Statistica Sinica 4(1994), 169-185

# ESTIMATION IN THE EXPONENTIAL FAMILY IN THE PRESENCE OF NUISANCE PARAMETERS: COMPROMISE BETWEEN BIAS AND PRECISION

Yue-Cune Chang and Kung-Yee Liang

*Academia Sinica and Johns Hopkins University*

*Abstract:* The estimation of the common odds ratio in one-to-one matched case-control studies is a typical example of the trade-off between bias and precision in public health research. Liang and Zeger (1988) proposed an estimator through estimating functions. An alternative approach motivated by reducing asymptotic MSE was presented by Kalish (1990). In this paper, a finite sample approach is conducted under a more general framework. Comparisons for pair-matched case-control studies are made among these three estimators in terms of bias, MSE, coverage probability, and length of confidence interval. Extension to the multidimensional case is also presented.

*Key words and phrases:* Bias, case-control study, conditional likelihood, empirical bayes, estimating functions, profile likelihood, simulation.

## 1. Introduction

Very often, when statisticians face the problem of making inferences in multiparameter statistical models, attention is usually focused on only one or two of the parameters, called structual parameters, the others being regarded as *nuisance* (or *incidental*) parameters necessary to characterize the scientific problem but of no intrinsic interest. To describe the situation in general, consider a sequence of independent random vectors $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3, \ldots$. The distribution of $\mathbf{Y}_i$ depends on $\beta$ and $\alpha_i$, where the value of $\beta$ is independent of $i$, while the value of $\alpha_i$ changes with $i$. The real-valued parameter $\beta$ is of interest while the values $\alpha_1, \alpha_2, \alpha_3, \ldots$ are regarded as nuisance parameters. In this setting Neyman and Scott (1948) showed that simultaneous estimation of $\beta$ and the $\alpha$'s via the maximum likelihood method fails to have the usual asymptotic properties. In particular, $\hat{\beta}$ could fail to be consistent. A typical example of this kind is the estimation of a common *odds ratio* in a series of $K$ $2 \times 2$ tables with sparse data. Here the nuisance parameters, $\alpha_i$, $i = 1, 2, \ldots, K$, are necessary to characterize strata effects. Table 1 presents the usual format for displaying data from a one-to-one matched case-control study, a cross-classification of case-control pairs by joint exposure status. Each frequency in Table 1 represents the number of pairs. For example, there are $b$ out of $K$ $(= a+b+c+d)$ pairs in which the case is exposed and the matched control is nonexposed.

Table 1. Cross-classification of the $K$ case-control pairs by joint exposure status

| | Control | | |
|---|---|---|---|
| Case | Exposed | Nonexposed | Total |
| Exposed | $a$ | $b$ | $a+b$ |
| Nonexposed | $c$ | $d$ | $c+d$ |
| Total | $a+c$ | $b+d$ | $K$ pairs |

A common solution of estimating the structual parameter, $\beta$, is the application of the conditional likelihood approach achieved by conditioning the data on the minimum sufficient statistics $T$ for nuisance parameters (Andersen (1970)). The merit of this approach is to focus the inference on a genuine likelihood which depends only on the parameter of interest, so the effects of nuisance parameters can be reduced. However, it is understandably frustrating for researchers to use only a subset of the data given the effort made to collect the whole data, especially when the subset is relatively small compared to the whole data set. A nice illustrative example is the one from a matched study of endometrial cancer and oral conjugated estrogen use reported in Schlesselman (1982). For this example the entries are $a = 12$, $b = 43$, $c = 7$, $d = 121$. Less than one-third of 183 pairs was used for the conditional estimate ($\hat{\beta}_S = \ln(b/c) = 1.82$, with s.e.$(\hat{\beta}_S) = (1/b + 1/c)^{1/2} = 0.4076$). An alternative approach is ignoring matching and using $\hat{\beta}_P = \ln[(a+b)(b+d)/((a+c)(c+d))]$ as an estimator for $\beta$. For the endometrial cancer example, $\hat{\beta}_P$ equals 1.31 with s.e.$(\hat{\beta}_P) = 0.291$. A primary reason that $\hat{\beta}_P$ has seldomly been used is that it is biased except when the matching (or stratifying) is indeed unnecessary (Breslow and Day (1980, p.271)). However, if the strata are indeed homogeneous, $\hat{\beta}_P$ would be preferred, due to its higher precision. The typical trade-off between bias and precision can be seen in this case.

It seems sensible to choose an estimator which combines the unbiased property of $\hat{\beta}_S$ and the higher precision of $\hat{\beta}_P$. Kalish (1990) proposed an optimal estimator within the family of weighted averages between $\hat{\beta}_S$ and $\hat{\beta}_P$ which minimizes the asymptotic mean squared error (AMSE), a criterion based solely on the large sample property. Instead of working on $\hat{\beta}_S$ and $\hat{\beta}_P$, Liang and Zeger (1988) used the corresponding estimating functions, $H_S(\beta) = b - ce^{\beta}$ and $H_P(\beta) = (a+b)(b+d) - (a+c)(c+d)e^{\beta}$, respectively. After establishing a heuristic criterion, they ended up with a weight, $W_{LZ} = bc/(ad)$. That is, the estimating function they used for estimating $\beta$ is

$$H_{LZ}(\beta) = (1 - W_{LZ})H_S(\beta) - W_{LZ}H_P(\beta)/K. \tag{1.1}$$

Both methods significantly reduce bias and improve precision simultaneously, compared to the method using either $\hat{\beta}_S$ or $\hat{\beta}_P$ alone. However, the estimator proposed by Kalish (1990), $\hat{\beta}_K$, is based on large sample theory. This is unsatisfactory, at least conceptually, as the problem one is facing is purely finite sample. On the other hand, the estimator presented by Liang and Zeger, $\hat{\beta}_{LZ}$, does not satisfy a crucial criterion, namely, it does not converge to $\beta_o$ (the true $\beta$ value) when $K \to \infty$.

In this paper, under a more general framework, we consider a family of weighted averages between two standardized estimating functions. One is obtained from the conditional likelihood approach, the other is derived from the profile likelihood method in which the heterogeneity among strata is ignored. An "optimal" weight was selected on the basis of the criterion proposed by Godambe (1960) and Godambe and Thompson (1974). We show, in §2.2 and §2.3, some desirable properties of the proposed estimator. In Section 3.1, application to the endometrial cancer study is presented. In Section 3.2, comparisons between this estimator, Liang and Zeger's estimator, and Kalish's estimator are made *via* simulations. Extension to the multidimensional case is presented in Section 4, followed by discussion.

## 2. The Proposed Method

### 2.1. Notations and estimating functions

We consider the problem of estimating $\beta$, the parameter of interest, in the presence of nuisance parameters, denoted by $\alpha_i$, $i = 1, 2, \ldots, K$. More specifically, suppose that there are $K$ independent vectors of observations $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_K$, where the $\mathbf{Y}_i$, each with dimension $r_i$, come from a relatively homogeneous stratum, while the strata are heterogeneous. It is assumed that given $\alpha_i$, the $i$th vector of observations $\mathbf{Y}_i$ has a joint density $f_i$ of the form

$$f_i(\boldsymbol{y}_i; \beta, \alpha_i) = \exp\{\beta S_i + \alpha_i T_i - \phi_i(\beta, \alpha_i)\}. \tag{2.1}$$

It is clear that $S_i$ and $T_i$ are sufficient statistics for $(\beta, \alpha_i)$ in the $i$th stratum. Furthermore, for fixed $\beta$, $T_i$ is a sufficient statistic for $\alpha_i$.

To estimate $\beta$, Lindsay (1982) suggests the use of a "conditional score" function to eliminate the nuisance parameters. Here, for the $i$th stratum, the conditional score is defined as

$$h_{iS}(\beta) = U_i - E(U_i \mid T_i; \beta),$$

which is the residual of the $\beta$-score, $U_i = \partial \log f_i/\partial\beta$, under its best prediction by a function of $T_i$. It can also be derived as the $\beta$-derivative of the conditional log likelihood of the data given $T_i$. For (2.1), we can rewrite the above conditional score function as

$$h_{iS}(\beta) = S_i - E(S_i \mid T_i; \beta),$$

which is functionally independent of $\alpha_i$. Therefore, an estimating function based on $K$ strata is

$$H_S(\beta) = \sum_{i=1}^{K} h_{iS}(\beta) = \sum_{i=1}^{K}[S_i - E(S_i \mid T_i; \beta)], \tag{2.2}$$

where the subindex "$S$" stands for "Stratified". When the strata are heterogeneous, this conditional score function has previously been shown to generate the optimal estimating equation for $\beta$ when the conditioning statistic is complete and sufficient for the nuisance parameters (Godambe (1976)), which is true in the exponential family considered here.

With homogeneous strata, i.e. $\alpha_i \equiv \alpha$, for all $i = 1, 2, \ldots, K$; or, more formally, we may assume that the $\alpha_i$'s are unobserved i.i.d. random variables with mean $\alpha$ and variance $\theta = 0$. The likelihood function reduces to

$$\prod_{i=1}^{K} f_i(y_i; \beta, \alpha) = \exp\Big\{\beta \sum S_i + \alpha \sum T_i - \sum \phi_i(\beta, \alpha)\Big\}.$$

Thus, $\sum S_i$ and $\sum T_i$ are sufficient statistics for $(\beta, \alpha)$ when $\theta = 0$. In this situation, the maximum likelihood estimate of $\beta$ is seen as the solution of the profile score equation, i.e.

$$H_P(\beta) = \sum_i \left[S_i - \frac{\partial \phi_i(\beta, \hat{\alpha}(\beta))}{\partial \beta}\right] = \sum_i h_{iP}(\beta) = 0, \tag{2.3}$$

where $\hat{\alpha}(\beta)$ is the maximum likelihood estimate of $\alpha$ given $\beta$. We note that for the one-to-one matched study mentioned in §1, $H_P(\beta)$ reduces to $(a+b)(b+d) - (a+c)(c+d)e^{\beta}$.

### 2.2. The proposed estimator

In this subsection, we propose an optimal estimator within the family of weighted averages of the standardized versions of $H_S(\beta)$ and $H_P(\beta)$. The standardization, namely, $H_S^*(\beta) = [E(-\frac{\partial H_S(\beta)}{\partial \beta})]^{-1} H_S(\beta)$ and $H_P^*(\beta) = [E(-\frac{\partial H_P(\beta)}{\partial \beta})]^{-1}$ $H_P(\beta)$, is necessary to insure that they are comparable; see for example (1.1). Moreover, we note that it is appropriate to allow the rescaling factors, $[E(-\frac{\partial H_S(\beta)}{\partial \beta})]^{-1}$ and $[E(-\frac{\partial H_P(\beta)}{\partial \beta})]^{-1}$ to depend on $\beta$, because, we believe that it is more similar to the original estimating function in spirit to compare the method of substitution of an estimator. The resulting estimating functions, $H_S^*(\beta)$ and $H_P^*(\beta)$, are the so called "standardized" estimating functions (see Godambe (1976) for reference). Thus, we have a family of estimating functions for $\beta$ defined as follows

$$H(W,\beta) = (1-W)H_S^*(\beta) + WH_P^*(\beta),$$

where $0 \le W \le 1$. Moreover, we choose $W^*$ to minimize

$$Q(W) = E[H^2(W,\beta)],$$

the expected mean square of $H(W,\beta)$, where the expectation is taken with respect to $\mathbf{Y}$ and $\alpha$.

The reason for choosing this criterion is that we eventually solve the estimating equation, $H(W,\beta)=0$, for $\beta$ to obtain an estimator. In other words, we want the value of $H(W,\beta)$ to cluster around 0, as much as possible (i.e. $E(H^2(W,\beta))$ should be as small as possible). Furthermore, it is important to note that both terms in the average have been standardized by their own scale.

After a straightforward calculation, it is easily seen that the optimal weight is

$$W^* = \frac{\left[E\left(-\frac{\partial H_S}{\partial\beta}\right)\right]^{-1} - \left[E\left(-\frac{\partial H_P}{\partial\beta}\right)\right]^{-1}}{\left[E\left(-\frac{\partial H_S}{\partial\beta}\right)\right]^{-1} - 2\left[E\left(-\frac{\partial H_P}{\partial\beta}\right)\right]^{-1} + \frac{E(H_P^2(\beta))}{[E(-\frac{\partial H_P(\beta)}{\partial\beta})]^2}}. \qquad (2.4)$$

The "optimal" estimator we propose, denoted by $\hat\beta$, is the one which satisfies the following estimating equation

$$H(W^*,\hat\beta) = (1-W^*)H_S^*(\hat\beta) + W^*H_P^*(\hat\beta) = 0.$$

Note that since $W^*$ depends on $\beta$ as well, an iteration procedure is needed which is outlined as follows

**Step 1.** Take $\hat\beta = \hat\beta_S$.

**Step 2.** Compute

$$\hat W^* = \frac{\left[\sum_i\left(-\frac{\partial h_{iS}}{\partial\beta}\Big|_{\beta=\hat\beta}\right)\right]^{-1} - \left[\sum_i\left(-\frac{\partial h_{iP}}{\partial\beta}\Big|_{\beta=\hat\beta}\right)\right]^{-1}}{\left[\sum_i\left(-\frac{\partial h_{iS}}{\partial\beta}\Big|_{\beta=\hat\beta}\right)\right]^{-1} - 2\left[\sum_i\left(-\frac{\partial h_{iP}}{\partial\beta}\Big|_{\beta=\hat\beta}\right)\right]^{-1} + \frac{(\sum_i h_{iP}(\hat\beta))^2}{[\sum_i(-\frac{\partial h_{iP}}{\partial\beta}|_{\beta=\hat\beta})]^2}}.$$

Note that $\hat W^*$ is restricted to the interval $[0,1]$.

**Step 3.** Update $\hat{\beta}$ by solving $\tilde{H}(\hat{W}^*, \beta) = 0$, where

$$\tilde{H}(\hat{W}^*, \beta) = (1 - \hat{W}^*)\Big[ - \frac{\partial H_S}{\partial \beta}\Big]^{-1} H_S(\beta) + \hat{W}^* \Big[ - \frac{\partial H_P}{\partial \beta}\Big]^{-1} H_P(\beta),$$

the empirical version of $H(\hat{W}^*, \beta)$.

**Step 4.** Repeat Steps 2 and 3, until convergence is evident.

The above iterative procedure has been used by Williams (1982) and Breslow (1984) for logistic and log-linear models with extra variation, respectively. It was also used by Liang and Waclawiw (1990) to extend Stein's estimating procedure through the use of estimating functions.

### 2.3. Some properties of $W^*$ and $\hat{\beta}$

For fixed $\theta = \mathrm{Var}(\alpha_i) > 0$, we note that both $E(\partial H_S/\partial \beta)$ and $E(\partial H_P/\partial \beta)$ are $O_p(K)$, while

$$E(H_P^2(\beta))/E^2(-\partial H_P(\beta)/\partial \beta) = O_p(1).$$

Thus $W^*$ in (2.4) approaches 0 as $K \to \infty$, i.e. $W^* = o_p(1)$. That is, $H(W^*, \beta)$ is dominated by $H_S(\beta)$ as $K$ increases. On the other hand, with fixed $K$, it can be shown easily (Patefield (1977), Liang (1984)) that

$$E(H_P^2(\beta)) = E(-\partial H_P/\partial \beta)$$

when $\theta = 0$. Consequently, $W^*$ approaches 1 and $\hat{\beta} \to \hat{\beta}_P$ in probability as $\theta \to 0$. In other words, when there is no or a little heterogeneity among strata, the optimal estimator of $\beta$ reduces to $\hat{\beta}_P$ as desired.

Regarding the large sample distribution of $\hat{\beta}$, we have

**Theorem 1.** *The two estimators $\hat{\beta}$ and $\hat{\beta}_S$ are asymptotically equivalent, i.e.*

$$\sqrt{K}(\hat{\beta} - \hat{\beta}_S) = o_p(1). \tag{2.5}$$

A sketch of the proof is given in Appendix I. The result of the following corollary can be seen easily as we note that $\sqrt{K}(\hat{\beta}_S - \beta_o) \to N(0, V_S^{-1})$ (Andersen (1970)).

**Corollary 1.** *As $K \to \infty$, one has*

$$\sqrt{K}(\hat{\beta} - \beta_o) \xrightarrow{d} N(0, V_S^{-1}),$$

*where*

$$V_S = \lim_{K \to \infty} E(-\partial H_S(\beta)/\partial \beta|_{\beta = \beta_o})/K.$$

## 3. Application to One-to-One Matched Case-Control Studies

For one-to-one matched case-control studies, one has

$$\tilde{H}(\hat{W},\beta) \propto (1-\hat{W})(b/c - e^{\beta}) + \hat{W}[(a+b)(b+d)/((a+c)(c+d)) - e^{\beta}].$$

Thus, for given $\hat{W}$,

$$e^{\hat{\beta}} = (1-\hat{W})b/c + \hat{W}(a+b)(b+d)/[(a+c)(c+d)], \tag{3.1}$$

a weighted combination between $e^{\hat{\beta}_S}$ and $e^{\hat{\beta}_P}$, where the weight $\hat{W}$ is given in Appendix II. As discussed in §2.2, iterations between (3.1) and $\hat{W}$ are employed until convergence is evident.

### 3.1. Applications to the endometrial cancer data

We now apply the methods discussed in the paper to the endometrial cancer and oral conjugated estrogen use example introduced in §1. Results are summarized in Table 2. The compromise between bias and precision is achieved in this example by using either $\hat{\beta}_K, \hat{\beta}_{LZ}$, or the proposed estimator $\hat{\beta}$. For example the estimated 95% confidence interval for the common odds ratio $e^{\beta}$ based on the proposed method ranges roughly from 3 to 10 rather than 3 to 14 by using $\hat{\beta}_S$ or 2 to 7 by using $\hat{\beta}_P$.

To examine how typical the above described pattern may be, a simulation was conducted in which data were generated from a distribution with the parameter values observed for the endometrial data. Thus, $\beta = \ln(P_{10}/P_{01}) = 1.82$, the log odds ratio; $\phi = \ln[P_{11}P_{00}/(P_{10}P_{01})] = 1.57$, a measure of heterogeneity across pairs, and $\gamma = P_{11} + P_{01} = 0.1$, the probability of exposure for a control (Liang and Zeger (1988), mistakenly used 0.9). Here

$$\begin{aligned} P_{lk} &= Pr(Y_{i1} = l, Y_{i2} = k) \qquad\qquad (l,k = 0,1) \\ &= \int e^{\alpha_i(l+k)+\beta l}(1+e^{\alpha_i+\beta})^{-1}(1+e^{\alpha_i})^{-1}dF(\alpha_i), \end{aligned}$$

where $F$ is the unspecified distribution for $\alpha_i$. The simulation shown in Table 3 reveals that $\hat{\beta}_P$ is subject to serious bias in this case, while its variance is less than one-third that of $\hat{\beta}_S$. The negative bias and increased precision, however, result in very poor coverage probabilities for $\hat{\beta}_P$. The nominal 2.5% lower and upper intervals for $\hat{\beta}_S$ have actual error rates of 0.6% and 3.0%, a result of high variability of $\hat{\beta}_S$. On the other hand, the three compromise estimators performed reasonably well regarding mean squared error. The proposed estimator has a slight edge in terms of averaged confidence interval length.

### 3.2. Simulation study

A simulation study was conducted to compare the finite-sample performances of $\hat{\beta}_S$, $\hat{\beta}_P$, $\hat{\beta}_{LZ}$, $\hat{\beta}_K$, and $\hat{\beta}$. For each value of $\mathbf{P} = (P_{11}, P_{10}, P_{01}, P_{00})$, which was chosen from a subset of all possible values investigated by Liang and Zeger (1988), 1,000 independent realizations of $\mathbf{T} = (a, b, c, d)$ were generated from a multinomial distribution with probability $\mathbf{P}$ and sample size $K = 60, 100$, and 200. To avoid numerical problems, we increased both $b$ and $c$ by 0.5 when either $b$ or $c$ was equal to zero.

The results for $\hat{\beta}_S$, $\hat{\beta}_P$, $\hat{\beta}_{LZ}$, and $\hat{\beta}_K$ are consistent with those in Liang and Zeger (1988) and Kalish (1990). For this reason, we focus on the comparison between $\hat{\beta}_{LZ}$, $\hat{\beta}_K$, and $\hat{\beta}$. Results from Tables 4 and 5 suggest that in general $\hat{\beta}_K$ and the proposed estimator $\hat{\beta}$ are comparable in terms of bias and mean squared errors, with both showing improvement over $\hat{\beta}_{LZ}$.

Tables 6 and 7 present, respectively, the lower and upper confidence limits for the true coverage probabilities of a nominal 95% confidence interval, respectively. Entries in these two tables are the observed probabilities of failing to cover the true $\beta$ minus the nominal probability (0.025), divided by 0.005, the standard deviation of the estimate based on 1,000 replications. All values are rounded to the nearest integer. Except when the true $\beta$ is zero, the proposed method appears to perform reasonably well relative to the other two procedures. These discrepancies are especially pronounced in the upper limit coverage when $\beta = 2$.

## 4. Extension to the Multidimensional Case

This section discusses briefly the extension of the results in §2 to the multidimensional case, i.e. $\boldsymbol{\beta}$ is a $q \times 1$ vector. This is common in matched case-control studies where investigators search for the joint effects of several risk factors. We adopt the same assumptions as in §2 except now that $\mathbf{S}_i$, and hence $\mathbf{H}_S$ and $\mathbf{H}_P$, are of dimension $q$. Consequently one needs to modify the optimality criterion for choosing $W^*$ by minimizing instead

$$Q(W) = Tr\{E[\mathbf{H}(W, \boldsymbol{\beta})\mathbf{H}'(W, \boldsymbol{\beta})]\},$$

where $Tr\{\cdot\}$ denotes the trace of a $q \times q$ matrix. Tedious, yet straightforward, calculation gives

$$W^* = \frac{Tr\Big\{\Big[E\Big(-\frac{\partial \mathbf{H}_S}{\partial \boldsymbol{\beta}}\Big)\Big]^{-1}\Big\} - Tr\Big\{\Big[E\Big(-\frac{\partial \mathbf{H}_p}{\partial \boldsymbol{\beta}}\Big)\Big]^{-1}\Big\}}{Tr\Big\{\Big[E\Big(-\frac{\partial \mathbf{H}_S}{\partial \boldsymbol{\beta}}\Big)\Big]^{-1}\Big\} - 2\,Tr\Big\{\Big[E\Big(-\frac{\partial \mathbf{H}_p}{\partial \boldsymbol{\beta}}\Big)\Big]^{-1}\Big\} + Tr\{E[\mathbf{H}_P^*\mathbf{H}_P^{*'}]\}}.$$

Again, it can be shown that (i) $W^*$ approaches 0 as $K \to \infty$ and (ii) $W^*$ approaches 1 as $\mathrm{Var}(\alpha_i) = \theta \to 0$.

It should be noted that there are two major issues for extending to the multivariate case. The first is the choice of weight, $W$, among real-valued, real vector-valued, and matrix weight. We used real-valued weight for simplicity. The second is in regard to the choice of optimal criteria, namely M-optimality (based on non-negative definite matrix), D-optimality (based on determinant), and T-optimality (based on trace). All these three optimality definitions refer to the variance-covariance matrix of a matrix-standardized unbiased statistical estimation function (see Chandrasekar & Kale (1984)). Our weight is optimal if one accepts the use of trace as a criterion.

## 5. Discussion

Stratification has been proved to be a useful tool in accounting for confounding effects in observational studies. Potential problems exist when many strata are uninformative due to either over stratification or sparse data. On the other hand, ignoring stratification in the analysis stage generally results in biased estimation. In this note we have introduced an estimating procedure which serves to compromise between bias and precision. This procedure has empirical Bayes flavor in that the weight $W^*$ is dictated by the degree of heterogeneity among strata which can be estimated empirically. We have presented, in §2.3, some statistical properties of the proposed method, which are intuitively desirable. Through simulations, we have demonstrated that the proposed method performs well in one-to-one matched case-control studies.

The notion of trade-off between bias and precision is well recognized in the statistical literature. We believe that the proposed method, which accomodates the estimating function method, is intuitive and likely to be useful in many applications. Work is currently under way in applying the same method to survival data from clinical trials and family studies, where one encounters a similar situation, namely, there are many strata which are sparse.

**Acknowledgements**

This work was supported in part by grant GM39261 from the National Institutes of Health.

Table 2. Estimates and standard errors based on $\hat{\beta}_S$, $\hat{\beta}_P$, $\hat{\beta}_{LZ}$, $\hat{\beta}_K$ and the proposed estimator $\hat{\beta}$ for the endometrial cancer data (Schlesselman (1982)).

| | $\hat{\beta}_S$ | $\hat{\beta}_P$ | $\hat{\beta}_{LZ}$ | $\hat{\beta}_K$ | $\hat{\beta}$ |
|---|---|---|---|---|---|
| | 1.815 | 1.311 | 1.674 | 1.686 | 1.670 |
| | (0.408) | (0.291) | (0.379) | (0.358) | (0.327) |
| Estimated $W$ | 0 | 1 | 0.207 | 0.257 | 0.255 |
| 95% confidence interval for $e^{\beta}$ | (2.76, 13.66) | (2.10, 6.56) | (2.54, 11.21) | (2.68, 10.89) | (2.80, 10.08) |

Table 3. Simulation results for comparing $\hat{\beta}_S$, $\hat{\beta}_P$, $\hat{\beta}_{LZ}$, $\hat{\beta}_K$, and $\hat{\beta}$, for a population similar to that of the endometrial cancer example where $\beta = 1.82$, $\phi = 1.57$, $\gamma = 0.1$, with the sample size $K = 183$ and 1000 replications.

| | $\hat{\beta}_S$ | $\hat{\beta}_P$ | $\hat{\beta}_{LZ}$ | $\hat{\beta}_K$ | $\hat{\beta}$ |
|---|---|---|---|---|---|
| $E(\hat{\beta})$ | 1.87 | 1.33 | 1.74 | 1.74 | 1.74 |
| $\mathrm{Var}(\hat{\beta})$ | 0.19 | 0.06 | 0.17 | 0.17 | 0.18 |
| bias(%) | 5.00 | −49.00 | −8.00 | −8.00 | −8.00 |
| MSE | 0.20 | 0.30 | 0.18 | 0.18 | 0.19 |
| Error rate (%) of nominal 2.5% lower C.I. | 0.60 | 0.00 | 0.10 | 0.50 | 0.50 |
| Error rate (%) of nominal 2.5% upper C.I. | 3.00 | 48.60 | 7.20 | 8.70 | 8.70 |
| Ratio of average confidence interval length to $\hat{\beta}$ | 1.21 | 0.79 | 1.12 | 1.04 | 1.00 |

Table 4. Bias (×100) in $\hat{\beta}_{LZ}$, $\hat{\beta}_K$, and $\hat{\beta}$.

| | | | Bias × 100 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\gamma = 0.1$ | | | $\gamma = 0.3$ | | | $\gamma = 0.5$ | | |
| $K$ | $\beta$ | $\phi$ | $\hat{\beta}_{LZ}$ | $\hat{\beta}_K$ | $\hat{\beta}$ | $\hat{\beta}_{LZ}$ | $\hat{\beta}_K$ | $\hat{\beta}$ | $\hat{\beta}_{LZ}$ | $\hat{\beta}_K$ | $\hat{\beta}$ |
| 60 | 0 | 0.00 | −2 | −2 | −2 | −1 | −1 | −1 | 2 | 2 | 2 |
| | | 0.25 | −3 | −3 | −3 | 1 | 1 | 1 | 0 | 0 | 0 |
| | | 1.00 | −4 | −4 | −4 | 2 | 2 | 2 | 1 | 1 | 1 |
| | 1 | 0.00 | 8 | 0 | 1 | 4 | −1 | 2 | 11 | 5 | 8 |
| | | 0.25 | 15 | 6 | 9 | 3 | −1 | 2 | 3 | −2 | 1 |
| | | 1.00 | 6 | −5 | −1 | −2 | −9 | −6 | −5 | −11 | −8 |
| | 2 | 0.00 | 23 | 12 | 17 | 15 | 7 | 11 | 14 | 3 | 8 |
| | | 0.25 | 13 | 3 | 6 | 8 | 1 | 5 | 7 | −2 | 3 |
| | | 1.00 | 6 | −6 | 0 | −8 | −13 | −8 | −1 | −9 | −2 |
| 100 | 0 | 0.00 | −1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | 0.25 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| | | 1.00 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 1 | 0.00 | 2 | −4 | −2 | 5 | 1 | 3 | 8 | 4 | 6 |
| | | 0.25 | 7 | 2 | 5 | −2 | −4 | −2 | −1 | −4 | −2 |
| | | 1.00 | 4 | −3 | 0 | −3 | −6 | −5 | −9 | −12 | −10 |
| | 2 | 0.00 | 16 | 10 | 13 | 9 | 4 | 7 | 8 | 1 | 4 |
| | | 0.25 | 9 | 3 | 7 | 1 | −2 | 1 | 0 | −5 | −2 |
| | | 1.00 | 2 | −3 | 1 | −16 | −13 | −11 | −4 | −5 | −2 |
| 200 | 0 | 0.00 | 2 | 2 | 2 | −1 | −1 | −1 | 0 | 0 | 0 |
| | | 0.25 | 3 | 2 | 2 | 1 | 1 | 1 | 0 | 0 | 0 |
| | | 1.00 | 0 | 0 | 0 | −1 | −1 | −1 | 1 | 1 | 1 |
| | 1 | 0.00 | 2 | −2 | 0 | 3 | 0 | 2 | 7 | 5 | 6 |
| | | 0.25 | 4 | 2 | 4 | −1 | −1 | 0 | −3 | −4 | −2 |
| | | 1.00 | 3 | 0 | 2 | −6 | −4 | −5 | −9 | −7 | −8 |
| | 2 | 0.00 | 13 | 11 | 13 | 4 | 2 | 3 | 6 | 2 | 4 |
| | | 0.25 | 1 | 1 | 2 | −4 | −3 | −2 | −3 | −4 | −2 |
| | | 1.00 | −4 | −1 | −1 | −18 | −9 | −10 | −11 | −3 | −4 |

Table 5. Mean squared errors for $\hat{\beta}_{LZ}$, $\hat{\beta}_K$, $\hat{\beta}$.

| | | | MSE × 100 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\gamma = 0.1$ | | | $\gamma = 0.3$ | | | $\gamma = 0.5$ | | |
| $K$ | $\beta$ | $\phi$ | $\hat{\beta}_{LZ}$ | $\hat{\beta}_K$ | $\hat{\beta}$ | $\hat{\beta}_{LZ}$ | $\hat{\beta}_K$ | $\hat{\beta}$ | $\hat{\beta}_{LZ}$ | $\hat{\beta}_K$ | $\hat{\beta}$ |
| 60 | 0 | 0.00 | 54 | 45 | 39 | 16 | 14 | 15 | 15 | 13 | 14 |
| | | 0.25 | 52 | 43 | 37 | 19 | 17 | 18 | 15 | 13 | 15 |
| | | 1.00 | 56 | 43 | 43 | 21 | 15 | 18 | 17 | 13 | 15 |
| | 1 | 0.00 | 37 | 32 | 31 | 16 | 15 | 15 | 21 | 19 | 20 |
| | | 0.25 | 42 | 33 | 34 | 20 | 19 | 20 | 18 | 17 | 18 |
| | | 1.00 | 39 | 33 | 34 | 24 | 25 | 24 | 21 | 22 | 21 |
| | 2 | 0.00 | 37 | 31 | 32 | 27 | 25 | 27 | 31 | 28 | 29 |
| | | 0.25 | 37 | 32 | 32 | 26 | 25 | 28 | 32 | 29 | 31 |
| | | 1.00 | 34 | 30 | 32 | 31 | 30 | 33 | 31 | 29 | 32 |
| 100 | 0 | 0.00 | 29 | 25 | 24 | 10 | 9 | 10 | 9 | 8 | 9 |
| | | 0.25 | 27 | 24 | 23 | 10 | 9 | 9 | 8 | 7 | 8 |
| | | 1.00 | 30 | 24 | 26 | 11 | 8 | 9 | 9 | 7 | 8 |
| | 1 | 0.00 | 22 | 20 | 20 | 9 | 9 | 9 | 11 | 11 | 11 |
| | | 0.25 | 23 | 20 | 21 | 10 | 11 | 10 | 10 | 10 | 10 |
| | | 1.00 | 25 | 22 | 23 | 13 | 15 | 14 | 12 | 14 | 13 |
| | 2 | 0.00 | 24 | 21 | 24 | 14 | 15 | 15 | 19 | 18 | 19 |
| | | 0.25 | 22 | 20 | 22 | 13 | 14 | 14 | 18 | 18 | 19 |
| | | 1.00 | 30 | 27 | 30 | 19 | 19 | 20 | 24 | 23 | 26 |
| 200 | 0 | 0.00 | 13 | 12 | 12 | 5 | 4 | 5 | 4 | 4 | 4 |
| | | 0.25 | 12 | 11 | 12 | 5 | 5 | 5 | 4 | 4 | 4 |
| | | 1.00 | 13 | 11 | 12 | 6 | 4 | 5 | 4 | 4 | 4 |
| | 1 | 0.00 | 9 | 9 | 9 | 5 | 5 | 5 | 5 | 5 | 5 |
| | | 0.25 | 10 | 10 | 10 | 5 | 5 | 5 | 5 | 5 | 5 |
| | | 1.00 | 11 | 11 | 11 | 7 | 8 | 7 | 6 | 7 | 7 |
| | 2 | 0.00 | 13 | 11 | 13 | 4 | 2 | 3 | 6 | 2 | 4 |
| | | 0.25 | 10 | 10 | 10 | 6 | 7 | 7 | 8 | 9 | 9 |
| | | 1.00 | 13 | 14 | 14 | 11 | 10 | 11 | 12 | 12 | 13 |

Table 6. Actual $\alpha$ level of nominal 2.5% lower confidence limits for $\hat{\beta}_{LZ}$, $\hat{\beta}_K$, and $\hat{\beta}$. Entries are the observed probabilities of failing to cover the true $\beta$ minus the nominal probability (.025), divided by 0.005, the standard deviation of the estimate based on 1,000 replications.

| | | | Transformed error rates for lower confidence limit | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\gamma = 0.1$ | | | $\gamma = 0.3$ | | | $\gamma = 0.5$ | | |
| $K$ | $\beta$ | $\phi$ | $\hat{\beta}_{LZ}$ | $\hat{\beta}_K$ | $\hat{\beta}$ | $\hat{\beta}_{LZ}$ | $\hat{\beta}_K$ | $\hat{\beta}$ | $\hat{\beta}_{LZ}$ | $\hat{\beta}_K$ | $\hat{\beta}$ |
| 60 | 0 | 0.00 | −1 | 0 | −3 | −1 | 0 | −4 | 0 | 0 | −2 |
| | | 0.25 | −1 | 0 | −4 | 1 | 1 | −2 | 0 | 0 | −1 |
| | | 1.00 | −2 | −1 | −3 | 0 | 0 | 1 | 0 | 1 | 1 |
| | 1 | 0.00 | −1 | −1 | −2 | −1 | −1 | −3 | 1 | 2 | −1 |
| | | 0.25 | −3 | −1 | 0 | −2 | −1 | 0 | −2 | −1 | −2 |
| | | 1.00 | −5 | −3 | −2 | −4 | −2 | 4 | −4 | −2 | 0 |
| | 2 | 0.00 | −2 | −1 | −4 | −2 | 0 | 10 | −4 | −2 | −3 |
| | | 0.25 | −2 | −3 | −5 | −3 | −1 | 12 | −4 | −1 | −2 |
| | | 1.00 | −4 | −4 | −5 | −5 | −5 | 14 | −5 | −4 | −2 |
| 100 | 0 | 0.00 | −1 | 0 | −3 | 0 | 0 | −2 | 1 | 1 | −1 |
| | | 0.25 | −1 | 0 | −3 | 0 | 0 | −2 | −1 | −1 | −2 |
| | | 1.00 | −1 | 0 | −2 | −1 | 0 | 0 | 0 | 1 | 0 |
| | 1 | 0.00 | 0 | 0 | −2 | 0 | 0 | −2 | 3 | 2 | −1 |
| | | 0.25 | −2 | −2 | −2 | −1 | 0 | −1 | 0 | 0 | −2 |
| | | 1.00 | −3 | −2 | 0 | −3 | −2 | 0 | −3 | −1 | 0 |
| | 2 | 0.00 | −3 | 0 | 3 | −2 | 0 | 0 | −3 | −1 | 0 |
| | | 0.25 | −4 | −3 | 2 | −3 | −1 | 0 | −4 | −1 | 0 |
| | | 1.00 | −5 | −5 | 10 | −5 | −5 | 2 | −5 | −4 | 9 |
| 200 | 0 | 0.00 | 1 | 1 | −2 | −1 | −1 | −2 | 0 | 0 | −1 |
| | | 0.25 | 0 | 0 | −3 | 2 | 2 | 0 | 1 | 1 | 0 |
| | | 1.00 | 0 | 0 | −1 | 0 | 0 | 0 | 1 | 1 | 2 |
| | 1 | 0.00 | −1 | −1 | −3 | 1 | 2 | −2 | 3 | 4 | −1 |
| | | 0.25 | 0 | 1 | 0 | −1 | 0 | −2 | −2 | −2 | −4 |
| | | 1.00 | −1 | 0 | 1 | −2 | −1 | 1 | −4 | −2 | −1 |
| | 2 | 0.00 | 2 | 5 | 4 | −1 | 0 | 0 | 0 | 0 | −1 |
| | | 0.25 | −4 | −2 | 0 | −2 | −1 | −1 | −2 | −1 | 0 |
| | | 1.00 | −4 | −3 | 3 | −5 | −3 | 2 | −4 | −2 | 7 |

Table 7. Actual $\alpha$ level of nominal 2.5% upper confidence limits for $\hat{\beta}_{LZ}$, $\hat{\beta}_K$, and $\hat{\beta}$. Entries are the observed probabilities of failing to cover the true $\beta$ minus the nominal probability (.025), divided by 0.005, the standard deviation of the estimate based on 1,000 replications.

| | | | Transformed error rates for upper confidence limit | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\gamma = 0.1$ | | | $\gamma = 0.3$ | | | $\gamma = 0.5$ | | |
| $K$ | $\beta$ | $\phi$ | $\hat{\beta}_{LZ}$ | $\hat{\beta}_K$ | $\hat{\beta}$ | $\hat{\beta}_{LZ}$ | $\hat{\beta}_K$ | $\hat{\beta}$ | $\hat{\beta}_{LZ}$ | $\hat{\beta}_K$ | $\hat{\beta}$ |
| 60 | 0 | 0.00 | −1 | 0 | 9 | 0 | 0 | 3 | 0 | 0 | 3 |
| | | 0.25 | 0 | 1 | 8 | 0 | 0 | 7 | 1 | 2 | 7 |
| | | 1.00 | −1 | 1 | 12 | −1 | 0 | 10 | −1 | 0 | 8 |
| | 1 | 0.00 | 0 | 1 | 0 | −1 | 1 | −1 | 0 | 1 | −1 |
| | | 0.25 | −1 | 0 | −1 | 0 | 3 | 0 | 0 | 4 | 0 |
| | | 1.00 | 1 | 6 | 4 | 3 | 16 | 8 | 3 | 15 | 8 |
| | 2 | 0.00 | −2 | 0 | −3 | −2 | 1 | −4 | −1 | 3 | −2 |
| | | 0.25 | −1 | 3 | −1 | 0 | 4 | −3 | 1 | 7 | −1 |
| | | 1.00 | 2 | 6 | 0 | 10 | 16 | 5 | 3 | 9 | 2 |
| 100 | 0 | 0.00 | 0 | 0 | 6 | 0 | 0 | 4 | 1 | 1 | 7 |
| | | 0.25 | −1 | 0 | 6 | 0 | 0 | 4 | 0 | 0 | 4 |
| | | 1.00 | −1 | −1 | 6 | −1 | −1 | 8 | 0 | 0 | 10 |
| | 1 | 0.00 | 1 | 3 | 0 | −1 | 0 | −2 | −2 | 0 | −3 |
| | | 0.25 | 0 | 2 | 0 | 1 | 4 | 1 | 1 | 4 | 0 |
| | | 1.00 | 0 | 5 | 1 | 5 | 18 | 10 | 5 | 17 | 8 |
| | 2 | 0.00 | −3 | −1 | −4 | −1 | 0 | −3 | 0 | 2 | −3 |
| | | 0.25 | 0 | 2 | −3 | 0 | 3 | −3 | 2 | 6 | −1 |
| | | 1.00 | 5 | 8 | 2 | 14 | 16 | 7 | 8 | 11 | 4 |
| 200 | 0 | 0.00 | 0 | 0 | 4 | 0 | 0 | 3 | 2 | 2 | 4 |
| | | 0.25 | −1 | −1 | 4 | −1 | −1 | 0 | 1 | 1 | 4 |
| | | 1.00 | −1 | −1 | 6 | 0 | 0 | 11 | 0 | 0 | 6 |
| | 1 | 0.00 | 0 | 2 | 0 | −1 | 1 | −2 | −3 | −2 | −4 |
| | | 0.25 | 0 | 0 | −2 | 2 | 4 | −1 | 3 | 6 | 0 |
| | | 1.00 | −1 | 4 | 0 | 7 | 12 | 10 | 10 | 15 | 11 |
| | 2 | 0.00 | −3 | −2 | −4 | −1 | 1 | −3 | −1 | 1 | −4 |
| | | 0.25 | 2 | 3 | −1 | 3 | 6 | −1 | 3 | 5 | −1 |
| | | 1.00 | 5 | 5 | 2 | 22 | 14 | 11 | 11 | 8 | 6 |

## Appendix

### (I) A sketch of the proof of Theorem 1

First we note, under some regularity conditions, that

$$\begin{aligned} A &= \sqrt{K}(H(W^*,\beta) - H_S(\beta)) \\ &= W^*\sqrt{K}(H_P^*(\beta) - H_S^*(\beta)) \\ &\xrightarrow{P} 0 \end{aligned} \tag{A.1}$$

as $K \to \infty$ since $\sqrt{K}W^* \to 0$ and both $H_P^*(\beta)$ and $H_S^*(\beta)$ are $O_p(1)$. Applying the Taylor expansion at $\hat\beta$ and $\hat\beta_S$ respectively, $A$ is approximated by

$$\begin{aligned} &\left[\frac{\partial H(W^*,\beta)}{\partial\beta}\right]\sqrt{K}(\beta-\hat\beta) - \left[\frac{\partial H_S(\beta)}{\partial\beta}\right]\sqrt{K}(\beta-\hat\beta_S) \\ =& \left[\frac{\partial H(W^*,\beta)}{\partial\beta}\right]\sqrt{K}(\hat\beta_S-\hat\beta) + \sqrt{K}(\beta-\hat\beta_S)\left[\frac{\partial H(W^*,\beta)}{\partial\beta} - \frac{\partial H_S(\beta)}{\partial\beta}\right] \\ =& \; B\cdot\sqrt{K}(\hat\beta_S-\hat\beta) + C. \end{aligned}$$

Using the fact that $\sqrt{K}(\beta-\hat\beta_S) = O_p(1)$, $W^* = o_p(1)$ and $\partial H_S(\beta)/\partial\beta = O_P(1)$, one has $B = O_p(1)$ and $C = o_p(1)$. Consequently, $\sqrt{K}(\hat\beta_S-\hat\beta) = o_p(1)$ by (A.1), and this completes the proof.

### (II) The exact terms of $E(-\partial H_S/\partial\beta), E(-\partial H_P/\partial\beta)$, and $E(H_P^2)$ in one-to-one matched study

Using the fact that $(a,b,c,d)$ has a multinomial distribution of size $K$ and cell probabilities $\mathbf{P} = (P_{11},P_{10},P_{01},P_{00})$, we can derive the exact values of $E(-\partial H_S/\partial\beta), E(-\partial H_P/\partial\beta)$, and $E(H_P^2)$. This in turn will provide an estimate of $W^*$ by replacing $\mathbf{P}$ with $\hat{\mathbf{P}} = (a/K, b/K, c/K, d/K)$. Straightforward but tedious calculation gives

$$\begin{aligned} E\left(-\frac{\partial H_S}{\partial\beta}\right) &= E\left(\frac{(b+c)e^\beta}{(1+e^\beta)^2}\right) \\ &= \frac{e^\beta}{(1+e^\beta)^2}[K(P_{01}+P_{10})] \\ &= \frac{P_{10}/P_{01}}{(P_{01}+P_{10})^2/P_{01}^2}[K(P_{10}+P_{01})] \\ &= \frac{K}{1/P_{10}+1/P_{01}}, \end{aligned}$$

$$E\left(-\frac{\partial H_P}{\partial\beta}\right) = E\left[\frac{[(a+b)(b+d)+(a+c)(c+d)]e^\beta}{(1+e^\beta)^2}\right]$$

$$\begin{aligned} &= \frac{e^{\beta}}{(1+e^{\beta})^2} E[(a+b)(b+d)+(a+c)(c+d)] \\ &= \frac{C_1 P_{10} P_{01}(P_{1+}P_{+0}+P_{+1}P_{0+})}{(P_{10}+P_{01})^2} + \frac{K}{1/P_{10}+1/P_{01}}, \end{aligned}$$

and

$$\begin{aligned} E(H_P^2(\beta)) &= E\left(\left[\frac{(a+b)(b+d)-(a+c)(c+d)e^{\beta}}{(1+e^{\beta})}\right]^2\right) \\ &= \frac{1}{(1+e^{\beta})^2} E\Big([(a+b)(b+d)-(a+c)(c+d)e^{\beta}]^2\Big) \\ &= \frac{N}{(1+e^{\beta})^2}, \end{aligned}$$

where

$$\begin{aligned} N = {} & C_3(P_{1+}P_{+0}-e^{\beta}P_{+1}P_{0+})^2 \\ & + C_2 \left( \begin{array}{c} P_{11}P_{+0}(P_{11}+P_{+0}) + P_{10}P_{+0}(6P_{1+}+P_{00}) \\ +e^{2\beta}(P_{11}P_{0+}(P_{11}+P_{0+}) + P_{01}P_{0+}(6P_{+1}+P_{00})) \\ -2e^{\beta}(P_{01}P_{+0}P_{1+} + P_{10}P_{0+}P_{+1} + P_{00}P_{1+}P_{+1} + P_{11}P_{+0}P_{0+}) \end{array} \right) \\ & + C_1 \left( \begin{array}{c} 3P_{11}P_{10} + 7P_{10}^2 + 3P_{10}P_{00} + P_{11}P_{00} \\ +e^{2\beta}(3P_{11}P_{01} + 7P_{01}^2 + 3P_{01}P_{00} + P_{11}P_{00}) \\ -2e^{\beta}(P_{11}P_{00} + P_{10}P_{01}) \end{array} \right) \\ & + K(P_{10}+e^{2\beta}P_{01}), \end{aligned}$$

and

$$\begin{aligned} P_{i+} &= P_{i1} + P_{i0}, \quad (i=0,1) \\ P_{+j} &= P_{1j} + P_{0j}, \quad (j=0,1) \\ C_3 &= K(K-1)(K-2)(K-3), \\ C_2 &= K(K-1)(K-2), \\ C_1 &= K(K-1). \end{aligned}$$

## References

Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *J. Roy. Statist. Soc. Ser.B* **32**, 283–301.

Breslow, N. E. (1984). Extra-Poisson variation in log-linear models. *Appl. Statist.* **33**, 38–44.

Breslow, N. E. and Day, N. E. (1980). *Statistical Methods in Cancer Research I: The Analysis of Case-Control Studies.* International Agency for Research in Cancer, Lyon, France.

Chandrasekar, B. and Kale, B. K. (1984). Unbiased statistical estimation functions for parameters in presence of nuisance parameters. *J. Statist. Plann. Inference* **9**, 45–54.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist.* **31**, 1208–1211.

Godambe, V. P. (1976). Conditional likelihood and unconditional optimum estimating equations. *Biometrika* **63**, 277–284.

Godambe, V. P. and Thompson, M. E. (1974). Estimating equations in the presence of a nuisance parameter. *Ann. Math. Statist.* **2**, 568–571.

Kalish, L. A. (1990). Reducing mean squared error in the analysis of pair-matched case-control studies. *Biometrics* **46**, 493–499.

Liang, K.-Y. (1984). The asymptotic efficiency of conditional likelihood methods. *Biometrika* **71**, 305–313.

Liang, K.-Y. and Zeger, S. L. (1988). On the use of concordant pairs in matched case-control studies. *Biometrics* **44**, 1145–1156.

Liang, K.-Y. and Waclawiw, M. A. (1990). Extension of the Stein estimating procedure through the use of estimating functions. *J. Amer. Statist. Assoc.* **85**, 435–440.

Lindsay, B. (1982). Conditional score functions: Some optimality results. *Biometrika* **69**, 503–512.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica* **16**, 1–32.

Patefield, W. M. (1977). On the maximized likelihood function. *Sankhyā Ser.B* **39**, 92–96.

Schlesselman, J. J. (1982). *Case-Control Studies: Design, Conduct, Analysis.* Oxford University Press, Oxford.

Williams, D. A. (1982). Extra-binomial variation in logistic linear models. *Appl. Statist.* **31**, 144–148.

Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan.
Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205, U.S.A.

(Received November 1992; accepted September 1993)