• • • • • • • • •

In their editorial, Spiegelman and Hertzmark (1) recommend an easy method to estimate risk and prevalence ratios, and they include SAS macros for performing the calculations. The method uses maximum likelihood when the correct binomial model converges and a Poisson model with a robust variance estimator when the correct model fails to converge. We agree completely with using maximum likelihood estimators (MLEs) when the model converges. However, one can do better than the Poisson model when the correct model fails to converge.

There are two deficiencies in the Poisson approximation. First, as the authors (1) mention, the estimates are not as efficient as those for the MLEs of the log-binomial model. Second, a point that the authors do not mention, estimates of probabilities obtained from the Poisson model can exceed 1 because the wrong model is being fit (2). In 2003, Deddens et al. (3) proposed using maximum likelihood even when the correct model fails to converge. Because of difficulties in computing MLEs when the model fails to converge, they proposed obtaining MLEs on a modified data set such that these MLEs could be made arbitrarily close to the MLEs for the original data set. Doing so retains the efficiency properties of maximum likelihood as well as assures that estimated probabilities will be between 0 and 1.

The reasons why the model fails to converge, the reasons why this method solves the convergence problems, and the accuracy of this method have been discussed in detail previously (2, 3). The method involves combining $c - 1$ copies of the original data set with one copy of the original data set with all values of the dependent variable $Y = 0,1$ interchanged. For the expanded data set, the solution should be very close to, but not exactly on, the boundary of the parameter space. The solution can be made as close to the original data set maximum likelihood solution as desired by increasing c. The standard error is then adjusted for the increase in sample size by multiplying by the square root of c. Simulations show that the method works well when $c = 1{,}000$. As a practical measure for large data sets, one should restrict the input data set to contain only the variables needed in the model.

Thus, the method that Deddens et al. (3) proposed is to fit the log-binomial model and use the results when the model converges. When the model does not converge, fit the log-binomial model on the modified data set. An SAS macro (3) was made available to other researchers (http://www.cdc.gov/niosh/ext-supp-mat/pr-sasmac). When we use the method

TABLE 1. Results for the Greenland data (4)

Method	Receptor	Stage2	Stage3
MLE* (original data)			
Estimate	1.5583	2.5382	5.8680
95% CI*	1.0487, 2.3155	1.1734, 5.4903	2.7458, 12.5406
Poisson approximation			
Estimate	1.6308	2.5207	5.9134
95% CI	1.0745, 2.4751	1.1663, 5.4479	2.7777, 17.5890
MLE (modified data), $c = 1{,}000$			
Estimate	1.5567	2.5235	5.8237
95% CI	1.0479, 2.3126	1.1699, 5.4432	2.7326, 12.4117
MLE (modified data), $c = 10{,}000$			
Estimate	1.5582	2.5367	5.8636
95% CI	1.0487, 2.3152	1.1730, 5.4855	2.7445, 12.5276

* MLE, maximum likelihood estimator; CI, confidence interval.

on Greenland's data (4), we naturally get the same answer as Spiegelman and Hertzmark (1) did, because the log-binomial model converges. (That the model converges is not surprising. Skov et al. (5) performed numerous simulations with categorical independent variables and had no convergence problems. With a continuous independent variable, however, the lack of convergence is much more common.) If we force SAS to fit the log-binomial model with $c = 1{,}000$ on the modified data set, as Spiegelman and Hertzmark did for the Poisson method on the original data set, the multivariate-adjusted risk ratios are 1.5567, 2.5235, and 5.8237 for receptor, stage2, and stage3, respectively (table 1). This is just one data set, but the results agree with those of many simulations that we have performed: the MLEs on the modified data set are closer to the MLEs on the original data set than are the Poisson-based estimates. As mentioned earlier, one can make them even closer by increasing c. For example, for $c = 10{,}000$, the estimated multivariate-adjusted risk ratios and 95 percent confidence intervals from the modified data set are almost the same as the maximum likelihood estimates and 95 percent confidence intervals for the original data set (table 1).

ACKNOWLEDGMENTS

Conflict of interest: none declared.

REFERENCES

1. Spiegelman D, Hertzmark E. Easy SAS calculations for risk or prevalence ratios and differences. Am J Epidemiol 2005; 162:199–200.
2. Deddens JA, Petersen MR. Re: ''Estimating the relative risk in cohort studies and clinical trials of common outcomes.'' (Letter). Am J Epidemiol 2004;159:213–14.
3. Deddens JA, Petersen MR, Lei X. Estimation of prevalence ratios when PROC GENMOD does not converge. Proceedings of the 28th Annual SAS Users Group International Conference, Seattle, Washington, March 30–April 2, 2003. (Paper 270-28). (http://www2.sas.com/proceedings/sugi28/270-28.pdf).
4. Greenland S. Model-based estimation of relative risks and other epidemiologic measures in studies of common outcomes and in case-control studies. Am J Epidemiol 2004;160: 301–5.
5. Skov T, Deddens J, Petersen MR, et al. Prevalence proportion ratios: estimation and hypothesis testing. Int J Epidemiol 1998;27:91–5.

Martin R. Petersen[1] and James A. Deddens[1,2] (e-mail: mrp1@cdc.gov)
[1] *National Institute for Occupational Safety and Health, Cincinnati, OH 45226-1998*
[2] *Department of Mathematical Sciences, University of Cincinnati, Cincinnati, OH 45215*

DOI: 10.1093/aje/kwj162; Advance Access publication May 17, 2006