

Modelagem matemática para o crescimento de tumores e sua associação aos níveis séricos de marcadores tumorais: uma análise do antígeno prostático específico em adenocarcinomas de próstata

Natália F. Reis,¹ Laércio L. Vendite²

DMA, IMECC – Unicamp, 13.083-859, Campinas/SP.

Resumo. O câncer é uma das principais causas de morte no mundo. Formado por um conjunto de mais de 100 doenças, sua complexa estrutura é palco para muitos estudos e pesquisas na atualidade. Devido a sua forte incidência, o câncer de próstata, juntamente ao seu Antígeno Prostático Específico, foi tratado de modo particular neste trabalho. Dentro da modelagem matemática de crescimento tumoral há um vasto e rico acervo literário. Pela lei universal de crescimento, o enfoque foi o ajuste de dados em pacientes com câncer de próstata, cuja evolução tumoral foi descrita pela lei e, em seguida ajustada ao método *leave-one-out*. Além do mais, comparou-se o ajuste de curvas de modelos matemáticos clássicos utilizando-se métodos numéricos. Em paralelo e de modo específico, analisou-se um conjunto de dados de pacientes diagnosticados com câncer de próstata, entre os quais associou-se o grau de correlação entre as principais variáveis, assim como suas relações em modelo de agrupamento não supervisionado.

Palavras-chave: *Câncer, câncer de próstata, antígeno prostático específico, métodos numéricos.*

¹n262894@dac.unicamp.br

²lvendite@unicamp.br

1. Introdução

A modelagem matemática, na maioria das vezes, fornece uma visão profunda, aproximada e real de alguns processos biológicos e auxilia na predição de crescimentos populacionais ao longo do tempo. Modelos matemáticos podem ser usados para representar o crescimento de tumores. Neste trabalho foi feita uma revisão bibliográfica de alguns modelos matemáticos clássicos, entre eles, os modelos de Malthus, Verhulst, Gompertz e Berterlanffy. Observamos como a mudança de parâmetros pode influenciar no processo de proliferação de células malignas, até que a massa tumoral atinja sua capacidade suporte.

Estudamos desde o processo de formação do núcleo necrótico do tumor, crescimento de massas tumorais, estadiamento clínico da doença, processos que desencadeiam a angiogênese, até à associação aos marcadores tumorais, que são macromoléculas presentes no tumor, denominadas proteínas, enzimas, antígenos, moléculas do sistema imune, receptores de membrana, hormônios, cromossomos, onco-genes e genes supressores no sangue ou em outros líquidos biológicos, cujo aparecimento e/ou alterações em suas concentrações estão relacionados com a gênese e o crescimento de células neoplásicas (P. C e Naoum, 2016).

Os marcadores podem ser úteis no manejo clínico de pacientes com câncer e auxiliam no diagnóstico, estadiamento, avaliação de resposta terapêutica, detecção de recidivas, prognóstico e desenvolvimento de novas modalidades de tratamento (Matos et al., 2005; Alonzo, 2005; Pacheco e Carvalho, 2002).

O foco deste trabalho é descrever o crescimento de tumores variando-se parâmetros nos modelos matemáticos e analisar de maneira específica o desenvolvimento do câncer de próstata através de dados pré-cirúrgicos de pacientes.

A metodologia desenvolvida busca revisar modelos e utilizar parâmetros da literatura sobre crescimento tumoral, simulando analítica e numericamente casos específicos. Também há um estudo e mineração de um conjunto de dados pré-cirúrgicos de pacientes diagnosticados com câncer de próstata, cedidos pelo Hospital das Clínicas da UNICAMP.

O objetivo é auxiliar a medicina nas análises clínicas, melhoramento de processos diagnósticos e prognósticos de cânceres, assim como, no reconhecimento de padrões e características do desenvolvimento de tumores.

O cenário apresentado é dividido da seguinte forma: simulações dos modelos de Gompertz e Verhulst utilizando-se parâmetros de Domingues (2010),

aplicação do método numérico de Euler para melhora na aproximação das curvas, ajuste de dados retirados de West (2018), aplicação da validação cruzada *leave-one-out* para aperfeiçoar na predição do modelo e análise e relação do Antígeno Prostático Específico (PSA) com as variáveis Estadiamento clínico, Escore de Gleason e idade, cujas informações foram retiradas da base de dados já mencionada anteriormente.

2. Objetivos

O objetivo do trabalho é descrever o crescimento de tumores, desde o processo de formação de células malignas até sua vascularização, tratando de maneira específica das características de desenvolvimento do câncer de próstata e associando-o a seu Antígeno prostático Específico (PSA). O processo busca relacionar um olhar interior e exterior do assunto tratado, obtendo caráter analítico diversificado sobre o mesmo tema.

3. Fundamentos matemáticos

Este estudo está fundamentado nos três tópicos a seguir: revisão de modelos matemáticos clássicos, análises e comparações entre simulações que descrevem o comportamento de tumores e tratamento de dados de pacientes diagnosticados com câncer de próstata.

3.1. Modelagem matemática

Segundo Bassanezi (1999), a modelagem matemática, focada no desenvolvimento de modelos, na atualidade, é arte de si mesma. Muito do que já foi produzido na matemática tem sido redirecionado a modelos e teorias, justificando-os através de aplicações.

A dinâmica das populações está ligada às variações numéricas no tempo de diversos parâmetros populacionais entre os quais, o tamanho da população, a taxa de sobrevivência, a taxa de mortalidade e a taxa de reprodução se destacam pela sua importância na descrição característica do meio.

Para descrever modelos de crescimento tumoral, utilizamos o modelo de Gompertz:

$$\frac{dN}{dt} = rN(t)\ln\left(\frac{k}{N(t)}\right) \quad (3.1)$$

em que $N(t)$ é a população tumoral, r é a taxa de crescimento intrínseco das células ($r > 0$) e k é a capacidade máxima que o tumor pode atingir com nutrientes disponíveis.

Sua solução é definida como:

$$N(t) = Ke^{-e^{rt} \ln\left(\frac{n_0}{k}\right)} \quad (3.2)$$

Com o desenvolvimento da angiogênese, $N(t)$ aumenta cada vez mais, de modo a escrevermos:

$$\lim_{t \rightarrow \infty} N(t) = K$$

Com isso é fácil verificar de maneira simples que o ponto de máximo global é $N = \frac{K}{e}$ e seu valor máximo é alcançado quando: $V_{max} = r \frac{K}{e}$. Por K estar ligado intimamente à quantidade de células tumorais $N(t)$, no instante t , consideremos que um tumor possua um limite de células que não pode ultrapassar o valor de 10^{13} células (Spencer et al., 2004).

Já no modelo logístico temos,

$$\frac{dN}{dt} = r_0 (\lambda - \alpha N) N = \lambda r_0 \left(1 - \frac{N}{K}\right) N \quad (3.3)$$

cuja solução é

$$N(t) = \frac{KN_0}{N_0 + (K - N_0)e^{\alpha t}} \quad (3.4)$$

Para $N < K$, a população cresce, pois $r_0 \left(1 - \frac{N}{K}\right) N > 0$. Para $N > K$, a população decresce, pois, $r_0 \left(1 - \frac{N}{K}\right) N < 0$. O parâmetro $K = \frac{\lambda}{\alpha}$ é a **capacidade suporte** ou **valor da saturação da população** ou ainda $\lambda = \alpha K$.

Para a simulação dos dados retirados de (Guiot et al., 2003; Yorke et al., 1993), foi realizada a técnica de validação cruzada denominada *leave-one-out* para avaliar a generalização do modelo apresentado.

Reconhecemos o crescimento de organismos através de características presentes na natureza. Pensar em escalas para analisar o crescimento de indivíduos não é algo tão familiar, no entanto, são fundamentais para descrever leis que regem o crescimento universal.

De acordo com West (2018), é possível redimensionar o tamanho de tumores utilizando essa lei. Usando a razão $r = \left(\frac{m}{M}\right)^{0,25}$, temos que m é a massa assintótica do táxon e r , a proporção de energia total, cujo objetivo é a manutenção celular.

Utilizando-se o tempo adimensional reescalado, temos:

$$\tau = \frac{1}{4}\alpha M^{-0,25}t - \ln\left(1 - \left(\frac{m_0}{M}\right)^{0,25}\right) = \alpha r_0 t - \ln(1 - r_0) \quad (3.5)$$

em que m_0 é a massa de nascimento tumoral e $r_0 = \left(\frac{m_0}{M}\right)^{0,25}$.

Além disso, $\alpha = 0,25am_0^{-0,25}$ e a é o parâmetro constante aproximado dentro do táxon.

A lei universal de crescimento é representada pela equação 3.6:

$$r = 1 - e^{-\tau} \quad (3.6)$$

As variáveis m_0 e M são as massas inicial e final do tumor e a , o parâmetro que associa as características tumorais à metástase e à invasão nos tecidos vizinhos.

3.2. Método numérico

3.2.1 Método de Euler

Também conhecido como método da reta tangente, o método de Euler é um método numérico simples para a solução de problemas de valor inicial (PVI). Segundo Atkinson et al. (2009), o PVI é dado da seguinte forma:

$$\begin{cases} y'(t) = (f, t(y)) \\ y(t_0) = y_0 \end{cases} \quad (3.7)$$

com $t_n = t_0 + n.h$ e $n = 0, 1, 2, \dots, m$, sendo h o espaçamento da malha e $m = \frac{b-a}{h}$, as subdivisões no intervalo $[t_0, t_m]$.

Seja a aproximação da derivada por uma equação de diferenças:

$$y'(t) \approx \frac{y(t+h) - y(t)}{h}$$

Aproximando o sistema 3.7, temos:

$$\frac{y(t+h) - y(t)}{h} = (f, t(y)) \quad (3.8)$$

Assim, fazendo $t = t_n$ e $y = y_n$, reescrevemos 3.8 da seguinte forma:

$$y_{n+1} = y_n + h.f(t_n, y_n)$$

Com isso é possível calcular a solução numérica y_n a partir de y_0 .

3.3. Ajuste de curvas

Considerando uma função conhecida $y = f(x)$, para um número finito de pontos discretos, $(x_0, y_0), (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, obtidos de forma experimental e correspondentes a um determinado intervalo, deseja-se obter valores $f(x_k)$, com $x_k \neq x_i$. Assim utiliza-se uma interpolação para encontrar $f(x_k)$ de maneira aproximada. Com isso, uma nova função θ , que interpola f , é criada de modo que:

$$(01) \quad (\forall x_i, x_0 \leq x_i \leq x_n)[\theta(x_i) = f(x_i)]$$

$$(02) \quad (\forall x \in [x_0, x_n])[\theta(x_i) \cong f(x_i)]$$

A escolha da função que mais se encaixa no conjunto de dados é de extrema importância na análise do comportamento de curvas. Devido à sua grande praticidade, os polinômios são utilizados para resolver problemas de interpolação, pois são fáceis na derivada, integração e cálculo do seu valor (Cláudio e Marins, 1989).

3.4. Análise de dados

A base de dados utilizada neste trabalho provém do Hospital das Clínicas da UNICAMP e conta com 530 instâncias de pacientes. O banco de dados apresenta informações pré-cirúrgicas, das quais podemos citar: Nível de PSA, Estadiamento Clínico, Idade e Escore de Gleason. O conjunto de dados cedidos a esta pesquisa possui 83 atributos relacionados à pacientes diagnosticados com câncer de próstata. Nessa análise foi utilizado o teste Qui-Quadrado e o agrupamento em *clusters*.

4. Metodologia e resultados

Algumas características definem um comportamento padrão do tumor, como sua capacidade suporte k , a quantidade inicial de células tumorais $N(0)$ e sua taxa de crescimento r . Domingues (2010) considera os seguintes parâmetros: $r = 0,0060$, $k = 10^{13}$ e $N(0) = 10^9$, em seu modelo matemático.

Com isso, foram realizadas simulações numéricas com estes parâmetros nos *softwares* Matlab e GNU Octave, onde comparamos as soluções analíticas dos modelos 3.1 e 3.2, em que a taxa de crescimento r foi modificada em ambas as figuras e o tempo t está definido em ano.

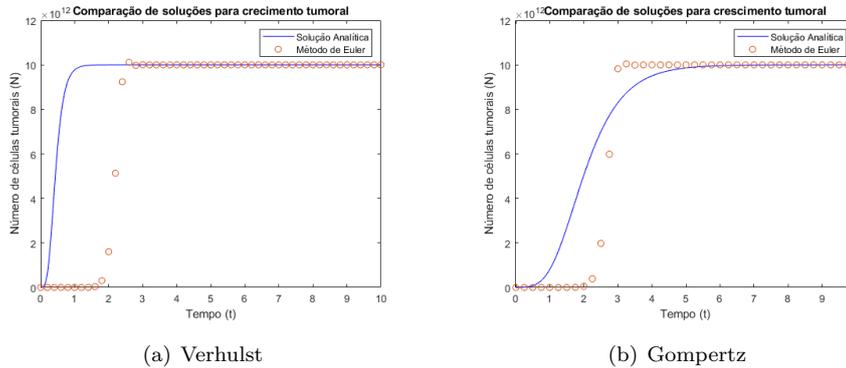


Figura 1: Modificação de parâmetros para comparação de curvas. Fonte: Comparação entre os métodos de Euler e as soluções exatas do modelo de Gompertz com parâmetros: $K = 10^{13}$, $r = 6$, $n_0 = n(0) = 10^9$, referentes a figura 1(a) e $K = 10^{13}$, $r = 1,3$, $n_0 = n(0) = 10^9$, de acordo com o gráfico em 1(b).

Os gráficos da figura 1 apresentados mostram uma aproximação pelo método de Euler, até $t = 10$. A diminuição do valor da taxa de crescimento, como mostra a segunda figura, faz com que a curva seja definida mais suavemente.

No entanto, para $t \geq 3$, no gráfico à esquerda, a aproximação numérica é eficiente apenas para descrever a capacidade suporte do tumor, caracterizando boas previsões a longo prazo, enquanto no gráfico à direita, a aproximação se torna mais eficiente no processo de crescimento acelerado.

Após isto, foi feito um ajuste em um conjunto de dados que analisa pacientes que desenvolveram câncer de próstata, retirados de (Guiot et al., 2003; Yorke et al., 1993). Os dados foram ajustados a uma curva polinomial de 3º grau, utilizando-se o Método dos Mínimos Quadrados (MMQ), como mostra a figura 2. Em seguida, os resultados desse ajuste são mostrados na tabela 1.

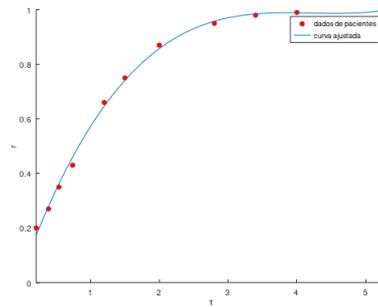


Figura 2: Ajuste polinomial de grau 3 aos dados de pacientes com câncer de próstata

Tabela 1: Valores referentes ao modelo de dados.

x	y	Erro²	SQReg	SQtot
0,21	0,20	0,2278	0,2541	0,2278
0,39	0,27	0,1659	0,1580	0,1659
0,54	0,35	0,1071	0,0999	0,1071
0,74	0,43	0,0612	0,0473	0,0612
1,2	0,66	0,0003	0,0009	0,0003
1,5	0,75	0,0053	0,0041	0,0053
2	0,87	0,0371	0,0324	0,0371
2,8	0,95	0,0744	0,0775	0,0744
3,4	0,98	0,0916	0,0944	0,0916
4	0,99	0,0978	0,0968	0,0978
5,3	1	0,1041	0,1035	0,1041

De fato, quando se analisa um modelo, o erro de estimativa aumenta com o número de variáveis, podendo não ser um bom preditor para grandes quantidades de dados. No nosso caso, analisamos uma quantidade de apenas 11 pontos para definirmos se o modelo é válido e suficientemente eficiente para descrever o comportamento de crescimento de tumores.

A princípio analisamos o coeficiente de determinação R^2 para verificarmos a medida de ajuste do modelo estudado. Obteve-se um bom desempenho de aproximadamente 0,9799, indicando que o modelo explica 97,99% da variância da variável dependente a partir das variáveis independentes.

Visualmente o modelo também mostrou um bom ajuste, no entanto, isso levantou a hipótese de um possível *overfitting*. Com isso, utilizamos a validação cruzada *Leave-one-out*.

Dividimos o conjunto de dados em um conjunto de treinamento, utilizando todos os valores numéricos, exceto um, referente ao conjunto de teste. Repetimos o processo 11 vezes e comparamos a raiz quadrada média dos erros com o RSME do modelo original.

O valor do RMSE referente ao total dos 11 dados originais foi igual a 0,01449, enquanto que a média dos RMSE dos demais modelos foi de 0,01589. Isso significa que o erro aumenta em apenas 10% do modelo referente à curva polinomial, indicando que não há grande *overfitting* em relação ao modelo inicial.

Após estudar o comportamento de crescimento de tumores, analisamos um conjunto de dados de 530 pacientes diagnosticados com câncer de próstata, cedidos pelos Hospital das Clínicas da UNICAMP. Primeiramente, utilizamos o teste Qui-Quadrado no ambiente Weka, para comparar as frequências entre os atributos.

Fizemos um filtro de variáveis para definir aquelas com um melhor nível de significância com a variável-alvo. Neste caso, definimos o PSA como a variável-alvo e a comparamos aos atributos citados na tabela 2.

Tabela 2: Análise de qui-quadrado.

Entrada	Qui-quadrado
Densidade	140628,366
GL-PR	3993,555
GL-BTR	3142,829
Estádio	2845,484

Após as análises, as variáveis que obedeceram um nível mínimo de significância com o PSA, foram: densidade, Escore de Gleason (GL-PR e GL-BTR).

Após retirar os dados faltantes, aplicamos novamente o teste qui-quadrado e obtivemos o resultado apresentado na tabela 3.

Tabela 3: Análise de qui-quadrado após ajuste dos dados.

Entrada	Qui-quadrado
Densidade	154.989
GL-PR	44.701
Estádio	31.953
Idade	23.842
GL-BTR	22.511
RM	16.235

Observamos que o PSA está ligado frequentemente à densidade, GL-PR, Estadiamento e Idade.

4.1. Nível de PSA e DPSA

Relacionando-se as variáveis densidade e nível de PSA, é possível estabelecer uma relação entre as duas. A densidade é expressa pela razão entre o nível de PSA, em ng/ml e o peso da próstata (g) ou (cm^3).

Tabela 4: Relação entre as densidades do PSA e seus níveis.

Densidade PSA	Total	PSA < 5ng/ml	$5 \leq \text{PSA} \leq 10$	PSA > 10 ng/ml
$\leq 0,15$ ng/ml/g	137	77	55	5
$> 0,15$ ng/ml/g	393	53	179	161

O PSA ≤ 10 ng/ml corresponde a 96,36% dos casos com densidades $\leq 0,15$ ng/ml/g e a 86,51% do PSA ≥ 5 , relacionados às densidades $> 0,15$ ng/ml/g. A média da densidade do PSA é igual a 0,35 ng/ml/g.

De fato, as densidades maiores que 0,15 ng/ml/g possuem uma média de PSA igual a 11,27 ng/ml na nossa análise. Considerando densidades menores ou iguais a 0,15 ng/mL/g, temos que a média do PSA é de aproximadamente 5,25 ng/ml.

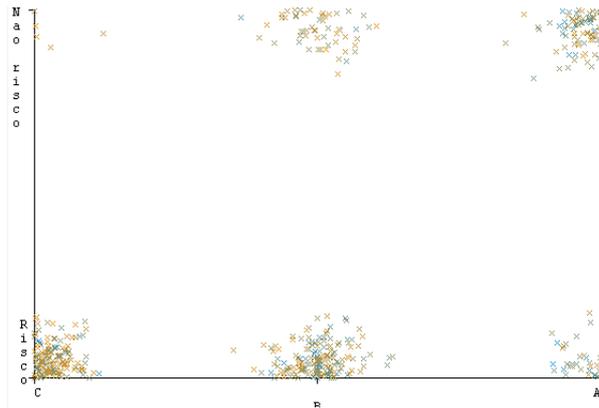
Associamos a densidade do DPSA $\leq 0,15$ ng/mL aos níveis de PSA estabelecidos:

Tabela 5: Relação entre a densidade de PSA e seus níveis séricos.

DPSA $\leq 0,15$	F	FR	Média PSA	DP
PSA ≤ 5	75	54,74%	3,36	1,33
PSA ≤ 6	94	68,61%	3,78	1,46
PSA ≤ 7	107	78,10%	4,11	1,63
PSA total	137	100%	5,17	2,68

Considerando a idade menor que 50 anos, estadiamento até T2b, Gleason até $3 + 4 = 7$) e DPSA menor ou igual a 0,15, temos que 66,7% delas se encaixam nesse padrão. Dos valores com densidades maiores que 0,15, 70% possuem pontuação de Gleason maior ou igual 7. Notamos também que 78,11% das densidades maiores que 0,15 possuem escore de Gleason maior que 7.

Dividimos o atributo densidade em: risco (valores maiores ou iguais a 0,15 ng/ml/g) e não risco (valores menores que 0,15 ng/ml/g) e reagrupamos o PSA em três intervalos: A (≤ 5), B ($5 < x \leq 10$) e C (> 10). A figura 3 mostra a disposição dos pontos e a formação dos *clusters*.

Figura 3: *Clusters* relacionando os níveis de PSA e densidade.

5. Conclusões

A modelagem matemática realizada com os dados apresentados, considerou uma taxa de crescimento bem baixa, o que pode ser relacionado ao crescimento lento do câncer de próstata. Nas simulações, o modelo de Gom-

pertz obteve uma melhor aproximação utilizando o método de Euler, que é um método mais simples de implementar.

Através dos dados experimentais utilizados, foi possível encontrar um bom ajuste, já que polinômios de graus não tão altos facilitam cálculos mais complexos. Pela validação cruzada *leave-one-out*, obtivemos informações adicionais sobre o ajuste do modelo. O RMSE referente ao total de pontos foi 0,01449 e a média dos demais modelos foi de 0,01589, indicando que não há *overfitting* significativo, o que é um bom ajuste, pois com a chegada de novos exemplos, a tendência é que o desempenho não seja tão satisfatório.

Na análise de dados, realizamos o teste qui-quadrado, cuja variável-alvo, Nível de PSA obteve melhor relação com as variáveis: densidade, pontuação de Gleason e estadiamento. Após retiramos os dados faltantes, a idade apareceu em seguida como uma correlação razoável.

Em alguns estudos, a densidade do PSA não é tão importante para medir o progresso do câncer de próstata e sua utilidade ainda é discutível. De fato, há falhas na ultrassonografia para determinar o volume da próstata e isso afeta diretamente na eficiência do exame. A DPSA é necessária para incrementar a acurácia do antígeno e também para solicitar ou a não a biópsia da próstata durante o rastreamento preventivo da neoplasia.

Por fim, analisando-se todas as informações referentes aos níveis de PSA e sua densidade, percebemos que 96,36% dos dados com PSA menor ou igual a 10 ng/ml possuem densidade menor que 0,15 ng/ml/g. Outra relação importante é que para densidades maiores que 0,15 ng/ml/g, 78,10% dos pacientes tem Escore de Gleason maior que 7.

Referências

- Alonzo, T. A. (2005). Standards for reporting prognostic tumor marker studies.
- Atkinson, K., Han, W., e Stewart, D. E. (2009). *Numerical solution of ordinary differential equations*. New Jersey: John Wiley & Sons.
- Bassanezi, R. C. (1999). Modelagem matemática uma disciplina emergente nos programas de formação de professores. *Biomatemática*, 9:9–22.
- Cláudio, D. M. e Marins, J. M. (1989). *Cálculo numérico computacional: teoria e prática*, volume 2. Atlas.

- Domingues, J. S. (2010). *Modelo matemático e computacional do surgimento da angiogênese em tumores e sua conexão com as células-tronco*. Tese de Doutorado, Dissertação de Mestrado, Centro Federal de Educação Tecnológica de Minas Gerais.
- Guiot, C., Degiorgis, P. G., Delsanto, P. P., Gabriele, P., e Deisboeck, T. S. (2003). Does tumor growth follow a “universal law”? *Journal of theoretical biology*, 225(2):147–151.
- Matos, L. L., Machado, L. N., Sugiyama, M. M., Bozzetti, R. M., e Pinhal, M. A. S. (2005). Tecnologia aplicada na detecção de marcadores tumorais. *Arquivos Médicos Do ABC*, 30(1).
- P. C. N. e Naoum, F. A. (2016). *Biologia médica do câncer humano*. Editora Mc Will, São Paulo.
- Pacheco, F. A. e Carvalho, M. E. P. M. G. C. (2002). Marcadores tumorais no câncer de pulmão: um caminho para a terapia biológica. *Jornal de Pneumologia*, 28:143–149.
- Spencer, S. L., Berryman, M. J., García, J. A., e Abbott, D. (2004). An ordinary differential equation model for the multistep transformation to cancer. *Journal of Theoretical Biology*, 231(4):515–524.
- West, G. (2018). *Scale: The universal laws of life, growth, and death in organisms, cities, and companies*. Penguin.
- Yorke, E. D., Fuks, Z., Norton, L., Whitmore, W., e Ling, C. C. (1993). Modeling the development of metastases from primary and locally recurrent tumors: comparison with a clinical data base for prostatic cancer. *Cancer research*, 53(13):2987–2993.

