STATISTICS IN MEDICINE
Statist. Med. 2001; **20**:2421–2428 (DOI: 10.1002/sim.902)

Data augmentation priors for Bayesian and semi-Bayes analyses of conditional-logistic and proportional-hazards regression

Sander Greenland[1,*] and Ronald Christensen[2]

[1]*Department of Epidemiology, UCLA School of Public Health and Department of Statistics, UCLA College of Letters and Science, Los Angeles, CA 90095-1772, U.S.A.*
[2]*Department of Statistics, University of New Mexico, Albuquerque, NM 87131, U.S.A.*

SUMMARY

Data augmentation priors have a long history in Bayesian data analysis. Formulae for such priors have been derived for generalized linear models, but their accuracy depends on two approximation steps. This note presents a method for using offsets as well as scaling factors to improve the accuracy of the approximations in logistic regression. This method produces an exceptionally simple form of data augmentation that allows it to be used with any standard package for conditional-logistic or proportional-hazards regression to perform Bayesian and semi-Bayes analyses of matched and survival data. The method is illustrated with an analysis of a matched case-control study of diet and breast cancer. Copyright © 2001 John Wiley & Sons, Ltd.

1. INTRODUCTION

Expressing prior information in the form of data augmenting the actual observations can be traced back to Laplace in the 18th century [1] and is now a well-established Bayesian technique [2, 3]. Such data augmentation priors (DAPs) are valuable in allowing approximate Bayesian analyses to be carried out with popular software packages, for in some fields (for example, epidemiology) few researchers will employ unfamiliar software. Although approximate formulae for DAPs have been presented for generalized-linear model coefficients [2], these formulae require special rescaling to ensure accurate results with logistic and Poisson regression [4]. The present paper presents a method that uses offsets as well as scaling factors to improve the approximations used to construct logistic-regression DAPs. This method yields especially simplified DAPs that can be extended to fit Bayesian and semi-Bayesian (mixed) conditional-logistic and proportional-hazards models with common software for ordinary fixed-effects modelling.

The method is illustrated with a study of 140 women with breast cancer (the cases) and 222 controls selected from sisters of cases [5], so the data are matched on sistership. The

*Correspondence to: Sander Greenland, Department of Epidemiology, UCLA School of Public Health, 22333 Swenson Drive, Topanga, CA 90290, U.S.A.

Received November 1999
Accepted October 2000

analysis concerns 87 diet-questionnaire items recorded in a 362×87 subject-diet matrix X, plus five confounders recorded in a 362×5 matrix W_0. Prior information includes an 87×35 diet-nutrient matrix Z compiled from external sources, which gives the amount of each of 35 food constituents (such as nutrients) in each diet item. For details about the study see Ursin *et al.* [5]. Because of the extreme data sparsity ($362/92 \doteq$ four subjects per covariate), use of prior information has a large impact on coefficient estimates. Witte *et al.* [6] and Greenland [7] analysed these data with hierarchical conditional-logistic models taking Z as a second-stage (prior) design matrix, using special programs. We here show how these results can be reproduced with ordinary software.

2. DATA AUGMENTATION FOR LOGISTIC AND COX REGRESSION

2.1. Unconditional logistic regression

Consider first data in N unstratified records (X, W, y, n, f) where X and W are $N \times J$ and $N \times K$ design matrices (one may contain a constant column), y and n are N-vectors of independent binomial counts ('successes') and totals, and f is an offset vector (that is, a covariate vector whose coefficient will be fixed at 1). Let $\pi(t)$ be the logistic transform $\pi(t) \equiv (1+e^{-t})^{-1}$; note that $1-\pi(t) = \pi(-t)$, $d\pi(t)/dt = \pi(t)\pi(-t)$, and the inverse transform is $\text{logit}(p) \equiv \ln\{p/(1-p)\}$. One mixed-effects logistic model is

$$E(y/n|f, X, W) = \pi(f + X\delta + W\theta) \tag{1}$$

where δ and θ are J and K vectors of coefficients with $\delta \sim \text{MVN}(\mu, T)$. If $W\theta$ is dropped, this is a fully Bayesian model; otherwise, the model is called semi-Bayes or partial Bayes [2, 8]. The logistic coefficients are interpretable as log-odds ratios for the association of success ($y = 1$) with a unit increase in the covariate [9].

A Bayesian analysis by data augmentation exploits the fact that a conjugate prior is proportional to the likelihood contribution from a set of prior data; those data can be read off the prior directly (reference [10], p. 53). For non-conjugate priors, one must find an approximating conjugate distribution. The following approximation is a refinement of one given by Bedrick *et al.* [2] with modifications based on offsetting (recentring) as well as rescaling the prior to simplify the augmenting data and improve accuracy.

We will approximate the δ prior with a product of J beta densities (the beta distribution being the conjugate prior for the binomial). With a scaling constant $s > 1$ to be discussed below, define $X_a \equiv T^{-1/2}/s$, $f_a \equiv -X_a\mu$, $v \equiv f_a + X_a\delta$ and $p \equiv \pi(v)$. Then

$$v = T^{-1/2}(\delta - \mu)/s \sim \text{MVN}(0, I/s^2)$$

The first-order Taylor expansion of the logistic transform $\pi(v_j)$ about 0 is

$$\pi(v_j) \doteq \pi(0) + \pi(0)\pi(-0)v_j = 1/2 + v_j/4$$

hence the logit-normal components $p_j = \pi(v_j)$ of p are independent and approximately normal, with

$$E(p_j) = 1/2 + E(v_j)/4 = 1/2$$

exactly, and

$$\text{var}(p_j) \doteq \text{var}(v_j)/4^2 = 1/16s^2$$

the accuracy of the approximation improving as the precision $\text{var}(v_j)^{-1} = s^2$ increases.

We next approximate the density of p_j by a beta density with the same mean and variance. A beta(a,a) density for p_j has $E(p_j) = a/2a = 1/2$ and $\text{var}(p_j) = a^2/(2a)^2(2a+1) = 1/4(2a+1)$; letting $a = 2s^2 - 1/2$ yields $\text{var}(p_j) = 1/16s^2$, as desired. Furthermore, letting $c = a - 1 = 2s^2 - 3/2$, the density for p_j is proportional to $p_j^{a-1}(1-p_j)^{a-1} = \pi^c(v_j)\pi^c(-v_j)$, a binomial likelihood contribution from c successes on $2c$ trials. This approximating density concentrates about $1/2$ and approaches normality as s^2 (and hence a) increases without bound.

The prior density for δ has now been approximated by a product of J densities that are proportional to the product over the components of $\pi(f_a + X_a\delta)^c$ and $\pi(-f_a - X_a\delta)^c$. Consequently, under model (1) with $\delta \sim \text{MVN}(\mu, T)$, the posterior density is approximately proportional to the unconditional logistic likelihood for the data augmented by the J pseudo-records $(X_a, W_a, cu, 2cu, f_a)$, where u is a J-vector of ones and W_a is a $J \times K$ matrix of zeros. Note that the number of added 'subjects' is $2cJ$, and so is a function only of the scaling constant s; it does *not* in any way measure the amount of information in the prior for δ.

Recentring the δ prior by f_a and rescaling by s concentrates the v_j priors between $-\ln(3)$ and $\ln(3)$, the domain over which $\pi(t)$ is nearly linear, and so improves the Taylor approximation; an application of Chebychev's inequality shows that taking $s \geqslant 10$ will ensure $\Pr\{-\ln(3) < v_j < \ln(3)\} > 0.99$, even if δ is not multivariate normal. Recentring and rescaling also improves the second (beta-to-normal) approximation by ensuring that the approximating beta density is symmetric and concentrated about $1/2$. The impact of rescaling may be seen by noting that when $s = 1$ the exact 95th percentile of p_j given normal δ is only the 90th percentile of the approximating beta(1.5,1.5) distribution, whereas when $s = 10$ the exact distribution for p_j and its beta(19.5,19.5) approximation have virtually the same 1st to 99th percentiles. The size of s is limited only by numeric precision.

Recentring by the prior offset f_a also greatly simplifies the form of the augmenting counts. This simplification can be important because some packages will truncate the input counts to integers. In the above method the augmenting counts become the arbitrary constants c and $2c$, so one can set s to ensure that c is an integer or truncation error is negligible (with $s = 50$ use of $c = 5000$ is more than adequate).

Unfortunately, some packages lack a provision for offsets. This lack can be accommodated by recognizing that fitting an offset f in a model is numerically equivalent to entering f as just another covariate, one however whose coefficient δ_f is given an independent prior with mean 1 and variance 0. This degenerate prior for δ_f will force the estimated (posterior) δ_f to be 1 with 0 variance, as needed. The impact of this degenerate δ_f prior can in turn be approximated by giving $\pi(\delta_f)$ a beta prior with mean $\pi(1)$ and very small variance. This prior can then be incorporated into the above DAP procedure by adding a pseudo-record (row) for δ_f in X_a that has a 1 in the offset column, 0 for the other covariates, and $\pi(1)10^6 = 731059$ successes out of 10^6 total. Of course, this column will not be needed if $f = 0$ (no offsets in the real data) and $\mu = 0$ (so that $f_a = 0$).

2.2. Conditional logistic regression

Consider next data subclassified into M strata, with D the $M \times N$ matrix of stratum indicators and no constant in X or W. Conditional logistic regression provides estimates of δ and θ in the model

$$E(y/n|D, f, X, W) = \pi(D\alpha + f + X\delta + W\theta) \tag{2}$$

using a likelihood conditional on the stratum margins $D'y$ and $D'(1 - y)$. That likelihood is free of the nuisance parameters α; see Breslow and Day [9] for details. Most commercial software accepts data of the form (X, W, y, f, d, r) where d is a vector of stratum identification numbers, and r is a vector of stratum repetition counts or prior weights; r is constant within strata. Because each record refers to only one subject, y is a vector of Bernoulli indicators while n is identically 1 and so is omitted. See, for example, Stata proc clogit [11].

A stratum with only two members is a matched pair. Consider a matched pair with one success and one failure, both with zero W-covariates, and a repetition count of r_m. The contribution of this pair to the conditional likelihood is

$$\left[\frac{\exp(f_1 + x_1\delta)}{\exp(f_0 + x_0\delta) + \exp(f_1 + x_1\delta)}\right]^{r_m} = \pi[f_1 - f_0 + (x_1 - x_0)\delta]^{r_m} \tag{3}$$

where f_1, x_1 and f_0, x_0 are the offset and the X-covariate row vector for the success and failure in the pair, respectively. This expression is proportional to the binomial-likelihood contribution from r_m successes in r_m trials under the logistic model $\pi[f_1 - f_0 + (x_1 - x_0)\delta]$ for the success probability. It follows that we can enter the DAP based on expression (3) into the conditional likelihood by augmenting the data with $2J$ pairs as follows. Let $g \equiv \max(d)$ be the largest real-data identification number, let $d_{a1} = (g + 1, \ldots, g + J)'$ and $d_{a0} = (g + J + 1, \ldots, g + 2J)'$, let X_a, W_a, f_a and c be as before, and let u and v be J vectors of ones and zeros. Then the augmenting data are the $4J$ pseudo-records

$$\begin{bmatrix} X_a & W_a & u & f_a & d_{a1} & cu \\ vv' & W_a & v & v & d_{a1} & cu \\ vv' & W_a & u & v & d_{a0} & cu \\ X_a & W_a & v & f_a & d_{a0} & cu \end{bmatrix} \tag{4}$$

Substituting these pair data into expression (3) shows that the first $2J$ records (J pairs) yield the J conditional-likelihood contributions in the vector $\pi(f_a + X_a\delta)^c$, and the last $2J$ records yield the J contributions in $\pi(-f_a - X_a\delta)^c$.

As in the unconditional-logistic case, if offsets are not allowed by the package they can be added as a covariate column whose coefficient is given a prior with mean 1, variance 0. The effect of this prior can then be approximated by augmenting the DAP matrix (4) with four pseudo-records composing two pairs with one success in each. One pair has offsets of 1 for the success and 0 for the failure, and a repetition count of $\pi(1)10^6$; the other has offsets of 0 for the success and 1 for the failure, and a repetition count of $\pi(-1)10^6$; all other covariates are zero.

2.3. *Proportional-hazards regression*

Consider next the stratified Cox proportional-hazards model

$$h_k(t|f,X,W) = \exp(f + X\delta + W\theta)h_{0k}(t) \tag{5}$$

with $h_{0k}(t)$ an unspecified baseline hazard function for stratum k. The algebraic equivalence [12] of the partial likelihood for (δ,θ) based on model (5) to the conditional-logistic likelihood for model (2) allows use of the above method in proportional-hazards regression. The augmenting data are identical to those in expression (4), with the case-indicator column corresponding to the failure indicator ($y=1$ if failure, 0 if censored). The data will also contain a time-on-test column, which may be augmented by a vector of $4J$ ones (or any other constant). If the original data are not stratified, they may be declared a single (first) stratum to the program, whence the augmenting data become an additional $2J$ strata.

3. APPLICATION TO THE EXAMPLE

In the example, a case is a 'success' and a control is a 'failure', D is the 362×140 matrix of matched-set (sistership) indicators, X is the 362×87 matrix of subject-specific diet information, W_0 is the 362×5 matrix of confounders, and y is the vector of case indicators ($1=$case, $0=$control). Because each pair is unique, r is a 362-vector of ones; because there is no offset, f is a 362-vector of zeros. Witte *et al.* [6] used a two-stage weighted least-squares (WLS) procedure to fit the hierarchical model

$$E(y|D,X,W_0) = \pi(D\alpha + X\beta + W_0\theta_0) \tag{6a}$$

$$\beta = Z\theta_1 + \delta, \quad \delta \sim \mathrm{MVN}(\mu, T) \tag{6b}$$

with T diagonal; here, δ is a vector of *residual* diet effects that remain after factoring out log-linear nutrient effects. The WLS algorithm has done well in simulation studies, including one based on the example data [13–15]. None the less, better performance has been obtained using a penalized-likelihood (PL) algorithm [14, 16], which is equivalent to profile posterior analysis [17] and which yields the exact posterior mode for δ.

Greenland [7] reanalysed the example via PL with $\mu=0$ and $T=0.125I$ (I the 87×87 identity matrix); the prior variance 0.125 was chosen because it translates into a 95 per cent prior interval for the odds ratio $\exp(\delta_j)$ of $\exp[\pm 1.96(0.125)^{1/2}] = 1/2,\ 2$. To approximate this PL analysis using data augmentation, first transform model (6) to the mixed-model form (2) by substituting the second-stage regression (6b) into the first stage (6a)

$$E(y|D,X,W) = \pi(D\alpha + X(Z\theta_1+\delta) + W_0\theta_0) = \pi(D\alpha + X\delta + W\theta) \tag{7}$$

where $W=[XZ\ \ W_0]$ and $\theta=(\theta_1',\theta_0')'$. Then add $2(87)=174$ pairs as $4(87)=348$ records in the DAP matrix (4). With $s=10$, $X_a = T^{-1/2}/10 = 0.08^{1/2}I$, W_a is an $87\times(35+5)$ matrix of zeros, u and v are 87-vectors of ones and zeros, and $c=2(100)-3/2=198.5$; because $f=0$ and $\mu=0$, the entire offset column is zero and so is omitted. There are now $362+348=710$ records, but the new data are of very simple form, mostly zero covariate entries, a constant

Table I. Selected odds-ratio estimates (coefficient antilogs) from conditional logistic regressions of breast cancer on 35 food constituents (fixed effects) and 87 diet items (mixed effects) fit to 140 cases and 222 matched controls. Approximate (Wald) 95 per cent intervals in parentheses; five potential confounders also forced into each model as fixed effects.

		Regression		
	Model (6ab) fit by PL*	Model (6ab) fit by DA† $s=10$	$s=1$	Model (6a) only fit by CML‡
Fixed effects (θ):				
protein	0.80 (0.32, 2.02)	0.79 (0.31, 2.00)	0.77 (0.14, 4.28)	
phytate	0.83 (0.38, 1.81)	0.87 (0.40, 1.90)	0.97 (0.24, 3.94)	
Mixed effects (β):				
tuna fish	1.66 (0.93, 2.94)	1.69 (0.95, 3.00)	2.06 (1.00, 4.25)	3.55 (1.14, 11.0)
white rice	0.88 (0.67, 1.17)	0.87 (0.66, 1.16)	0.73 (0.51, 1.04)	0.36 (0.19, 0.72)
white wine	1.23 (0.80, 1.90)	1.25 (0.81, 1.93)	1.75 (0.92, 3.35)	7.57 (1.68, 34.1)
ice cream	1.12 (0.66, 1.92)	1.15 (0.67, 1.96)	1.57 (0.73, 3.37)	6.27 (1.55, 25.4)

*Penalized likelihood (PL), using prior covariance matrix of $0.125I$ for the residual diet effects δ.
†Data augmentation approximations to PL analysis, using scaling constant s; $s=1$ corresponds to no rescaling.
‡Conditional maximum likelihood (no prior).

repetition-count column r_a of 198.5, and an identifier column $d_a=(d'_{a1}, d'_{a1}, d'_{a0}, d'_{a0})'$ whose elements range from 141 to $141+174=315$. The DAP estimate of the mixed coefficient vector β is then computed from the estimates for model (7) as $\tilde{\beta}=Z\tilde{\theta}_1+\tilde{\delta}$.

Table I presents odds-ratio estimates $\exp(\tilde{\beta}_j)$ from the PL and DAP analyses, including those with the largest disparity. Also shown are estimates from unconstrained conditional maximum-likelihood (CML) fitting of model (6a) to the unaugmented data, which are quite wild (as one would expect given the number of parameters). For every covariate, the difference between the PL and DAP results with $s=10$ were statistically and scientifically trivial, especially in comparison to their differences from the CML (ordinary non-Bayesian) estimates. In contrast, the DAP analysis with no rescaling ($s=1$) poorly approximates the PL analysis, reflecting the poor approximation of the beta(3/2,3/2) density to the actual logit-normal density of the p_j.

The approximate 95 per cent posterior intervals in the table were computed by the Wald method (that is, based on a normal approximation to the augmented-likelihood profile for β_j) as $\exp(\tilde{\beta}_j \pm 1.96\tilde{\sigma}_j)$, where $\tilde{\sigma}_j^2$ is the jth diagonal element of

$$\tilde{\text{cov}}(\tilde{\beta})=\tilde{\text{cov}}(\tilde{\delta})+Z\tilde{\text{cov}}(\tilde{\theta}_1,\tilde{\delta})+\tilde{\text{cov}}(\tilde{\delta},\tilde{\theta}_1)Z'+Z\,\tilde{\text{cov}}(\tilde{\theta}_1,\tilde{\theta}_1)Z' \tag{8}$$

and the $\tilde{\text{cov}}$ are blocks from the inverse negative second derivative of the conditional log-likelihood at $(\tilde{\delta},\tilde{\theta})$. For comparability, the PL intervals were also computed by the Wald method, taking $\tilde{\sigma}_j^2$ as the jth diagonal element of the inverse negative of the second derivative of the log posterior distribution at $\tilde{\beta}$. This type of interval performed acceptably in the simulations cited earlier, though one can do better using PL-ratio or Monte Carlo methods [4, 17–19]. Unlike empirical Bayes procedures, T is here specified *a priori* as in classical Bayesian analyses, and so no correction for estimating T is needed; see Greenland [7, 8, 13] for arguments and simulations favouring prespecification of T in small-sample and sparse-data epidemiologic analysis.

4. DISCUSSION

Data augmentation allows approximate Bayesian analysis of conditional-logistic and proportional-hazards regression with standard maximum-likelihood fitting programs with rescaled DAP input; the outputted 'MLEs' and 'standard errors' will be identical to penalized-likelihood approximate marginal posterior modes and standard deviations. Its rapidity also facilitates direct computation of profile posterior limits [17], which can provide even better approximations to the exact marginal posterior limits [4]. These and other Bayesian methods have returned to general acceptability over the last three decades, though issues remain. One of us has discussed some of the issues from an epidemiologic perspective in other articles [4, 7, 8, 20, 21]; we here recap those concerning prior specification and approximation accuracy.

Much of the Bayesian literature employs so-called non-informative prior specifications, which can generate procedures with desirable classical frequentist properties (for example, posterior intervals with good coverage properties for all parameter values) [19]. In many epidemiologic contexts, however (including the examples given here and elsewhere [4, 8, 21]), such priors are scientifically absurd; like ordinary maximum likelihood estimates, estimates based on non-informative priors preserve global performance at the cost of poor performance where accuracy is most important, that is, in neighbourhoods of the null value such as $\|\beta\| < 2$ or $\|\beta\| < 1$. This problem is most acute in 'data dredging' studies like the present example, for which epidemiologists vehemently reject classical multiple-comparisons procedures [22, 23], but in which numerous random artefacts should be expected. Here, hierarchical Bayesian analyses can provide at least a cautionary counterpoint to the abundance of 'findings' often seen in maximum-likelihood and even exact results [7, 8, 13, 15, 24], one that is more acceptable to epidemiologists than standard multiple-comparisons adjustments [23, 25].

Much of the modern Bayesian literature also focuses on Monte Carlo methods for computing posterior distributions [18, 19]. As valuable as these methods may be in many applications, in our epidemiologic experiences to date the difference between results from those methods and results from even very crude approximations has been an order of magnitude below statistical uncertainty, and trivial in comparison to the total uncertainty in typical problems (that is, after informally considering effects of uncontrolled confounders, measurement error and selection bias in addition to random error); for example, see Greenland [4]. In comparison, the benefits of seeing the sensitivity of results to prior information can be dramatic, even if one is limited to crude approximations (as in Table I, where the apparent risks associated with white wine are spectacularly sensitive to the prior). It thus seems that simple Bayesian extensions to popular software remain valuable, insofar as they allow more investigators to examine Bayesian as well as frequentist results.

ACKNOWLEDGEMENTS

The authors thank Robert Haile and John Witte for the example data.

REFERENCES

1. Good IJ. *Good Thinking*. University of Minnesota Press: Minneapolis, 1983; chapter 9.
2. Bedrick EJ, Christensen R, Johnson W. A new perspective on generalized linear models. *Journal of the American Statistical Association* 1996; **91**:1450–1460.

3. Bedrick EJ, Christensen R, Johnson W. Bayesian binomial regression: predicting survival at a trauma center. *American Statistician* 1997; **51**:211–218.
4. Greenland S. Putting background information about relative risks into conjugate prior distributions. *Biometrics* 2001; **57** (in press).
5. Ursin G, Aragaki CC, Paganini-Hill A, Siemiatycki J, Thompson WD, Haile RW. Oral contraceptives and premenopausal bilateral breast cancer: a case-control study. *Epidemiology* 1992; **3**:414–419.
6. Witte JS, Greenland S, Haile RW, Bird CL. Hierarchical regression analysis applied to a study of multiple dietary exposures and breast cancer. *Epidemiology* 1994; **5**:612–621.
7. Greenland S. When should epidemiologic regressions use random coefficients? *Biometrics* 2000; **56**:915–921.
8. Greenland S. A semi-Bayes approach to the analysis of correlated associations, with an application to an occupational cancer-mortality study. *Statistics in Medicine* 1992; **11**:219–230.
9. Breslow NE, Day NE. *Statistical Methods in Cancer Research. Vol. I. The Analysis of Case-Control Studies*. IARC: Lyon, 1980.
10. Cox DR, Hinkley DV. *Theoretical Statistics*. Chapman and Hall: New York, 1974.
11. Statacorp. *Stata Statistical Software Release 5*. Stata Press: College Station, TX, 1997.
12. Prentice RL, Breslow NE. Retrospective studies and failure time models. *Biometrika* 1978; **65**:153–158.
13. Greenland S. Methods for epidemiologic analyses of multiple exposures: A review and comparative study of maximum-likelihood, preliminary-testing, and empirical-Bayes regression. *Statistics in Medicine* 1993; **12**: 717–736.
14. Greenland S. Second-stage least squares versus penalized quasi-likelihood for fitting hierarchical models in epidemiologic analyses. *Statistics in Medicine* 1997; **16**:515–526.
15. Witte JS, Greenland S. Simulation study of hierarchical regression. *Statistics in Medicine* 1996; **15**:1161–1170.
16. Breslow NE, Leroux B, Platt R. Approximate hierarchical modelling of discrete data in epidemiologic research. *Statistical Methods in Medical Research* 1998; **7**:49–62.
17. Leonard T, Hsu JSJ. *Bayesian Methods*. Cambridge University Press: Cambridge, 1999.
18. Gelman A, Carlin JB, Stern HA, Rubin DB. *Bayesian Data Analysis*. Chapman and Hall: New York, 1995.
19. Carlin BP, Louis TA. *Bayes and Empirical-Bayes Data Analysis*. Chapman and Hall: New York, 1996.
20. Greenland S. Probability logic and probabilistic induction. *Epidemiology* 1998; **9**:322–332.
21. Greenland S. Small-sample bias and corrections for conditional maximum-likelihood odds-ratio estimators. *Biostatistics* 2000; **1**:113–122.
22. Rothman KJ. No adjustments for multiple comparisons are needed. *Epidemiology* 1990; **1**:43–46.
23. Savitz DA, Olshan AF. Multiple comparisons and related issues in the interpretation of epidemiologic data. *American Journal of Epidemiology* 1995; **142**:904–908.
24. Greenland S, Robins JM. Empirical-Bayes adjustments for multiple comparisons are sometimes useful. *Epidemiology* 1991; **1**:43–46.
25. Poole C. Multiple comparisons? No problem! *Epidemiology* 1991; **2**:241–243.

