

Estudo Comparativo de Métodos para Estimação da Razão de Prevalência

Rodrigo Citton Padilha dos Reis - DEs, UFMG¹ ²

José Luiz Padilha da Silva - DEs, UFMG¹ ³

Enrico Antonio Colosimo - DEs, UFMG

Aluísio Jardim Dornellas de Barros - Dep. Med. Social, UFPel

Resumo: *Uma estimativa da razão de prevalência, ajustada para variáveis confundidoras, pode ser obtida do odds ratio do modelo de regressão logística. Para conjuntos de dados em que a resposta é comum (não rara), o odds ratio superestima esta estimativa. Assim, é interessante a escolha de modelos que estimem diretamente a razão de prevalência. Apresentamos alguns modelos e métodos (log-binomial, Poisson, Cox e Mantel-Haenszel) para estimação pontual e intervalar da razão de prevalência. Realizamos um breve estudo de simulação para comparar estes métodos. Os modelos log-binomial e de Poisson com variância robusta obtiveram resultados bastante satisfatórios.*

Palavras-chave: *modelo logístico, modelo log-binomial, modelo Poisson, odds ratio, razão de prevalência, variância robusta.*

1 Introdução

Em estudos de corte transversal com dados de saúde, quando a variável resposta é de natureza binária, a associação entre exposição e resposta é medida pela razão de prevalência (RP). Quando é necessário ajustar para potenciais variáveis de confusão, um modelo muito utilizado na prática é o de regressão logística. Este modelo produz estimativa do *odds ratio* (OR), frequentemente interpretado como uma estimativa da RP. A utilização do OR como aproximação para razão de prevalência é recomendada apenas quando a doença é “rara” (prevalência < 10%) (1). Como medida de associação não há nada de errado com o *odds ratio* (2), porém sua interpretação é menos intuitiva que a interpretação da RP, e quanto menos rara é a doença mais o OR superestima a RP. Diferentes abordagens foram propostas para converter o *odds ratio* na razão de prevalência (3,4). Contudo estes métodos se mostraram viesados, ou não podem

¹Agradecimento à FAPEMIG apoio financeiro.

²Contato: rodrigocpdosreis@gmail.com

³Contato: jlpadilha@yahoo.com.br

ser aplicados a conjuntos de dados com variáveis contínuas. É necessário, então, um modelo estatístico que estime a RP em vez do OR e ainda preserve as vantagens da regressão logística.

Recentemente, o estudo de modelos que estimam diretamente a razão de prevalência vem recebendo grande atenção na literatura epidemiológica (2, 5, 6). Um método para estimação da razão de prevalência, quando todas as covariáveis são categóricas, é o método de Mantel-Haenszel (7). Entretanto, na presença de covariáveis contínuas o método de Mantel-Haenszel não funciona. Wacholder (8) sugeriu a utilização de um modelo linear generalizado (MLG) (9) com distribuição binomial e função de ligação logarítmica. Lee e Chia (10) observaram que o modelo de Cox (11), geralmente usado na análise de dados de sobrevivência, utilizando a proposta de Breslow (12), em estudos transversais atribuindo tempos de falha iguais para todos indivíduos, estima a RP. Barros e Hirakata (2) ainda sugerem a utilização da regressão de Poisson para estimar a razão de prevalência. Este modelo tem resultados idênticos ao modelo de Cox citado anteriormente.

Neste trabalho realizamos um breve estudo de simulação para comparar os métodos citados anteriormente que visam estimar a razão de prevalência.

2 Metodologia

Para comparação dos diferentes métodos para estimar a razão de prevalência utilizamos os dados do artigo de McNutt *et al.* (5). Para isso, estimamos as razões de prevalência ajustadas pontualmente e por intervalo de confiança de 95% através do método de Mantel-Haenszel, dos modelos log-binomial, de Poisson (com e sem variância robusta) e Cox com variância robusta. Também utilizamos o modelo logístico, que estima o OR ajustado, e estimamos a RP bruta.

Utilizando o mesmo raciocínio de McNutt *et al.* (5), realizamos um estudo de simulação. Geramos amostras de tamanho 500, com 50% de expostos. A razão de prevalência verdadeira foi fixada em 2,00. No grupo de expostos ($E = 1$), o confundimento está presente ($C = 1$) em 60% dos indivíduos, e para estes a prevalência da doença é 0,80, enquanto que nos indivíduos que o confundimento está ausente ($C = 0$), a prevalência da doença é 0,40. No grupo de não-expostos ($E = 0$), o confundimento está presente em 40% dos indivíduos, e para estes a prevalência da doença é 0,40, e para os indivíduos com confundimento ausente, a prevalência da doença é 0,20. Geramos a variável resposta de distribuições binomiais conforme o esquema apresentado a seguir

$$(D_i|E_i = 1, C_i = 1) \sim Bin(m = 1, p = 0,8),$$

$$(D_i|E_i = 1, C_i = 0) \sim Bin(m = 1, p = 0,4),$$

$$(D_i|E_i = 0, C_i = 1) \sim Bin(m = 1, p = 0,4),$$

$$(D_i|E_i = 0, C_i = 0) \sim Bin(m = 1, p = 0,2),$$

Realizamos 1.000 replicações. Para cada réplica estimamos razão de prevalência ajustada e seu intervalo de confiança de 95% utilizando o método de Mantel-Haenszel, modelos log-binomial,

Poisson (sem e com variância robusta), Cox com variância robusta e logístico. Foi utilizada a linguagem R 2.9.0 (13).

3 Resultados

Os resultados das simulações são apresentados na Tabela 1. O viés observado (Viés) é a média das 1.000 estimativas pontuais menos a RP verdadeira. O erro quadrático médio observado (EQM) é o viés elevado a segunda potência mais a variância das 1.000 estimativas pontuais. Observou-se baixo viés para o método de Mantel-Haenszel, modelos log-binomial, Poisson e Cox. O EQM também foi baixo, com exceção do modelo logístico. As taxas de coberturas dos ICs de 95% estimados dos modelos log-binomial, de Poisson e Cox, estes dois últimos utilizando variância robusta, estão bem próximas ao nível de confiança nominal.

Tabela 1: Resultados das simulações para comparar as estimativas pontuais e intervalares da RP ajustada.

Método	Med. RP	Viés	EQM	Cob. IC 95%	Int. quant. 95%
Mantel-Haenszel	1,9891	0,0081	0,0499	99,4	1,62; 2,49
Log-binomial	1,9908	0,0087	0,0496	95,4	1,62; 2,49
Poisson (sem var. robusta)	1,9909	0,0081	0,0506	99,2	1,62; 2,49
Poisson (com var. robusta)	1,9909	0,0081	0,0506	95,4	1,62; 2,49
Cox (com var. robusta)	1,9909	0,0081	0,0506	95,4	1,62; 2,49
Logístico	4,0562	2,1439	5,3493	5,4	2,71; 6,07

O modelo logístico se mostrou inapropriado para a estimação da RP. Para este modelo foi observado um alto viés e conseqüentemente um alto EQM e baixa cobertura do IC de 95%.

Observamos que o modelo log-binomial apresentou problemas em relação ao algoritmo de estimação para algumas situações. Para alguns conjuntos de dados o algoritmo não converge, e o pacote estatístico solicita melhores valores iniciais para os parâmetros. Este é um problema conhecido na literatura (2,6,14) e não foi do objetivo deste trabalho tratar uma possível solução para este problema. Apenas fornecemos outros valores de entrada para o algoritmo, nos casos em que não houve convergência.

4 Conclusões

Nesse estudo comparamos modelos que estimam a razão de prevalência medindo associação entre exposição e uma resposta, na presença de uma variável de confundimento. Comparando os resultados com aqueles obtidos por McNutt *et al.* (5), encontramos valores ligeiramente diferentes para o método de Mantel-Haenszel nas estimativas intervalares. Para os modelos log-binomial e de Poisson (sem variância robusta) os resultados parecem idênticos aos encontrados por (5). Observamos o mesmo comportamento com as estimativas dos modelos de Poisson e Cox com variância robusta. Quanto ao modelo logístico as estimativas dos *odds ratio* ajustados e razão de prevalência bruta são idênticas as encontradas por McNutt *et al.* (5).

Como tem sido observado na literatura, o uso de *odds ratio* para estimar a razão de prevalência, apropriado para estudos quando a resposta é rara, pode levar a conclusões muito enganosas quando a resposta em questão é comum. Como vimos nos cenários considerados aqui, o modelo logístico não apresentou bons resultados. Como consequência de sua utilização nessas situações a superestimação da RP pode acarretar efeitos adversos no processo de decisão clínica.

Os métodos vistos aqui (log-binomial e Poisson com variância robusta, além do modelo de Cox com variância robusta) constituem boas alternativas para a estimação da razão e prevalência em estudos com resposta comum.

Referências

- 1 GREENLAND, S. Interpretation and choice of effect measures in epidemiologic measures of common outcomes and in case-control studies. *American Journal of Epidemiology*, v. 125, p. 761–768, 1987.
- 2 BARROS, M. V. L.; HIRAKATA, V. N. Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. *BMC Medical Research Methodology*, v. 3, p. 21, 2003.
- 3 ZHANG, J.; YU, K. F. What's relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *Journal of American Medical Association*, v. 280, p. 1690–1691, 1998.
- 4 GREENLAND, S. Model-based estimation of relative risks and other epidemiologic measures of common outcomes and in case-control studies. *American Journal of Epidemiology*, v. 160, p. 301–305, 2004.
- 5 MCNUTT, L. A.; WU, C.; XUE, X.; HAFNER, J. P. Estimating relative risk in cohort studies and clinical trials of common events. *American Journal of Epidemiology*, v. 157, p. 940–943, 2003.
- 6 PETERSEN, M. R.; DEDDENS, J. A. A comparison of two methods for estimating prevalence ratios. *BMC Medical Research Methodology*, v. 8, p. 9, 2008.
- 7 MANTEL, N.; HAENZEL, W. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, v. 22, p. 719–748, 1959.
- 8 WACHOLDER, S. Binomial regression in glim, estimating risk ratios and risk differences. *American Journal of Epidemiology*, v. 123, p. 174–184, 1986.
- 9 MCCULLOUGH, P.; NELDER, J. *Generalized Linear Models*. 2nd. ed. New York: Wiley, 1989.
- 10 LEE, J.; CHIA, K. S. Estimation of prevalence rate ratios for cross sectional data: an example in occupational epidemiology (letter). *British Journal of Industrial Medicine*, v. 50, p. 861–862, 1993.
- 11 COX, D. R. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, B*, v. 34, p. 187–220, 1972.
- 12 BRESLOW, N. Covariance analysis of censored survival data. *Biometrics*, v. 30, p. 89–99, 1974.

13 R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2009. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org>>.

14 BLIZZARD, L.; HOSMER, D. W. Parameter estimation and goodness-of-fit in log binomial regression. *Biometrical Journal*, v. 48, p. 5–22, 2006.