Approaches for estimating prevalence ratios

J A Deddens,[1,2] M R Petersen[1]

[1] National Institute for Occupational Safety and Health, Cincinnati, Ohio, USA;
[2] Department of Mathematical Sciences, University of Cincinnati, Cincinnati, Ohio, USA

Correspondence to:
James A Deddens, National Institute for Occupational Safety and Health, Mail Stop R15, 4676 Columbia Parkway, Cincinnati, OH 45226, USA; jad0@cdc.gov

Recently there has been much interest in estimating the prevalence (risk, proportion or probability) ratio instead of the odds ratio, especially in occupational health studies involving common outcomes (for example, with prevalence rates above 10%). For example, if 80 out of 100 exposed subjects have a particular disease and 50 out of 100 non-exposed subjects have the disease, then the odds ratio (OR) is (80/20)/(50/50) = 4. However, the prevalence ratio (PR) is (80/100)/(50/100) = 1.6. The latter indicates that the exposed subjects are only 1.6 times as likely to have the disease as the non-exposed subjects, and this is the number in which most people would be interested. There is considerable literature on the advantages and disadvantages of OR versus PR (see Greenland,[1] Stromberg,[2] Axelson *et al*[3] and others). In this article we will review the existing methods and give examples and recommendations on how to estimate the PR.

The most common method of modelling binomial (no/yes or 0/1) health outcomes today is logistic regression. In logistic regression one models the probability of the binomial outcome (Y = 1) of interest as:

$$P(Y = 1 | X_1, X_2, \ldots, X_k) = e^{X\beta}/(1+e^{X\beta})$$

where $X\beta = \beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_kX_k$. Then $\exp(\beta_1)$ = OR for a 1 unit increase in X_1 adjusted for all other variables in the model. Logistic regression yields maximum likelihood estimates (MLEs) of the OR (adjusted for other covariates). If the adjusted OR is the parameter of interest, then these MLEs are generally considered the best estimators available. The adjusted OR can also be used to estimate the adjusted PR, but this should only be done for a rare disease (eg, one with a prevalence of 10% or less). This, together with the fact that logistic regression is currently available in most standard statistical software packages has led to its wide popularity. However, there have not been any good alternatives for estimating the adjusted PR if the outcome is not rare.

Historically if investigators wanted to estimate the adjusted risk ratio they used the Mantel-Haenzsel procedure which works as long as the covariates X_j are all categorical. However, if some of the covariates are continuous the Mantel-Haenzsel procedure does not work, so researchers sought other methods. Wacholder[4] was the first to suggest using generalised linear models with the binomial distribution and the log link via the software package GLIM. Schouten *et al*[5] suggested modifying the data in such a way that the OR from logistic analysis for the modified data is an estimate of the PR for the original data. Lee and others[6 7] recommended using the Cox proportional hazard model to estimate the PR. This method yields partial likelihood estimates of linear model coefficients except for the intercept, which is not estimated. Zocchetti *et al*[8] in a letter to the editor suggested using generalised linear models with the log link. Skov *et al*[9] introduced the term "log-binomial". Skov *et al*,[9] Deddens *et al*[10] and Barros and Hirakata[11] showed that the Cox method yields estimated standard deviations which are too large, which leads to low power for Wald tests. Barros and Hirakata[11] observed that the Cox model and Poisson regression estimates are identical.

LOG BINOMIAL MODEL

Wacholder,[4] Zocchetti *et al*[8] and Skov *et al*[9] recommended using the log-binomial model, which directly models the PR. If for each combination of independent variables, the dependent variable has a binomial distribution with the logarithm of its probability being linearly related to the independent variables, then the log-binomial is the correct model, and MLEs of the parameters and PR can be directly obtained. In the log-binomial, one models the probability of the binomial outcome (Y = 1) of interest as:

$$P(Y = 1 | X_1, X_2, \ldots, X_k) = e^{X\beta}$$

where $X\beta = \beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_kX_k$. Then $\exp(\beta_1)$ = PR for a 1 unit increase in X_1, adjusted for the other covariates. Since $P(Y = 1 | X_1, X_2, \ldots, X_k)$ is a probability and hence must be between 0 and 1, the log-binomial model imposes a restriction on the set of parameters β, namely that $X\beta \leqslant 0$. By definition, substituting the MLE for β will satisfy this restriction for all X in the data set. For many situations (especially with quantitative covariates), the MLE will be on the boundary of this restricted parameter space. When using software packages, the model might fail to converge either because of bad starting values, for example, starting values not in the restricted parameter space, which can be easily fixed by specifying better starting values, or because the maximum likelihood solution occurs on the boundary of the parameter space. In the latter situation, the derivative of the likelihood at its maximum may not be 0. Thus standard software packages which maximise the likelihood by finding the point at which the derivative is equal to 0 (namely Newton's method) may not find the solution. We should mention that all popular software packages (eg, SAS, STATA, etc)

use some form of Newton's method for finding the MLE. R and S-plus have more sophisticated methods for solving constrained maximum likelihood estimation.

Deddens *et al*[10] developed a straightforward method to modify the data set and to use the same software that had failed on the original data set, in order to get approximate MLEs. In addition, they supplied macros for SAS users to easily obtain the approximate estimates, standard errors and log-likelihoods. This method, called the COPY method, involves using the log-binomial model when it converges, and, when the log-binomial does not converge, using MLEs from a new expanded data set that contains c−1 copies of the original data and one copy of the original data with the dependent variable values interchanged (1's changed to 0's and 0's changed to 1's). For any finite c, the solution is no longer on the boundary, and, as c becomes large, the MLE estimates for this modified data set approach the MLEs for the original data set. The larger c is, the better the approximation, but the slower the macro executes. The number c should be at least 100. In this paper, c = 1000 was used. In order to obtain the correct approximate standard error, one simply multiples the standard error from the expanded data set by the square root of c. In order to obtain the approximate correct log-likelihood one simply divides the log-likelihood from the expanded data set by c, in order to account for the fact that the expanded data set contains c times as many observations as the original data set. Lumley *et al*[12] pointed out that this is equivalent to creating a new data set which consists of one copy of the original data set (with weight 999) and one copy of the original data set with the dependent variable values interchanged (with weight 1), and then performing a weighted log-binomial regression. In fact, use of the weights 0.999 and 0.001 eliminates the need to adjust the standard error and likelihood. This weighted log-binomial approach has the advantage that it can also be used to compute likelihood ratio confidence intervals. Lumley *et al*[12] recommend using non-linear least squares when the log-binomial does not converge, since the MLE might be sensitive to outliers and model misspecification.

ROBUST POISSON MODEL

One can think of a sample of binomial data (0 or 1) as being approximately Poisson, where the probability of a value of 2 or greater is low enough that no values greater than 1 occurred in the obtained sample. If the logarithm of the Poisson parameter is assumed to be linearly related to a set of independent variables, then the exponentiation of any coefficient of the model will yield an estimate of a ratio of Poisson parameters, adjusted for the other covariates. Because the observed data consist of only 0's and 1's, this ratio can be used as an approximation to the adjusted PR. The estimates thus obtained are identical to those obtained from Cox proportional hazard regression, if one assumes equal follow-up times for all subjects and handles ties properly. Thus, Poisson regression has standard errors which are too large. Poisson regression will be discussed in this paper rather than Cox proportional hazard regression because Poisson regression also estimates an intercept.

Barros and Hirakata[11] used robust variance estimation in STATA to solve the large variance problem for Poisson regression. They concluded that the best adjustment was to use a sandwich estimator of the variance. They compared this robust method to the Mantel-Haenszel method in situations where all covariates were categorical. One can obtain the non-intercept robust solution using PROC PHREG in SAS with the COVS option. Independently, Zou[13] showed how to use PROC GENMOD with the REPEATED option in SAS to obtain the robust Poisson estimates.

METHODS BASED ON ODDS RATIOS

Thompson *et al*[14] discussed direct estimation of the PR from the OR, weighted averages of these estimated PRs over strata, and the stratified Mantel-Haenszel estimate. They recommended using either the proportional hazards (ie, Poisson) or the log-binomial method, even though, at that time, the former was the non-robust version.

Zhang and Yu[15] suggested another method to directly convert an OR into a PR. McNutt *et al*[16] showed this method is fairly biased when adjusted for other covariates.

The Schouten *et al* method,[5] mentioned earlier, uses logistic software to estimate the PR. Skov *et al*[9] showed that the method was inferior to the log-binomial method, primarily because it could yield estimates of probability which were greater than 1 for points in the data set.

VALID AND INVALID ESTIMATES OF PROBABILITIES

Unlike logistic regression, the log-binomial and robust Poisson methods can yield estimated probabilities greater than 1 for covariate values not in the (convex hull of the) data used to estimate the model. However, despite what one might find in the literature, MLEs for log-binomial models cannot yield estimates of probabilities outside the range of 0–1 for covariate values in the (convex hull of the) data used to fit the model. Thus the log-binomial method, with or without the COPY method approximation, cannot yield estimates of probabilities outside this range. It is known[17,18] that the robust Poisson can yield such invalid probability estimates for covariate values in the original data file, especially when the log-binomial model does not converge. Although the Cox (or robust Cox) method does not estimate an intercept, and hence does not formally estimate probabilities, it also can be considered to yield estimated probabilities greater than 1 because its solution is identical to the Poisson solution.

In order to estimate the PR from a generalised linear model, the log link is assumed to be the correct link function. For each combination of independent variables, the distribution of the 0/1

Table 1 Data for example 1

	Males		Females	
Treatment	Same	Better	Same	Better
Active	16	12	12	16
Placebo	19	7	21	5

dependent variable is binomial. Both the log-binomial method and the Poisson method are generalised linear models with a log link function. However, the Poisson model treats this distribution as Poisson (which can take on integer values greater than 1), while the log-binomial model correctly treats it as binomial.

EXAMPLE 1

Example 1 is from the PROC FREQ documentation in SAS and involves a clinical trial for treatment of migraine headaches with 54 male and 54 female patients, with an active and placebo treatment. Using a log-binomial model to estimate the PR of treatment to placebo, a binomial variable for migraine headache status (better = 1, same = 0) was modelled on treatment adjusted for gender (table 1). Notice that the overall prevalence is 37% (40/108). Since this is not a rare outcome, one should be estimating the PR instead of the OR. The results are shown in table 2. Due to the high (37%) prevalence the estimated PR is quite different than the estimated OR. For this example, there is very little difference between the M-H, log-binomial, and robust Poisson estimates for the PR. The corresponding 95% confidence intervals are also quite similar. Log-binomial gives a slightly larger estimate of PR, while M-H and robust Poisson give virtually identical estimates of PR. We consider the log-binomial estimate to be the "best" estimate since it maximises the correct binomial likelihood. Because there were no estimated probabilities greater than 1 for the robust Poisson for the data in this example, values of the log-likelihood can be compared. The binomial log likelihood is equal to −66.7020 for the log-binomial solution, which is slightly larger than the value of −66.7262 for the robust Poisson solution.

EXAMPLE 2

Example 2 comes from a book on logistic regression by Paul Allison but it is also available online (http://ftp.sas.com/samples/A55770). These data are from a study relating death penalty (death = 1, life in prison = 0) to defendant race (two levels), victim race (two levels), crime seriousness (quantitative scale) and culpability (quantitative scale) (table 3). Although this was used by Allison to illustrate logistic analysis, the prevalence of death was 0.34, so ORs would not be good estimates of PRs. This example will illustrate another issue with using the log-binomial model to estimate PRs instead of ORs. The quantitative variable culpability is linear in the log-odds but not in the log-probability. Thus one can use culpability in logistic regression, but in the log-binomial model one needs to introduce a quadratic term. In theory, if a variable is linear in logistic regression then it is not linear in the log-binomial model, and visa versa. Thus one should always test for linearity if quantitative variables are used in the model. In this example, the quadratic term in culpability is significant in the log-binomial model ($p<0.001$) and for the robust Poisson method ($p<0.001$), but it is not significant in logistic regression ($p = 0.1246$ with the likelihood ratio test). Of course this could also happen the other way (significant in logistic, but not in log-binomial). The estimates for the robust Poisson are larger in absolute value than for the log-binomial, and the same is true for the standard errors. With the robust Poisson, five of 147 estimated probabilities were greater than 1, and the largest was 1.28. Since the log-binomial model and the logistic model are fundamentally different, the notions of linearity, confounding, and interaction are not equivalent between logistic regression and log-binomial regression. For this reason, it is impossible to develop methods that convert adjusted ORs into adjusted PRs.

Table 2 Comparison of the logistic, Mantel-Haenszel, Poisson, robust Poisson and log-binomial methods for treatment of headache adjusted for gender

Method	Estimates (SE)	Ratio	Wald 95% CI	Wald p value
Logistic	1.2060 (0.4244)	OR = 3.340	1.454 to 7.674	0.0045
M-H	NA	PR = 2.167	1.232 to 3.809	NA
Log-binomial	0.7837 (0.2859)	PR = 2.190	1.250 to 3.834	0.0061
Cox/Poisson	0.7732 (0.3450)	PR = 2.167	1.102 to 4.261	0.0250
Robust Poisson	0.7732 (0.2866)	PR = 2.167	1.236 to 3.799	0.0070

NA, not applicable; OR, odds ratio; PR, prevalence ratio.

EXAMPLE 3

Example 3 comes from the SAS PROC LOGISTIC documentation. It is a study of transient reflex vaso-constriction of the digit skin. A binomial variable for constriction (constricted = 1, not constricted = 0) was modelled on the logarithm of air rate and the logarithm of air volume using 39 trials. Because the prevalence of constriction was 20 out of 39 (0.51), ORs would not be good estimates of PRs. The results are given in table 4. Because PROC GENMOD did not converge for the log-binomial model due to the MLE being on the boundary of the parameter space, the COPY method was used to obtain the estimates for the log-binomial model. The p values for the log-binomial method were similar to those of the robust Poisson method. As expected, the standard errors of the log-binomial were smaller than those of the robust Poisson. In addition, the estimated coefficients for the robust Poisson are somewhat higher than those for the log-binomial for the logarithm of rate, and nearly twice as high for the logarithm of volume. With the robust Poisson solution, three of 39 estimated probabilities were greater than 1, and the largest was 1.82. Again we consider the solution to the log-binomial regression using the COPY method the "best" solution since it yields an approximate solution to the correct binomial likelihood. No quadratic terms in either log(rate) or log(volume)

Table 3 Comparison of log-binomial and robust Poisson methods for analysis of death penalty associated with covariates*

Independent variable	Log prevalence ratio estimate† (SE)		p Value	
	Log-binomial	Robust Poisson	Log-binomial	Robust Poisson
Black defendant	0.3152 (0.1367)	0.5935 (0.1992)	0.0224	0.0029
White victim	0.1219 (0.1078)	0.3173 (0.2061)	0.2288	0.1238
Serious	−0.0010 (0.0174)	0.0023 (0.0352)	0.9305	0.9475
Culpability	1.8062 (0.2750)	1.9223 (0.4453)	<0.001	<0.001
Culpability squared	−0.2006 (0.0308)	−0.2158 (0.0624)	<0.001	<0.001

*Wald tests were used for the robust Poisson method, and likelihood ratio tests were used for the log-binomial method. The latter were obtained by fitting a model without the effect being tested, and calculating minus twice the difference in log likelihoods to get a χ^2 test statistic and p value. In case the log-binomial did not converge, the COPY method approximation was used.
†The intercept estimate was −4.4445 for the log-binomial method and −4.9193 for the robust Poisson method. Of the 147 probability estimates, five were greater than unity for the robust Poisson method, and the largest was 1.28.

are significant in the log-binomial model using the COPY method, robust Poisson regression or logistic regression.

SAS CODE

All analyses were performed using SAS. The log-binomial method was performed with PROC GENMOD using the binomial distribution and the log link. The SAS code for fitting the log binomial model is:

```
PROC GENMOD DATA = ONE;
MODEL Y = X1 X2 … Xk/D = BIN LINK = LOG
INTERCEPT = −4 LRCI;
```

It might be necessary to specify an initial value for the intercept (eg, −4) to ensure a valid starting value. The robust Poisson method was performed with PROC GENMOD using the Poisson distribution. The SAS code for fitting the robust Poisson model is:

```
PROC GENMOD DATA = ONE; CLASS ID;
MODEL Y = X1 X2 … Xk/D = POISSON;
REPEATED SUBJECT = ID; RUN;
```

Here ID is a unique identifying variable for each observation. Use of the repeated statement (even though there are no repeated measures per subject) causes SAS to use the robust sandwich variance estimator. When GENMOD did not converge for the log-binomial model, the COPY portion of the COPY method from Deddens *et al*[10] was used with c = 1000 copies. The SAS code for implementing the COPY method (for c = 1000) using the weight statement is:

```
DATA ONE; SET ONE; W = .999;
DATA TWO; SET ONE; Y = 1−Y; W = .001
DATA THREE; SET ONE TWO;
PROC GENMOD DATA = THREE;
WEIGHT W;
MODEL Y = X1 X2 … Xk/D = BIN LINK = LOG
INTERCEPT = −4 LRCI;
```

Table 4 Comparison of the log-binomial method and the robust Poisson method for vaso-constriction associated with the logarithm of rate and logarithm of volume of inspired air*

Independent variable	Log PR estimate† (SE)		p Value	
	Log-binomial	Robust Poisson	Log-binomial	Robust Poisson
Log(rate)	1.3132 (0.3362)	1.5578 (0.4270)	<0.001	<0.001
Log(volume)	0.7715 (0.1960)	1.4614 (0.3510)	<0.001	0.0000

*Wald tests were used for the robust Poisson method, and likelihood ratio tests were used for the log-binomial method. The latter were obtained by fitting a model without the effect being tested, and calculating minus twice the difference in log likelihoods to get a χ^2 test statistic and p value. In case the log-binomial did not converge, the COPY method approximation was used.
†The intercept estimate was −1.5147 for the log-binomial method and −1.8311 for the robust Poisson method. Of the 39 probability estimates, three were greater than unity for the robust Poisson method, and the largest was 1.82.

DISCUSSION

If one wants to estimate the adjusted PR in a study with a common outcome, one should not use logistic regression. When the independent variables are all categorical, as they are in example 1, MLEs are usually easy to obtain using the log-binomial model.[9] If there are continuous covariates and the log-binomial model does not converge (for example, as in examples 2 and 3), then one can use either the COPY method[10] to get approximate MLEs, the robust Poisson method or non-linear least squares.[12] For examples 2 and 3, the COPY method and the robust Poisson methods yield estimates which are quite different, and the robust Poisson yields probability estimates which are greater than 1. Thus the decision on which method to use is important. Maximum likelihood estimation and likelihood ratio tests are held in high regard by statisticians. Using the log-binomial model allows one to say that both were used

When the log-binomial model does not converge on the original data, the main advantage of using the robust Poisson method is that it always converges and is fairly easy to use. The main disadvantages are that it will usually yield larger standard errors than the COPY method and will often produce estimated probabilities greater than 1 for observations in the original data file. The main advantage of the non-linear least square method is that it may be more robust to outliers and model misspecification than the COPY method.[12] The advantage of using the COPY method is that it does find a very good approximation of the true log-binomial MLE, and use of the likelihood ratio test will give the best test of hypotheses. (Likelihood ratio tests require fitting of two models and subtraction of the log-likelihoods to get p values.) The main disadvantage of using the COPY method is that it does require some data manipulation, it may on rare occasions yield invalid Wald confidence intervals, and it may be more sensitive to outliers and model misspecification. It should be emphasised that the log-binomial model is often correct, and that the problem is with standard statistical software which will not converge when the maximum likelihood solution is on the boundary of the parameter space. Hopefully, statistical software developers will improve their algorithms in the future.

Deddens *et al*[10] recommend using use the log-binomial results when the model converges and the COPY method when it does not. Spiegelman and Hertzmark[19] recommend using the log-binomial model when it converges but replacing it with the

Key points

- For common outcomes, the odds ratio (logistic regression) is not a good approximation of the prevalence ratio (log-binomial).
- Standard software does not always converge when using the log-binomial model.
- In such cases, the COPY method can be used to find the approximate maximum likelihood solution to the log-binomial model.
- Robust Poisson and non-linear least squares methods may be less affected by model misspecification.
- One should always check the fit of continuous covariates in the log-binomial or logistic regression model.

robust Poisson method when the log-binomial model does not converge. Lumley *et al*[12] recommend using the log-binomial model when it does converge, but using non-linear least squares when the log-binomial model does not converge because the non-convergence might be due to a few outlying observations that do not fit the log-binomial model or because the model has been misspecified. In this case they state that the MLE produces a more misleading estimator than an estimator that allows for a few observations to have predicted values greater than 1.

Maximum likelihood estimators are consistent, tend to have small variances, and are asymptotically unbiased and efficient. In addition, the likelihood ratio test is generally considered to perform better than the Wald test. The likelihood ratio test does not make sense for the robust Poisson method, because the likelihoods are the same for Poisson and robust Poisson since the solutions are identical. Thus one must use the Wald test and Wald confidence intervals with the robust Poisson method. With the log-binomial method, the Wald test is usually reasonable, but we have found data for which it performed badly. Use of the likelihood ratio test avoids this problem. If one uses the weighted log-binomial approach, then one can also obtain likelihood ratio confidence intervals (since GENMOD does compute them). One can then perform likelihood ratio tests of individual parameters by simply using the likelihood ratio confidence intervals.

It is important to remember that when modelling one should always investigate whether the model fits the data. One should always look for possible interactions and for a quantitative covariate one should always check for linearity. The presence of a quadratic (or interaction) term means that there is not a single adjusted PR, but rather that the adjusted PR depends on the level of the quantitative variable. For example 2, we do not obtain a single PR for culpability since the model contains quadratic terms, but we do obtain a single adjusted PR for black defendant adjusted for quadratic culpability. A quantitative variable could be linear in a logistic regression but non-linear in a log-binomial regression (as in example 2) or vice versa.

In logistic regression the regression coefficients when modelling $Y = 0$ are merely the negatives of the coefficients when modelling $Y = 1$. This is definitely NOT true when using the log-binomial model to estimate the PR. We should also mention that Blizzard and Hosmer[20] have recently developed goodness-of-fit statistics for log-binomial models similar to those for logistic regression.

CONCLUSION

If one wants to estimate the adjusted PR in a study with a common outcome, one should not use logistic regression. Instead one should use the log-binomial model. If the log-binomial model does not converge then one can use either the COPY method to obtain MLEs, or the robust Poisson method, or non-linear least squares to estimate the adjusted PR. Deddens *et al*[10] recommend using the COPY method since it produces MLEs. Spiegelman *et al*[19] recommend the robust Poisson method. Lumley *et al*[11] argue that in the case of non-convergence, the log-binomial MLE is insufficiently robust to model misspecification and that other estimators are often preferable. As shown with the real data used in this study, the results can be quite different depending on which method is used.

Key references

4. **Wacholder S.** *Am J Epidemiol* 1986;**123**:174–84. This was the first article to recommend using generalised linear models with the binomial distribution and log link to estimate the PR.
6. **Lee J.** *Int J Epidemiol* 1994;**23**:201–3. This was the first article to recommend using the Cox proportional hazard model to estimate the PR. Unfortunately the standard error is not correct.
9. **Skov T,** Deddens J, Petersen MR, *et al*. *Int J Epidemiol* 1998;**27**:91–5. This was the first article to compare the log-binomial model to the Cox proportional hazard model.
10. **Deddens JA,** Petersen MR, Lei X. In: *Proceedings of the 28th Annual SAS Users Group International Conference, March 30–April 2, 2003* (http://www2.sas.com/proceedings/sugi28/270-28.pdf). This was the first article to describe when the log-binomial model fails to converge. This article introduced the COPY method as a way to find the approximate maximum likelihood solution when the log-binomial model fails to converge.
11. **Barros AJ,** Hirakata VN. *BMC Med Res Methodol* 2003;**3**:21 (http://www.biomedcentral.com/1471-2288/3/21). This was the first article to recommend using the robust variance estimator for the Cox proportional hazard model when estimating the PR. This article also discussed the equivalence of the Cox and Poisson models.
12. **Lumley T,** Kronmal R, Ma S. *UW Biostatistics Working Paper Series*, paper 293, 2006 (http://www.bepress.com/uwbiostat/paper293/). This article recommends the use of non-linear least squares over the COPY method when the log-binomial model does not converge, due to the lack of robustness of the log-binomial model when the model fails to converge.
13. **Zou G.** *Am J Epidemiol* 2004;**159**:702–6. This article showed how to use the robust variance estimator with Poisson regression to estimate the PR.
16. **McNutt LA,** Wu C, Xue X, *et al*. *Am J Epidemiol* 2003;**157**:940–3. This article showed that the widely cited article of Zhang and Yu using the adjusted OR to estimate the adjusted PR was not valid.
20. **Blizzard L,** Hosmer D. *Biom J* 2006;**48**:5–22. This article introduced goodness-of-fit statistics for the log-binomial model.

Thus the decision on which method to use is very important.

Competing interests: None.

Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the National Institute for Occupational Safety and Health.

REFERENCES

1. **Greenland S.** Interpretation and choice of effect measures in epidemiologic studies. *Am J Epidemiol* 1987;**125**:761–8.
2. **Stromberg U.** Prevalence odds ratio v prevalence ratio. *Occup Environ Med* 1994;**51**:143–4.
3. **Axelson O,** Fredricksson M, Ekberg K. Use of the prevalence ratio v the prevalence odds ratio as a measure of risk in cross sectional studies. *Occup Environ Med* 1994;**51**:574.
4. **Wacholder S.** Binomial regression in GLIM, estimating risk ratios and risk differences. *Am J Epidemiol* 1986;**123**:174–84.
5. **Schouten EG,** Dekker JM, Kok FJ, *et al*. Risk ratio and rate ratio estimation in case-cohort designs: hypertension and cardiovascular mortality. *Stat Med* 1993;**12**:1733–45.
6. **Lee J.** Odds ratio or relative risk for cross-sectional data? *Int J Epidemiol* 1994;**23**:201–3.
7. **Lee J,** Chia KS. Estimation of prevalence rate ratios for cross-sectional data: an example in occupational epidemiology. *Br J Ind Med* 1993;**50**:861–2.
8. **Zocchetti C,** Consonni D, Bertazzi P. RE: Estimation of prevalence rate ratios from cross-sectional data (letter). *Int J Epidemiol* 1995;**24**:1064–1105.
9. **Skov T,** Deddens J, Petersen MR, *et al*. Prevalence proportion ratios: estimation and hypothesis testing. *Int J Epidemiol* 1998;**27**:91–5.
10. **Deddens JA,** Petersen MR, Lei X. Estimation of prevalence ratios when PROC GENMOD does not converge. In: *Proceedings of the 28th Annual SAS Users Group International Conference*, *March 30–April 2, 2003*.Paper 270-28. Cary, NC: SAS Institute Inc, 2003. Available from http://www2.sas.com/proceedings/sugi28/270-28.pdf (accessed 17 April 2008).
11. **Lumley T,** Kronmal R, Ma S. Relative risk regression in medical research: models, contrasts, estimators, and algorithms. *UW Biostatistics Working Paper Series*, paper 293, 2006. Available from http://www.bepress.com/uwbiostat/paper293/ (accessed 17 April 2008).
12. **Barros AJ,** Hirakata VN. Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. *BMC Med Res Methodol* 2003;**3**:21. Available at http://www.biomedcentral.com/1471-2288/3/21 (accessed 18 April 2008).
13. **Zou G.** A modified Poisson regression approach to prospective studies with binary data. *Am J Epidemiol* 2004;**159**:702–6.
14. **Thompson ML,** Myers JE, Kriebel D. Prevalence odds ratio or prevalence ratio in the analysis of cross sectional data: what is to be done? *Occup Environ Med* 1998;**55**:272–7.
15. **Zhang J,** Yu K. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *JAMA* 1998;**280**:1690–1.
16. **McNutt LA,** Wu C, Xue X, *et al*. Estimating the relative risk in cohort studies and clinical trials of common outcomes. *Am J Epidemiol* 2003;**157**:940–3.
17. **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.
18. **Petersen MR,** Deddens JA. RE: Easy SAS calculations for risk or prevalence ratios and differences (letter). *Am J Epidemiol* 2006;**163**:1157–63.
19. **Spiegelman D,** Hertzmark E. Easy SAS calculations for risk or prevalence ratios and differences. *Am J Epidemiol* 2005;**162**:199–200.
20. **Blizzard L,** Hosmer D. Parameter estimation and goodness-of-fit in log binomial regression. *Biom J* 2006;**48**:5–22.

QUESTIONS (SEE ANSWERS ON P 481)

Which statements are true and which are false?

(1) Logistic regression should always be used if the response outcome variable is binary.

(2) The best measure of risk is the odds ratio.

Please select the best answer to the following multiple choice questions:

(3) For observations in the original data file, which of the following will always yield predicted probabilities between 0 and 1?

(a) Schouten's method
(b) Log-binomial (maximum likelihood) method
(c) Lee's proportional hazard method
(d) Robust Poisson method

(4) Maximum likelihood estimators for the log-binomial model

(a) exist only where the derivative of the likelihood is zero.
(b) can be found using the robust Poisson method.
(c) can be found exactly or approximately using the log-binomial method.
(d) can be found using Lee's proportional hazard method.