

Estimativa de Incerteza na Idade Média de Aquisição de Infecção em Modelos Epidêmicos¹

M. AMAKU², Instituto de Física, USP, SP, Brasil

R.S. AZEVEDO, A.L.F. YU, LIM01 HC/FMUSP e Departamento de Patologia, Faculdade de Medicina, USP, SP, Brasil

R.M. CASTRO, Instituto de Física, USP, SP, Brasil

L.M. de ALMEIDA, LIM01 HC/FMUSP e Departamento de Patologia da Faculdade de Medicina, USP, Núcleo de Estudos de Saúde Coletiva, UFRJ, RJ, Brasil

F.A.B. COUTINHO, LIM01 HC/FMUSP e Departamento de Patologia da Faculdade de Medicina, USP, SP, Brasil.

Resumo. Através de simulações de Monte Carlo, estimamos a incerteza estatística da idade média de aquisição de infecção, um parâmetro epidemiológico importante no modelamento matemático de doenças infecciosas. Também estimamos o intervalo de confiança para a força de infecção por meio de propagação de erros. Estimativas de incerteza para a força de infecção e para a idade média são informações importantes na comparação e caracterização de situações epidemiológicas, como mostramos nos exemplos para varicela e hepatite A.

1. Introdução

O modelamento matemático de doenças infecciosas tem se mostrado uma ferramenta auxiliar no planejamento de estratégias de controle de doenças e na avaliação de seu impacto. Um dos parâmetros relevantes na área epidemiológica é a *idade média de aquisição de infecção*, relacionada ao conceito de *risco* de adquirir infecção, uma “quantidade” (se pudermos entender risco como tal) difícil de ser mensurada. Uma estimativa desse risco é feita através da *força de infecção*, definida como a taxa *per capita* em que indivíduos suscetíveis adquirem infecção [2], por unidade de tempo, estimada ou através de registros de casos notificados ou por meio de dados de soroprevalência obtidos em levantamentos sorológicos.

¹M. Amaku agradece o apoio financeiro da Fapesp (Processo 97/03679-0)

²amaku@if.usp.br

Pode-se aproximar a idade média como sendo o inverso da força de infecção média [2]. No entanto, essa é uma aproximação que pode apresentar baixa acurácia, dependendo do comportamento da força de infecção em função da idade.

A idade média de aquisição de infecção permite que se caracterize a incidência de uma infecção estudada na população. Uma maior incidência de casos em idades mais precoces ou mais avançadas pode estar relacionada a diversos fatores, como condições sanitárias, idade de entrada na escola, número de pessoas por domicílio, entre outras variáveis. Melhoras nas condições sanitárias e a introdução de um programa de vacinação podem levar, por exemplo, a mudanças na idade média de infecção, o que demonstra a importância deste parâmetro na avaliação dos efeitos de um programa de controle ambiental e/ou de uma estratégia vacinal.

Diferentes padrões de transmissão de infecção levam provavelmente a diferentes idades médias de aquisição de infecção, o que nos conduz ao problema estatístico de comparar tais idades. Em outras palavras, a principal motivação do presente trabalho é, dados dois ou mais valores de idades médias de infecção, estudar sob que condições podemos considerá-los (in)distinguíveis estatisticamente. Queremos, então, saber, com base nas informações contidas nos dados epidemiológicos, qual o intervalo de confiança associado a cada valor de idade média, o que nos permitiria compará-los. Para fazer tal comparação, é importante obter uma estimativa acurada da idade média de infecção e de sua incerteza, de modo a evitar uma inferência enganosa. Estamos tratando o termo *incerteza* no seu sentido estatístico [10], relacionado ao intervalo de confiança de uma dada grandeza.

Este trabalho visa a demonstrar a aplicabilidade do Método de Monte Carlo [8] na estimativa da incerteza estatística da idade média de infecção. Para ilustrar as potencialidades do método, incluímos dois exemplos de aplicação, um para a varicela e outro para a hepatite A (Subseção 2.3).

2. O Modelo

2.1. O Modelo Catalítico

Modelos catalíticos são assim chamados em analogia a modelos de estudo de reações cinéticas envolvendo enzimas em Química. A utilização de tais modelos para determinar a taxa em que suscetíveis adquirem infecção foi proposta inicialmente por Muench (citado por Grenfell e Anderson [7]), em 1959.

Definindo $S^+(a)$ como a proporção de indivíduos soropositivos com idade a , pode-se deduzir o modelo catalítico para a força de infecção $\lambda(a)$ (veja, por exemplo, [7])

$$\lambda(a) = \frac{1}{1 - S^+(a)} \frac{dS^+(a)}{da}. \quad (2.1)$$

A equação (2.1) nos permite estimar, uma vez conhecida a função de soroprevalência $S^+(a)$, a força de infecção em função da idade. Sendo $x(a)$ a proporção de indivíduos suscetíveis de idade a , $\lambda(a)x(a)$ é um termo que se relaciona à passagem de indivíduos suscetíveis para a condição de infectados e que, portanto, de-

termina qual será a idade média de infecção. Assim sendo, a idade média de infecção pode ser definida como [3]

$$\bar{a} = \frac{\int_0^L a\lambda(a)x(a)da}{\int_0^L \lambda(a)x(a)da}, \quad (2.2)$$

onde L é a idade máxima considerada nos cálculos, lembrando que $x(a)$ pode ser tomado como $1 - S^+(a)$.

2.2. Estimativa de Incerteza através do Método de Monte Carlo

A proporção de indivíduos soropositivos, $S^+(a)$, pode ser estimada através do Método de Gauss-Marquardt [4, 9, 5], utilizando-se uma modificação de uma função proposta por Farrington [6]

$$S^+(a, \mathbf{k}) = 1 - \exp\left\{\frac{k_1}{k_2^2}[(k_2 a + 1)e^{-k_2 a} - 1]\right\}, \quad (2.3)$$

onde $\mathbf{k} = [k_1 \ k_2]^T$ é o vetor de parâmetros do ajuste. O método de Gauss-Marquardt também fornece uma estimativa da matriz de covariância \mathbf{V} dos parâmetros

$$\mathbf{V} = \begin{pmatrix} \sigma_{k_1}^2 & cov(k_1, k_2) \\ cov(k_1, k_2) & \sigma_{k_2}^2 \end{pmatrix},$$

onde $\sigma_{k_i}^2$ ($i = 1, 2$) é a variância do parâmetro k_i ($i = 1, 2$) e $cov(k_1, k_2)$ a covariância entre os dois parâmetros. Os parâmetros ajustados serão indicados como \hat{k}_1 e \hat{k}_2 .

O modelo catalítico (equação (2.1)) nos dá a expressão correspondente para a força de infecção

$$\lambda(a, \mathbf{k}) = k_1 a e^{-k_2 a}.$$

Da teoria de propagação de erros [4], nós estimamos a variância da força de infecção, para uma dada idade a , como

$$\sigma_\lambda^2(a) = \left(\frac{\partial\lambda(a, \mathbf{k})}{\partial k_1}\right)^2 \sigma_{k_1}^2 + \left(\frac{\partial\lambda(a, \mathbf{k})}{\partial k_2}\right)^2 \sigma_{k_2}^2 + 2\frac{\partial\lambda(a, \mathbf{k})}{\partial k_1}\frac{\partial\lambda(a, \mathbf{k})}{\partial k_2}cov(k_1, k_2). \quad (2.4)$$

Por conseguinte, é possível, para uma idade a , estimar um intervalo de confiança padrão de 68% com aproximação normal, que é equivalente ao intervalo $(\lambda(a) - \sigma_\lambda(a); \lambda(a) + \sigma_\lambda(a))$.

O coeficiente de correlação entre k_1 e k_2 é definido como

$$\rho = \frac{cov(k_1, k_2)}{\sigma_{k_1} \sigma_{k_2}}, \quad (2.5)$$

e sempre obedece a $-1 \leq \rho \leq 1$. Se $\rho = 0$, k_1 e k_2 são não-correlacionados, se $\rho > 0$, há uma correlação positiva entre os parâmetros e, se $\rho < 0$, uma correlação negativa.

Tendo uma estimativa da matriz de covariância \mathbf{V} do ajuste, podemos gerar aleatoriamente, com base nas distribuições dos parâmetros, pares de valores (k_1, k_2) e calcular a idade média de infecção \bar{a} correspondente.

O algoritmo de Monte Carlo adotado baseia-se na geração de números aleatórios para k_1 e k_2 e subsequente substituição desses valores na equação (2.2) para a idade média de infecção. Foi suposto que tanto k_1 quanto k_2 apresentam distribuição normal com médias \hat{k}_1 e \hat{k}_2 e desvios padrões σ_{k_1} e σ_{k_2} , respectivamente. Levando em conta a covariância $cov(k_1, k_2)$, o algoritmo para a geração de pares de valores (k_1, k_2) é o seguinte

$$k_1 = \hat{k}_1 + \sigma_{k_1} \cdot g_1() \quad (2.6)$$

$$k_2 = \hat{k}_2 + \sigma_{k_2} \cdot [\rho \cdot g_1() + \sqrt{1 - \rho^2} \cdot g_2()], \quad (2.7)$$

onde $g_1()$ e $g_2()$ são dois diferentes geradores de números aleatórios gaussianos com média 0 e variância 1. As rotinas computacionais utilizadas para a geração de números aleatórios estão descritas em Press *et al.* [9].

Podemos verificar a consistência do algoritmo acima, testando algumas possibilidades para ρ : se k_1 e k_2 apresentam uma correlação positiva perfeita ($\rho = 1$), o mesmo gerador de aleatórios $g_1()$ é usado para ambos k_1 e k_2 e o algoritmo para k_2 é $k_2 = \hat{k}_2 + \sigma_{k_2} \cdot g_1()$; se k_1 e k_2 apresentam uma correlação negativa perfeita ($\rho = -1$), o mesmo gerador de aleatórios $g_1()$ é adotado para ambos os parâmetros, mas um sinal negativo, relacionado à correlação negativa, aparece no algoritmo para k_2 , sendo $k_2 = \hat{k}_2 - \sigma_{k_2} \cdot g_1()$; se ambos os parâmetros não são correlacionados ($\rho = 0$), diferentes geradores de números aleatórios são adotados para cada parâmetro, e os valores para k_1 e k_2 são gerados independentemente, com o algoritmo k_2 dado por $k_2 = \hat{k}_2 + \sigma_{k_2} \cdot g_2()$.

2.3. Aplicações para Varicela e Hepatite A

Para demonstrar as potencialidades do método, nós escolhemos duas doenças infecciosas, varicela e hepatite A, com modos de transmissão característicos, que podem ser descritos, de modo simplista, como segue.

A infecção pelo vírus da varicela leva a sintomas como febre e aparecimento de lesões de pele. É um vírus que se transmite por vias aéreas ou por contato direto com o fluido secretado das lesões. Simplificadamente, podemos dizer que a transmissão da varicela se dá por via direta com contato pessoa-a-pessoa. Por outro lado, a infecção por vírus de hepatite A é adquirida principalmente por via fecal-oral, seja por contato pessoa-a-pessoa, seja por ingestão de água ou comida contaminada [12]. De modo distinto à varicela, a hepatite A apresenta uma via de transmissão indireta fortemente relacionada às condições ambientais, além da via direta de transmissão.

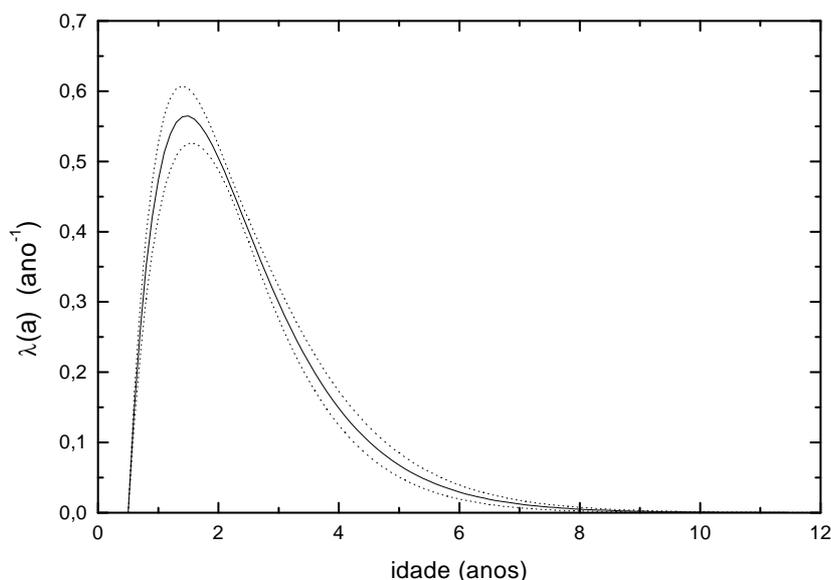


Figura 1: Força de infecção em função da idade para a varicela, referente ao levantamento sorológico feito em São Paulo, em 1992 [11]. As linhas pontilhadas indicam o intervalo de confiança de 68% para $\lambda(a)$.

Analisamos os dados de soroprevalência para a varicela — a partir de levantamentos realizados em 1992 — em duas comunidades [11]: crianças de São Paulo, uma cidade com cerca de 10 milhões de habitantes na época, 97% dos quais morando na área urbana, e de Caieiras, uma pequena cidade nos arredores de São Paulo, com cerca de 30000 habitantes em 1990, 90% na área urbana. Em Caieiras, crianças vivem tipicamente dentro de casa até os 3 anos de idade, quando começam a freqüentar os jardins de infância. Em São Paulo, até os 4 anos de idade, as crianças costumam ficar em creches e berçários e, após essa idade, começam a freqüentar os jardins de infância. Assim, nestas duas comunidades, as crianças apresentam um comportamento social diferente [11].

Ajustamos curvas de soroprevalência, de acordo com a expressão (2.3), e, com o auxílio do modelo catalítico (equação (2.1)), estimamos a força de infecção natural, como se observa nas Figuras 1 e 2, juntamente com a aproximação gaussiana para o intervalo de confiança de 68%.

Para Caieiras, obtivemos uma idade média de $\bar{a} = (4,07 \pm 0,47)$ anos e, para São Paulo, $\bar{a} = (1,940 \pm 0,083)$ anos. Para estimar as incertezas, foram gerados 10000 pares de valores (k_1, k_2) , de acordo com o algoritmo dado pelas expressões (2.6) e (2.7). Para cada par (k_1, k_2) , calculamos um valor \bar{a} . Os desvios padrões

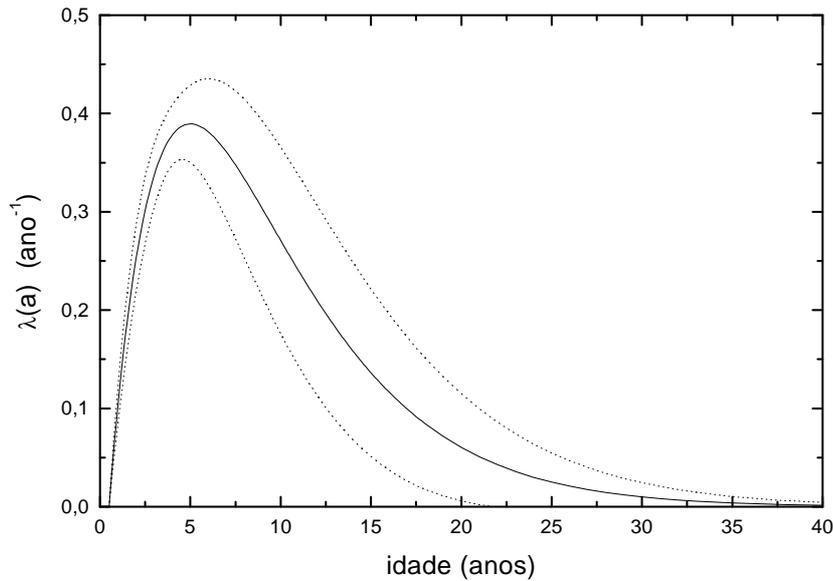


Figura 2: Força de infecção em função da idade para a varicela, referente ao levantamento sorológico feito em Caieiras, em 1992 [11]. As linhas pontilhadas indicam o intervalo de confiança de 68% para $\lambda(a)$.

das distribuições dos valores \bar{a} foram tomados como estimativas de incerteza. Como podemos notar, a incerteza relativa é maior para Caieiras quando comparada a São Paulo, o que está diretamente relacionado à maior dispersão para $\lambda(a)$, observada na Figura 2.

O algoritmo de Monte Carlo foi também aplicado aos dados obtidos de um levantamento de soroprevalência de hepatite A em Duque de Caxias, Rio de Janeiro, como parte do Programa de Descontaminação da Baía da Guanabara [1]. A amostra de dados foi dividida em duas categorias de acordo com a presença ou ausência de vala negra (esgoto a céu aberto) nas proximidades do domicílio e, para estas duas categorias, estimamos a idade média de infecção e sua incerteza. As curvas para $\lambda(a)$ relativas a esses dados estão apresentadas na Figura 3.

No caso em que não há vala negra nos arredores, a idade média é $\bar{a} = (11,90 \pm 0,59)$ anos, ao passo que, para a situação com vala negra, temos $\bar{a} = (8,61 \pm 0,36)$ anos. Podemos ver que essas idades são distintas, o que reflete a diferença nas condições sanitárias. Portanto, melhorias nas condições sanitárias levam provavelmente a mudanças na idade média de infecção, o que significa que \bar{a} pode ser um parâmetro adequado para avaliar tais mudanças.

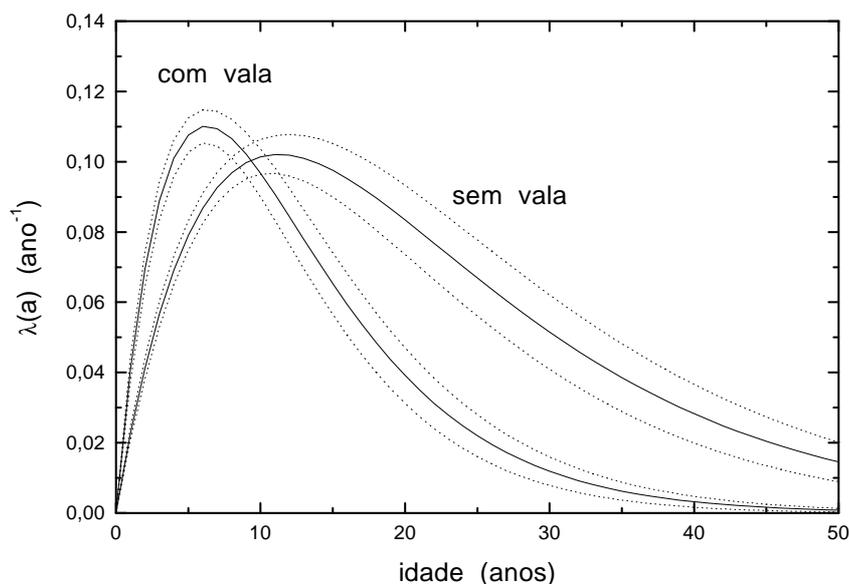


Figura 3: Força de infecção em função da idade para a hepatite A, referente ao levantamento sorológico feito em Duque de Caxias. São mostradas as curvas para as duas categorias analisadas, com e sem a presença de vala nas proximidades dos domicílios [1]. As linhas pontilhadas indicam o intervalo de confiança de 68% para $\lambda(a)$.

3. Comentários Finais

Os exemplos aqui apresentados têm, de uma certa forma, um caráter didático, uma vez que nossa intenção é demonstrar as potencialidades do método. Qualitativamente, seria de se esperar que, *a priori*, os resultados se comportassem como se comportaram, mas a quantificação da incerteza estatística da idade média de infecção é uma informação que nos habilita a comparar valores com um embasamento maior.

Uma aplicação interessante do método seria verificar a incerteza da idade média em modelos incluindo vacinação. Neste caso, poder-se-ia comparar as idades médias de infecção antes e após o procedimento vacinal.

A técnica de propagação de erros (Subseção 2.2) foi utilizada para se estimar a incerteza da força de infecção. Poderíamos também, através de propagação de erros, estimar a incerteza na idade média de infecção, uma vez que esta é também uma função do vetor de parâmetros \mathbf{k} . No entanto, teríamos de avaliar numericamente não apenas as integrais da equação (2.2) mas também derivadas parciais (como as

encontradas na expressão (2.4)).

Um outro aspecto que merece ser mencionado refere-se à idade máxima de integração, L . A escolha natural para o valor de L seria a expectativa de vida humana. Todavia, adotamos L como sendo a idade máxima observada nos dados soropidemiológicos, evitando extrapolações.

Dos resultados que obtivemos para a idade média de infecção da varicela nas duas comunidades estudadas, podemos observar que diferentes idades médias de infecção estão relacionadas a diferentes padrões de interação entre as crianças. As incertezas estimadas nos possibilitaram verificar se há diferenças entre as idades médias. Por outro lado, as simulações para a hepatite A nos forneceram informações que permitiram concluir que diferenças nas condições sanitárias estão associadas a idades de infecção distintas, na comunidade analisada.

Também merece atenção o fato de que a incerteza da idade média de infecção está diretamente relacionada ao intervalo de confiança da força de infecção dependente da idade.

Embora tenhamos desenvolvido uma simulação de Monte Carlo para o caso específico em que os parâmetros são estimados através da função de soroprevalência (2.3), o algoritmo aqui apresentado pode também ser aplicado a outras funções. O método de Monte Carlo também pode ser utilizado em conjunto com modelos epidêmicos para avaliar as incertezas de outras variáveis epidemiológicas, desde que se tenha informações sobre as distribuições estatísticas desses parâmetros.

Abstract. Monte Carlo simulations have been carried out in order to estimate the statistical uncertainty of the average age of acquisition of infection, an important epidemiological parameter in the mathematical modelling of infectious diseases. We have also estimated the confidence intervals for the age-dependent force of infection by means of propagation of errors. Uncertainty estimates for both the force of infection and the average age at infection are auxiliary tools for comparing and characterizing distinct epidemiological situations, as we show in the applications for varicella and hepatitis A.

Referências

- [1] L.M. Almeida, M. Amaku, G.L. Werneck, R.S. Azevedo e E. Massad, Force of infection by age and average age of infection of hepatitis A in Duque de Caxias, Rio de Janeiro, Brazil, 1997, em “9th International Congress on Infectious Diseases”, Abstract Book, p.191, Buenos Aires, Argentina, 2000.
- [2] R.M. Anderson e R.M. May, “Infectious Diseases of Humans: Dynamics and Control”, Oxford University Press, Oxford, 1991.
- [3] R.M. Anderson e R.M. May, Vaccination against rubella and measles: quantitative investigations of different policies, *J. Hyg., Camb.* **90** (1983), 259-325.
- [4] P.R. Bevington, “Data Reduction and Error Analysis for the Physical Sciences”, McGraw-Hill, New York, 1969.

- [5] N.R. Draper e H. Smith, “Applied Regression Analysis”, Wiley, New York, 1966.
- [6] C.P. Farrington, Modelling forces of infection for measles, mumps and rubella, *Stat. Med.* **9** (1990), 953-967.
- [7] B.T. Grenfell e R.M. Anderson, The estimation of age-related rates of infection from case notifications and serological data, *J. Hyg., Camb.* **95** (1985), 419-436.
- [8] J.M. Hammersley e D. C. Handscomb, “Monte Carlo Methods”, Methuen, London, 1964.
- [9] W.H. Press, S.A. Teukolsky, W.T. Vetterling e B. P. Flannery, “Numerical Recipes in C: The Art of Scientific Computing”, 2.ed., Cambridge University Press, Cambridge, 1996.
- [10] V.R. Vanin e P. Gouffon, “Tópicos Avançados em Tratamento Estatístico de Dados em Física Experimental”, edição preliminar, Instituto de Física da USP - LAL, São Paulo, 1996.
- [11] A.L.F. Yu, M. Amaku, J.M. Costa, C.S. Pannutti, V.A.U.F. Souza, D.M.T. Zanetta, M.N. Burattini, E. Massad e R.S. Azevedo, Varicella transmission in two samples of children with different social behavior in the State of São Paulo, Brazil, em “9th International Congress on Infectious Diseases”, Abstract Book, p.16, Buenos Aires, Argentina, 2000.
- [12] Centers for Disease Control and Prevention (CDCP), Prevention of hepatitis A through active or passive immunization: recommendations of the advisory committee on immunization practices (ACIP), *MMWR* **48**, No. RR-12 (1999), 1-37.

