References

1. Cox HS, Orozco JD, Male R, Ruesch-Gerdes S, Falzon D, Small I, et al. Multidrug-resistant tuberculosis in central Asia. Emerg Infect Dis. 2004;10:865–72.
2. Aerts A, Habouzit M, Mschiladze L, Malakmadze N, Sadradze N, Menteshashvili O, et al. Pulmonary tuberculosis in prisons of the ex-USSR state Georgia: results of a nationwide prevalence survey among sentenced inmates. Int J Tuberc Lung Dis. 2000;4:1104–10.
3. Pfyffer GE, Strassle A, van Gorkum T, Portaels F, Rigouts L, Mathieu C, et al. Multidrug-resistant tuberculosis in prison inmates, Azerbaijan. Emerg Infect Dis. 2001;7:855–61.
4. Balabanova Y, Fedorin I, Kuznetsov S, Graham C, Ruddy M, Atun R, et al. Antimicrobial prescribing patterns for respiratory diseases including tuberculosis in Russia: a possible role in drug resistance? J Antimicrob Chemother. 2004;54:673–9.
5. Pardini M, Varaine F, Iona E, Arzumanian E, Checchi F, Oggioni MR, et al. Cetylpyridinium chloride is useful for isolation of Mycobacterium tuberculosis from sputa subjected to long-term storage. J Clin Microbiol. 2005;43:442–4.
6. Pfyffer GE, Brown-Elliott BA, Wallace RJ, Jr. Mycobacterium: general characteristics, isolation and staining procedures. In: Murray PR, Baron EJ, Jorgensen JH, Pfaller MA, Yolken RH, editors. Manual of clinical microbiology, 8th edition. Washington: American Society for Microbiology; 2003. p. 532–59.
7. National Committee for Clinical Laboratory Standards. Susceptibility testing of mycobacteria, nocardia, and other aerobic actinomycetes; approved standards. Vol. 23, no. 18. M24-A. Wayne (PA): National Committee for Clinical Laboratory Standards; 2003.
8. Inderlied CB, Pfyffer GE. Susceptibility test methods: mycobacteria. In: Murray PR, Baron EJ, Jorgensen JH, Pfaller MA, Yolken RH, editors. Manual of clinical microbiology, 8th edition. Washington: American Society for Microbiology; 2003. p. 1149–77.
9. Inderlied CB, Salfinger M. Antimicrobial agents and susceptibility tests. In: Murray PR, Baron EJ, Pfaller MA, Tenover FC, Yolken RH, editors. Manual of clinical microbiology, 7th edition. Washington: American Society for Microbiology; 1999. p. 1601–23.
10. World Health Organization. Guidelines for establishing DOTS-plus pilot projects for the management of multidrug-resistant tuberculosis (MDR-TB) 2000. [cited 2005 Jan 18]. WHO/CDS/TB/2000.279. Available from http://www.who.int/gtb/publications/dotsplus/dotspluspilot-2000-279/english/contents/html

Address for correspondence: Lanfranco Fattorini, Dipartimento di Malattie Infettive, Parassitarie e Immunomediate, Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Rome, Italy; fax: 39-6-49387112; email: fattolan@iss.it

Endogeneity in Logistic Regression Models

To the Editor: Ethelberg et al. (1) report on a study of the determinants of hemolytic uremic syndrome resulting from Shiga toxin–producing *Escherichia coli*. The dataset is relatively small, and the authors use stepwise logistic regression models to detect small differences. This indicates that the authors were aware of the limitations of the statistical power of the study. Despite this, the study has an analytic flaw that seriously reduces the statistical power of the study.

An often overlooked problem in building statistical models is that of endogeneity, a term arising from econometric analysis, in which the value of one independent variable is dependent on the value of other predictor variables. Because of this endogeneity, significant correlation can exist between the unobserved factors contributing to both the endogenous independent variable and the dependent variable, which results in biased estimators (incorrect regression coefficients) (2). Additionally, the correlation between the dependent variables can create significant multicollinearity, which violates the assumptions of standard regression models and results in inefficient estimators. This problem is shown by model-generated coefficient standard errors that are larger than true standard errors, which biases the interpretation towards the null hypothesis and increases the like-lihood of a type II error. As a result, the power of the test of significance for an independent variable X_1 is reduced by a factor of $(1-r^2_{(1|2,3,....)})$, where $r_{(1|2,3,....)}$ is defined as the multiple correlation coefficient for the model $X_1 = f(X_2,X_3,…)$, and all X_i are independent variables in the larger model (3,4).

The results of this study clearly show that the presence of bloody diarrhea is an endogenous variable in the model showing predictors of hemolytic uremic syndrome, in that the diarrhea is shown to be predicted by, and therefore strongly correlated with, several other variables used to predict hemolytic uremic syndrome. Similarly, Shiga toxin 1 and 2 (*stx1, stx2*) genes are expected to be key predictors of the presence of bloody diarrhea, independent of strain, due to the known biochemical effects of that toxin (5,6). Because the strain is in part determined by the presence of these toxins, including both strain and genotype in the model means that the standard errors for variables for the Shiga-containing strains and bloody diarrhea symptom are likely to be too high, and hence the significance levels (p values) obtained from the regression models are higher than the true probability because of a type I error.

This flaw is a particular problem with studies that use a conditional stepwise technique for including or excluding variables. The authors note that they excluded variables from the final model if the significance in initial models for those variables was less than an α level (p value) of 0.05. Given the inefficiencies due to the endogeneity of bloody diarrhea, as well as those that may result from other collinearities significant predictors were likely excluded from the study, although this cannot be confirmed from the data presented.

The problems associated with the endogeneity of bloody diarrhea can be overcome by a number of approaches.

For example, the simultaneous equations approach, such as that outlined by Greene (7), would have used predicted values of bloody diarrhea from the first stage of the model as instrumental variables for the actual value in the model for hemolytic uremic syndrome. Structural equations approaches, such as those suggested by Greenland (8), would also be appropriate. However, bloody diarrhea is not the only endogenous variable in their models, and extensive modeling would be necessary to isolate the independent effects of the various predictor variables. Given the small sample size, this may not be possible.

The underlying problem in the study is the theoretical specifications for the model, in which genotypes, strains, and symptoms are mixed, despite reasonable expectations that differences in 1 level may predict differences in another. For example, the authors' data demonstrate that all O157 strains contain the stx_2 gene and have higher rates of causing hemolytic uremic syndrome and bloody diarrhea. This calls into question the decision to build an analytic model combining 3 distinct levels of analysis. Such a model depends on the independence of the variables to gain unbiased, efficient estimators. The model of the relationships one would develop from a theoretical perspective would predict the opposite (Figure). We expect that the genotypes (by definition) will predict the strain, and that strains have a differential effect on symptoms. The high level of inter-variable correlation due to these relationships, coupled with the decision to exclude variables based on likely inefficient p values, raises questions concerning the reliability of the results and conclusions. In particular, the conclusions that strains O157 and O111 are not predictors of hemolytic uremic syndrome deserve to be revisited; other excluded variables may also be significant predictors when considered under an appropriate model. These problems point to the need to ensure proper specification of analytic models and to demonstrate due regard for the underlying assumptions of statistical models used.

George Avery*

*University of Minnesota, Duluth, Minnesota, USA

References

1. Ethelberg S, Olson KEP, Schuetz F, Jensen C, Schiellerup P, Engberg J, et al. Virulence factors for hemolytic uremic syndrome, Denmark. Emerg Infect Dis. 2004;10:842–7.
2. Dowd B, Town R. Does X really cause Y? Washington: Academy Health; 2002.
3. Hsieh F, Bloch D, Larsen M. A simple method of sample size calculation for linear and logistic regression. Stat Med. 1998;17:1623–34.
4. Menard S. Applied logistic regression analysis, 2nd ed. Thousand Oaks (CA): Sage Publications: 2002. p. 75–8.
5. Blackall DP, Marques MB. Hemolytic uremic syndrome revisited: Shiga toxin, factor H, and fibrin generation. Am J Clin Pathol. 2004;121 (Suppl):S81–8.
6. Harrison LM, van Haaften WC, Tesh VL. Regulation of proinflammatory cytokine expression by Shiga toxin 1 and/or lipopolysaccharides in the human monocytic cell line THP-1. Infect Immun. 2004;72:2618–27.
7. Greene W. Gender economics courses in liberal arts colleges: further results. J Econ Ed. 1998;29:291–300.
8. Greenland S, Brumback B. An overview of relations among causal modelling methods. Int J Epidemiol. 2002;31:1030–7.

Address for correspondence: George Avery, 1207 Ordean Ct., BohH 320, University of Minnesota Duluth, Duluth, MN 55812, USA; fax: 218-726-7186; email: aver0042@umn.edu
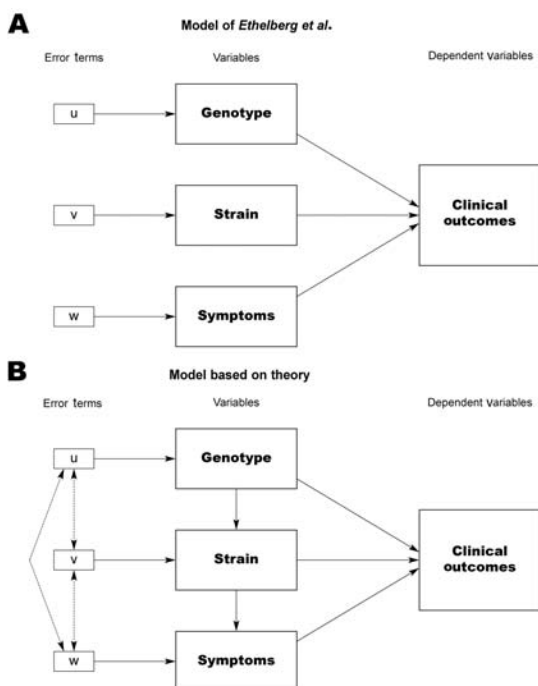
Figure. Model for determining virulence factors for hemolytic uremic syndrome

In response: We appreciate Avery's interest (1) in our article (2), although we believe the critique of the methods is largely based on misunderstandings. We developed a model for the risk of progression to hemolytic uremic syndrome (HUS) containing 3 variables: whether the infecting Shiga toxin–producing *Escherichia coli* isolate had the stx_2 gene, age of the patient, and occurrence of bloody diarrhea. The critique relates to the fact that bloody diarrhea and stx_2 are not independent, since we showed that stx_2 was strongly associated with progression to HUS (odds ratio [OR] =

18.9) and also weakly associated with development of bloody diarrhea (OR = 2.5) (2). Avery uses the term endogeneity as it is used in econometric analyses; however, the term "intermediary variable," i.e., a factor in the causal pathway leading from exposure to disease, is more frequently used in epidemiology. In this context, we chose to consider bloody diarrhea as a potential confounder (3). A confounder is a risk factor but is also independently associated with the exposure variable of interest and is not regarded as part of the causal pathway (see online Figure at http://www.cdc. gov/ ncidod/EID/vol 11no03/05-0071-G.htm). Bloody diarrhea may act as a confounder if patients with bloody stools are treated differently by the examining physicians or if, for instance, unknown virulence factors contribute to the risk of having bloody stools.

A second line of critique of our methods apparently develops from the idea that virulence factors determine the serogroup. This idea, however, is a biological misconception. In fact, virulence genes and serogroup are independent at the genetic level, and an important point of our article is that HUS is determined by the virulence gene composition of the strain rather than the serogroup.

Regardless of the status of the bloody diarrhea variable, excluding it from the model doesn't change the conclusions of the article. A revised model contains only the significant variables age and stx_2 (Table). Serotype O157 is still not an independent predictor of HUS, and this result is robust.

Steen Ethelberg* and Kåre Mølbak*
*Statens Serum Institut, Copenhagen, Denmark

References

1. Avery G. Endogeneity in logistic regression models. Emerg Infect Dis. 2005;11: 499–500..
2. Ethelberg S, Olsen KE, Scheutz F, Jensen C, Schiellerup P, Enberg J, et al. Virulence factors for hemolytic uremic syndrome, Denmark. Emerg Infect Dis. 2004;10: 842–7.
3. Griffin PM, Mead PS, Sivapalasingam S. *Escherichia coli* O157:H7 and other enterohaemorrhagic *E. coli*. In: Blaser MJ, Smith PD, Ravdin JI, Greenberg HB, Guerrant RL, editors. Infections of the gastrointestinal tract. Philadelphia: Lippincott Williams & Wilkins; 2002. p. 627–42.

Address for correspondence: Steen Ethelberg, Department of Bacteriology, Mycology and Parasitology, Statens Serum Institut, Artillerivej 5, DK-2300 Copenhagen S, Denmark; fax: 45-3268-8238; email: set@ssi.dk

Rectal Lymphogranuloma Venereum, France

To the Editor: Lymphogranuloma venereum (LGV), a sexually transmitted disease (STD) caused by *Chlamydia trachomatis* serovars L1, L2, or L3, is prevalent in tropical areas but occurs sporadically in the western world, where most cases are imported (1). LVG commonly causes inflammation and swelling of the inguinal lymph nodes, but it can also involve the rectum and cause acute proctitis, particularly among men who have sex with men. However, LGV serovars of *C. trachomatis* remain a rare cause of acute proctitis, which is most frequently caused by *Neisseria gonorrhoeae* or by non-LGV *C. trachomatis* (2). In 1981, in a group of 96 men who have sex with men with symptoms suggestive of proctitis in the United States, Quinn et al. found that 3 of 14 *C. trachomatis* infections were caused by LGV serovar L2 (3). In France, 2 cases of rectal LGV were reported in an STD clinic in Paris from 1981 to 1986 (4). In 2003, an outbreak of 15 rectal LGV cases was reported among men who have sex with men in Rotterdam; 13 were HIV-infected, and all reported unprotected sex in neighboring countries, including Belgium, France, and the United Kingdom (5). At the same time, a rise in *C. trachomatis* proctitis (diagnosed by using polymerase chain reaction [PCR]; [Cobas Amplicor Roche Diagnostic System, Meylan, France]) was detected in 3 laboratories in Paris and in the *C. trachomatis* national reference center located in Bordeaux. To identify the serovars of these *C. trachomatis* spp., all stored rectal specimens were analyzed by using a nested *omp*1 PCR-restriction fragment length polymorphism assay. The amplified DNA product was digested by restriction enzymes. Analysis of digested DNA was performed by elec-

Table. Risk factors for HUS among 343 STEC patients, Denmark 1997–2003, comparison of models with and without bloody diarrhea as a variable*

Determinant	No. of patients	No. (%) with HUS	Original model, OR (95% CI)	New model, OR (95% CI)
eae				
Negative	111	0 (0.0)		
Positive	232	21 (9.1)	NI	NI
stx_2				
Negative	159	1 (0.6)	1	1
Positive	184	20 (10.9)	18.9 (2.4–146)	24.6 (3.2–187)
Age				
≥8 y	178	3 (1.7)	1	1
≤7 y	165	18 (10.9)	11.4 (3.2–41.3)	9.7 (2.7–34.1)
Bloody diarrhea				
No	218	6 (2.8)		
Yes	125	15 (12.0)	4.5 (1.6–12.7)	EX
O157				
No	262	10 (3.8)		
Yes	81	11 (13.6)	NS	NS

*HUS, hemolytic uremic syndrome; STEC, Shiga toxin–producing *Escherichia coli*; OR, odds ratio; CI, confidence interval; NI, not included (test not appropriate); NS, not significant; EX, excluded from model.