

## **Comparação de modelos que estimam a razão de prevalência, como alternativa para a regressão logística em estudos transversais**

**Luis Fernando Lisboa** – Hospital Israelita Albert Einstein – Av. Albert Einstein, 627-701 – Morumbi – São Paulo, SP

**E-mail:** [luislisboa@einstein.br](mailto:luislisboa@einstein.br)

### **Resumo:**

Em estudos de corte transversal com desfecho dicotômico, a associação entre exposição e desfecho pode ser estimada por modelos de regressão logística, porém, quando o desfecho não é raro, ou quando é necessário ajustar para potenciais variáveis, a *odds ratio* (OR) não é uma boa estimativa para a razão de prevalência (RP).

Foram analisados os fatores associados a prevalência de obstrução das vias aéreas superiores (VAS) pós-extubação em crianças internadas no CTI pediátrico de um hospital privado. As RP's foram obtidas pelo método de estratificação de Mantel-Haenzel. As estimativas ajustadas foram obtidas pelas regressões de Cox e de Poisson com variância robusta, e regressão log-binomial. As OR's ajustadas foram obtidas pela regressão logística.

A prevalência de obstrução das VAS foi de 38,7%. As co-variáveis uso de *cuff* e tempo de VMI apresentaram associação significativa.

Os modelos de Cox/Poisson com variância robusta e o log-binomial ofereceram estimativas mais consistentes comparados à regressão logística. Os modelos de Cox/Poisson com variância robusta apresentaram estimativas próximas comparadas às da estratificação de Mantel-Haenzel para as co-variáveis mais prevalentes, enquanto o log-binomial apresentou estimativas próximas para as co-variáveis menos prevalentes. O modelo log-binomial apresentou intervalos de confiança de menor amplitude e nível descritivo mais conservador.

## **Introdução**

Em estudos de corte transversal, como a exposição e o desfecho são observados simultaneamente, não é possível determinar a duração do desfecho (no caso de uma doença, por exemplo), por isso, esse tipo de estudo também é chamado de estudo de prevalência [1]. Em estudos de corte transversal com desfecho dicotômico, a associação entre exposição e desfecho pode ser estimada por modelos de regressão logística, porém, quando o desfecho não é raro, ou quando é necessário ajustar para potenciais variáveis de confusão ou analisar a interação entre as variáveis, a razão de chances ou *odds ratio* não é uma boa estimativa para a razão de prevalência [2]. Até pouco tempo atrás poucos pacotes estatísticos ofereciam modelos que pudessem estimar diretamente a razão de prevalência, porém, agora que a maioria dos pacotes oferece opções mais amigáveis para utilização de modelos lineares generalizados, que incluem os modelos de regressão linear, logística, Poisson, log-binomial, entre outros, a persistência em continuar a usar a regressão logística para nesses casos não é justificável.

Nos dias de hoje, várias alternativas tem sido propostas para analisar dados de estudos de corte transversal, e parece que ainda não há um consenso, o que sugere que a escolha do modelo depende das características dos dados a serem analisados.

O objetivo desse estudo foi estimar a razão de prevalência (RP) pelos modelos de regressão de Cox, Poisson com variância robusta, e regressão log-binomial, comparar com *odds ratios* (OR) obtidos pela regressão logística.

## **Método**

Como exemplo, foi analisado a prevalência de obstrução das vias aéreas superiores (VAS) pós-extubação em 75 crianças internadas no CTI pediátrico de um hospital privado no período entre 2007 e 2009. As variáveis consideradas foram idade da criança, tamanho da cânula, uso do *cuff*, uso de corticóide e tempo de uso de ventilação mecânica intermitente (VMI).

Foram estimados como medidas de associação a razão de prevalência (RP) bruta, obtidos pela estratificação de Mantel-Haenszel, os *odds Ratio* (OR) ajustados, estimados pelo modelo de regressão logística, e as razões de prevalência múltiplas, estimadas pelos modelos de Cox, Poisson e log-binomial. Todas as medidas foram estimadas com seus respectivos intervalos de 95% de confiança.

A regressão logística tem sido usada amplamente em estudos epidemiológicos com desfechos binários para obter estimativas não enviesadas do OR ajustado para uma ou mais variáveis de confusão.

O modelo de regressão de Cox geralmente é usado para analisar o tempo até a ocorrência de um evento [2-3]. A regressão de Cox estima a razão de risco que expressa como a razão de risco depende das co-variáveis. O modelo pode ser expresso na forma

$$h(t) = h_0(t) \exp(\beta_1 z_1 + \dots + \beta_k z_k)$$

Onde,  $h_0(t)$  é a função de risco base do período,  $z_i$  são as co-variáveis e  $\beta_i$  são os coeficientes para as  $k$  co-variáveis. A função  $h_0(t)$  é tratada como uma função de perturbação, mas na verdade o modelo de Cox não a estima. Quando é atribuído um período de risco constante a todos os pacientes, a razão de risco estimada é igual à RP, porém a variância dos coeficientes tende a ser superestimada, que resulta em intervalos de confiança maiores. Para contornar essa situação, foi utilizado o método de variância robusta [3].

Os modelos de regressão de Poisson geralmente são utilizados em estudos longitudinais, onde o resultado é uma contagem de um desfecho ao longo do tempo [2]. O modelo pode ser expresso pela fórmula

$$\log\left(\frac{n}{t}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

Onde  $n$  é a contagem de eventos para cada paciente,  $t$  é o tempo de ocorrência,  $X_i$  são as co-variáveis, e  $\beta_i$  são os coeficientes para as  $k$  co-variáveis, que expressam o logaritmo do risco relativo. Da mesma forma que o modelo de Cox, se considerar o tempo constante para todos os pacientes, o modelo estima a RP, e o método de variância robusta pode contornar o problema de variância superestimada [3].

O modelo de regressão log-binomial é um modelo linear generalizado onde a função de ligação é o logaritmo da proporção em estudo e a distribuição do erro é binomial. Esse modelo pode ser expresso pela fórmula

$$\log(\pi) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

Onde  $\pi$  é a proporção de pessoas com o desfecho de interesse, e  $X_i$  são as co-variáveis. No entanto, pode ocorrer problema de convergência das estimativas. Geralmente esse problema é devido ao método de Newton-Raphson, usado para encontrar um mínimo ou um máximo dessa função, que pode não ser capaz de encontrar a estimativa de máxima verossimilhança quando a solução está nas margens do restrito intervalo do parâmetro [3]. Problemas de convergência no processo de estimação são comuns quando o modelo contém uma co-variável contínua, ou quando as co-variáveis apresentam mais de duas categorias, ou ainda quando a prevalência do desfecho é alta [2]. Petersen & Deddens (2003) sugeriram o método COPY, que expande os dados em  $c-1$  cópias das observações originais e 1 cópia dos dados originais com os valores dos desfechos invertidos, para contornar o problema de convergência [4].

Todas as análises foram realizadas com o auxílio do software *SAS System for Windows* (Statistical Analysis System), versão 9.1.3. SAS Institute Inc, (c) 2002-2003 by SAS Institute Inc., Cary, NC, USA. [5]

## Resultados

A prevalência de obstrução das VAS nas crianças avaliadas foi de 38,7%. Com exceção da co-variável uso de corticóide, todas as demais apresentam mais de duas categorias (tabela 1). Na análise univariada foi identificada associação significativa para o uso do *cuff*, uso de corticóide e tempo de VMI. Os OR estimados pela regressão logística múltipla apresentaram valores superestimados para uso do *cuff*, especialmente, para a categoria “Sim-desinsuflado” (RP = 3,10 (1,72:5,58) e OR = 91,22 (4,22:1000)), e também para tempo de VMI, para a categoria “acima de 48h” (RP = 3,06 (1,42:6,61) e OR = 8,37 (1,88:37,34)). A co-variável uso de corticóide não manteve a significância estatística no contexto múltiplo em nenhum dos modelos. As estimativas obtidas pelo modelo de regressão e Cox e de Poisson apresentaram resultados idênticos.

O modelo log-binomial não convergiu, porque a maioria das co-variáveis tinha mais de dois fatores. Foi aplicado o método o método COPY, através da macro disponível para o software SAS. A macro foi executada com mil, dez mil e cem mil cópias, e o modelo convergiu com mil cópias. As estimativas obtidas só apresentaram diferença à partir da terceira casa decimal para a maioria das co-variáveis. Os resultados apresentados na tabela 1 são referentes ao modelo executado com cem mil cópias.

De modo geral, as estimativas obtidas pelos modelos de Cox ou Poisson foram as que mais se aproximaram da RP estimadas pela estratificação de Mantel-Haenszel, porém os intervalos de confiança estimados pelo modelo log-binomial apresentaram menor amplitude. Os níveis descritivos obtidos no modelo log-binomial foram mais conservadores.

**Tabela 1:** Comparação da estratificação de Mantel-Haenszel (M-H) com os modelos de regressão logística, Cox / Poisson com variância robusta e log-binomial para a análise de prevalência de obstrução das VAS pós-extubação em 75 crianças internadas no CTI pediátrico.

Variáveis	Obstrução de VAS (n, %)		M-H (bruto)				Logística				Cox / Poisson				log-binomial			
	Sim	Não	RP	Inf	Sup	p	OR	Inf	Sup	p	RP	Inf	Sup	p	RP	Inf	Sup	p
<b>Idade</b>																		
Abaixo de 12 meses	11 45,8%	13 54,2%	ref	-	-	0,106	ref	-	-	0,452	ref	-	-	0,640	ref	-	-	0,382
12 a 23 meses	9 56,3%	7 43,8%	1,23	0,67	2,27		0,43	0,05	3,82		0,84	0,46	1,54		1,00	1,00	1,00	
24 meses a 7 anos	4 23,5%	13 76,5%	0,51	0,20	1,34		0,11	0,01	1,70		0,54	0,19	1,51		0,52	0,21	1,30	
7 anos ou superior	5 27,8%	13 72,2%	0,61	0,26	1,44		0,11	0,00	4,82		0,38	0,07	2,02		0,16	0,04	0,65	
<b>Tamanho da cânula</b>																		
3.0 - 3.5	9 47,4%	10 52,6%	ref	-	-	0,300	ref	-	-	0,663	ref	-	-	0,299	ref	-	-	0,173
4.0 - 4.5	14 42,4%	19 57,6%	0,90	0,48	1,66		0,43	0,07	2,62		0,63	0,35	1,15		0,66	0,38	1,17	
5.0 - 5.5	1 10,0%	9 90,0%	0,21	0,03	1,44		0,22	0,01	6,48		0,41	0,07	2,44		0,48	0,07	3,56	
>= 6	5 38,5%	8 61,5%	0,81	0,35	1,87		0,67	0,02	23,77		1,20	0,25	5,79		3,35	0,94	11,93	
<b>Cuff</b>																		
Não	9 29,0%	22 71,0%	ref	-	-	0,006	ref	-	-	0,016	ref	-	-	0,001	ref	-	-	0,048
SIM-desinsuflado	9 90,0%	1 10,0%	3,10	1,72	5,58		91,22	4,22	1000		3,30	1,71	6,38		2,09	1,22	3,60	
Sim-com pressão	11 32,4%	23 67,7%	1,11	0,54	2,32		5,99	0,67	53,68		1,85	0,94	3,66		2,09	1,22	3,60	
<b>Uso de corticóide</b>																		
Não	11 25,6%	32 74,4%	ref	-	-	0,009	ref	-	-	0,494	ref	-	-	0,178	ref	-	-	0,176
Sim	18 56,3%	14 43,8%	2,20	1,21	3,98		1,68	0,38	7,42		1,54	0,82	2,88		1,50	0,85	2,64	
<b>Tempo de VMI, horas</b>																		
01-24hs	6 20,0%	24 80,0%	ref	-	-	0,001	ref	-	-	0,006	ref	-	-	0,001	ref	-	-	0,005
25-48hs	4 28,6%	10 71,4%	1,43	0,48	4,27		0,66	0,08	5,34		0,92	0,32	2,62		1,47	0,50	4,27	
Acima de 48hs	19 61,3%	12 38,7%	3,06	1,42	6,61		8,37	1,88	37,34		2,54	1,16	5,54		2,60	1,29	5,23	

RP = razão de prevalência ; OR = odds ratio ; Inf Sup = limites inferior e superior para IC95% ; p = nível descritivo

## **Conclusões**

Embora os resultados obtidos pelos modelos tenham sido parecidos quanto ao nível descritivo, as estimativas apresentaram diferenças consideráveis. Os modelos de Cox/Poisson com variância robusta e o modelo log-binomial ofereceram estimativas mais consistentes comparados à regressão logística, considerando a amplitude dos intervalos de confiança. Baseados nos dados analisados, as estimativas obtidas pelos modelos de Cox/Poisson com variância robusta apresentaram estimativas próximas às obtidas pela estratificação de Mantel-Haenszel para as co-variáveis mais prevalentes, enquanto as estimativas obtidas pelo modelo log-binomial apresentaram estimativas próximas às obtidas pela estratificação de Mantel-Haenszel para as co-variáveis menos prevalentes. De modo geral, as estimativas obtidas pelo modelo log-binomial apresentaram intervalos de confiança de menor amplitude, e nível descritivo mais conservador. Esses resultados reforçam as conclusões de outros autores que defendem que os modelos de Cox/Poisson com variância robusta estimam melhor a RP para desfechos muito freqüentes, enquanto o modelo log-binomial apresenta melhor desempenho em casos onde a prevalência é intermediária [3-4]. Esse estudo permitiu ilustrar que os resultados podem diferir dependendo da escolha do método, por isso é importante analisar a natureza e a freqüência do evento e das exposições estudadas.

## **Referência Bibliográfica**

[1] – GORDIS, L. **Epidemiologia**. 2. ed. Rio de Janeiro. Revinter, 2004. 302p.

[2] – BARROS A; HIRAKATA V. **Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio**, *BMC Medical Research Methodology* 2003;3:21.

[3] – COUTINHO, L.M.S. ; SCAZUFCA, M.; MENEZES, P.R. **Métodos para estimar razão de prevalência em estudos de corte transversal**. *Rev. Saúde Pública*. 2008, vol.42, n.6, pp. 992-998.

[4] – PETERSEN, M.R.; DEDDENS, J.A. **A revised SAS macro for maximum likelihood estimation of prevalence ratios using the COPY method**. *Occup. Environ. Med.*, September 1, 2009; 66(9): 639 - 639.

[5] – SAS Institute Inc: **SAS/STAT® User's Guide, Version 9.1.3. 2002-2008**

[<http://support.sas.com/onlinedoc/913/docMainpage.jsp>]. Cary, NC: SAS Institute Inc