



Revista de Saúde Pública

ISSN: 0034-8910

revsp@usp.br

Universidade de São Paulo  
Brasil

Zangiacomi Martinez, Edson; Ruffi no-Netto, Antonio; Achcar, Jorge Alberto; Casale  
Aragon, Davi

Modelagem Bayesiana do risco de infecção tuberculosa para estudos com perdas de  
seguimento

Revista de Saúde Pública, vol. 42, núm. 6, diciembre, 2008, pp. 999-1004

Universidade de São Paulo  
São Paulo, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=67240171004>

- Como citar este artigo
- Número completo
- Mais artigos
- Home da revista no Redalyc

re<sup>o</sup>alyc.org

Sistema de Informação Científica

Rede de Revistas Científicas da América Latina, Caribe, Espanha e Portugal

Projeto acadêmico sem fins lucrativos desenvolvido no âmbito da iniciativa Acesso Aberto

Edson Zangiacomi Martinez<sup>I</sup>

Antonio Ruffino-Netto<sup>I</sup>

Jorge Alberto Achcar<sup>I</sup>

Davi Casale Aragon<sup>II</sup>

# Modelagem Bayesiana do risco de infecção tuberculosa para estudos com perdas de seguimento

## Bayesian model for the risk of tuberculosis infection for studies with individuals lost to follow-up

---

### RESUMO

**OBJETIVO:** Desenvolver um modelo estatístico baseado em métodos Bayesianos para estimar o risco de infecção tuberculosa em estudos com perdas de seguimento, comparando-o com um modelo clássico determinístico.

**MÉTODOS:** O modelo estocástico proposto é baseado em um algoritmo de amostradores de Gibbs, utilizando as informações de perdas de seguimento ao final de um estudo longitudinal. Para simular o número desconhecido de indivíduos reatores ao final do estudo e perdas de seguimento, mas não reatores no tempo inicial, uma variável latente foi introduzida no novo modelo. Apresenta-se um exercício de aplicação de ambos os modelos para comparação das estimativas geradas.

**RESULTADOS:** As estimativas pontuais fornecidas por ambos os modelos são próximas, mas o modelo Bayesiano apresentou a vantagem de trazer os intervalos de credibilidade como medidas da variabilidade amostral dos parâmetros estimados.

**CONCLUSÕES:** O modelo Bayesiano pode ser útil em estudos longitudinais com baixa adesão ao seguimento.

**DESCRITORES:** Tuberculose, epidemiologia. Fatores de Risco. Técnicas de Estimativa. Teorema de Bayes. Modelos Matemáticos.

<sup>I</sup> Departamento de Medicina Social.  
Faculdade de Medicina de Ribeirão Preto  
(FMRP). Universidade de São Paulo.  
Ribeirão Preto, SP, Brasil

<sup>II</sup> Programa de Pós-Graduação em Saúde na  
Comunidade. FMRP. Ribeirão Preto, SP,  
Brasil

#### Correspondência | Correspondence:

Edson Zangiacomi Martinez  
Faculdade de Medicina de Ribeirão Preto  
Departamento de Medicina Social  
Universidade de São Paulo - USP  
Av. Bandeirantes, 3900  
14049-900 Ribeirão Preto, SP, Brasil  
E-mail: edson@fmrp.usp.br

Recebido: 20/8/2007

Revisado: 9/5/2008

Aprovado: 2/6/2008

---

## ABSTRACT

**OBJECTIVE:** To develop a statistical model based on Bayesian methods to estimate the risk of tuberculosis infection in studies including individuals lost to follow-up, and to compare it with a classic deterministic model.

**METHODS:** The proposed stochastic model is based on a Gibbs sampling algorithm that uses information of lost to follow-up at the end of a longitudinal study. For simulating the unknown number of reactors at the end of the study and lost to follow-up, but not reactors at time 0, a latent variable was introduced in the new model. An exercise of application of both models in the comparison of the estimates of interest was presented.

**RESULTS:** The point estimates obtained from both models are near identical; however, the Bayesian model allowed the estimation of credible intervals as measures of precision of the estimated parameters.

**CONCLUSIONS:** The Bayesian model can be valuable in longitudinal studies with low adherence to follow-up.

**DESCRIPTORS:** Tuberculosis. Epidemiology. Risk factors. Estimation techniques. Bayes' theorem. Mathematical models.

---

## INTRODUÇÃO

A tuberculose é um importante problema de saúde pública, com cerca de nove milhões anuais de novos casos e responsável por dois milhões de óbitos a cada ano.<sup>10</sup> A magnitude da doença está associada à desigualdade social (e portanto a pobreza, má distribuição de renda, más condições de alimentação, transporte e de trabalho), à urbanização acelerada, bem como ao envelhecimento da população, movimentos migratórios e ao advento da Aids. No Brasil, o coeficiente de incidência foi de 42 por 100.000 em 2005. O País ocupa a 16.<sup>a</sup> posição em um *ranking* de 22 países onde se estima que ocorram 80% dos casos de tuberculose do mundo todo.<sup>21</sup> Em um país de alta prevalência de tuberculose como o Brasil, as ações para o diagnóstico precoce e seu efetivo tratamento são de grande importância.

Os modelos epidemiométricos são ferramentas úteis para compreensão da dinâmica da tuberculose, podendo nortear políticas públicas de rastreamento e explicar o impacto de outras infecções, por exemplo, pelo HIV. Estes modelos utilizam conhecimento da biologia da propagação das doenças, buscando quantificar elementos da dinâmica da transmissão do agente infeccioso, para evidenciar efeitos de fatores demográficos e epidemiológicos sobre o comportamento da transmissão. Na literatura médica, vários autores têm proposto modelos determinísticos para descrever a dinâmica da tuberculose pulmonar.<sup>3,5,8,12,19</sup> Em especial, tratando-se da estimação do risco anual da infecção tuberculosa ( $\theta$ ), Ruffino-Netto & Arantes<sup>13</sup> propuseram o uso da expressão  $\theta = (1/t) \ln (A_0/A_t)$ , onde  $A_0$  é o número inicial de indivíduos não infectados no início da contagem de

tempo (tempo  $t=0$ ) e  $A_t$  é o número de indivíduos não infectados ou que permanecem virgens de infecção, decorrido certo tempo  $t$  (em anos). Este modelo é ainda bastante utilizado em estudos da epidemiologia da tuberculose, como no trabalho de Kusano et al<sup>11</sup> sobre o risco anual da infecção tuberculosa na população escolar e residente no Distrito Federal, em 1997.

Para obter uma estimativa do risco da infecção tuberculosa em um intervalo de tempo, Ruffino-Netto<sup>14</sup> propôs um modelo baseado em equações diferenciais, capaz de operar na situação onde há perdas de seguimento. O presente artigo teve por objetivo desenvolver um modelo baseado em um método Bayesiano<sup>6,7,17</sup> de inferência estatística, comparando-o ao modelo de Ruffino-Netto.<sup>14</sup> Para simular o número desconhecido de indivíduos reatores ao final do estudo e perdas de seguimento, mas não reatores no tempo inicial, uma variável latente foi introduzida no novo modelo.<sup>18</sup> O modelo proposto tem a vantagem de apresentar medidas para a variabilidade amostral (na forma de intervalos de credibilidade) para os parâmetros de interesse.

## MÉTODOS

### Estimação pelo modelo determinístico

Suponhamos que em um dado momento, efetuamos a prova tuberculínica em  $K$  indivíduos e encontramos que  $R_0$  são reatores e  $N_0$  são não reatores à prova. Portanto,  $K = R_0 + N_0$ . Posteriormente, no tempo  $t$ , repetimos o mesmo levantamento e encontramos

dentre os  $N_0$  não reatores iniciais  $N$  indivíduos que permanecem na categoria de não reator,  $I$  indivíduos que apresentam conversão tuberculínica, e  $Pn$  indivíduos são perdidos de seguimento. Notar que  $N + I + Pn = N_0$ . Dentre os  $R_0$  reatores iniciais,  $R$  continuam reatores e  $Pr$  são perdidos de seguimento. Temos por objetivo encontrar  $\theta$ , o risco de infecção tuberculosa entre estes indivíduos, levando em consideração que uma parcela foi perdida de seguimento.

Considerando ainda as taxas de transferência  $\lambda_1$ ,  $\lambda_2$  e  $\lambda_3$ , de um estado a outro,<sup>15</sup> sendo  $\lambda_1$  a taxa de transferência do estado  $N$  ao estado  $Pn$ ,  $\lambda_2$  a taxa de transferência de  $I$  a  $Pi$ , e  $\lambda_3$  a taxa de transferência de  $R$  a  $Pr$ . Aqui,  $Pi$  é o número (desconhecido) de indivíduos infectados e perdidos no tempo  $t$ . Vamos considerar que estas taxas permanecem constantes dentro do intervalo de tempo entre os dois levantamentos. Para encontrar  $\theta$ , Ruffino-Netto<sup>14</sup> utiliza dois pressupostos. O primeiro tem por base o conhecimento que a reversão tuberculínica é um fato cuja ocorrência, além de pouco freqüente, envolve alguns anos após o organismo previamente infectado conseguir esterilizar o bacilo de Koch nele existente. O segundo considera que a infecção primária (sem complicações) não constitui, por si só, um fator que aumente a mortalidade e/ou leve ao maior “desaparecimento” das pessoas durante um estudo longitudinal. Portanto, assumimos que as taxas de transferência entre os estados de  $Pn$  a  $I$ , de  $I$  a  $N$ , de  $Pn$  a  $N$ , de  $Pi$  a  $I$  e de  $Pr$  a  $R$  são iguais a zero. Assumimos ainda que  $\lambda_2$  é igual a  $\lambda_3$ , e, por simplicidade, denotaremos estas duas taxas de transferência por  $\lambda_2$ . Notar que o risco de infecção  $\theta$  é a taxa de transferência do estado  $N$  ao  $I$ .

Utilizando equações diferenciais lineares, Ruffino-Netto<sup>14</sup> apresentou o seguinte sistema de três equações:

$$N = N_0 \exp[-t(\lambda_1 + \theta)]$$

$$I = \frac{N_0 \theta}{\lambda_2 - \lambda_1 - \theta} \{ \exp[-t(\lambda_1 + \theta)] - \exp(-\lambda_2 t) \} \quad e$$

$$R = R_0 \exp(-\lambda_2 t)$$

Considerar, por exemplo, que efetuamos a prova tuberculínica em 1.100 indivíduos, sendo  $N_0 = 1000$  não reatores à prova e  $R_0 = 100$  reatores. Digamos que, no tempo  $t$ ,  $N = 880$  indivíduos são não reatores (ou seja,

continuem virgens de infecção e não perdidos de seguimento),  $I = 10$  apresentam conversão tuberculínica e  $Pn = 110$  são perdidos de seguimento. Sabemos ainda que, dentre os 100 reatores iniciais,  $R = 90$  permanecem reatores e  $Pr = 10$  são perdidos de seguimento no tempo  $t$ . Considerando  $t = 1$  ano, das equações apresentadas, temos  $\lambda_1 = 0,1166$ ,  $\lambda_2 = 0,1054$  e  $\theta = 0,0112$ . Portanto, temos um risco de infecção igual a 1,12% ao ano, entre estes indivíduos.

### Proposta de um modelo Bayesiano

Considerando  $T_0$  uma variável aleatória associada ao resultado da prova tuberculínica no tempo inicial, assumindo 0 (zero) se não reator e 1 se reator, e  $T_1$  uma variável aleatória associada ao resultado da prova tuberculínica no tempo  $t$ , novamente assumindo 0 se não reagente e 1 se reagente. Para uma abordagem estocástica do problema, consideramos o risco de infecção  $\theta$  como a probabilidade condicional  $P(T_1 = 1 | T_0 = 0)$  de um indivíduo ser infectado antes do tempo  $t$ , dado que era livre da infecção no tempo inicial. Por outro lado, assumimos que a probabilidade condicional  $P(T_1 = 1 | T_0 = 1)$  de um indivíduo ser reator ao tempo  $t$ , dado que já era reator no tempo inicial é igual a 1. Assim,  $P(T_1 = 0 | T_0 = 0) = 1 - \theta$  e  $P(T_1 = 0 | T_0 = 1) = 0$ . Consideramos ainda a probabilidade  $P(T_0 = 1) = \beta$ . Para discriminar os indivíduos perdidos de seguimento dos presentes ao final do estudo, consideramos ainda a variável  $S$ , que assume 0 para os indivíduos perdidos de seguimento e 1 caso contrário. Assim, nesta abordagem, temos que  $\lambda_1 = P(S = 0 | T_1 = 0, T_0 = 0)$  e  $\lambda_2 = P(S = 0 | T_1 = 1)$ , ou seja,  $\lambda_1$  é a probabilidade de perda de seguimento, dado que o indivíduo é não reator no início e no final do estudo, e  $\lambda_2$  é a probabilidade de perda de seguimento, dado que o indivíduo é reator ao final do estudo.

A Tabela 1 apresenta as freqüências absolutas encontradas no cruzamento de  $S$ ,  $T_0$  e  $T_1$ , considerando a notação introduzida na Seção anterior. Assumindo a não ocorrência da reversão tuberculínica dentro do seguimento, notar que a freqüência associada à situação onde  $S = 1$ ,  $T_1 = 0$  e  $T_0 = 1$  é necessariamente igual a zero. Com base em regras de probabilidade condicional, encontramos que a chance de termos um indivíduo simultaneamente presente ao final do estudo ( $S=1$ ), reator ao final do estudo ( $T_1 = 1$ ), mas não reator no início ( $T_0 = 0$ ) é

**Tabela 1.** Freqüências absolutas encontradas no cruzamento de  $S$ ,  $T_0$  e  $T_1$ .

	Pessoas presentes ao final do estudo ( $S = 1$ )		Perdas de seguimento ( $S = 0$ )	
	Não reatores ao final do estudo ( $T_1 = 0$ )	Reatores ao final do estudo ( $T_1 = 1$ )		Total
Não reatores no início do estudo ( $T_0 = 0$ )	$N$	$I$	$Pn$	$N_0$
Reatores no início do estudo ( $T_0 = 1$ )	0	$R$	$Pr$	$R_0$
Total	$N$	$I + R$	$Pn + Pr$	$K$

$$\begin{aligned}
P(S = 1, T_1 = 1, T_0 = 0) \\
&= P(S = 1 | T_1 = 1, T_0 = 0) P(T_1 = 1, T_0 = 0) \\
&= P(S = 1 | T_1 = 1, T_0 = 0) P(T_1 = 1 | T_0 = 0) P(T_0 = 0) \\
&= (1 - \lambda_1) \theta (1 - \beta).
\end{aligned}$$

As probabilidades associadas às demais combinações entre  $S$ ,  $T_0$  e  $T_1$  são exibidas na Tabela 2.

Seja o vetor de parâmetros de interesse  $\xi = (\theta, \lambda_1, \lambda_2, \beta)$ . Utilizaremos então uma abordagem Bayesiana para a estimação destas quantidades. Para isto, considere no modelo a variável  $Pi$ , o número de indivíduos reatores ao final do estudo e perdidos de seguimento, mas não reatores no tempo inicial. Esta variável não pode ser observada diretamente da amostra, mas podemos atribuir a ela uma distribuição de probabilidade binomial, assim definida:

$$Pi | Pn, \xi \sim \text{Binomial} \left( Pn, \frac{\lambda_2 \theta (1 - \beta)}{\lambda_2 \theta (1 - \beta) + \lambda_1 (1 - \theta) (1 - \beta)} \right)$$

A função de verossimilhança para  $\xi$  é dada pelo produto entre cada uma das probabilidades exibidas na Tabela 2 (exceto as probabilidades da última coluna, associadas aos totais), elevadas às contagens apresentadas na Tabela 1, ou seja,

$$\begin{aligned}
L(\xi) &= \lambda_1^{Pn - Pi} (1 - \lambda_1)^N \lambda_2^{Pi} (1 - \lambda_2)^{I + R} \\
&\times \theta^{I + Pi} (1 - \theta)^{N + Pn - Pi} \beta^{Pr + R} (1 - \beta)^{N_0}
\end{aligned}$$

Esta função tem por objetivo a caracterização da distribuição conjunta dos dados amostrais. Dois tipos de informação são considerados pelo método Bayesiano: a proveniente do conhecimento do pesquisador e aquela contida nos dados observados, aqui representa por  $L(\xi)$ . Se, por um lado, o método tradicional de inferência estatística baseia-se somente nos dados amostrais, os modelos Bayesianos representam o conhecimento do pesquisador por uma distribuição de probabilidades que será calibrada pela observação dos dados. A informação vinda do conhecimento do pesquisador é chamada de *a priori*, enquanto aquela já calibrada pela observação dos dados é chamada *a posteriori*. Dado que as quantidades  $\theta$ ,  $\lambda_1$ ,  $\lambda_2$  e  $\beta$  são interpretadas como probabilidades, é conveniente atribuir a estes parâmetros uma distribuição de probabilidade beta, dada a capacidade desta distribuição assumir valores dentro

do intervalo de zero a um. Desta forma, assumindo as distribuições *a priori* independentes  $\theta \sim \text{Beta}(a_\theta, b_\theta)$ ,  $\beta \sim \text{Beta}(a_\beta, b_\beta)$ ,  $\lambda_1 \sim \text{Beta}(a_1, b_1)$  e  $\lambda_2 \sim \text{Beta}(a_2, b_2)$ , a distribuição *a posteriori* é dada pelo produto entre  $L(\xi)$  e a distribuição conjunta *a priori* de  $\theta$ ,  $\lambda_1$ ,  $\lambda_2$  e  $\beta$ . As distribuições condicionais *a posteriori* são

$$\theta | I, N, Pn, Pi, a_\theta, b_\theta \sim \text{Beta}(I + Pi + a_\theta, N + Pn - Pi + b_\theta),$$

$$\beta | Pr, R, N_0, a_\beta, b_\beta \sim \text{Beta}(R_0 + a_\beta, N_0 + b_\beta),$$

$$\lambda_1 | N, Pn, Pi, a_1, b_1 \sim \text{Beta}(Pn - Pi + a_1, N + b_1) \text{ e}$$

$$\lambda_2 | I, R, Pi, a_2, b_2 \sim \text{Beta}(Pi + a_2, I + R + b_2).$$

A estimação Bayesiana de  $\beta$  independe da variável latente  $Pi$ . Considerando a estimativa Bayesiana de cada parâmetro dada pela média da sua respectiva distribuição condicional *a posteriori*, temos que  $\beta$  é estimado por  $R_0 + a_\beta$  dividido por  $K + a_\beta + b_\beta$ . Os demais parâmetros,  $\theta$ ,  $\lambda_1$  e  $\lambda_2$ , não podem ser assim diretamente estimados, dado que as suas distribuições condicionais *a posteriori* dependem de  $Pi$ . Por outro lado, podemos simular as distribuições destes parâmetros por meio do algoritmo de amostradores de Gibbs (Casella & George),<sup>2</sup> facilmente implementado em programas computacionais como o *WinBugs*.<sup>16</sup>

## RESULTADOS

Seja o exemplo apresentado da prova tuberculínica em Métodos, de aplicação da prova tuberculínica em 1.100 indivíduos, sendo  $N_0 = 1000$ ,  $R_0 = 100$ ,  $N = 880$ ,  $I = 10$ ,  $Pn = 110$ ,  $R = 90$ ,  $Pr = 10$  e  $t = 1$ . Para o modelo Bayesiano proposto, buscamos especificar distribuições *a priori* não-informativas, ou seja, distribuições de probabilidade que têm por intenção trazer pouco poder explanatório ao parâmetro desconhecido de interesse. Assim, consideramos  $a_\theta = b_\theta = a_\beta = b_\beta = a_1 = b_1 = a_2 = b_2 = 0,5$  (ver motivação em Box & Tiao<sup>1</sup>). Usando o programa de computador *WinBugs*,<sup>16</sup> inicialmente foram geradas 10 mil amostras utilizando o algoritmo proposto, que foram descartadas para eliminar o efeito dos valores iniciais (*burn-in samples*). A seguir, 500 mil amostras foram geradas e selecionamos valores em saltos tamanho 10 para a obtenção de observações independentes. A convergência do algoritmo foi verificada graficamente e usando métodos usuais, como o critério de Gelman & Rubin.<sup>9</sup>

**Tabela 2.** Probabilidades associadas às combinações entre as variáveis  $S$ ,  $T_0$  e  $T_1$ .

	Pessoas presentes ao final do estudo ( $S = 1$ )		Perdas de seguimento	
	Não reatores ao final do estudo ( $T_1 = 0$ )	Reatores ao final do estudo ( $T_1 = 1$ )	( $S = 0$ )	Total
Não reatores no início do estudo ( $T_0 = 0$ )	$(1 - \lambda_1) (1 - \theta) (1 - \beta)$	$(1 - \lambda_2) \theta (1 - \beta)$	$\lambda_2 \theta (1 - \beta) + \lambda_1 (1 - \theta) (1 - \beta)$	$1 - \beta$
Reatores no início do estudo ( $T_0 = 1$ )	0	$(1 - \lambda_2) \beta$	$\lambda_2 \beta$	$\beta$

**Tabela 3.** Estimativas fornecidas pelos modelos determinístico e Bayesiano.

Parâmetro	Modelo determinístico	Modelo Bayesiano	
		Estimativa <i>a posteriori</i>	Intervalo de credibilidade 95%
$\lambda_1$	11,7%	11,0%	9,1%;13,1%
$\lambda_2$	10,5%	10,5%	5,2%;17,1%
$\theta$	1,10%	1,2%	0,5%;2,0%
$\beta$	-	9,1%	7,4%;10,9%

A Tabela 3 compara as estimativas Bayesianas com as fornecidas pelo modelo determinístico. Os intervalos de credibilidade 95% são versões Bayesianas dos intervalos de confiança gerados pelos métodos tradicionais de inferência estatística. As estimativas pontuais fornecidas por ambos os modelos são próximas, mas o modelo Bayesiano tem a vantagem de trazer os intervalos de credibilidade como medidas da variabilidade amostral dos parâmetros estimados.

Para analisar os efeitos da especificação das distribuições *a priori* sobre os nossos resultados, conduzimos uma análise de sensibilidade, onde estimativas Bayesianas foram obtidas para várias possibilidades de  $a_p$ ,  $b_p$ ,  $a_\beta$ ,  $b_\beta$ ,  $a_1$ ,  $b_1$ ,  $a_2$  e  $b_2$ . Observamos que as médias das distribuições *a posteriori* não foram muito diferentes daquelas apresentadas na Tabela 3, quando as médias das distribuições *a priori* foram próximas àquelas encontradas pelo modelo determinístico. No entanto, as estimativas dos parâmetros sofrem efeitos mais expressivos quando as médias das distribuições *a priori* são maiores que 50%.

Para verificar o desempenho do modelo Bayesiano em uma situação onde o número de perda de seguimento é maior, simulamos um conjunto de dados onde  $\lambda_1 = 25\%$ ,  $\lambda_2 = 30\%$  e  $\theta = 1,12\%$ . Considerando  $K = 1.100$ , temos  $N = 742$ ,  $I = 8$ ,  $Pn = 250$ ,  $R = 70$  e  $Pr = 30$ . As soluções para o modelo determinístico são  $\lambda_1 = 28,7\%$ ,  $\lambda_2 = 35,7\%$  e  $\theta = 1,11\%$ . Estimativas Bayesianas para  $\lambda_1$ ,  $\lambda_2$  e  $\theta$  são, respectivamente, 24,9% (intervalo de credibilidade 95% [IC 95%] 7,4% a 10,9%), 30,2% (IC 95% 21,7% a 39,6%) e 1,18% (IC 95% 0,54% a 2,2%).

## DISCUSSÃO

O uso de modelos matemáticos no campo das doenças infecciosas está inserido em uma das mais antigas linhas epidemiológicas de investigação, sendo seus primórdios brevemente revisados por Werneck,<sup>20</sup> que menciona os trabalhos pioneiros de Bernoulli e Farr, publicados nos séculos XVIII e XIX, respectivamente. No âmbito da tuberculose, o trabalho de Ruffino-Netto<sup>15</sup> conceituou os “estados” na história natural da doença – “virgem de

infecção”, “infectados” e “doentes” – propondo um modelo matemático para a descrição dos seus inter-relacionamentos e o cálculo do risco de infecção. Em um estudo anterior, Ruffino-Netto<sup>14</sup> já discutia a possibilidade de calcular o risco de infecção tuberculosa considerando perdas de seguimento. No presente artigo, apresentamos uma releitura desta proposta, agora sob uma perspectiva estocástica, onde um algoritmo Bayesiano permite estimativas intervalares do risco de infecção.

A obtenção de estimativas de máxima verossimilhança para os parâmetros do modelo a partir da expressão apresentada para  $L(\xi)$  não é simples, dada a presença da variável latente  $P_i$ . Considerando que as perdas de seguimento trazem uma estrutura de dados incompletos, uma maneira de obter estimativas de máxima verossimilhança é por meio do algoritmo EM (*Expectation-Maximization algorithm*).<sup>4</sup> No entanto, sua implementação traz algumas complexidades algébricas, cuja discussão extrapolaria os objetivos do presente estudo.

Ainda que o modelo Bayesiano necessite de um algoritmo Monte Carlo para a simulação das distribuições *a posteriori* dos parâmetros de interesse, procedimento este portador de alguma complexidade matemática, acreditamos que a proposta aqui apresentada não seja desprovida de parcimônia, quanto à sua formulação. Os procedimentos Bayesianos vêm tornado-se cada vez mais frequentes nas publicações da área da saúde,<sup>7,17</sup> e suas aplicações são facilitadas por programas computacionais como o *WinBugs*,<sup>16</sup> que não requer amplos conhecimentos de programação para a implementação do modelo.

Em conclusão, o modelo Bayesiano proposto é uma alternativa eficiente na estimação do risco de infecção tuberculosa considerando perdas de seguimento, sendo especialmente útil em estudos longitudinais com baixa adesão ao seguimento. Dado que o método Bayesiano permite a incorporação do conhecimento prévio de especialistas, há ainda a possibilidade de obter estimativas de risco mais precisas especificando distribuições *a priori* para os parâmetros de interesse de acordo com estas informações.

## REFERÊNCIAS

1. Box GEP, Tiao GC. Bayesian inference in statistical analysis. Boston: Addison Wesley; 1973.
2. Casella G, George EL. Explaining the Gibbs sampler. *Am Statistician*. 1992;46(3):167-74. DOI: 10.2307/2685208
3. de la Pava E, Salguero B, Alzate A. A mathematical model of the annual risk of tuberculosis infection in Cali, Colombia. *Rev Panam Salud Publica*. 2002;11(3):166-71. DOI: 10.1590/S1020-49892002000300005
4. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM Algorithm. *J R Statist Soc B*. 1977;39:1-38.
5. Diekmann O, Heesterbeek JAP. Mathematical epidemiology of infectious diseases: Model building, analysis and interpretation. New York: Wiley; 2000.
6. Etzioni RD, Kadane JB. Bayesian statistical methods in public health and medicine. *Annu Rev Public Health*. 1995;16:23-41. DOI: 10.1146/annurev.pu.16.050195.000323
7. Freedman L. Bayesian statistical methods. *BMJ*. 1996;313(7057):569-70.
8. Garcia A, Maccario J, Richardson S. Modelling the annual risk of tuberculosis infection. *Int J Epidemiol*. 1997;26(1):190-203. DOI: 10.1093/ije/26.1.190
9. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Statist Scien*. 1992;7(4):457-511. DOI: 10.1214/ss/1177011136
10. Harries AD, Dye C. Tuberculosis. *Ann Trop Med Parasitol*. 2006;100(5-6):415-31. DOI: 10.1179/136485906X91477
11. Kusano MSE, Mendes IJM, Alves ED, Assis MCM. Risco anual da infecção tuberculosa no Distrito Federal (Brasil). *Rev Bras Epidemiol*. 2005;8(3):262-71. DOI: 10.1590/S1415-790X2005000300008
12. Montesinos-López OA, Hernández-Suárez CM. Modelos matemáticos para enfermedades infecciosas. *Salud Publica Mexico*. 2007;49(3):218-26. DOI: 10.1590/S0036-36342007000300007
13. Ruffino-Netto A, Arantes GR. Risco de infecção em município do interior do Estado de São Paulo e suas aplicações. *Rev Saude Publica*. 1976;10(2):143-9. DOI: 10.1590/S0034-89101976000200002
14. Ruffino-Netto A. Cálculo do risco de infecção tuberculosa levando em consideração pessoas perdidas de seguimento. *Rev Divisão Nac Tuberculose*. 1976;20(80):383-90.
15. Ruffino-Netto A. Modelos epidemiométricos em tuberculose: definição de “estados” e risco de infecção. *Rev Saude Publica*. 1977;11(2):188-98. DOI: 10.1590/S0034-89101977000200004
16. Spiegelhalter DJ, Thomas A, Best N, Lunn D. WinBUGS user manual: version 1.4. Cambridge: MRC Biostatistics Unit; 2001.
17. Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. An introduction to Bayesian methods in health technology assessment. *BMJ*. 1999;319(7208):508-12.
18. Tanner M, Wong W. The calculation of posterior distributions by data augmentation. *J Amer Statist Assoc*. 1987;82(398):528-50. DOI: 10.2307/2289457
19. Waaler H, Geser A, Andersen S. The use of mathematical models in the study of the epidemiology of tuberculosis. *Am J Public Health*. 1962;52:1002-13.
20. Werneck GL. Comentário: primórdios da investigação sobre a dinâmica de transmissão da tuberculose no Brasil. *Rev Saude Publica*. 2006;40(2):218-9. DOI: 10.1590/S0034-89102006000200005
21. World Health Organization. Global tuberculosis control: surveillance, planning, financing. Geneva; 2004.