

Suggested citation for this article: Lukrafka JL, Zavascki AP, Barcellos N, Fuchs SC.

Determining risk factors for infection with influenza A (H5N1) [letter]. *Emerg Infect Dis.* 2007 Jun; [Epub ahead of print]

Determining Risk Factors for Infection with Influenza A (H5N1)

To the Editor: Novel antigenic subtypes of influenza viruses have been introduced periodically into the human population, resulting in large-scale global outbreaks (1). Highly pathogenic avian influenza (H5N1) viruses reemerged in 2003. Since then, they have reached endemic levels among poultry in several Southeast Asian countries, and across Asia, they have caused nearly 300 human infections, with a high rate of mortality (1,2). The results of many studies, including those for one recently conducted by Dinh et al. (3), have been published in an effort to identify the source(s) and modes of transmission of influenza A (H5N1) to humans and to guide the control and prevention of influenza infection.

Although new data regarding influenza A (H5N1) are urgently required, scientific rigor must be maintained during research and analysis to prevent misidentification of exposures as a risk factor for the disease and to prevent creation of iatrogenic panic among the exposed population and the scientific community (4). One point of scientific rigor that must be maintained is the use of adequate statistical analysis. The multivariate model in the study by Dinh et al. (3) was constructed by using a backward, stepwise variable selection strategy, in which variables with $p < 0.20$ were included in the initial model. However, such a strategy has resulted in a first model and subsequent steps with far more than 10 variables per outcome (e.g., 28 persons with avian flu), resulting in model overfitting (i.e., a statistical model that is too complex for the amount of data), which could result in imprecise estimates or spurious associations (5).

We believe that scientific methods must be meticulously applied when planning, executing, analyzing, and interpreting the results of influenza (H5N1) studies to prevent identification of false risk factors for acquiring infection.

Janice Luisa Lukrafka,* Alexandre Prehn Zavascki,* Nêmora Barcellos,*
and Sandra Costa Fuchs*

*Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

References

1. de Jong MD, Hien TT. Avian influenza A (H5N1). J Clin Virol. 2006;35:2–13. [Medline](#)
2. World Health Organization. Epidemic and Pandemic Alert and Response: Confirmed Human Cases of Avian Influenza A(H5N1) [cited 2007 Apr 23]. Available from http://www.who.int/csr/disease/avian_influenza/country/en/index.html
3. Dinh PN, Long HT, Tien NTK, Hien NT, Mai LTQ, Phong LH, et al. Risk factors for human infection with avian influenza A H5N1, Vietnam, 2004. Emerg Infect Dis. 2006;12:1841–7. [Medline](#)
4. Bonneux L, Van Damme W. An iatrogenic pandemic of panic. BMJ. 2006;332:786–8. [Medline](#)
5. Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. Ann Intern Med. 1993;118:201–10. [Medline](#)

Address for correspondence: Janice Luisa Lukrafka, Medical Sciences Postgraduate Program, Universidade Federal do Rio Grande do Sul, 2400 Ramiro Barcelos St, 90035-903 Porto Alegre, RS Brazil; email: jllukrafka@pop.com.br

Suggested citation for this article: Horby P. Determining risk factors for infection with influenza A (H5N1) [response]. Emerg Infect Dis. 2007 Jun; [Epub ahead of print]

In Response: Lukrafka et al. (1) warn against the dangers of overfitting a regression model when the number of outcomes is <10 per variable, “which could result in imprecise estimates or spurious associations.” This warning is valid, but it is equally important to consider the relative merits of multiple analysis options given the data available, the difficulties in collecting the data, and the objective of the study. The objective of our study (2) was to explore possible risk factors for human infection with influenza A (H5N1) rather than to test an explicit a priori hypothesis or to obtain precise estimates of risk. We were limited to a finite number of

cases, and had we slavishly followed criteria to avoid overfitting, we would not have run a regression model at all because we could have included only 2 variables, for which a stratified analysis would have been preferable. The regression model was run to confirm that the variables identified in the bivariate analysis retained their importance in the context of other variables; it was not intended to confirm or refute an a priori hypothesis, to be a predictive model, or to obtain precise and adjusted measures of risk. Despite the sample size limitations, we felt that looking at independence in a multivariable analysis was still valuable.

We explicitly acknowledge the limitations imposed by a small study size and were cautious in our interpretation, stating that the findings are the “basis for formulating new hypotheses.” The wide confidence intervals clearly indicate the low level of precision. The 3 variables in the final regression model were all statistically significant in bivariate analysis, and we do not believe they are spurious associations arising solely from an overfitted regression model.

Peter Horby*

*National Institute for Infectious and Tropical Diseases, Hanoi, Vietnam

References

1. Lukrafka JL, Zavascki AP, Barcellos N, Fuchs SC. Determining risk factors for infection with influenza A (H5N1) [letter]. *Emerg Infect Dis.* 2007 Jun; [Epub ahead of print]
2. Dinh PN, Long HT, Tien NTK, Hien NT, Mai LTQ, Phong LH, et al. Risk factors for human infection with avian influenza A H5N1, Vietnam, 2004. *Emerg Infect Dis.* 2006;12:1841–7. [Medline](#)

Address for correspondence: Peter Horby, National Institute for Infectious and Tropical Diseases, 78 Giai Phong St, Hanoi, Vietnam; email: peter.horby@gmail.com