelo de regressão de Cox**: uma aplicação a dados de pacientes com a doença de chagas. Monografia (Especialização em *Data Science* e *Big Data*) – Programa de Especialização em *Data Science* e *Big Data*. Departamento de Estatística. Universidade Federal do Paraná. Curitiba (PR), 2020.

SCHNEIDER, IJC *et al.* Neoplasias relacionadas ao tabaco: análise de sobrevivência e risco de óbito de dados populacionais de Florianópolis, SC. **Revista de Saúde Pública**, v. 56, n. 16, 2022. <https://doi.org/10.11606/s1518-8787.2022056003651>

SCHUMACHER, M; SCHULGEN, G. **Methodik Klinischer Studien**. Berlin (Deutschland): Springer, 2008.

SHEHADEH, I *et al.* Abdomen aberto ou fechado pós-laparotomia para controle de sepse de foco intra-abdominal severa: uma análise de sobrevivência. **Revista do Colégio Brasileiro de Cirurgiões**, v. 51, e20243595, 2024.

SILVA, TL; MOREIRA, IA. Evaluation of anthropometric and nutritional profile of garbage collectors in training for half marathon. **Revista Brasileira de Nutrição Esportiva**, v. 5, n. 28, p. 342-348, 2011.

SOLOAGA, AO. **Atletas olímpicos brasileiros apresentam maior longevidade que a população geral?** Um estudo de coorte retrospectivo. Dissertação (Mestrado em Ciências do Movimento – programa de Pós-graduação em Ciências do Movimento. Instituto Integrado de Saúde. Universidade Federal de Mato Grosso do Sul. Campo Grande (MS), 2021.

SOUSA, JF. **Efeitos do treinamento intervalado de alta intensidade na biomecânica da corrida.** Dissertação (Mestrado em Engenharia Biomédica) – Programa de Pós-graduação em Engenharia Biomédica. Faculdade de Engenharia Elétrica. Universidade Federal de Uberlândia. Uberlândia (MG), 2023.

SOUZA, RG; GOLGHER, AB; SILVA, BFA. Determinantes de reincidência prisional em Santa Catarina utilizando a análise de sobrevivência. **Nova Economia**, v. 34, n. 3, e8027, 2024.

THERNEAU, TM; GRAMBSCH, PM. **Modeling survival data: extending the Cox model.** New York (USA): Springer, 2000.

TOMAZELA, SMO. **Avaliação de desempenho de modelos de credit score ajustados por análise de sobrevivência.** Dissertação (Mestrado em Ciências Estatística) – Instituto de Matemática e Estatística. Universidade de São Paulo. São Paulo, 2007.

VEDOVELLI, KS *et al.* Análise comparativa da pisada durante a marcha entre atletas de diferentes modalidades esportivas. **Revista Brasileira de Fisiologia do Exercício**, v. 16, n. 1, p. 2-9, 2017.

VIEIRA, TMBG. **Estudo epidemiológico sobre transplante renal de dador de coração parado** – fatores de risco e análise de sobrevivência. Dissertação (Mestrado em Estatística e Investigação Operacional) – Departamento de Estatística e Investigação Operacional. Faculdade de Ciências. Universidade de Lisboa. Lisboa (Portugal), 2024.