INVITED EDITORIAL NOTE

Easy SAS Calculations for Risk or Prevalence Ratios and Differences

We would like to make the readership aware that risk or prevalence ratios and differences, when they are the parameter of interest, can be directly calculated by using SAS software (SAS Institute, Inc., Cary, North Carolina). There is no longer any good justification for fitting logistic regression models and estimating odds ratios when the odds ratio is not a good approximation of the risk or prevalence ratio. Instead, SAS PROC GENMOD's log-binomial regression (1) capability can be used for estimation and inference about the parameter of interest. Here is an example of the code required to analyze the breast cancer survival data discussed by Greenland (2):

```
proc genmod descending;
model death=receptor stage2 stage3/dist=bin link=log;
estimate 'RR receptor low vs. high' receptor 1/exp;
estimate 'RR stage2 vs stage1' stage2 1/exp;
estimate 'RR stage 3 vs stage1' stage3 1/exp;
```

from which the multivariate-adjusted risk ratios are 1.5583 (95 percent confidence interval: 1.0487, 2.3155), 2.5382 (95 percent confidence interval: 1.1734, 5.4903), and 5.8680 (95 percent confidence interval: 2.7458, 12.5406) for receptor, stage2, and stage3, respectively. The results from the SAS output are given without rounding to allow replication by the reader.

There are times when the log-binomial model fails to converge. It is well known that the log-binomial model is less numerically stable than the logistic model. When this is the case, the analyst may use SAS PROC GENMOD's Poisson regression capability with the robust variance (3, 4), as follows:

```
proc genmod;
class id;
model death=receptor stage2 stage3/dist=poisson link=log;
repeated subject=id/type=ind;
estimate 'RR receptor low vs. high' receptor 1/exp;
estimate 'RR stage2 vs stage1' stage2 1/exp;
estimate 'RR stage 3 vs stage1' stage3 1/exp;
```

from which the multivariate-adjusted risk ratios are 1.6308 (95 percent confidence interval: 1.0745, 2.4751), 2.5207 (95 percent confidence interval: 1.1663, 5.4479), and 5.9134 (95 percent confidence interval: 2.7777, 17.5890) for receptor, stage2, and stage3, respectively. Note that, *on average*, the modified Poisson estimates are valid but not fully efficient when compared with these log-binomial maximum likelihood estimators. In this particular example, the theoretical efficiency of the log-binomial maximum likelihood estimates is clearly evident.

By replacing `link=log` with `link=identity` in the MODEL statement, multivariate-adjusted risk (prevalence) differences are obtained as follows:

```
proc genmod descending;
model death=receptor stage2 stage3/dist=bin link=identity;
```

from which the multivariate-adjusted risk differences are 0.1613 (95 percent confidence interval: 0.0069, 0.3158), 0.1492 (95 percent confidence interval: 0.0367, 0.2618), and 0.5723 (95 percent confidence interval: 0.3842, 0.7604) for receptor, stage2, and stage3, respectively. If this binomial model for the risk difference fails to converge, the modified Poisson approach can be used as above, again replacing `link=log` with `link=identity`:

Am J Epidemiol 2005;162:199–200

```
proc genmod;
class id;
model death=receptor stage2 stage3/dist=poisson link=identity;
repeated subject=id/type=ind;
```

As noted previously, these modified Poisson risk differences will be valid, but they tend to be less efficient than their binomial maximum-likelihood-based counterparts.

A well-documented, user-friendly SAS macro, %RELRISK8, has been developed that automates this computational and analytic approach. The modified Poisson estimates are used to start the iterations to obtain the log-binomial maximum likelihood estimates. These are the final estimates if convergence of the binomial likelihood is not obtained. The macro can be downloaded from the first author's website (http://www.hsph.harvard.edu/faculty/spiegelman/relrisk8.html).

REFERENCES

1. Wacholder S. Binomial regression in GLIM: estimating risk ratios and risk differences. Am J Epidemiol 1986;123: 174–84.
2. 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.
3. Huber PJ. The behavior of maximum likelihood estimates under non-standard conditions. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Vol 1. Berkeley, CA: University of California Press, 1967:221–33.
4. Zou G. A modified Poisson regression approach to prospective studies with binary data. Am J Epidemiol 2004;159:702–6.

Donna Spiegelman
Editor, American Journal of Epidemiology
Departments of Epidemiology and Biostatistics
Harvard School of Public Health
Boston, MA 02115

Ellen Hertzmark
Department of Epidemiology
Harvard School of Public Health
Boston, MA 02115