STATISTICS IN MEDICINE
Statist. Med. **18**, 643–654 (1999)

GENERALIZED LINEAR MIXED MODELS FOR META-ANALYSIS

ROBERT W. PLATT[1]*, BRIAN G. LEROUX[2] AND NORMAN BRESLOW[2]

[1] *McGill University, Montreal Children's Hospital Research Institute, 2300 Tupper St., Rm. PT-309, Montreal, QC H3H 1P3 Canada*
[2] *Department of Biostatistics, University of Washington, Seattle WA 98195, U.S.A.*

SUMMARY

We examine two strategies for meta-analysis of a series of 2×2 tables with the odds ratio modelled as a linear combination of study level covariates and random effects representing between-study variation. Penalized quasi-likelihood (PQL), an approximate inference technique for generalized linear mixed models, and a linear model fitted by weighted least squares to the observed log-odds ratios are used to estimate regression coefficients and dispersion parameters. Simulation results demonstrate that both methods perform adequate approximate inference under many conditions, but that neither method works well in the presence of highly sparse data. Under certain conditions with small cell frequencies the PQL method provides better inference. Copyright © 1999 John Wiley & Sons, Ltd.

1. INTRODUCTION

Meta-analysis has become an essential tool of the medical researcher for synthesizing the results of many small studies of the same problem. It enables researchers to generate summary conclusions and to better understand sources of between-study variability, which are often substantial. For example, the study of ulcer treatment by Sacks *et al.*[1] cited by Efron[2] shows crude odds ratios for recurrent bleeding after treatment for ulcer that range from 0 to ∞ over 41 studies.

This paper focuses on the particular situation where each study can be summarized by a 2×2 table, as occurs in the meta-analysis of case-control and cohort studies in which both the outcome and primary exposure are dichotomous. Several researchers have proposed models for these data which augment the hypergeometric generalized linear model[3] by unobserved normally distributed random effects to account for unexplained between-study variation. DerSimonian and Laird[4] fitted a normal theory mixed model to empirical transforms of the odds ratios. Rubin[5] cited the need to account for systematic between-study differences in odds ratios, which can be done using covariates in a regression approach. Berkey *et al.*[6] developed odds ratio and relative risk regression models using an empirical transformation and mixed-model estimation techniques

* Correspondence to: Robert W. Platt, McGill University, Montreal Children's Hospital Research Institute, 2300 Tupper St., Rm. PT-309, Montreal, QC H3H 1P3 Canada. E-mail: robertp@epid.lan.mcgill.ca

Contract/grant sponsor: United States Public Health Service
Contract/grant number: CA40644

CCC 0277–6715/99/060643–12$17.50

Received December 1997
Accepted June 1998

due to Morris.[7] Breslow and Clayton[8] proposed penalized quasi-likelihood (PQL) for approximate estimation in generalized linear mixed models (GLMM) and demonstrated its use with an odds ratio regression problem. Other methods proposed for analysis of GLMMs for these data include approximate maximum likelihood implemented with the EM algorithm,[9] Bayes methods for the normally distributed empirical transform of the odds ratio[10] and for the binomial likelihood[11] and empirical Bayes methods for combining likelihoods.[2]

In this paper, we consider PQL as a possible alternative to the widely used empirical transform method. PQL is a simple and easily implemented method which has been shown to give satisfactory approximate results in a variety of problems. Of particular interest is the analysis of sparse data. The empirical transform method requires that 1/2 be added to zero cells to avoid singularities. This can bias the odds ratio estimator when the disease or exposure is rare or when the studies are finely stratified and there are small frequencies in some cells. We examine the performance of PQL and the empirical transform for simulated and real data, in an effort to determine the situations where one or the other method might be preferred.

2. GENERALIZED LINEAR MIXED MODELS FOR REGRESSION ANALYSIS OF THE LOG-ODDS RATIO

Consider K independent studies yielding 2×2 tables of the form (a_k, b_k, c_k, d_k), $k = 1, \ldots, K$ where a_k and b_k are the numbers of occurrences (D) and non-occurrences ($\bar{D}$) of the disease in the treated group (E) and c_k and d_k are the numbers of occurrences and non-occurrences in the untreated group ($\bar{E}$). Define the marginal totals by $n_{k1} = a_k + c_k, n_{k2} = b_k + d_k$, $m_{k1} = a_k + b_k$, $m_{k2} = c_k + d_k$, and the grand total by $N_k = n_{k1} + n_{k2} = m_{k1} + m_{k2}$. Define the odds ratio in the kth table

$$\psi_k = \frac{P_k(D \mid E)}{P_k(\bar{D} \mid E)} \Big/ \frac{P_k(D \mid \bar{E})}{P_k(\bar{D} \mid \bar{E})}$$

and suppose $\theta_k = \log \psi_k$ satisfies

$$\theta_k = X_k \beta + u_k \tag{1}$$

where X_k is a vector of study-level covariates, β is a vector of regression coefficients and u_k is a normal random variate with variance σ^2, included to account for unexplained study to study variation. Under the usual paired binomial sampling model, and conditional on the marginal totals and u_k, cell frequency a_k has a non-central hypergeometric distribution with odds ratio ψ_k. Estimating this by the observed odds ratio $\hat{\psi}_k = (a_k d_k)/(b_k c_k)$, the empirical transform method[6] approximates the distribution of $\hat{\theta}_k = \log(\hat{\psi}_k)$ by a normal distribution with mean θ_k and variance $\sigma^2 + \sigma_k^2$, where σ_k^2 is the study-specific sampling error variance of $\hat{\theta}_k$. Model parameters are estimated using linear mixed model techniques due to Morris.[7] The value of σ^2 is estimated using all the data. There are multiple possibilities for estimating σ_k. We use the method of Berkey *et al.*[6] and designed to reduce the correlation between odds ratio and variance estimates:

$$\hat{\sigma}_k^2 = \left[(a_k + c_k) \left(\sum_{k=1}^{K} (a_k/(a_k + c_k))/K \right) \right]^{-1} + \left[(a_k + c_k) \left(1 - \sum_{k=1}^{K} (a_k/(a_k + c_k))/K \right) \right]^{-1}$$

$$+ \left[(b_k + d_k) \left(\sum_{k=1}^{K} (b_k/(b_k + d_k))/K \right) \right]^{-1} + \left[(b_k + d_k) \left(1 - \sum_{k=1}^{K} (b_k/(b_k + d_k))/K \right) \right]^{-1}.$$

This compares to the usual

$$\hat{\sigma}_k^2 = \frac{1}{a_k} + \frac{1}{b_k} + \frac{1}{c_k} + \frac{1}{d_k}$$

but effectively replaces a_k with the number of treated cases that would have been observed in the kth study if the kth study had the mean proportion of treated cases, and similarly for b_k, c_k and d_k. Berkey *et al.*[6] used an analogous estimate of σ_k^2 for cohort studies and demonstrated via simulation that the alternative estimator had improved bias and variance properties compared to the usual estimator based only on the kth table.

PQL combines the features of standard generalized linear model (GLM) algorithms with the REML method for fitting mixed linear models. While PQL estimates have been found to perform well for some responses which are approximately normally distributed[8,12] (large Poisson counts or binomial proportions with large denominators, for example), their adequacy has yet to be determined for other practical problems, in particular for meta-analysis of the log-odds ratio.

The hypergeometric model is most conveniently parameterized in terms of the odds ratio (1) unlike other GLMs which are parameterized in terms of the mean. To use the GLM framework it is necessary to evaluate the mean μ_k and variance v_k of the observed frequency a_k in the upper left cell of each 2×2 table in terms of the odds ratio ψ_k. Exact computation of these quantities can be very difficult when the marginal totals are large. Although efficient methods for exact evaluation of the mean and variance[13,14] have recently been developed which are effective for all but the largest tables, a simple approximation suggested by McCullagh[15] is adequate for most practical purposes.[16] Here, the mean μ_k and variance v_k are determined as the joint solutions to the equations

$$e^{\theta_k} = \frac{\mu_k(\mu_k + n_{k1} - m_{k1}) + v_k}{(m_{k1} - \mu_k)(n_{k2} - \mu_k) + v_k} \tag{2}$$

and

$$v_k = \frac{N_k}{N_k - 1}\left\{\frac{1}{\mu_k} + \frac{1}{(m_{k1} - \mu_k)} + \frac{1}{(n_{k2} - \mu_k)} + \frac{1}{(\mu_k + n_{k1} - m_{k1})}\right\}^{-1}. \tag{3}$$

Saddlepoint approximations are also reasonable for use in practical situations.[17] In this study we use the McCullagh approximation (2), as it is fast and accurate and easily implemented in the PQL framework. At each step of the PQL algorithm, the usual GLM adjusted dependent variate[18] $Y = (Y_1, \ldots, Y_K)$ is calculated as

$$Y = \hat{\eta} + (a - \hat{\mu})\frac{d\hat{\eta}}{d\hat{\mu}} \tag{4}$$

where $\hat{\mu}$ is the current vector of fitted values obtained by solving equation (2), $\hat{\eta}$ is the current linear predictor $X\beta + u$, $d\hat{\eta}_k/d\hat{\mu}_k = v_k^{-1}$ and $a = (a_1, \ldots, a_k)$. The covariance matrix of Y is approximated by

$$\hat{V} = \hat{W}^{-1} + \hat{D} \tag{5}$$

where $\hat{D}$ is the (diagonal) random effects covariance matrix evaluated at the current estimates for the variance parameters, and $\hat{W}$ is the diagonal matrix of GLM weights v_k. Updated estimates of

the fixed effects β and random effect u are obtained through solution of the (mixed model) equations as follows:

$$\hat{\beta} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}Y \tag{6}$$

and

$$\hat{u} = \hat{D}\hat{V}^{-1}(Y - X\hat{\beta}). \tag{7}$$

A new estimate of the variance parameter σ is obtained from the Fisher scoring step:

$$\hat{\sigma}^{\text{new}} = \hat{\sigma}^{\text{old}} + I^{-1}s \tag{8}$$

where s is the score vector and I is the expected (Fisher) information based on the REML likelihood for Y. The variance parameter is restricted to its natural range ($\sigma \geqslant 0$); estimates outside of this range are set at the appropriate boundary value. *S*-plus programs for this method are given in the Appendix. Note that the McCullagh equations (2) and (3) are only satisfied at covergence of the algorithm.

Convergence can be a problem for PQL, especially with regard to the variance parameter. Halving the Newton step (8) when jumps in the variance component were too large reduced convergence problems significantly. In rare cases when convergence problems still occurred, varying the starting value of the variance parameter estimate improved convergence properties.

Like other methods using the distribution of table values conditional on the marginal totals,[2] the PQL method eliminates tables with a marginal total of zero. These tables provide no information about the odds ratio under the hypergeometric model, because the distribution is completely non-informative. This is in contrast to methods that use prior information about the baseline probabilities[11] or correct data when cell values are zero[6] to be able to use these tables. Conditioning on the marginal total involves factoring the likelihood into two components: the likelihood for a_k given the marginal totals and the likelihood for the marginal totals. The conditional approach uses only the first component of the likelihood. This causes some loss of information about the odds ratio. The magnitude of this information loss has been the subject of considerable discussion and debate.[19,20] This potential loss of information can lead to inefficiency when it is substantial. However, the conditional method does not require distributional assumptions about the marginal totals, and is less biased when these assumptions are incorrect.

3. AN EXAMPLE: ULCER TREATMENT

Efron[2] applied empirical Bayes methods for combining likelihoods to a set of 41 studies on stomach ulcers that was originally examined by Sacks *et al.*[1] The data consist of 41 randomized trials of a new surgical treatment to prevent recurrence of stomach ulcers which were conducted between 1980 and 1989. We analysed Efron's data, which included 39 tables, as one (study 41) had a marginal total zero and one (study 40) was removed due to extreme values using the following model:

$$\theta_k = \log(\psi_k) = \beta_0 + u_k \tag{9}$$

where β_0 is the true overall log-odds ratio for the data and the u_k are independent identically distributed random effects. The results of analysis using PQL were $\hat{\beta}_0 = -1{\cdot}141 \pm 0{\cdot}218$ and $\hat{\sigma} = 1{\cdot}070 \pm 0{\cdot}315$. The corresponding empirical transform results were $\hat{\beta}_0 = -1{\cdot}270 \pm 0{\cdot}240$ and $\hat{\sigma} = 1{\cdot}274$ (Berkey *et al.*'s[6] program does not provide standard errors for σ). Efron's results,

$\hat{\beta}_0 = -1{\cdot}22 \pm 0{\cdot}260$ and $\hat{\sigma} = 1{\cdot}190 \pm 0{\cdot}310$, were intermediate between the PQL and empirical transform estimates.

Next we considered Morris's[21] correction of Efron's tables. Morris noted that in the original publication one of Efron's tables (study 4) did not appear, while three others (studies 24, 32 and 33) had minor differences in table values. The deletion of study 4 left 38 useful tables. Results of analysis using PQL were $\hat{\beta}_0 = -1{\cdot}211 \pm 0{\cdot}223$ and $\hat{\sigma} = 1{\cdot}079 \pm 0{\cdot}329$, and using empirical transform, $\hat{\beta}_0 = -1{\cdot}296 \pm 0{\cdot}245$ and $\hat{\sigma} = 1{\cdot}303$. Reference to the original data suggested the use of stratification according to the type of bleeding ulcer classified as ulcers with: (i) bleeding or spurting blood vessels (12 tables); (ii) visible but non-bleeding blood vessels (13 tables); (iii) no visible blood vessel (5 tables); or (iv) unspecified bleeding type (8 tables). With the model

$$\theta_k = \log(\psi_k) = \beta_0 + \beta_2 x_{2k} + \beta_3 x_{3k} + \beta_4 x_{4k} + u_k \tag{10}$$

where x_i is an indicator for stratum i, the PQL estimates were $\hat{\beta}_0 = -2{\cdot}304 \pm 0{\cdot}388$, $\hat{\beta}_2 = 1{\cdot}154 \pm 0{\cdot}489$, $\hat{\beta}_3 = 2{\cdot}036 \pm 0{\cdot}670$ and $\hat{\beta}_4 = 1{\cdot}967 \pm 0{\cdot}566$. This suggests that the surgical treatment was most effective for ulcers with bleeding or spurting blood vessels, and was ineffective for ulcers with no visible blood vessel or unspecified bleeding type. Inclusion of the covariate reduced the estimated variance component to $\hat{\sigma} = 0{\cdot}831$. The empirical transform results for the fixed effects were similar, $\hat{\beta}_0 = -2{\cdot}437 \pm 0{\cdot}393$, $\hat{\beta}_2 = 1{\cdot}176 \pm 0{\cdot}520$, $\hat{\beta}_3 = 2{\cdot}009 \pm 0{\cdot}726$ and $\hat{\beta}_4 = 2{\cdot}032 \pm 0{\cdot}615$, while $\hat{\sigma} = 1{\cdot}038$.

Sample sizes in the 38 tables that we used from Sacks *et al.*[1] ranged from 11 to 142, with a median sample size of 41·5 and interquartile range of 31. The spread in sample sizes was reasonably low. To determine whether any of the studies had an extreme influence on the meta-analysis, the analysis with no covariate was re-run 38 times with each study in turn deleted from the sample. The resulting PQL estimates $\hat{\beta}_0$ ranged from $-1{\cdot}128$ to $-1{\cdot}266$ and $\hat{\sigma}$ ranged from 0·982 to 1·118. There was no evidence for one study having undue influence on the results.

4. SIMULATION STUDIES

A simulation study was conducted to examine the performance of the two methods for data sets having features similar to the data on ulcers. For each parameter setting we simulated 500 meta-analyses of 39 studies with the 39 sample sizes m_{k1} and m_{k2} taken from the ulcer data set as listed in Sacks[1] (including study 40). The model that generated the data for the kth study was

$$\theta_k = \beta_0 + u_k \tag{11}$$

for $k = 1, \ldots, 39$, with $u_k \sim \mathrm{N}(0, \sigma^2)$, where $\sigma^2 = 0{\cdot}0$, $1{\cdot}0$ or $2{\cdot}0$, and $\beta_0 = 0$ or 2. For the kth table, the number of diseased people in the control group c_k was generated from a binomial distribution with sample size m_{k2} and success probability $p = p_k$ equal to the observed proportion of recurrences in the control group in that table. The number of diseased people in the treatment group a_k was generated from a binomial distribution with sample size m_{k1} and success probability $p_{k1} = p_k\psi_k/(1 - p_k + p_k\psi_k)$ where $\psi_k = \exp(\theta_k)$ the odds ratio. Tables with a zero marginal total were eliminated from the analysis because they provide no information about the odds ratio.

A second set of simulations used the same configuration, except the baseline probabilities were set equal to one-half the proportion of recurrences in the control group to simulate sparser data with more empty cells. The third set of simulation runs used the same set-up but reduced the

Table I. Bias of parameter estimates using the PQL and empirical transform methods based on 500 simulated data sets from a model with no covariate and sample sizes and baseline probabilities taken from Sacks' study of recurrent bleeding and ulcers

β_0	σ	Bias (β_0)		RMSE (β_0)		Coverage probability of 95% CI		SE ($\hat{\beta}_{0(\mathrm{PQL})}$)	
		PQL	ET	PQL	ET	PQL	ET	Estimated	Observed
0·0	0·0	0·003	−0·000	0·119	0·124	0·968	0·960	0·119	0·106
2·0	0·0	−0·001	0·072	0·133	0·148	0·972	0·926	0·133	0·124
0·0	1·0	−0·000	−0·020	0·202	0·216	0·966	0·954	0·202	0·189
2·0	1·0	−0·103	−0·036	0·224	0·208	0·922	0·950	0·213	0·207
0·0	2·0	−0·006	−0·021	0·329	0·331	0·946	0·944	0·328	0·325
2·0	2·0	−0·264	−0·287	0·427	0·426	0·846	0·836	0·336	0·347

Table II. Bias of parameter estimates using the PQL and empirical transform methods based on 500 simulated data sets from a model with no covariate. Sample sizes are taken from Sacks *et al.* and baseline probabilities are reduced by half

β_0	σ	Bias (β_0)		RMSE (β_0)		Coverage probability of 95% CI		SE ($\hat{\beta}_{0(\mathrm{PQL})}$)	
		PQL	ET	PQL	ET	PQL	ET	Estimated	Observed
0·0	0·0	−0·002	−0·002	0·134	0·145	0·968	0·970	0·134	0·127
2·0	0·0	−0·003	0·125	0·129	0·182	0·954	0·830	0·129	0·127
0·0	1·0	0·127	0·132	0·244	0·260	0·898	0·910	0·208	0·201
2·0	1·0	−0·006	0·144	0·205	0·256	0·956	0·896	0·205	0·197
0·0	2·0	0·250	0·312	0·414	0·457	0·878	0·844	0·330	0·332
2·0	2·0	−0·020	0·085	0·326	0·334	0·952	0·942	0·326	0·322

sample sizes in each table by half, to simulate meta-analysis with smaller studies. A fourth set of simulations used the covariate model

$$\theta_k = \beta_0 + \beta_1 x_k + u_k \tag{12}$$

where $\beta_0 = 0$, $\beta_1 = 1$ or 2, and x_k was generated from a normal distribution with mean 0 and variance 1. This situation corresponds to a modifier of the odds ratio that is clinically relevant.

Results of the simulation studies are presented in Tables I–V. Over 99·5 per cent of PQL and empirical transform runs converged for each model. Table I gives the bias, mean square error and confidence interval coverage probability for the fixed effects parameter β_0 for each method. Both estimates of the intercept were severely biased for large values of the intercept and variance parameter. The mean square error of the estimate of the intercept was similar for the two methods, and both methods had adequate coverage probabilities except for $\beta_0 = \sigma = 2$. The standard error of $\hat{\beta}_0$ was well estimated by the PQL method.

Table II gives results for the simulations where the baseline probabilities were reduced by half from those observed in Sacks.[1] In this setting we observed substantial bias in estimates of β_0 for both methods when there was a large random effect and the true $\beta_0 = 0$. When $\beta_0 = 2$, the PQL estimates of β_0 were nearly unbiased, while bias persisted in the empirical transform estimates. In contrast to Table I, the bias was in general positive. Mean square errors of the fixed effect

Table III. Bias of parameter estimates using the PQL and empirical transform methods based on 500 simulated data sets from a model with no covariate. Sample sizes are one half those found in Sacks *et al.*

β_0	σ	Bias (β_0)		RMSE (β_0)		Coverage probability of 95% CI		SE ($\hat{\beta}_{0(\mathrm{PQL})}$)	
		PQL	ET	PQL	ET	PQL	ET	Estimated	Observed
0·0	0·0	0·001	0·003	0·169	0·175	0·980	0·962	0·170	0·154
2·0	0·0	0·013	0·105	0·192	0·208	0·974	0·932	0·191	0·177
0·0	1·0	0·024	0·027	0·237	0·245	0·942	0·934	0·236	0·238
2·0	1·0	−0·112	−0·038	0·275	0·235	0·916	0·924	0·252	0·248
0·0	2·0	0·071	0·099	0·354	0·354	0·954	0·942	0·347	0·328
2·0	2·0	−0·226	−0·244	0·425	0·403	0·906	0·876	0·360	0·366

coefficient were consistently lower for PQL estimates than for empirical transform estimates, and coverage probability for the 95 per cent confidence interval was closer to 95 per cent for PQL estimates than for empirical transform estimates in most cases. In the situation with a large variance component and no ($\beta_0 = 0$) fixed effect, the coverage probabilities were below 95 per cent. Again, in this setting, estimated standard errors of the fixed effects parameters were similar to the observed standard errors based on the simulations.

Bias in the estimated fixed effects can be explained by the occurrence of zero cells. For example, in Table I, there was a high frequency (23–29 per cent) of zeros in cell c (the number of control failures) under $\beta_0 = 2$ when σ was positive. Such data were associated with negative biases in $\hat{\beta}_0$ for both methods. When the baseline probability was reduced, zero cells tended to occur frequently in both cells c and a (the number of treatment successes) under $\beta_0 = 0$. This was associated with positive biases in $\hat{\beta}_0$ for both methods; the bias increased in direct proportion with σ and with the frequency of tables with $a = 0$. The biases were substantially worse for the empirical transform method. Interestingly, even though $a = 0$ infrequently when $\beta_0 = 2$, the empirical transform method suffered from a positive bias in β_0.

Table III gives results for the simulation runs where the sample size per table was reduced by half from those observed in Sacks.[1] In this setting the observed bias in estimates of β_0 for both methods was similar to that seen in the simulations using the full sample size. Mean square errors of the fixed effect coefficient were generally lower for PQL estimates than for empirical transform estimates, and coverage probability for the 95 per cent confidence interval was close to 95 per cent for both PQL and empirical transform estimates except when $\beta_0 = \sigma = 2$. Again, in this setting, estimated standard errors of the fixed effects parameters were similar to the observed standard errors based on the simulations.

When a covariate was present both methods provided adequate estimation of the relevant parameters except in the case of very large fixed effects ($\beta_0 = 2$) and very large random effects ($\sigma = 2$) (Table IV). Both biases and mean square errors for the intercept were similar for the two methods, while empirical transform estimates of the regression coefficient were somewhat more biased than those using PQL, although mean square errors were nearly identical.

We also studied the estimated standard deviations of the random effects for the simulation runs using the covariate (Table V). The results without the covariate and with the reduced sample size followed a similar pattern. PQL estimates of the random effect standard deviation were negatively biased except when σ was equal to zero, while the empirical transform estimates were positively

Table IV. Mean values of parameter estimates using the PQL and empirical transform methods based on 500 simulated data sets from a model with one covariate. The true value of β_1 was 1·0

β_0	σ	Bias (β_0)		RMSE (β_0)		Bias (β_1)		RMSE (β_1)	
		PQL	ET	PQL	ET	PQL	ET	PQL	ET
0·0	0·0	−0·000	0·014	0·127	0·129	−0·001	0·014	0·141	0·121
2·0	0·0	−0·004	0·020	0·142	0·139	0·012	−0·041	0·148	0·135
0·0	1·0	−0·018	−0·014	0·210	0·213	−0·020	−0·017	0·216	0·200
2·0	1·0	−0·091	−0·084	0·239	0·225	−0·030	−0·098	0·223	0·217
0·0	2·0	−0·005	−0·000	0·336	0·327	−0·097	−0·138	0·347	0·339
2·0	2·0	−0·222	−0·295	0·407	0·427	−0·094	−0·192	0·348	0·347

Table V. Bias in values of estimated random effect standard deviations based on 500 simulated data sets

β_0	σ	Bias ($\hat{\sigma}$)	
		PQL	ET
0·0	0·0	0·122	0·237
2·0	0·0	0·148	0·283
0·0	1·0	−0·070	0·020
2·0	1·0	−0·084	0·051
0·0	2·0	−0·236	−0·194
2·0	2·0	−0·275	−0·343

biased for moderate ($\sigma = 1$) random effects but negatively biased for larger random effects, where the biases were similar between the two methods.

To determine whether the differences observed in the parameter estimates were method-specific or data-specific, we used a paired t-test on the results of the first set of simulation runs for β_0 and σ. All differences between empirical transform and PQL methods were highly statistically significant, indicating that the differences here were not data-specific (that is, not due to simulation error).

5. DISCUSSION

Random effects models for meta analysis of the odds ratio provide a useful method of accounting for unexplained between-study variation. Regression models help to explain the dependence on known study-level covariates.

In our analysis of the Sacks–Efron–Morris data, the empirical Bayes likelihood estimates of both the fixed and random effects were intermediate between the PQL estimates and the empirical transform estimates. This is in keeping with previous work and indicates that for this particular example the empirical Bayes estimates may be closest to the true parameters.

Computer simulation showed that PQL estimates of the regression parameters were nearly unbiased, except when the data in the tables were sparse or the study-to-study variation was

large. Parameters were slightly underestimated in many cases, but the bias was extremely small relative to the predicted standard deviation. Estimates of regression parameters based on linear models for the transformed odds ratio were also very close to the true values, except when the data were extremely sparse and many cell entries were zero. In these cases positive regression coefficients were badly overestimated and coverage probabilities of 95 per cent confidence intervals were well below 95 per cent. Given the substantial bias using either method when the data are sparse, one should exercise extreme caution when analysing a set of 2×2 tables that have many cells equal to zero.

Given the availability of efficient software for linear mixed models and the relative simplicity of the approach, it is reasonable to recommend the use of the transformation method when the data are not sparse. For smaller numbers in the tables, we recommend PQL, since the fixed effects estimates have been demonstrated to be less biased. Implementation of PQL is only somewhat more complex than that for linear models. Appendix A provides S-plus code for implementing PQL estimation of random effects models for the log-odds ratio. A similar set could easily be developed for SAS based on the GLIMMIX macro.[23]

APPENDIX: PROGRAM LISTING

The S-plus function `hgpql` requires the input vectors (one observation for each table) y the number of exposed cases, $n1$ the number of exposed subjects, $m1$ the number of cases and n the total number of subjects, as well as a matrix x of covariates (including an intercept). The function calls `hgam` to calculate the expectation of y using the McCullagh approximation.

```
hgpql <- function (y, x, n1, m1, n, alpha.start = 0,
sigma.start = 0·5, b.start = 0, maxit = c(5, 3, 3), tol = 0·0001)
{
    conv <- 1
    options (digits = 4)
    p <- ncol (x)
    k <- nrow (x)
    if (length(maxit) == 1)
        maxit <- rep(maxit, 3)
    if (length(alpha.start) == 1)
        alpha.start <- rep(alpha.start, p)
    if (length(b.start) == 1)
        b.start <- rep(b.start, k)
    alpha <- alpha.start
    alpha.out <- alpha
    random <- 1:k
    z <- lm(y~(-1 + factor(random)), x = T)$x
    d <- diag(rep(1, length(unique(random))))
    dzt <- d%*% t(z)
    zdzt <- z%*% dzt
    sigma <- sigma.start
    theta <- sigma^2
    b <- b.start
    v <- (m1 * (n - m1) * n1 * (n - n1))/(n * n * (n - 1))
    i <- 1
    repeat
```



```
# Iterate until convergence
        j <- 1
        repeat {
# Iterate on alpha and b
          fix.pred <- as.vector(x %*% alpha)
          lin.pred <- fix.pred + b
          or <- exp(lin.pred)
          fit <- hgam(n1, m1, n, or, v)
          v <- n/((n - 1)*(1/fit + 1/(m1 - fit) + 1/(n1 - fit) + 1/(n - m1 - n1 + fit)))
          if (any(is.na(fit))) {
              conv <- 0
              break
          }
          deta.dmu <- 1/v      # link derivative = 1/cond. variance
          w.inv <- deta.dmu      # inverse weight matrix
          v.inv <- solve(diag(w.inv) + theta * zdzt)
          Y <- lin.pred + (y - fit) * deta.dmu
          resid <- Y - fix.pred
          info.alpha <- t(x) %*% v.inv %*% x
          new.alpha <- as.vector(solve(info.alpha)%*%t(x)%*%v.inv%*%Y)
          delta <- new.alpha - alpha
          alpha <- new.alpha      #print(t(alpha))
          b <- as.vector(z %*% as.vector(theta*dzt%*%v.inv%*%resid))
          if (max(alpha) < tol)
              break
          if (j >= maxit [2] || (j > 1 && max(abs(delta/alpha)) < tol)) break
          j <- j + 1
        }
# end iteration on alpha and b
        alpha.out <- rbind(alpha.out, t(alpha))
        j <- 1
        repeat {
# Iterate on theta
          v.inv <- solve(diag(w.inv) + theta * zdzt)
          info.alpha <- t(x) %*% v.inv %*% x
          p <- v.inv - v.inv%*%x%*%solve(info.alpha)%*%t(x)%*%v.inv
          u <- as.vector(v.inv %*% resid)
          pzdzt <- p %*% zdzt
          score.theta <- 0.5*(t(u)%*%zdzt%*%u - sum(diag(pzdzt))
          info.theta <- 0.5 * sum(diag(pzdzt %*% pzdzt))
          delta <- as.vector(score.theta/info.theta)
          repeat {
            if (theta + delta >= 0 & abs(delta) < 1) break
            delta <- as.vector(score.theta/
                (info.theta + 2^(iter2)*info.theta))
            if (iter2 > 9) break
            iter2 <- iter2 + 1
          }
          theta <- max(0.00001, theta + delta)
          if (max(theta) < tol)
                break
          if (j >= maxit [3] || (j > 1 && abs(delta/theta) < tol))
                break
          print(theta)
          j <- j + 1
```



```
        }
# end iteration on theta
        sigma <- rbind(sigma, sqrt(theta))
        delta <- c(abs((alpha.out[i, ] -
            alpha.out [i - 1, ])/alpha.out [i - 1, ]),
            abs((sigma [i, ] - sigma [i - 1, ])/sigma [i - 1, ]))
        for (k in 1 : length (delta))
          {delta [k] <- ifelse(is.na(delta[k]), 0, delta[k])}
        if (i >= maxit [1] || (i > 1 && max(delta) < tol)) break
        i <- i + 1
      }
      alpha.out1 <- rbind(alpha.out, sqrt(diag(solve(info.alpha))))
      return(list(Fixed = alpha.out[length(alpha.out[, 1]), ],
          SEFixed = sqrt(diag(solve(info.alpha))),
          Var.Comp = (sigma[length(sigma[, 1]),
          SERandom = ifelse(sigma[length(sigma[, 1]), ] < 0·005,
              NA, sqrt(0·5 * theta * (1/info.theta))),
Iter = i, Converged = conv, Fitted = or, Random = b))
}

hgam <- function(N1, M1, n, or, v)
{
    a <- ifelse(or ! = 1, (or - 1), or)
    b <- (M1 + N1) * a + n
    c <- M1 * N1 * or
    R <- sqrt(b * b - 4 * a * c)
    fv <- sqrt(1 - (4 * a * a * v)/(R * R))
    fv <- (b - R * fv)/(2 * a)
    fv <- ifelse(or = = 1, (N1 * M1)/n, fv)
    return(fv)
}
```

ACKNOWLEDGEMENTS

This research was supported in part by United States Public Health Service Grant CA40644.

REFERENCES

1. Sacks, H. S., Chalmers, T. C., Blum, A. L., Berrier, J. and Pagano, D. 'Endoscopic hemostasis, an effective therapy for bleeding ulcers', *Journal of the American Medical Association*, **264**, 494–499 (1990).
2. Efron, B. 'Empirical Bayes methods for combining likelihoods', *Journal of the American Statistical Association*, **91**, 538–550 (1996).
3. Breslow, N. E. 'Regression analysis of the log odds ratio: a method for retrospective studies', *Biometrics*, **32**, 409–416 (1976).
4. DerSimonian, R. and Laird, N. 'Meta-analysis in clinical trials', *Controlled Clinical Trials*, **7**, 177–188 (1986).
5. Rubin, D. 'A new perspective', in Wachter, K. and Straf, M. (eds), *The Future of Meta-Analysis*, Russell Sage Foundation, New York, 1990, chapter 14, pp. 155–165.
6. Berkey, C., Hoaglin, D., Mosteller, F. and Colditz, G. 'A random effects regression model for meta-analysis', *Statistics in Medicine*, **14**, 395–411 (1995).
7. Morris, C. N. 'Parametric empirical Bayes inference: theory and applications', *Journal of the American Statistical Association*, **78**, 47–55 (1983).
8. Breslow, N. E. and Clayton, D. G. 'Approximate inference in generalized linear mixed models', *Journal of the American Statistical Association*, **88**, 9–25 (1993).

9. Raghunathan, T. and Ii, Y. 'Analysis of binary data from a multicentre clinical trial', *Biometrika*, **80**, 127–139 (1993).
10. Carlin, J. B. 'Meta-analysis for 2×2 tables', *Statistics in Medicine*, **11**, 141–158 (1992).
11. Smith, T. C., Spiegelhalter, D. J. and Thomas, A. 'Bayesian approaches to random-effects meta-analysis: A comparative study', *Statistics in Medicine*, **14**, 2685–2699 (1995).
12. Leroux, B. G., Lei, X. and Breslow, N. E. 'Estimation of disease rates in small areas: a new mixed model for spatial dependence', in Halloran, M. E. and Berry, D. (eds), *Epidemiology*, *Environment and Clinical Trials*, Springer, in press (1998)
13. Satten, G. and Kupper, L. 'Continued fraction representation for exact cell counts of a 2×2 table; a rapid and exact method for conditional maximum likelihood estimation', *Biometrics*, **46**, 217–223 (1990).
14. Liao, J. 'An algorithm for the mean and variance of the noncentral hypergeometric distribution', *Biometrics*, **48**, 889–892 (1992).
15. McCullagh, P. 'On the elimination of nuisance parameters in the proportional odds model', *Journal of the Royal Statistical Society*, *Series B*, **46**, 250–256 (1984).
16. Breslow, N. E. and Cologne, J. 'Methods of estimation in log odds ratio regression models', *Biometrics*, **42**, 949–954 (1986).
17. Platt, R. W. 'Estimation using saddlepoint approximations and modified profile likelihood in log odds ratio regression analysis', *Communications in Statistics*: *Simulation and Computation*, **27**, in press (1998).
18. McCullagh, P. and Nelder, J. *Generalized Linear Models*, Chapman and Hall, London, 1989.
19. Plackett, R. L. 'The marginal totals of a 2×2 table', *Biometrika*, **64**, 37–42 (1977).
20. Reid, N. 'The roles of conditioning in inference', *Statistical Science*, **10**, 138–157 (1995).
21. Morris, C. N. 'Comment on "Empirical Bayes methods for combining likelihoods"', *Journal of the American Statistical Association*, **91**, 555–558 (1996).
22. Breslow, N. E., Leroux, B. G. and Platt, R. W. 'Approximate hierarchical modeling of discrete data in epidemiology', *Statistical Methods in Medical Research*, **4**, 49–62 (1998).
23. Wolfinger, R. and O'Connell, M. 'Generalized linear mixed models: A pseudo-likelihood approach', *Journal of Statistical Computation and Simulation*, **48**, 233–243 (1993).

