STATISTICS IN MEDICINE
Statist. Med. 2001; **20**:3375–3390 (DOI: 10.1002/sim.890)

Estimation for paired binomial data with application to radiation therapy

Jeremy M. G. Taylor[1,*,†], Robert E. Weiss[2], Wenzhi Li[3], Chiu-Hsieh Hsu[1] and Rafal Suwinski[4]

[1]*Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, U.S.A.*
[2]*Department of Biostatistics, UCLA School of Public Health, Los Angeles CA 90095, U.S.A.*
[3]*Department of Statistics, Stanford University, Stanford, CA 94305, U.S.A.*
[4]*Department of Radiation Oncology, Oncology Center Gliwice, Poland*

SUMMARY

We compare and contrast several different methods for estimating the effect of treatment when responses are paired binomial observations. The ratio of binomial probabilities is the parameter of interest, while the binomial probabilities are nuisance parameters which may vary between pairs. The application is a meta-analysis of the treatment of rectal cancer, with observations in each study indicating the number of recurrences of the cancer in each of two groups, one with radiation therapy and one without. The ratio of the probabilities of recurrence in the radiation to non-radiation groups is of substantive interest, and is modelled as a logistic or complementary log-log function of an unknown linear combination of the covariates. The three methods we consider are maximum likelihood, a Bayesian approach and an approach based on estimating equations. For the MLE and Bayesian approach the potentially large number of nuisance parameters are estimated together with the parameters of interest, whereas for the estimating equation approach only the parameters of interest are estimated. A simulation study is performed to compare the methods and evaluate the impact of overdispersion. Copyright © 2001 John Wiley & Sons, Ltd.

1. INTRODUCTION

In this article we contrast three methods for estimating regression coefficients in a relative risk regression model from a set of paired binomial responses, $(Y_{i1}, N_{i1}), (Y_{i2}, N_{i2})$, $i = 1, \ldots, I$. The methods are applied to a data set in which the aim is to estimate the effect of radiation dose and overall treatment time on reducing the incidence of recurrence from metastases following the treatment of rectal cancer. The three methods are maximum likelihood, a weighted estimating equation approach and a Bayesian scheme from a hierarchical model. The parameter estimates from the estimating equation approach will be driven primarily by the mean structure of

*Correspondence to: Jeremy M. G. Taylor, Department of Biostatistics, School of Public Health, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109-2029, U.S.A.
†E-mail: jmgt@umich.edu

Contract/grant sponsor: NIH; contract/grant numbers: CA16042, GM50011

Received August 1998
Accepted November 2000

the assumed model, whereas the MLE and Bayes approaches will use the full distributional assumptions in the assumed model. Thus we might expect the estimating approach to be more robust to misspecification of the variance and other aspects of the model. An obvious question is how does the efficiency of the weighted estimating equation scheme compare with the other two approaches. We might suspect that it may be less efficient particularly when there is a small number of pairs. The efficiency of the estimator will also depend on the choice of weights. Another fact which will be relevant if N_{i1} and N_{i2} are small and I large is the inconsistency of the maximum likelihood estimate. For a simple special case it is known that the maximum likelihood estimate of the relative risk is inconsistent as the number of pairs (I) tends to infinity [1]. This inconsistency is a result of the number of nuisance parameters tending to infinity. The magnitude of the bias will tend to be larger if N_{i1} and N_{i2} are small and may be of no consequence if N_{i1} and N_{i2} are large. In contrast, the estimating equation scheme is consistent [2] as $I \to \infty$. Another question is how are the various estimators affected by extra-binomial variation.

The motivation for this work came from a study of the effect of radiation on rectal cancer [3]. The results from the comparisons of surgery versus surgery plus radiation have been quite variable, even amongst large randomized trials. One major reason for these differences has been the diverse range of radiation prescriptions used in these studies, ranging from a single small dose up to 6.5 weeks of daily treatment with a considerably larger dose. Understanding whether the radiation prescription is responsible for the variation in the effect of radiation is of considerable importance and could lead to more rational choice of fractionation protocol. Several other potential problems exist in comparing these studies; these include different eligibility requirements, different surgical techniques and different methods of reporting patterns of failure. However, within any one study we would expect similarity in these factors.

The specific data we use in this article consists of pairs of binomial counts; one count is the number of recurrences following surgery for rectal cancer, and the second count is the number of recurrences in which the treatment was radiation followed by surgery. Each pair consists of data from a particular institution or clinical trial. In a few cases there was a cluster of three studies rather than a pair. For simplicity we will develop the methods assuming pairs, and indicate how we handle larger clusters later. There were 23 studies in total, 13 of which, and particularly the larger studies, are randomized clinical trials. These data were obtained by an exhaustive search of the rectal carcinoma literature and are shown in Table I. Details of the inclusion criteria, the assignment of covariate values, the determination of the recurrence rate, and many other factors are explained in Suwinski *et al.* [3]. It is important that these data be modelled as pairs (or triples), because there is likely to be considerable between-cluster variation in the patient population, the surgical technique and the exact definition of the outcome, but much less within-cluster variation in these factors. For example, it can be easily seen from Table I that there is considerable variation in the observed recurrence rate between the surgical arms of the various studies. Part of this variation could be due to sampling variability, as some of the studies are quite small. The recurrence rates will in effect be nuisance parameters in the model, as they are allowed to vary between clusters. We also see from the table that the addition of radiation tended to lower the recurrence rate, but the amount by which it is lowered is quite variable. The covariates of interest were the total radiation dose (D) and overall treatment time (T) in the group of patients treated with radiation as shown in Table I. From simple radiobiological principles we would expect the highest total dose delivered in the shortest overall treatment time to be the most effective at

Table I. Recurrence rates in rectal cancer studies.

Study	Surgery			Radiation + surgery				
	Y_{i1}	N_{i1}	Y_{i1}/N_{i1}	Y_{i2}	N_{i2}	Y_{i2}/N_{i2}	Dose	Time
1*	118	275	0.43	125	277	0.45	6.25	1
1*				128	272	0.47	20.0	12
2*	2	36	0.06	2	34	0.06	6.25	1
3*	22	138	0.16	11	120	0.09	18.75	5
4*	32	87	0.37	27	93	0.29	20.0	12
5*	29	75	0.39	8	68	0.12	23.3	5
6	21	70	0.30	4	37	0.11	27.1	5
7*	131	557	0.24	41	453	0.09	31.3	5
8*	105	347	0.30	55	337	0.16	31.3	6
9	62	144	0.43	27	209	0.13	31.9	6
10	34	81	0.42	2	28	0.07	31.9	6
11	35	135	0.26	2	38	0.05	32.4	12
11				5	71	0.07	45.0	30
12	41	226	0.18	8	189	0.04	37.5	8
13*	29	127	0.23	19	129	0.15	30.8	24
14*	40	465	0.09	37	435	0.09	30.8	24
15*	33	175	0.19	14	166	0.08	35.4	21
16*	21	106	0.20	8	64	0.13	35.4	21
17*	16	34	0.47	5	34	0.15	40.0	28
18	24	78	0.31	3	32	0.09	41.4	30
19	13	89	0.15	2	36	0.06	44.1	30
20	7	83	0.08	8	61	0.13	45.0	30
21*	6	64	0.09	8	78	0.10	25.0	17
22	18	41	0.44	3	40	0.08	30.0	19
23	19	103	0.18	6	75	0.08	30.0	19

*Randomized trial.

decreasing the recurrence rate. Examination of Y_{i2}/N_{i2} compared to Y_{i1}/N_{i1} in Table I does suggest that this is supported by the data.

Radiobiology concepts [4] provide the basis for the statistical models we use. A simple biological model for rectal cancer is that the cancer consists of a core of tumour cells which can be detected and possibly also some undetectable metastatic cells within the pelvic region but outside this core region. Let M be the number of such undetectable metastatic cells if they exist. Recurrences following the therapy are generally thought to be a result of growth of these metastatic cells. The patient has such metastatic cells with probability π. The patient will be cured if all tumour cells are removed or killed by the therapy. Surgery removes the central core so cures the patients if there are no metastases. Then the probability of cure for the surgery only group is $1-\pi$. Radiation therapy is given to a larger volume so is effective in killing metastatic cells as well as cells in the central core. For a patient who does have metastases assume that the probability that the radiation does not kill all these metastatic cells is a function $\psi(D,T)$ of covariates D and T. Then the probability of cure for the surgery plus radiation group is $1-\pi+\pi(1-\psi(D,T))=1-\pi\psi(D,T)$. Thus the probability of recurrence is reduced from π to $\pi\psi(D,T)$ by the addition of radiation to the surgery, and the proportion by which it is reduced is $\psi(D,T)$ which depends on the covariates. The ratio of binomial

probabilities of recurrence $P(\text{recurrence}|\text{radiation} + \text{surgery})/P(\text{recurrence}|\text{surgery})$ is thus the important radiobiological quantity, which we model as a function of the covariates.

A simple form for $\psi(D,T)$ can be derived from a standard radiobiological model [5] in which the expected surviving fraction (S) of tumour cells following radiation of dose D given in overall time T is $S = \exp(\beta_1 D + \beta_2 T)$, where $\beta_1 < 0$ and $\beta_2 > 0$. This model is obviously an approximation because excessively large values of T would make $S > 1$, which is nonsensical, however it does have the sensible property that if there is no radiation then $S = 1$. In our data, T does not take on excessive values. Assuming independence between cells and a binomial distribution for the number of surviving cells gives $\psi(D,T) = 1 - (1 - S)^M$. Making the reasonable assumption that for the typical course of radiation therapy that S is very small, then $\psi(D,T)$ can be approximated by $\psi(D,T) = 1 - \exp(-SM)$ or $\log(-\log(1 - \psi(D,T))) = \log(M) + \beta_1 D + \beta_2 T$. We note that this has the form of a generalized linear model with complementary log-log link. An alternative model to use for $\psi(D,T)$ is a logistic function. In practice the difference in shape between the logistic and complementary log-log links functions will likely be of little consequence compared to the other aspects of heterogeneity in the data for this application. In the data analysis we consider both a logistic function and a complementary log-log link for $\psi(D,T)$, but in the simulation we use only the logistic function.

The motivation we give for the forms of the link functions do not allow for between-person variation in M, β_1 and β_2. Such heterogeneity may change the shape of the link function, as was shown for a special case in a similar model [6]. Another aspect of the model is that $D = T = 0$ does not lead to $\psi(D,T) = 1$, but this is not a real concern because such points are outside the region of interest and as we will see, the estimates of the intercept parameter tend to be large so any problems with this limit are of small consequence.

The radiobiological considerations we described above motivate the relative risk as the quantity of primary interest, and suggest the form of the statistical model. An implicit assumption is that the amount by which the radiation lowers the recurrence rate can be explained by D and T. There may be other unmeasured factors which contribute to the between-pair variation in relative risk, resulting in extra-binomial overdispersion in the observed values of Y or more generally lack-of-fit of the hypothesized model.

The paper is organized as follows: Section 2 describes the model and the three methods of estimation; Section 3 describes the analysis of the radiation therapy data; Section 4 describes the results of a simulation study including an evaluation of the effect of overdispersion, and Section 5 contains a brief discussion.

2. MODEL ESTIMATION AND INFERENCE

2.1. Notation and the model

Let Y_{ij} denote the number of recurrences of rectal cancer for patients in study i and group j, $i = 1,\ldots,I$, $j = 1,2$, where $j = 1$ denotes the surgery only group and $j = 2$ denotes the radiation plus surgery group. Let N_{ij} denote the number of patients in group ij. We assume

$$Y_{i1} \sim \text{Binomial}(N_{i1}, \pi_i) \tag{1}$$

$$Y_{i2} \sim \text{Binomial}(N_{i2}, \psi_i \pi_i) \tag{2}$$

Thus ψ_i measures the effectiveness of the radiation therapy for study i. For the logistic link we assume

$$\log\left(\frac{\psi_i}{1-\psi_i}\right) = \beta_0 + \beta_1 D_i + \beta_2 T_i \tag{3}$$

and for the complementary log-log link we assume

$$\log(-\log(1-\psi_i)) = \beta_0 + \beta_1 D_i + \beta_2 T_i \tag{4}$$

where D_i is the total dose for the ith study and T_i is the overall treatment time for the ith study. Define $\pi = (\pi_1, \ldots, \pi_I)$ and $\beta = (\beta_0, \beta_1, \beta_2)$. The parameters of interest are the β's, with the π_i's being nuisance parameters; thus there are a large number of nuisance parameters compared to the number of β's.

2.2. Maximum likelihood and least squares approaches

Ignoring additive constants, the log-likelihood ($l_i = l_i(\pi, \beta|Y)$) contribution for pair i is

$$l_i = (Y_{i1} + Y_{i2})\log(\pi_i) + Y_{i2}\log(\psi_i) + (N_{i1} - Y_{i1})\log(1-\pi_i) + (N_{i2} - Y_{i2})\log(1-\pi_i\psi_i)$$

This can be maximized to give estimates of all $I+3$ parameters [7]. As with all the estimation schemes to be described, care must be taken to ensure that $\pi_i\psi_i$ is not larger than one. The variances of the parameter estimates can be calculated from the observed information matrix in the standard way; we refer to this as the binomial variance. It is well known for this model that the maximum likelihood estimate is inconsistent for β as the number of pairs (I) tends to infinity [1] because of the increasing number of nuisance parameters.

A slightly more convenient approach for parameter estimation is to perform iteratively reweighted least squares. We obtain estimates of the β's and π's by minimizing

$$\sum_i [v_{i1}(Y_{i1} - N_{i1}\pi_i)^2 + v_{i2}(Y_{i2} - N_{i2}\pi_i\psi_i)^2]$$

with respect to π and β where the weights, which are iteratively updated, are given by

$$v_{i1}^{-1} = N_{i1}\pi_i(1-\pi_i) \tag{5}$$

$$v_{i2}^{-1} = N_{i2}\psi_i\pi_i(1-\psi_i\pi_i) \tag{6}$$

This approach gives the same point estimates as maximum likelihood [8, 9]. The standard errors from this least squares approach are based on the usual residual sum of squares statistics, and may differ from standard errors derived from the information matrix of the maximum likelihood estimates if there is overdispersion in the data.

2.3. Bayesian estimation for a hierarchical model

In this formulation the likelihood is given as above and we specify prior distributions for the parameters π and β. We use Markov chain Monte Carlo methods to obtain estimates

Table II. Parameter estimates for radiation therapy data.

Method	β_0 (intercept)		β_1 (dose)		β_2 (time)	
	Estimate	SE	Estimate	SE	Estimate	SE
Logistic link						
MLE*	3.83	0.78	−0.152	0.026	0.0803	0.0177
Estimating equations†	5.16	1.95	−0.199	0.065	0.0951	0.0296
Bayes*	3.91	0.78	−0.154	0.026	0.0792	0.0181
Bayes OD*	3.32	1.00	−0.142	0.041	0.0838	0.0343
Complementary log–log link						
MLE*	2.23	0.54	−0.105	0.019	0.0559	0.0124
Estimating equations†	2.58	1.02	−0.118	0.034	0.0608	0.0174
Bayes*	2.27	0.50	−0.106	0.017	0.0545	0.0120
Bayes OD*	1.93	0.64	−0.103	0.028	0.0631	0.0249

*23 clusters.
†25 pairs.

of the posterior distribution of the parameters. We used BUGS [10] for Table II and our own code for the simulation study. We use independent beta(a, b) priors for each of the π_i's and an essentially flat prior for the β's. We let $a = 2$, $b = 3$ or $a = b = 1$, both of which represent minimal prior information. We perform Gibbs sampling, sampling from each π_i separately, followed by the β's. *A posteriori*, the π_i's are independent given β, and with a beta prior, they have a conditional density which is a very high dimensional polynomial which we draw from using a Metropolis–Hastings (MH) step [11]. The posterior of β is not in a simple form and also requires an MH step. The proposal jump density for β is a normal density with covariance matrix set equal to the inverse of the 3×3 observed information matrix evaluated at the MLE, setting the π_i's to the MLEs. The proposal density for each π_i is normal, with standard deviation set equal to a rough estimate of the posterior variance $(Y_{i1}/N_{i1})(1 - Y_{i1}/N_{i1})/N_{i1}$. Whenever Y_{i1} equals 0 (or N_{i1}) we set $Y_{i1} = 0.1$ (or $N_{i1} - 0.1$) to calculate the standard deviation. When a jump proposes a new π_i outside the interval $(0, 1)$, we reject the proposal and keep the previous value. We discard the first 1000 draws and take the next 5000 to compute the means and standard errors of all the parameters.

The Bayesian approach is based on the same likelihood as the MLE but it should have fewer problems with the large number of nuisance parameters because it integrates over the nuisance parameters rather than conditions on estimates of them. Further, estimates of variability are averaged over the known uncertainties.

2.4. Estimating equations

A third alternative in which we avoid explicit estimation of the large number of nuisance parameters is based on estimating equations. The method consists of finding the vector β which solves the following equations:

$$\sum_{i=1}^{I} w_i \left(\psi_i \frac{Y_{i1}}{N_{i1}} - \frac{Y_{i2}}{N_{i2}} \right) = 0$$

$$\sum_{i=1}^{I} w_i D_i \left(\psi_i \frac{Y_{i1}}{N_{i1}} - \frac{Y_{i2}}{N_{i2}} \right) = 0$$

$$\sum_{i=1}^{I} w_i T_i \left(\psi_i \frac{Y_{i1}}{N_{i1}} - \frac{Y_{i2}}{N_{i2}} \right) = 0 \tag{7}$$

where w_i are weights. Denote the solution by $\hat{\beta}$ and the corresponding estimate of ψ_i by $\hat{\psi}_i$. The reason that this approach works is because $(\psi_i Y_{i1}/N_{i1} - Y_{i2}/N_{i2})$ has expectation zero under the model. We used $w_i = (N_{i1}^{-1} + N_{i2}^{-1})^{-1}$ to give more emphasis to pairs in which both N_{i1} and N_{i2} were large. We refer to them as Mantel–Haenszel weights. This choice of weights will be discussed later. Three estimating equations are needed because there are three parameters in β. The choice of the factors D_i and T_i in the second and third equations is by analogy with usual normal estimating equations in regression. We note that a solution to the estimating equations in (7) is not guaranteed to exist. For example, if $\frac{Y_{i2}}{N_{i2}}$ is greater than or equal to $\frac{Y_{i1}}{N_{i1}}$ for all i then a solution with $\psi_i < 1$ could not be found. A related method to the above estimation scheme for the case $N_{i1} = N_{i2} = 1$ has been suggested [2, 12], but these authors assumed a log-linear model for ψ_i rather than a logistic or complementary log-log model and thus avoid the restriction of ψ_i less than 1.

We used standard delta method techniques to obtain the variance of $\hat{\beta}$. Denote the set of estimating equations (7) by the vector equation $\sum_{i=1}^{I} U_i(\beta) = 0$. Then the large I asymptotic variance of $(I^{1/2}\hat{\beta})$ is

$$V^{-1} \mathrm{var}\left[I^{-1/2} \sum U_i(\beta) \right] (V^{-1})^{\mathrm{T}} \tag{8}$$

where

$$V_{jk} = \frac{1}{I} \sum_{i=1}^{I} \partial U_{ij}(\beta)/\partial \beta_k \tag{9}$$

and $U_{ij}(\beta)$ is the jth element of U_i.

For $\mathrm{var}[I^{-1/2} \sum U_i(\beta)]$ we use either $I^{-1} \sum \mathrm{var}(U_i(\beta))$ evaluated at $\hat{\beta}$ where $\mathrm{var}(U_i(\beta))$ is calculated from the binomial variance of Y_{i1} or Y_{i2} or we use $(I-1)^{-1} \sum_{i=1}^{I} U_i(\beta) U_i^{\mathrm{T}}(\beta)$ evaluated at $\hat{\beta}$. The second form is analogous to the variance in a GEE estimator [13, 14], it is appropriate in the large I situation and is likely to be more robust to model misspecification. We label the first form for the variance as binomial and the second form as robust. For numerical calculation of the variance we replace V_{jk} by its expected value, which requires estimates of the π_i's as well as the β's. For π_i we use

$$(1-\alpha) Y_{i1}/N_{i1} + \alpha Y_{i2}/\psi_i N_{i2} \tag{10}$$

where

$$\alpha = [1 + (1 - \psi_i \pi_i) N_{i1} / (\psi_i N_{2i} (1 - \pi_i))]^{-1} \tag{11}$$

which is the value of α which minimizes the variance of π_i in equation (10) for known ψ_i. We iterate between equation (10) and equation (11) with ψ_i set at $\hat{\psi}_i$ to obtain an estimate

of π_i to use in equation (9). Alternative methods of obtaining a value for each π_i to use in the variance formula could be used. In our experience the choice of π_i made little difference to the results.

There has been much research for the case of no covariates including various suggestions regarding the choice of weights [2, 7, 15–18]. Tarone *et al.* [17] considered fixed I, large N_{ij} asymptotics and demonstrated that using the Mantel–Haenszel weights could be inefficient compared to the MLE in situations where the π_i's are homogeneous. Some weights give inconsistent estimates as I tends to infinity. This can occur if the weights depend on the observed responses Y_{i1} and Y_{i2}. The Mantel–Haenszel weights $w_i = (N_{i1}^{-1} + N_{i2}^{-1})^{-1}$ were suggested by Nurminen [15]. They are known to give consistent estimates both in the large I asymptotics and in the fixed I, large N_{ij} asymptotics [19]. These weights minimize the variance of $\hat{\psi}$ in the special case in which all the π_i's are equal and there are no covariates. However, they may suffer some efficiency loss in small I situations compared to the MLE [18].

The optimal set of weights is

$$w_i = (\psi(1 - \pi_i)N_{i1}^{-1} + (1 - \psi\pi_i)N_{i2}^{-1})^{-1} \tag{12}$$

From the practical point of view the optimal weights are not so useful because they depend on the π's, the unknown nuisance parameters. We will investigate the performance of these various weighting schemes in a simulation study.

In the case of covariates the optimal estimating equations are of the form

$$\sum_{i=1}^{I} E\left(\frac{\partial g_i}{\partial \beta_j}\right) V_i^{-1} g_i$$

where $g_i = (\psi_i \frac{Y_{i1}}{N_{i1}} - \frac{Y_{i2}}{N_{i2}})$ and $V_i = \mathrm{var}(g_i)$. For both link functions this depends on the unknown nuisance parameters so is less appealing than the simpler form in equation (7).

2.5. *Confidence intervals*

There are a number of ways a 95 per cent confidence or probability interval could be constructed for each of the three methods. For simplicity in this article we use the form 'estimate ± 1.96 SE'. For the Bayesian scheme estimate and SE are taken to be the posterior mean and standard deviation as estimated from the posterior samples.

3. DATA ANALYSIS

Table II gives estimates and standard errors for β obtained from the three approaches for the logistic and the complementary log-log models. For the Bayesian approach we set $a = 2$ and $b = 3$. For β, we used a $\mathrm{N}(0, 10^6)$ prior for the coefficients; this is indistinguishable from a flat prior. The standard errors are based on the binomial variance for the MLE and on the robust variance for the estimating equation approach. The two triplet studies are easily handled in the MLE and Bayesian approach. For the estimating equation approach we split the surgery-only group into two approximately equal half sized studies and regard them as independent, that is we reformulate the data to have 25 pairs. The Bayes overdispersed (OD) estimates are discussed in Section 5.

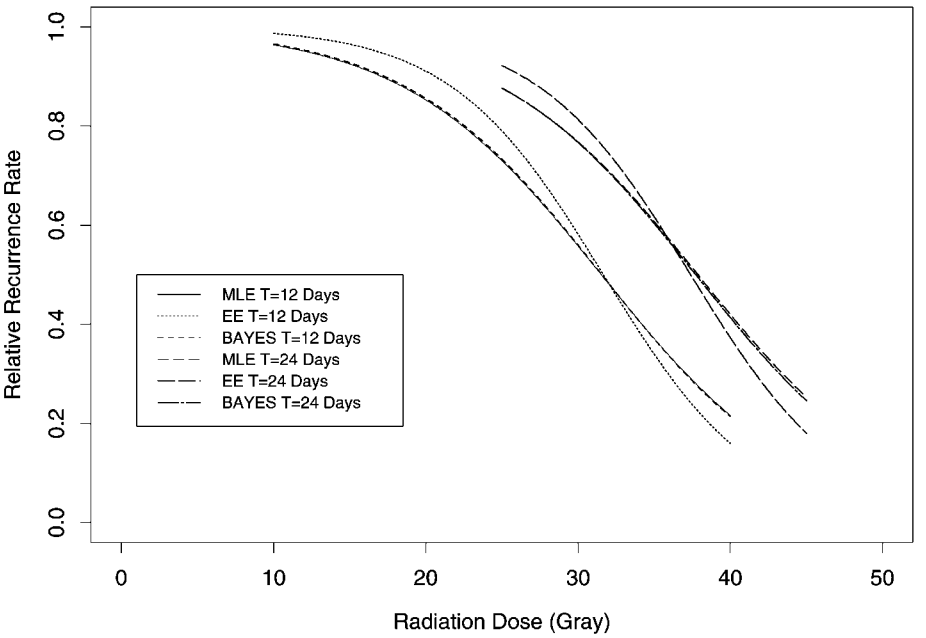

Figure 1. Relative recurrence rate for radiation plus surgery compared to surgery. The plot shows the predicted relative decrease in recurrence probability versus dose at two fixed treatment times. Lines are based on the logistic model parameter estimates in Table II for the estimating equations, the MLE and Bayes. The lines are restricted to the range of doses which are considered plausible for each treatment time.

For the logistic model the MLE and Bayes estimates and SEs are very similar and appear to differ somewhat from the estimating equation results. For the complementary log-log model there is a smaller difference between the MLE and estimating equation results. From all the estimates in Table II it would be concluded that dose and time do significantly affect the benefit of radiation, and the best treatment with radiation would consist of the largest possible dose in the shortest overall time. There are of course other restrictions concerned with logistics and the potential for side-effects which also put restrictions on the dose and treatment time. Figure 1 shows the estimated relative recurrence rate for the logistic model as a function of dose, with time fixed at either 12 (left and lower lines) or 24 days (right and upper lines). The various lines are from the estimating equation parameter estimates, the MLE and Bayes. From all the curves we see the benefit of increasing the dose; we also see that shorter treatment times lead to lower recurrence rates. The apparent difference between the estimating equation and MLE logistic model results is not as great as might appear from the parameter estimates in Table II. The figure shows they give similar predicted curves. Furthermore, as can be seen from the data in Table I, the design points for dose and time are strongly correlated. This leads to substantial correlation between the estimates of β_1 and β_2 (correlation equals -0.61 for the MLE in the logistic model) suggesting a ridge type of region in the likelihood surface.

The sets of parameter estimates in Table II differ between the logistic model and the complementary log-log models because they have different interpretations, however Figure 2

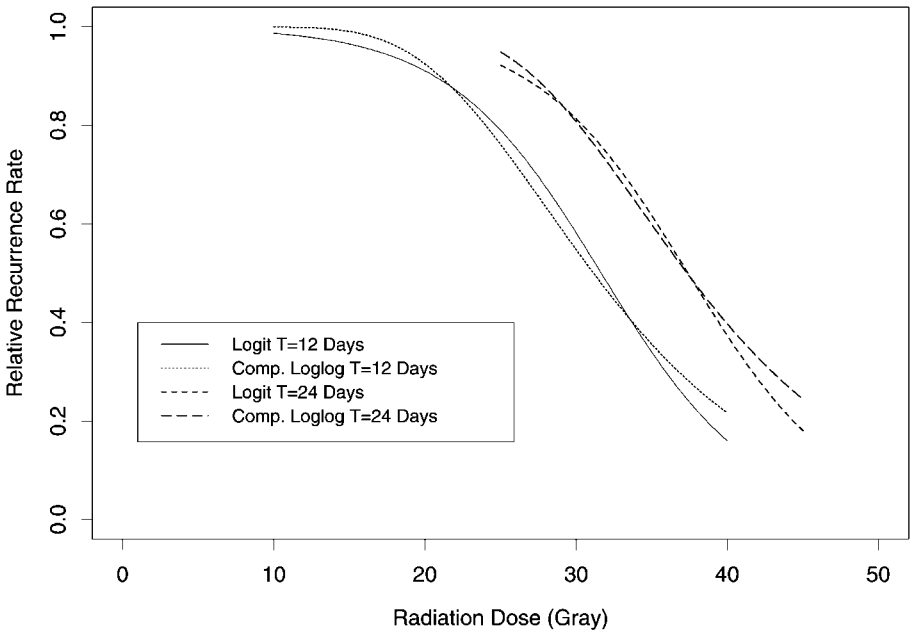

Figure 2. Relative recurrence rate for radiation compared to surgery for the estimating equation approach, comparing logistic and complementary log-log link.

shows that the predicted curves are similar in the region where observations are logistically possible. Figure 2 shows the predictions for the logistic and the complementary log-log links for the EE approach. Analogous plots for ML and Bayes were similar and are not shown.

Robust standard errors from the estimating equation approach are larger than those of the Bayesian approach and those derived from the information matrix for the MLE. The reason for this is overdispersion of the binomial responses, even after allowing the covariates to account for some of the variation. In particular for the MLE, the average value of the squared standardized residual $((1/22)\sum_i \sum_j N_{ij}(Y_{ij} - \hat{Y}_{ij})^2/\hat{Y}_{ij}(N_{ij} - \hat{Y}_{ij}))$ is 2.1 for the logistic model and 2.0 for the complementary log-log model, both of which are much larger than the value 1 which you would expect if the observations were binomial.

Another useful quantity in radiation oncology is the ratio of the regression coefficients. For example, $-\beta_2/\beta_1$ can be interpreted as the increase in dose necessary to balance an increase in treatment time by one day to give the same expected outcome. The ratios calculated from the eight estimates in Table II are all between 0.47 and 0.62, which are similar to published values found for head and neck cancer [20].

Careful examination of the data in Table I shows that $(Y_{i2}/N_{i2})/(Y_{i1}/N_{i1})$ tends to be smaller for the non-randomized studies, suggesting some sort of publication bias for these data. However, it can also be seen that the non-randomized studies tend to be smaller, and tend to have larger doses in shorter times, which is exactly the scenario in which we would expect $(Y_{i2}/N_{i2})/(Y_{i1}/N_{i1})$ to be smaller. Formal inclusion of study type in the model suggests it is of borderline statistical significance (using MLE with the logistic model). It is also clear

from examination of observed and expected values of ψ_i that study 20 is an outlier. For the purposes of the data analysis illustration in this article these issues have been ignored.

4. SIMULATION STUDY

The data analysis suggested a number of potential differences between the various methods, particularly in the standard errors and particularly between the estimating equation methods and the likelihood based methods. We undertook a simulation study to address these and a number of other issues concerned with the various methods. We focus on the logistic model. One issue is how practically important is the lack of consistency of the MLE. A second issue concerns the efficiency of the point estimates from the estimating equation approach compared to the MLE and the Bayesian approach, and how this is influenced by the choice of weights. A third issue is the effect of overdispersion on the various estimators and their standard errors. In addition all of the above may depend on the π_i's, N_{i1} and N_{i2}. We first consider the simple no covariate case to investigate some of the more theoretical issues and then consider a two covariate situation to investigate issues raised by the data analysis. Unless otherwise indicated we use a beta$(1,1)$ prior for π_i and a flat prior for β in the Bayesian approach.

4.1. No covariate case

In simulating the observations we considered $I=32$ or $I=4$, we considered $\pi_i=0.5$ (homogeneous π) for all i or $\pi_i=0.2,\ 0.4,\ 0.6,\ 0.8$ each for one quarter of the pairs (heterogeneous π). For N_{ij} we considered two balanced cases, $N_{ij}=3$ or 10 (denoted by 3 or 10) for all ij and an unbalanced case $(N_{i1},N_{i2})=(10,10),\ (10,100),\ (100,10),\ (100,100)$ each for a quarter of the pairs (denoted by unbalanced). We generated ψ_i from two different schemes either $\psi_i=0.75$ or $\psi_i\sim$ beta$(9,3)$ (denoted by overdispersed). A beta$(9,3)$ random variable has mean 0.75 and standard deviation 0.120. All results are based on 500 replications. In the tables the columns labelled mean are the average of the 500 point estimates and the column labelled SD is the standard deviation of these 500 numbers.

Table III shows the results of comparing different weighting schemes in the estimating equation approach with the MLE. To generate the data, we fixed ψ_i at 0.75. We show results for bias, efficiency and coverage rate of two different 95 per cent confidence intervals, derived from either binomial or robust standard errors. In the table MH refers to using Mantel–Haenszel weights, opt(true) refers to using equation (12) with the true values of the parameters and opt(est) refers to using equation (12) with estimated values of the parameters. The results for $w_i=1$ and $w_i=$ opt(true) are omitted in the configurations where these weights are constant and the same as the MH weights. We see that the only estimator which consistently shows bias is the estimating equation approach with estimated optimal weights. Although the MLE is not consistent as I tends to infinity, this inconsistency is not manifested as any real bias in this simulation study even for $N=3$. A second finding is that using constant weights can clearly be less efficient than other schemes. All the other schemes have comparable efficiency except that the MLE is less efficient for the $N=3$ case. The opt(true) weights are included as a benchmark of the best achievable efficiency, even though it is not a method which can be used for real data. We see no practical difference in efficiency between using these weights and using the MH weights. The coverage rate of the MLE is adequate except in the $N=3$

Table III. Monte Carlo results, comparisons of MLE and estimating equation approach, effect of different weights.

I	π_i	N	Method	Mean	SD	Coverage rate	
						Binomial	Robust
32	0.5	Unbalanced	Estimating equations ($w_i = \text{MH}$)	0.751	0.036	94.2	93.2
32	0.5	Unbalanced	Estimating equations ($w_i = 1$)	0.749	0.052	94.0	92.4
32	0.5	Unbalanced	Estimating equations ($w_i = \text{opt(est)}$)	0.753	0.037	93.6	93.0
32	0.5	Unbalanced	Estimating equations ($w_i = \text{opt(true)}$)	0.751	0.036	94.6	92.6
32	0.5	Unbalanced	MLE	0.751	0.036	94.4	
32	Heterogeneous	Unbalanced	Estimating equations ($w_i = \text{MH}$)	0.750	0.025	95.2	93.8
32	Heterogeneous	Unbalanced	Estimating equations ($w_i = 1$)	0.752	0.047	93.4	92.4
32	Heterogeneous	Unbalanced	Estimating equations ($w_i = \text{opt(est)}$)	0.754	0.024	93.8	91.0
32	Heterogeneous	Unbalanced	Estimating equations ($w_i = \text{opt(true)}$)	0.750	0.025	95.2	92.8
32	Heterogeneous	Unbalanced	MLE	0.750	0.024	94.6	
32	0.5	10	Estimating equations ($w_i = \text{MH}$)	0.751	0.071	94.0	93.0
32	0.5	10	Estimating equations ($w_i = \text{opt(est)}$)	0.771	0.067	95.0	93.8
32	0.5	10	Estimating equations ($w_i = \text{opt(true)}$)	0.751	0.071	94.0	93.0
32	0.5	10	MLE	0.749	0.071	93.2	
32	Heterogeneous	10	Estimating equations ($w_i = \text{MH}$)	0.755	0.062	94.8	96.0
32	Heterogeneous	10	Estimating equations ($w_i = \text{opt(est)}$)	0.781	0.058	92.6	91.8
32	Heterogeneous	10	Estimating equations ($w_i = \text{opt(true)}$)	0.755	0.062	94.8	96.0
32	Heterogeneous	10	MLE	0.752	0.067	91.4	
32	0.5	3	Estimating equations ($w_i = \text{MH}$)	0.760	0.126	94.2	94.4
32	0.5	3	Estimating equations ($w_i = \text{opt(est)}$)	0.829	0.113	92.8	93.6
32	0.5	3	MLE	0.735	0.147	84.2	
32	Heterogeneous	3	Estimating equations ($w_i = \text{MH}$)	0.761	0.116	94.6	96.4
32	Heterogeneous	3	Estimating equations ($w_i = \text{opt(est)}$)	0.829	0.098	87.0	86.7
32	Heterogeneous	3	MLE	0.743	0.126	80.6	
4	0.5	10	Estimating equations ($w_i = \text{MH}$)	0.767	0.207	95.0	85.6
4	0.5	10	Estimating equations ($w_i = \text{opt(est)}$)	0.782	0.200	95.6	85.0
4	0.5	10	MLE	0.771	0.212	93.6	
4	Heterogeneous	10	Estimating equations ($w_i = \text{MH}$)	0.770	0.192	94.8	86.8
4	Heterogeneous	10	Estimating equations ($w_i = \text{opt(est)}$)	0.780	0.173	94.6	81.4
4	Heterogeneous	10	MLE	0.770	0.182	92.6	

case; also in this case the MLE is less efficient. There is little difference in the coverage rates of the binomial and the robust based confidence intervals for the estimating equation method, and both appear adequate, except for $I = 4$ when the robust confidence intervals appear to be too narrow.

Table IV shows the results for the comparison of the Mantel–Haenszel estimating equation method with the MLE, with the particular emphasis on the effect of overdispersion. All results here are for $I = 32$. The results for bias and efficiency are not shown because there was no evidence of bias or differences in efficiency. The results show that the standard errors based on the binomial distribution can lead to reduced coverage rates when there is overdispersion; this is particularly the case if π_i is heterogeneous and N is unbalanced. The coverage rates of the robust confidence intervals are less affected by overdispersion.

Table IV. Monte Carlo results, comparisons of coverage rates of MLE and estimating equation approach, effect of overdispersion.

ψ_i	π_i	Design N	Estimating equations Binomial	Estimating equations Robust	MLE Binomial
0.75	0.5	10	94.0	93.0	93.2
0.75	Heterogeneous	10	94.8	96.0	91.4
0.75	0.5	Unbalanced	94.2	93.2	94.4
0.75	Heterogeneous	Unbalanced	95.2	93.8	94.6
$\beta(9,3)$	0.5	10	93.0	93.8	92.0
$\beta(9,3)$	Heterogeneous	10	91.2	94.0	87.2
$\beta(9,3)$	0.5	Unbalanced	85.8	93.0	85.8
$\beta(9,3)$	Heterogeneous	Unbalanced	74.2	90.4	66.4

Table V shows the results for the comparison of the MLE, Bayesian method and the estimating equations method. For the estimating equation approach the robust standard errors are used. Because of the computational intensity, only a few scenarios were considered here. The results show very little bias except for the Bayesian scheme for small I. Other than this the Bayesian and the MLE results are very similar. There is very little difference between the efficiency of the three methods except for small I where the Bayesian scheme is less efficient. The effect of overdispersion leads to inadequate coverage rates of the MLE and Bayesian schemes, but not in general for the estimating equation approach. The relative merits of the three schemes appear not to be influenced by whether the π_i's are homogeneous or heterogeneous and by whether N is unbalanced or balanced. The case $N = 1000$ is included to emphasize the poor performance of the MLE and Bayes procedures for this model in the case of overdispersion. Changing the prior distribution on π_i and β in the Bayesian scheme had no real effect.

4.2. Two covariate case

In this part of the study we used two covariates with design points given by the real data in Table I. The true value of the β's are given by the MLE results in line 1 of Table II. The true value of the π's were either the MLE results (labelled Heterogeneous) or 0.26 for all i. We generated data from both a correct model and an overdispersed model. For the overdispersion case, values of ψ_i were generated from a beta$(a(i),3)$ where $a(i) = 3 \times \exp(X_i\beta)$. All results are based on 1000 replications. We experienced a few cases for which the estimating equations solution or the MLE were abnormally large. To avoid the effect of these outliers on the summary of the findings we used median and interquartile range, instead of mean and SD to express the bias and efficiency. The results are given in Table VI. Unlike the no covariate case we do see more differences between the EE and MLE approaches. For the no overdispersion case the MLE is more efficient. Using MH weights is more efficient than unweighted estimating equations, but somewhat surprisingly gives worse coverage rate of confidence intervals. Overdispersion has less effect on properties of the EE approach, but does result in a loss of efficiency and poor coverage rate for the MLE approach.

Table V. Monte Carlo results, comparisons of MLE, Bayesian and estimating equation approach.

I	ψ_i	π_i	N	Method	Mean	SD	Coverage
32	0.75	Heterogeneous	Unbalanced	Bayes	0.745	0.023	93.8
32	0.75	Heterogeneous	Unbalanced	MLE	0.748	0.023	94.4
32	0.75	Heterogeneous	Unbalanced	Estimating equations (MH)	0.749	0.025	93.2
32	$\beta(9,3)$	Heterogeneous	Unbalanced	Bayes	0.757	0.044	73.8
32	$\beta(9,3)$	Heterogeneous	Unbalanced	MLE	0.761	0.044	72.8
32	$\beta(9,3)$	Heterogeneous	Unbalanced	Estimating equations (MH)	0.752	0.042	92.2
32	0.75	Heterogeneous	10	Bayes	0.754	0.072	88.2
32	0.75	Heterogeneous	10	MLE	0.759	0.070	92.2
32	0.75	Heterogeneous	10	Estimating equations (MH)	0.754	0.068	94.8
32	$\beta(9,3)$	Heterogeneous	10	Bayes	0.751	0.075	88.0
32	$\beta(9,3)$	Heterogeneous	10	MLE	0.752	0.074	91.2
32	$\beta(9,3)$	Heterogeneous	10	Estimating equations (MH)	0.749	0.070	91.2
32	0.75	0.5	Unbalanced	Bayes	0.740	0.034	93.8
32	0.75	0.5	Unbalanced	MLE	0.751	0.035	95.2
32	0.75	0.5	Unbalanced	Estimating equations (MH)	0.751	0.035	92.6
32	$\beta(9,3)$	0.5	Unbalanced	Bayes	0.746	0.045	87.0
32	$\beta(9,3)$	0.5	Unbalanced	MLE	0.758	0.045	87.8
32	$\beta(9,3)$	0.5	Unbalanced	Estimating equations (MH)	0.752	0.045	94.4
4	0.75	Heterogeneous	Unbalanced	Bayes	0.781	0.090	85.6
4	0.75	Heterogeneous	Unbalanced	MLE	0.754	0.068	95.0
4	0.75	Heterogeneous	Unbalanced	Estimating equations (MH)	0.754	0.071	77.2
4	$\beta(9,3)$	Heterogeneous	Unbalanced	Bayes	0.772	0.146	59.0
4	$\beta(9,3)$	Heterogeneous	Unbalanced	MLE	0.745	0.128	67.4
4	$\beta(9,3)$	Heterogeneous	Unbalanced	Estimating equations (MH)	0.741	0.123	59.8
32	0.75	Heterogeneous	1000	Bayes	0.750	0.0059	93.8
32	0.75	Heterogeneous	1000	MLE	0.750	0.0059	95.8
32	0.75	Heterogeneous	1000	Estimating equations (MH)	0.750	0.0064	93.8
32	$\beta(9,3)$	Heterogeneous	1000	Bayes	0.758	0.027	30.6
32	$\beta(9,3)$	Heterogeneous	1000	MLE	0.758	0.027	30.2
32	$\beta(9,3)$	Heterogeneous	1000	Estimating equations (MH)	0.750	0.024	94.4
32	0.75	Heterogeneous	Unbalanced	Bayes, Prior($\pi_i \sim$ beta$(2,3)$)	0.745	0.024	94.2
32	$\beta(9,3)$	Heterogeneous	Unbalanced	Bayes, Prior($\pi_i \sim$ beta$(2,3)$)	0.753	0.046	67.4
32	0.75	Heterogeneous	Unbalanced	Bayes, Prior($\beta \sim$ N$(0,1)$)	0.746	0.025	92.6
32	$\beta(9,3)$	Heterogeneous	Unbalanced	Bayes, Prior($\beta \sim$ N$(0,1)$)	0.756	0.044	71.2

5. DISCUSSION

The results in this paper show that the estimating approach with Mantel–Haenszel weights is a useful method of analysing data when the response is the ratio of binomial probabilities in the case of a large number of strata. The MLE is inconsistent in this case, although there did not appear to be any meaningful bias in our study. Large I and small N would be needed for the bias to be substantial. The estimating equation approach appears to handle overdispersion better than the two likelihood based approaches. However, in fairness to these two methods,

Table VI. Monte Carlo results for two covariate case. Comparisons of MLE and estimating equation (EE) approach, effect of different weights and overdispersion.

I	π_i	Over dispersion	Method	β_1 Median	IQR	95 per cent CI	β_2 Median	IQR	95 per cent CI
	True value			−0.152			0.0803		
25	Heterogeneous	No	EE ($w_i = 1$)	−0.148	0.063	91.5	0.079	0.038	93.9
25	Heterogeneous	No	EE ($w_i = \mathrm{MH}$)	−0.146	0.053	84.7	0.075	0.029	88.8
25	Heterogeneous	No	MLE	−0.146	0.040	93.1	0.074	0.027	93.9
25	0.26	No	EE ($w_i = 1$)	−0.148	0.080	91.2	0.080	0.039	94.7
25	0.26	No	EE ($w_i = \mathrm{MH}$)	−0.145	0.058	88.4	0.074	0.024	92.3
25	0.26	No	MLE	−0.151	0.042	94.3	0.073	0.022	93.9
25	Heterogeneous	Yes	EE ($w_i = 1$)	−0.153	0.069	94.0	0.080	0.051	92.8
25	Heterogeneous	Yes	EE ($w_i = \mathrm{MH}$)	−0.149	0.064	90.9	0.075	0.051	90.0
25	Heterogeneous	Yes	MLE	−0.154	0.071	79.2	0.077	0.052	69.1
25	0.26	Yes	EE ($w_i = 1$)	−0.151	0.081	92.5	0.081	0.047	94.9
25	0.26	Yes	EE ($w_i = \mathrm{MH}$)	−0.151	0.067	92.2	0.076	0.050	87.9
25	0.26	Yes	MLE	−0.158	0.076	74.5	0.078	0.051	68.1

they could be adapted to allow for overdispersion. For example, using robust standard errors or standard errors from the least squares approach mentioned in Section 2.2 would achieve this for the MLE. The model used in the Bayesian approach could be extended. The results labelled Bayes OD in Table II are for an overdispersed version of each model. In these models each ψ_i is assumed to have a beta$(b\mu_i, b)$ distribution, where μ_i is chosen so that the beta has the correct mean as given by equations (3) and (4). The parameter b allows for overdispersion. There is little information in this data with regards to b, consequently we decide to fix b, and we started with an initial choice of $b = 3$. Exploration of the prior SD of ψ_i given $\mu_i = 0.5$ and b in the range of one to ten suggested that the choice did not matter greatly, and that $b = 1$ was perhaps too small and $b = 10$ was perhaps too large, and so we kept our original choice of $b = 3$. We note that other ways of determining overdispersion are possible. We see that including overdispersion, while changing the point estimates, more importantly increases the posterior standard deviations to a value more consistent with the estimating equation values.

ACKNOWLEDGEMENTS

This work was supported by NIH grants CA16042 and GM50011. The authors thank Xiaowei Yang for computing assistance. The majority of this research was performed while the authors were at UCLA.

REFERENCES

1. Greenland S, Robins JM. Estimation of a common effect parameter from sparse follow-up data. *Biometrics* 1985; **41**:55–68.
2. Greenland S. Modelling risk ratios from matched cohort data: an estimating equation approach. *Applied Statistics* 1994; **43**:223–232.
3. Suwinski R, Taylor JMG, Withers HR. Rapid growth of microscopic rectal cancer as a determinant of response to preoperative radiation therapy. *International Journal of Radiation Oncology Biology, Physics* 1998; **42**: 943–951.

4. Withers HR, Peters LJ, Taylor JMG. Dose response relationship for radiation therapy of subclinical disease. *International Journal of Radiation Oncology Biology*, *Physics* 1995; **31**:353–359.
5. Thames HD, Hendry JH. *Fractionation in Radiotherapy*. Taylor and Francis: London, 1987.
6. Chappell R, Nondahl DM, Fowler JF. Modelling dose and local control in radiotherapy. *Journal of the American Statistical Association* 1995; **90**:829–838.
7. Gart J. Approximate tests and interval estimation of common relative risk in the combination of 2*2 tables. *Biometrika* 1985; **72**:673–677.
8. Jennrich RI. Asymptotic properties of non-linear least squares estimators. *Annals of Mathematical Statistics* 1969; **40**:633–643.
9. Jennrich RI, Moore RH. Maximum likelihood estimation by means of nonlinear least squares. *American Statistical Association Proceedings of the Statistical Computing Section* 1975; **1**:57–65.
10. Thomas A, Spiegelhalter DJ, Gilks WR. BUGS: a program to perform Bayesian inference using Gibbs sampling. In *Bayesian Statistics 4*, Bernardo JM, Berger JO, Dawid AP, Smith AFM (ed.). Clarenden Press: Oxford, U.K., 1992; 837–842.
11. Carlin BP, Louis TA. *Bayes and Empirical Bayes Methods for Data Analysis.* Chapman and Hall: London, 1996.
12. Stijnen T, van Houwelingen HC. Relative risk, risk difference and rate difference models for sparse stratified data: a pseudo likelihood approach. *Statistics in Medicine* 1993; **12**:2285–2303.
13. Royall RM. Model robust confidence intervals using maximum likelihood estimators. *International Statistical Review* 1986; **54**:221–226.
14. Liang KY, Zeger SL. Longitudinal data analysis using generalised linear models. *Biometrika* 1986; **73**:13–22.
15. Nurminen M. Asymptotic efficiency of general noniterative estimators of common relative risk. *Biometrika* 1981; **68**:525–530.
16. Tarone RE. On summary estimators of relative risk. *Journal of Chronic Diseases* 1981; **34**:463–468.
17. Tarone RE, Gart JJ, Hauck WW. On the asymptotic inefficiency of certain noniterative estimators of a common relative risk or odds ratio. *Biometrika* 1983; **70**:519–522.
18. Sato T. A risk ratio estimator in both sparse-data and large-strata limiting models. *Bulletin of the Biometric Society of Japan* 1989; **10**:99–103.
19. Sato T. Estimation of common risk ratio in stratified case-control studies. *Statistics in Medicine* 1992; **11**: 1599–1605.
20. Withers HR, Taylor JMG, Maciejewski B. The hazard of accelerated tumour clonogen repopulation during radiotherapy. *Acta Oncologica* 1988; **27**:131–146.

