

Conditional Logistic Regression with Sandwich Estimators: Application to a Meta-Analysis

Michael P. Fay,^{1,*} Barry I. Graubard,¹
Laurence S. Freedman,¹ and Douglas N. Midthune²

¹Biometry Branch, Division of Cancer Prevention and Control,
National Cancer Institute, Executive Plaza North, Suite 344,
6130 Executive Boulevard, MSC 7354, Bethesda, Maryland 20892-7354, U.S.A.

²Information Management Services, Inc., 12501 Prosperity Drive,
Suite 200, Silver Spring, Maryland 20902, U.S.A.

SUMMARY

Motivated by a meta-analysis of animal experiments on the effect of dietary fat and total caloric intake on mammary tumorigenesis, we explore the use of sandwich estimators of variance with conditional logistic regression. Classical conditional logistic regression assumes that the parameters are fixed effects across all clusters, while the sandwich estimator gives appropriate inferences for either fixed effects or random effects. However, inference using the standard Wald test with the sandwich estimator requires that each parameter is estimated using information from a large number of clusters. Since our example violates this condition, we introduce two modifications to the standard Wald test. First, we reduce the bias of the empirical variance estimator (the middle of the sandwich) by using standardized residuals. Second, we approximately account for the variance of these estimators by using the t -distribution instead of the normal distribution, where the degrees of freedom are estimated using Satterthwaite's approximation. Through simulations, we show that these sandwich estimators perform almost as well as classical estimators when the true effects are fixed and much better than the classical estimators when the true effects are random. We achieve simulated nominal coverage for these sandwich estimators even when some parameters are estimated from a small number of clusters.

1. Introduction

In this paper, we discuss a meta-analysis of animal experiments on dietary fat, calorie intake, and mammary tumors. We analyze these data using conditional logistic regression with sandwich (robust) variance estimators for clustered data. Although both conditional logistic regression and sandwich estimators are well known in the literature, we know of no detailed study of their combined use. We address small sample issues by using standardized residuals to estimate the middle of the sandwich estimator and using a t -distribution for the resulting Wald test with degrees of freedom estimated by Satterthwaite's approximation. We show through simulation and our meta-analysis that this combination of methods is useful in a practical situation.

In this paper, we re-examine the analyses of Table 5 in Freedman, Clifford, and Messina (1990) that model the effect of dietary fat and total caloric intake on mammary tumor incidence for two data sets, 43 experiments on Sprague-Dawley rats fed corn oil, and 17 experiments on mice breed for spontaneous tumors. Separate analyses were performed on the two data sets because the total caloric intake, and hence fat calorie intake, varies considerably between the species. Each experiment comprises two or more groups of animals receiving essentially the same treatment except for diets,

* Corresponding author's email address: m7f@helix.nih.gov

Key words: Generalized estimating equations; Random effects; Robust variance; Satterthwaite's formula; Standardized residuals.

which differ in the amount of fat and/or total calories fed to the animals. The primary response variable is the number of animals developing a mammary tumor in each group.

As an example, consider three experiments on Sprague-Dawley rats reported by Carroll and Khor (1970). For each experiment (or set in the terminology of Freedman et al., 1990), Carroll and Khor divided about 60 rats into 2 equal groups, a low-fat group that was fed 0.5% corn oil by weight and a high-fat group that was fed 20% corn oil by weight. At 50 days of age, all rats were given a single dose of a carcinogen, 7,12-dimethylbenz(α)anthracene (DMBA). From day 48 to 51 of age, all rats were fed a standard commercial diet to minimize the effect of the experimental diets on the absorption of the DMBA. The three experiments were identical except for the doses of DMBA, which were 1 mg, 2.5 mg, and 5 mg, and the follow-up times, which were 4 months after DMBA administration for the experiments with doses of 2.5 mg and 5 mg and 6 months after DMBA for the experiment with a dose of 1 mg. The experiment with a smaller dose was followed longer because the tumor incidence was smaller. In Carroll and Khor's (1970) experiments, the total caloric intake was not regulated or reported. Since most rats will self-regulate their intake to eat a constant amount of calories regardless of the fat content of the diet, we assumed the total caloric intake was constant within each of these three experiments.

Other experiments on the Sprague-Dawley rats differed in the type of carcinogen or dose, the amount of follow-up time, and the diets given each of the groups. There were typically 20–30 rats in each group and 2–4 groups per experiment. There was a total of 104 groups altogether for the 43 experiments on the Sprague-Dawley rats. Of these 43 experiments, 7 approximately fixed the total number of calories, 1 approximately fixed the fat calories, 3 varied both values, and 32 varied the fat calories but did not report the total caloric intake. For these latter experiments, we assumed the total caloric intake was constant within each experiment, as we did with the data from Carroll and Khor (1970). More details on the mice experiments are given in Section 4.

Freedman et al. (1990) assumed for both data sets that the data come from an unconditional logistic model

$$\text{logit}(p_{ij}) = \alpha_i + z'_{ij}\beta, \quad (1)$$

where p_{ij} is the expected proportion of responses (e.g., animals with tumors) in the j th group of the i th experiment, z_{ij} is the $p \times 1$ vector of corresponding covariates, α_i is the intercept parameter for the i th experiment, and β is the fixed diet parameter vector. The intercept terms are fixed nuisance parameters. For Freedman et al. (1990), z_{ij} consists of two terms, $(TCAL)_{ij}$, the average total calories consumed per animal per day, and $(FCAL)_{ij}$, the average total calories of fat consumed per animal per day for the j th group of the i th experiment. For those experiments with unrecorded total caloric intake, Freedman et al. (1990) imputed the average value for the species. For the Sprague-Dawley rats fed corn oil, the value of the imputed total calories does not affect the analysis. In fact, if for each of the experiments with unrecorded total calories the total calories were known and constant within the experiment, we would obtain the same results because the constant value will be absorbed into the intercept parameter. Although there is a nuisance parameter corresponding to each cluster (i.e., experiment), the analysis is asymptotically correct as the expected number of responses per group goes to infinity. Since there are 20–30 animals per group, Freedman et al.'s analysis is reasonable in this respect.

We wish to improve on Freedman et al.'s analysis by allowing heterogeneity of the diet effects, β . Keep in mind two aspects of the data that are important in motivating Freedman et al.'s analysis. First, the between-experiment variability is large; although within an experiment the animals are treated the same with respect to initiation of the tumors (e.g., given the same type and dose of carcinogen) and are followed for the same length of time, these and other factors vary considerably between experiments. This is why Freedman et al. (1990) modeled an intercept term for each experiment. Second, there are two primary effects to be estimated, the effect of dietary fat and of total calorie intake. Because of the way the experiments are designed, both effects cannot be estimated for each experiment. The primary effects within an experiment can be estimated only in experiments where both fat calories and total calories vary and, even then, more than two groups are required (see Freedman, 1994). This second reason is why we cannot use a traditional mixed-effects model.

Consider a traditional mixed-effects model, which is fixed in the intercept parameters but random in the diet parameters,

$$\text{logit}(p_{ij}) = \alpha_i + z'_{ij}(\beta + b_i), \quad (2)$$

where b_i are $p \times$ zero. Traditional Laird and Ware, as explained above experiment. Of course numerical methods (1997). However, effects distributions require the specification of the

A second way marginal model properly account. The sandwich estimator (White, 1982; Binder) applied to our data

This model may be mentioned above

These two traditional Liang, and Albert marginal model equation (2), where

They show that, β in the mixed-effects averaged parameter Zeger et al. (1988) values.

In this paper, diet parameters. (2). Our model the heterogeneity heterogeneity in correction for degrees of the residuals total caloric intake total calories are meta-analysis) is each experiment

In Section 2, standardized residuals the degrees of freedom analyses of this present simulation sandwich estimator we reanalyze the

2. The Model

2.1 Conditional Sandwich estimator

the mean is correct present, then the sandwich estimator (see White, 1982) logistic regression

where b_i are $p \times 1$ random effects vectors identically distributed from a distribution with mean zero. Traditional estimation methods for estimating the random effects may not be applied (see Laird and Ware, 1982; Stiratelli, Laird, and Ware, 1984; DerSimonian and Laird, 1986) because, as explained above, it is not possible to estimate the random effects parameters b_i from each experiment. Of course, even without estimating the b_i , we can still obtain an estimate for β by using numerical methods to integrate the likelihood over the random effects distribution (see McCulloch, 1997). However, a problem with applying this model is the dependence of the results on a random effects distribution that is unknown. Nonparametric versions of random-effects models that do not require the specification of the random-effects distribution (see Butler and Louis, 1992) also require estimation of the primary effects from each experiment and cannot be applied to our data.

A second way to model heterogeneity in the diet parameters is through a marginal model. A marginal model does not explicitly parameterize the heterogeneity of the diet parameters but properly accounts for this heterogeneity through the use of the sandwich estimator of variance. The sandwich estimator requires a fixed number of parameters and a large number of clusters (see White, 1982; Binder, 1983; Royall, 1986; Zeger and Liang, 1986). Thus, a traditional marginal model applied to our data would have all parameters common to each experiment even the intercepts,

$$\text{logit}(p_{ij}) = \alpha + z'_{ij}\beta^* \quad (3)$$

This model may be impractical for our data because of the large between-experiment variability mentioned above.

These two traditional ways of modeling heterogeneity of parameters are compared in Zeger, Liang, and Albert (1988). They call the mixed-effects model a subject-specific model and the marginal model a population-averaged model. They consider a simpler mixed-effects model than equation (2), where the number of fixed-effects does not increase with the number of clusters, e.g.,

$$\text{logit}(p_{ij}) = \alpha + z'_{ij}(\beta + b_i) \quad (4)$$

They show that, even though the expected value of b_i is zero for all i , the subject-specific parameter β in the mixed-effects model [equation (4)] does not estimate the same effect as the population-averaged parameter β^* in the marginal model [equation (3)]. Although these effects are different, Zeger et al. (1988) show that tests on both models applied to the same data give similar significance values.

In this paper, we allow for heterogeneity in the true effects for both the intercept terms and the diet parameters. So, for example, the true data could come from a process described by equation (2). Our model starts with the model of Freedman et al. (1990) [equation (1)] and accounts for the heterogeneity in the intercept parameters by conditioning out these parameters and for the heterogeneity in the diet parameters by using the sandwich estimator of variance. We study a correction for degrees of freedom in the middle part of the sandwich estimator using standardization of the residuals and Satterthwaite's approximation. This correction may be important since the total caloric intake effect is estimated from the small percentage of the experiments that varied total calories among groups. An added benefit of using a conditional model (not needed for our meta-analysis) is that we no longer require the assumption of a large number of animals within each experimental group, as with the unconditional model.

In Section 2, we introduce the sandwich estimator for conditional logistic regression using the standardized residuals. We create an approximate Wald test using the t -distribution and estimating the degrees of freedom with Satterthwaite's approximation. (A PC executable program to perform analyses of this type is available at <http://dcp.nci.nih.gov/BB/software.html>.) In Section 3, we present simulations to examine the performance of the conditional logistic regression with the sandwich estimator of variance when the true model is a random effects one. Finally, in Section 4, we reanalyze the data from Freedman et al. (1990) using these methods.

2. The Model

2.1 Conditional Logistic Regression with Sandwich Estimator of Variance

Sandwich estimators of variance are important because they are consistent when the model for the mean is correctly specified (Liang and Zeger, 1986). For example, if there is overdispersion present, then the variance estimator is still consistent. In fact, even if the mean is misspecified, the sandwich estimator of the variance consistently estimates the variance of the misspecified model (see White, 1982). In this section, we apply the sandwich estimator of the variance to conditional logistic regression.

Let Y_{ij} and m_{ij} represent the number of positive responses (i.e., animals with tumors) and the number of possible responses (i.e., animals), respectively, for the j th group in the i th cluster (i.e., experiment). Letting $p_{ij} = E(Y_{ij})/m_{ij}$, the unconditional logistic model is given in equation (1). Since the vector $m_i = [m_{i1} \cdots m_{in_i}]$ is fixed, the sufficient statistic for α_i is $a_i = \sum_{j=1}^{n_i} y_{ij}$, the number of positive responses in the i th cluster. Thus, we condition on the a_i to obtain the conditional likelihood that does not depend on the α_i (see the Appendix). Within the i th cluster, the sufficient statistic for β is the $p \times 1$ vector $t_i \equiv \sum_{j=1}^{n_i} z_{ij} y_{ij}$. We review in the Appendix that the score statistic for β for the conditional logistic regression may be written as

$$U(\beta) = \sum_{i=1}^K [t_i - E(t_i | \beta)],$$

where $E(t_i | \beta)$ is defined explicitly in the Appendix. We solve for β using a Newton-Raphson algorithm. In the algorithm, if $\tilde{\beta}$ is the current estimate, then the subsequent estimate $\hat{\beta}$ is

$$\hat{\beta} \approx \tilde{\beta} - \left(\frac{\partial U(\beta)}{\partial \beta'} \right)_{\beta=\tilde{\beta}}^{-1} U(\tilde{\beta}). \tag{5}$$

By using a similar Taylor series expansion but replacing $\tilde{\beta}$ with β and by taking the covariance of both sides, we obtain the sandwich estimator of variance as a function of β , i.e.,

$$\begin{aligned} \text{cov}(\hat{\beta}) &\approx \left(\frac{\partial U(\beta)}{\partial \beta'} \right)^{-1} \text{cov}(U(\beta)) \left(\frac{\partial U(\beta)}{\partial \beta'} \right)^{-1} \\ &= \left(\sum_{i=1}^K D_i \right)^{-1} \left(\sum_{i=1}^K \text{cov}(t_i) \right) \left(\sum_{i=1}^K D_i \right)^{-1}, \end{aligned} \tag{6}$$

where

$$\frac{\partial U(\beta)}{\partial \beta'} = \sum_{i=1}^K D_i$$

and D_i is given explicitly in the Appendix. Let \hat{D}_i be D_i with β replaced by $\hat{\beta}$. To calculate \hat{D}_i and $\hat{\beta}$, we use the numerical methods given in Gail, Lubin, and Rubinstein (1981). The classical variance estimator for conditional logistic regression is $\text{cov}(\hat{\beta}) = (\sum_{i=1}^K \hat{D}_i)^{-1}$, which can be related to equation (6) by substituting \hat{D}_i for D_i and $\sum_{i=1}^K \hat{D}_i$ for $\sum_{i=1}^K \text{cov}(t_i)$. The standard sandwich estimator of variance also uses \hat{D}_i for D_i but estimates $\sum_{i=1}^K \text{cov}(t_i)$ with a sum of empirical variance estimators,

$$\sum_{i=1}^K (t_i - \hat{E}(t_i)) (t_i - \hat{E}(t_i))',$$

where $\hat{E}(t_i) = E(t_i | \hat{\beta})$. Each estimator $(t_i - \hat{E}(t_i))(t_i - \hat{E}(t_i))'$ is a quite variable estimator of $\text{cov}(t_i)$. However, the sum of the estimators is a fairly good estimator of the sum of the covariances as long as $K \gg p$ and none of the elements of β is estimated primarily by a small number of clusters (see Diggle, Liang, and Zeger, 1994, pp. 71-72, for a similar discussion with respect to generalized least squares estimators). However, this latter situation is not met in our data example. Most of the experiments in the meta-analysis fix the total number of calories while varying the percentage of fat in the diet. For these experiments, the element of t_i corresponding to the total number of calories will be exactly equal to the corresponding element of $E(t_i | \hat{\beta})$ regardless of the value of $\hat{\beta}$; therefore, the estimator of the corresponding element of the sum of covariances depends only on those experiments that vary total caloric intake. Thus, even when $K \gg p$, the middle of the sandwich estimator may be quite unstable and have substantial bias.

To correct for these two problems, we perform two modifications to the usual application of the Wald test associated with a sandwich estimator. These modifications are analogous to the modifications to the Z -test to obtain the t -test. We first correct the bias of the variance estimator (or at least partially correct it), then we use the t -distribution instead of the normal distribution to account for the variability of the variance estimator.

2.2 Bias of Empirical

Similar to the way squared model (see C can show that the er expectations as pK

Use a Taylor series e

where $D' = [D_1 \ D_2$ the above equation,

where I_n is an $n \times n$ V is a block diagona

Note that C is sim $\hat{E}(t_i)'$ by the recipr

This correction does equally. Recall the experiments estimat and $K^* \ll K$. Idea inflated by approxin

A second method standardized residua element of C , using cluster is given by

where $\text{diag}(M)$ is for numerical problems

2.2 Bias of Empirical Variance Estimator

Similar to the way that residuals from a simple linear model or residuals from a nonlinear least squared model (see Carroll and Ruppert, 1988, pp. 32-33) can give biased estimators of variance, we can show that the empirical variance estimators are biased. First write the t_i 's and their estimated expectations as $pK \times 1$ vectors,

$$t - \hat{E}(t) = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_K \end{bmatrix} - \begin{bmatrix} \hat{E}(t_1) \\ \hat{E}(t_2) \\ \vdots \\ \hat{E}(t_K) \end{bmatrix}$$

Use a Taylor series expansion about β to approximate $\hat{E}(t)$, so that

$$t - \hat{E}(t) \approx t - E(t) + D(\hat{\beta} - \beta),$$

where $D' = [D_1 \ D_2 \ \dots \ D_K]$. Then we use equation (5) with $\hat{\beta}$ replaced by β to estimate $\hat{\beta}$ in the above equation, i.e.,

$$\begin{aligned} t - \hat{E}(t) &\approx t - E(t) - D \left(\sum_{i=1}^K D_i \right)^{-1} \sum_{i=1}^K (t_i - E(t_i)) \\ &= \left\{ I_{pK} - D \left(\sum_{i=1}^K D_i \right)^{-1} [I_p \ I_p \ \dots \ I_p] \right\} \{t - E(t)\} \\ &= C \{t - E(t)\}, \end{aligned}$$

where I_n is an $n \times n$ identity matrix. If the model is correctly specified, then $E([t - E(t)][t - E(t)]') = V$ is a block diagonal matrix with elements D_1, \dots, D_K , and

$$E([t - \hat{E}(t)][t - \hat{E}(t)]') \approx CVC' = CV.$$

Note that C is singular. A simple correction would be to multiply each value of $(t_i - \hat{E}(t_i))(t_i - \hat{E}(t_i))'$ by the reciprocal of the average of the diagonal elements of C so that

$$\frac{pK}{\text{trace}(C)} = \frac{pK}{pK - p} = \frac{K}{K - 1}.$$

This correction does not suit our needs because it increases the variance estimators of all clusters equally. Recall the problem of only a few clusters estimating one of the parameters (e.g., a few experiments estimating the parameter for total calories). Say there were only K^* of those clusters and $K^* \ll K$. Ideally, we would like the empirical variance estimators for those clusters to be inflated by approximately $K^*/(K^* - 1)$.

A second method to account for bias is to standardize the residuals $t_i - \hat{E}(t_i)$. We form the standardized residuals by dividing each element by the square-root of the accompanying diagonal element of C , using \hat{D}_i for D_i . In other words, the vector of standardized residuals for the i th cluster is given by

$$S_i^* = \text{diag} \left\{ I_p - \hat{D}_i \left(\sum_{i=1}^K \hat{D}_i \right)^{-1} \right\}^{-1/2} (t_i - \hat{E}(t_i)), \quad (7)$$

where $\text{diag}(M)$ is formed by replacing all off-diagonal elements of M with zeros. There will be some numerical problems with taking the square-root inverse of

$$\text{diag} \left\{ I_p - \hat{D}_i \left(\sum_{i=1}^K \hat{D}_i \right)^{-1} \right\} \quad (8)$$

animals with tumors) and
 jth group in the ith cluster
 model is given in equation
 tic for α_i is $a_i = \sum_{j=1}^{n_i} y_{ij}$,
 on on the a_i to obtain the
 dix). Within the ith cluster,
 review in the Appendix that
 written as

3 using a Newton-Raphson
 sequent estimate $\hat{\beta}$ is

(5)

by taking the covariance of
 of β , i.e.,

Material may be protected by copyright. U.S. Code

(6)

lated by $\hat{\beta}$. To calculate \hat{D}_i
 imstein (1981). The classical
 $(\hat{\beta})^{-1}$, which can be related
 . The standard sandwich
 h sum of empirical variance

a quite variable estimator of
 of the sum of the covariances
 by a small number of clusters
 n with respect to generalized
 our data example. Most of
 while varying the percentage
 dding to the total number of
 $\hat{\beta}$) regardless of the value of
 of covariances depends only
 n $K \gg p$, the middle of the

to the usual application of
 ations are analogous to the
 bias of the variance estimator
 d of the normal distribution

if some of the diagonal elements equal zero or are very near zero. For example, we would obtain a diagonal element of zero for the i th experiment if all experiments held one covariate constant (e.g., total caloric intake) except the i th experiment, which holds the other covariates constant. To guard against this problem, we recommend a modification; if there are diagonal elements of equation (8) that are less than 0.01 times the maximum diagonal value (for that cluster), set them equal to 0.01 times the maximum diagonal value. We use this modification for all analyses using standardized residuals in this paper. Our corrected estimator of the $\text{cov}(t_i)$ is $S_i^* S_i^{*'}.$

2.3 Accounting for Variability of Variance Estimator

A Wald test may be performed on each parameter by comparing

$$T_j = \frac{\hat{\beta}_j}{\sqrt{\widehat{\text{var}}(\hat{\beta}_j)}} \tag{9}$$

to a standard normal distribution, where $\hat{\beta}_j$ is the j th element of $\hat{\beta}$ and $\widehat{\text{var}}(\hat{\beta}_j)$ is the (j, j) th element of the sandwich estimator of the variance using the standardized residuals. This estimator may perform poorly since the variance of $\widehat{\text{var}}(\hat{\beta}_j)$ may be large; therefore, we use a t -distribution for $T_j.$

To motivate the t -distribution, consider first the ideal situation where $p = 1$ and the S_i^* are independent and distributed normally with mean 0 and variance $V_0.$ Then from equation (6), we get

$$\text{cov}(\hat{\beta}) \approx \left(\sum D_i\right)^{-1} (KV_0) \left(\sum D_i\right)^{-1}$$

and, under the null hypothesis of $\beta = 0,$ $K^{-1/2}(\sum D_i)\hat{\beta}$ is approximately normally distributed with mean 0 and variance $V_0,$ where the unmarked summations are from $i = 1$ to $K.$ Also, $V_0^{-1} \sum S_i^{*2}$ is distributed as a chi-squared distribution with K degrees of freedom. If we can assume that $K^{-1/2}(\sum D_i)\hat{\beta}$ and $\sum S_i^{*2}$ are independent, then under the null hypothesis,

$$T^2 = K \left(K^{-1/2} \left(\sum D_i\right) \hat{\beta}\right)' \left(\sum S_i^* S_i^{*'}\right)^{-1} \left(K^{-1/2} \left(\sum D_i\right) \hat{\beta}\right) = \hat{\beta}' (\widehat{\text{cov}}(\hat{\beta}))^{-1} \hat{\beta}$$

is distributed as an F -distribution with 1 and K degrees of freedom or T is distributed as a t -distribution with K degrees of freedom.

Now we relax this ideal situation slightly by allowing the variance to change between clusters, i.e., $S_i^* \sim N(0, V_i).$ In this case we estimate the degrees of freedom by Satterthwaite's formula, which models the distribution of $\hat{V}_0^{-1} \sum S_i^{*2}$ with the χ^2 distribution with the best fit of moments (see Cochran, 1977). To obtain Satterthwaite's formula, we first assume that there exist constants \hat{V}_0 and d such that

$$\frac{1}{\hat{V}_0} \sum S_i^{*2} \sim \chi_d^2.$$

If the V_i are known, using the fact that $S_i^{*2}/V_i \sim \chi_1^2,$ we solve for the constants by equating the first two moments of $\hat{V}_0^{-1} \sum S_i^{*2}$ to those of a chi-squared random variable with d degrees of freedom. Thus, $\hat{V}_0 = d^{-1} \sum V_i$ and

$$d = \frac{(\sum V_i)^2}{\sum V_i^2}.$$

The value d must be estimated from the data. Satterthwaite's formula applied to our problem estimates d with

$$\hat{d} = \frac{\left(\sum_{i=1}^K (S_i^*)^2\right)^2}{\sum_{i=1}^K (S_i^*)^4}.$$

This estimator of degrees of freedom is in general biased downward, leading to a conservative test when the normality assumptions hold. We make no correction for this bias because the distribution of the residuals may be far from normal.

Finally, consider a time. We form freedom for testing to a t -distribution

where S_{ij}^* is the j We have only d using linear combi

where L is full ra $(z'_{ij} L^{-1})(L\beta) = z'_j$

3. Simulation

Because the sand standardized resid several simulations the Introduction. compare the classi variance with the the previous sectio estimator. The diff

Altogether, we sets of simulations random and fixed 4 different designs hold for all designs every group having response for each g

logit

where α_i are the i $b'_i = [b_{i1} \ b_{i2}]$ is the correlation zero an effects, and $z'_{ij} = [$ As throughout the

We can make any

For Designs A, type. Between the Design B: $K = 40,$

Cluster
(K/4)
(K/4)
(K/4)
(K/4)

Finally, consider the case where $p > 1$. The Wald test of equation (9) tests one element at a time. We form our approximate t -tests in this same manner, estimating a separate degree of freedom for testing each element of $\hat{\beta}$. We assume that T_j of equation (9) is distributed according to a t -distribution with degrees of freedom

$$\hat{d}_j = \frac{\left(\sum_{i=1}^K (S_{ij}^*)^2\right)^2}{\sum_{i=1}^K (S_{ij}^*)^4},$$

where S_{ij}^* is the j th element of S_i^* .

We have only discussed testing for $\beta_j = 0$ versus $\beta_j \neq 0$. We can test any linear hypothesis by using linear combinations of the covariates. For example, suppose we wish to test

(9)

$$H_0: L\beta \equiv \theta = \begin{bmatrix} 0 \\ \theta_2 \end{bmatrix}$$

$$H_1: L\beta \equiv \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix},$$

where L is full rank, θ_1 is $r \times 1$ and θ_2 is $(p - r) \times 1$. To perform this test, we rewrite $z'_{ij}\beta$ as $(z'_{ij}L^{-1})(L\beta) = z'^*_{ij}\theta$, then proceed as with the usual test, replacing z'_{ij} for z'^*_{ij} .

3. Simulation

Because the sandwich estimators have good properties only asymptotically and the use of standardized residuals with the t -distribution for the Wald test relies on approximations, we ran several simulations. The simulations model a simplified version of the meta-analysis mentioned in the Introduction. We let the true process that generates the data be a random-effects model. We compare the classical conditional logistic regression to the one that uses the sandwich estimator of variance with the standardized residuals and the t -distribution for the Wald test as described in the previous section. Both methods estimate a population-averaged parameter β^* with the same estimator. The difference between the methods is the variance estimators and the coverage.

Altogether, we performed 24 sets of simulations. Each set consists of 1000 simulations. The 24 sets of simulations were done by taking all combinations of 6 different cases (Cases 1-6) defining the random and fixed effects of the true treatment parameters and 4 designs (Designs A-D) defining 4 different designs of the meta-analysis. We first discuss the general form of the simulations that hold for all designs and cases. We have K clusters (i.e., experiments) with 2 groups in each cluster, every group having $m_{ij} = m = 30$ Bernoulli responses (i.e., animals). The true probability of a response for each group is given by the logistic model

$$\text{logit}(p_{ij}) = \alpha_i + z'_{ij}(\beta + b_i) = \alpha_i + z_{ij1}(\beta_1 + b_{i1}) + z_{ij2}(\beta_2 + b_{i2}), \quad (10)$$

where α_i are the intercepts that come from a normal distribution with mean α and variance σ_α^2 , $b'_i = [b_{i1} \ b_{i2}]$ is the vector of random effects that comes from a normal distribution with mean and correlation zero and variances σ_1^2 and σ_2^2 , $\beta' = [\beta_1 \ \beta_2]$ is the vector of subject-specific treatment effects, and $z'_{ij} = [z_{ij1} \ z_{ij2}]$ is the vector of the two covariates (i.e., dietary fat and total calories). As throughout the paper, the index i represents clusters and j represents groups within a cluster. We can make any of the parameters constant across clusters by setting its variance equal to zero.

For Designs A, B, and C, we have four different types of clusters with equal numbers of each type. Between these three designs, we change the total number of clusters (Design A: $K = 20$, Design B: $K = 40$, Design C: $K = 80$.) The designs are listed in Table 1. In Design D, for 37 out

Table 1
Designs A, B, and C

Clusters	Group 1		Group 2	
	z_{i11}	z_{i12}	z_{i21}	z_{i22}
(K/4)	0.5	0.0	0.5	1.0
(K/4)	0.0	0.5	1.0	0.5
(K/4)	0.0	0.0	1.0	1.0
(K/4)	0.0	1.0	1.0	0.0

Table 2
Design D

Clusters	Group 1		Group 2	
	z_{i11}	z_{i12}	z_{i21}	z_{i22}
37	0.50	0.0	0.50	1.0
1	0.25	0.5	0.75	0.5
1	0.25	0.0	0.75	1.0
1	0.25	1.0	0.75	0.0

of 40 of the clusters, we have $z_{i11} = z_{i21}$, i.e., the factor is constant between the two groups within these clusters. Thus, only three of the clusters estimate the parameter for β_1^* . The details of Design D are listed in Table 2.

For each of the cases, we test different true values for the subject-specific parameters. These values are listed in Table 3. For all cases, we have nonconstant effects on the intercept, where $\sigma_\alpha^2 = 1$. The values of α are chosen to avoid probabilities of response within any group that are very close to 0 or 1. The treatment (e.g., fat calories and total calories) effects are varied between cases. All cases may be described in terms of fixed treatment effects versus random treatment effects and null hypotheses (i.e., $\beta = 0$) versus alternative hypotheses (i.e., $\beta \neq 0$). The values for the parameters are of the same order as the effects from the 104 groups of Sprague-Dawley rats from 43 clusters mentioned in the Introduction (although the designs are greatly simplified).

The expected value of our population-averaged estimator β^* is not equal to the subject-specific parameter β (see Zeger et al., 1988). The population-averaged interpretation is the one applicable to our meta-analysis because we are not interested in each subject-specific (i.e., experiment-specific) effect in the meta-analysis but in the overall population-averaged effect. Unfortunately, we do not know the value of β^* under the simulation, and no exact closed-form expression of β^* is available.

To estimate β^* , we perform another simulation, letting the number of clusters grow large. Note that β^* is approximately equal (perhaps exactly equal) to β when either $\sigma_k^2 = 0, k = 1, 2$, or $\beta = 0$ (see the approximation given in Zeger et al., 1988, p. 1054). So we only need to simulate Cases 5 and 6. We run four sets of simulations, Cases 5 and 6 at each of two designs, one representing Designs A, B, and C and one representing Design D. For each of these four sets of simulations, we let the number of clusters equal 2000 in the proportions of Tables 1 and 2. We let the number of simulations for each set be 2000. The mean estimates for β^* for each of the four sets are

	Case 5	Case 6
Designs A, B, and C	[4.094, 2.363]	[3.946, 2.298]
Design D	[4.152, 2.352]	[4.032, 2.246]

We use these values to center the confidence intervals to calculate the estimated coverage below.

Before considering the coverage of the simulations, consider first the variance estimators. Table 4 lists the average of the estimates of the variance for Designs A and D. (The results for Designs B and C are similar to Design A.) We list four different variances:

Table 3
Cases

Case	Treatment effect	Hypothesis	α	σ_α^2	β_1	σ_1^2	β_2	σ_2^2
1	Fixed	Null	0.0	1.0	0.0	0.0	0.0	0.0
2	Fixed	Alternative	-3.3	1.0	4.2	0.0	2.4	0.0
3	Random	Null	0.0	1.0	0.0	0.5	0.0	0.5
4	Random	Null	0.0	1.0	0.0	1.0	0.0	2.0
5	Random	Alternative	-3.3	1.0	4.2	0.5	2.4	0.5
6	Random	Alternative	-3.3	1.0	4.2	1.0	2.4	2.0

Case	β_1	β_2
1	0.0	0.0
2	0.0	0.0
3	0.0	0.0
4	0.1	0.1
5	0.1	0.1
6	0.2	0.2

1. The simulation
2. The classical logistic regression
3. The sandwich estimator
4. The sandwich estimator

As expected in specified, the classical estimator to be close to the true parameter underestimates the variance. This underestimates half the size of the true variance. To estimate the variances, we see Design D since it has more residuals are closer than the classical estimates for β_1^* . standardized residuals are correct for this case as discussed in Section 2.1.

We list the simulated coverage are given by β for each case. The simulated coverage value. The classical estimator variance estimator formula. Recall that if the coverage is 95%, we expect the coverage to be 0.967. The coverage effect, but the coverage is not conservative, particularly for Design D.

Table 4
Average of the estimates of the variance for the parameters

Group 2		Design A				Design D			
		Sim.	Class.	Sandwich unstand. resid.	Sandwich stand. resid.	Sim.	Class.	Sandwich unstand. resid.	Sandwich stand. resid.
z _{i21}	z _{i22}								
0.50	1.0								
0.75	0.5								
0.75	1.0								
0.75	0.0								
Case									
1	β_1	0.0204	0.0218	0.0190	0.0204	0.4930	0.4518	0.2904	0.4487
	β_2	0.0226	0.0218	0.0191	0.0206	0.0081	0.0083	0.0080	0.0082
2	β_1	0.0652	0.0703	0.0628	0.0691	0.7087	0.6779	0.4235	0.7203
	β_2	0.0560	0.0554	0.0489	0.0537	0.0130	0.0124	0.0118	0.0121
3	β_1	0.0737	0.0224	0.0648	0.0699	1.1077	0.4700	0.6354	0.9952
	β_2	0.0752	0.0225	0.0654	0.0706	0.0209	0.0085	0.0196	0.0201
4	β_1	0.1450	0.0234	0.1234	0.1335	2.2520	0.5176	1.3946	2.2551
	β_2	0.1517	0.0234	0.1364	0.1476	0.0460	0.0089	0.0459	0.0473
5	β_1	0.1428	0.0718	0.1257	0.1405	1.3942	0.7021	0.7168	1.2616
	β_2	0.1172	0.0564	0.0979	0.1075	0.0256	0.0126	0.0244	0.0252
6	β_1	0.2576	0.0730	0.2248	0.2525	3.1164	0.7859	1.4022	2.8291
	β_2	0.2482	0.0576	0.1913	0.2103	0.0578	0.0128	0.0546	0.0562

1. The simulated variance is the sample variance of the 1000 different parameter estimates.
2. The classical variance is the average of the variances calculated from the standard conditional logistic regression model.
3. The sandwich variance with unstandardized residuals is the average of the sandwich variance estimates using $\Sigma (t_i - \hat{E}(t_i))(t_i - \hat{E}(t_i))'$ for the middle of the sandwich estimator.
4. The sandwich variance with standardized residuals is the average of the sandwich variance estimates using $\Sigma S_i^* S_i^{*'} / S_i^*$ for the middle of the sandwich estimator.

As expected in the fixed treatment cases (Cases 1 and 2) where the classical variance is correctly specified, the classical estimate is approximately equal to the simulated estimate, which we expect to be close to the true variance. For the random effects models (Cases 3-6), the classical variance underestimates the variance of the parameter estimates as measured by the simulated variance. This underestimation is severe; in most of the random effects cases, the classical variance is under half the size of the simulated variance. The sandwich estimators with the standardized residuals estimate the variance well for both the fixed and random effects cases. Comparing the two sandwich variances, we see the standardization is particularly needed for estimating the variance of β_1 in Design D since in this design β_1 is estimated from only three clusters.

Although the mean of the variance estimates for the sandwich estimate with standardized residuals are close to the simulated values, the sample variance of those estimates is much larger than the classical ones. For example, in Design A, Case 6, the sample variance of the variance estimates for β_1^* is 0.0004 for the classical variance and 0.0165 for the sandwich estimate with standardized residuals. Because the variance of the variance estimators is large, it is important to correct for this by using the *t*-distribution to calculate the distribution of the Wald statistic as discussed in Section 2.3.

We list the simulated coverage of the nominal 95% confidence intervals in Table 5. The true values are given by β for Cases 1-4 and by the simulated β^* mentioned previously for Cases 5 and 6. The simulated coverage is the percent of time the calculated confidence interval contained the true value. The classical intervals are formed from the normal approximation using the classical variance. The sandwich intervals are formed using the standardized residuals in the middle of the sandwich variance estimate and the *t*-distribution with degrees of freedom estimated by Satterthwaite's formula. Recall each case-design combination was simulated 1000 times so that, if the true coverage is 95%, we expect over 99% of Monte Carlo simulations to have simulated coverage between 0.931 and 0.967. The simulated coverage of the classical intervals appears correct when there is a fixed effect, but the coverage appears to be liberal when there are random effects, with very poor coverage when there is a large variance on the random effects. In contrast, the sandwich intervals appear conservative, particularly for designs A and B.

stant between the two groups
parameter for β_1^* . The details
ject-specific parameters. These
effects on the intercept, where
onse within any group that are
ories) effects are varied between
ffects versus random treatment
ses (i.e., $\beta \neq 0$). The values for
roups of Sprague-Dawley rats
are greatly simplified).
not equal to the subject-specific
pretation is the one applicable
specific (i.e., experiment-specific)
effect. Unfortunately, we do not
ing expression of β^* is available.
ber of clusters grow large. Note
either $\sigma_k^2 = 0, k = 1, 2$, or $\beta = 0$
only need to simulate Cases 5
f two designs, one representing
the four sets of simulations, we
and 2. We let the number of
each of the four sets are

Case 6
946, 2.298]
082, 2.246].

the estimated coverage below.
the variance estimators. Table 4
d D. (The results for Designs B

β_1	σ_1^2	β_2	σ_2^2
0.0	0.0	0.0	0.0
0.2	0.0	2.4	0.0
0.0	0.5	0.0	0.5
0.0	1.0	0.0	2.0
0.2	0.5	2.4	0.5
0.2	1.0	2.4	2.0

Table 5
*Simulated coverage for 95% confidence intervals (sandwich method
 uses standardized residuals and Satterthwaite's approximation)*

Case		Design A		Design B		Design C		Design D	
		Classical	Sandwich	Classical	Sandwich	Classical	Sandwich	Classical	Sandwich
1	β_1	0.950	0.971	0.950	0.955	0.947	0.960	0.951	0.951
	β_2	0.952	0.964	0.953	0.971	0.943	0.951	0.949	0.964
2	β_1	0.961	0.985	0.947	0.971	0.952	0.967	0.943	0.953
	β_2	0.953	0.980	0.945	0.961	0.946	0.964	0.942	0.957
3	β_1	0.721	0.972	0.705	0.952	0.745	0.963	0.803	0.949
	β_2	0.720	0.964	0.744	0.978	0.717	0.952	0.776	0.958
4	β_1	0.561	0.973	0.570	0.970	0.593	0.966	0.639	0.966
	β_2	0.573	0.972	0.541	0.965	0.551	0.970	0.621	0.960
5	β_1	0.829	0.986	0.832	0.978	0.827	0.970	0.842	0.951
	β_2	0.819	0.978	0.828	0.963	0.827	0.969	0.835	0.958
6	β_1	0.717	0.986	0.699	0.985	0.687	0.973	0.672	0.965
	β_2	0.664	0.977	0.643	0.975	0.685	0.958	0.612	0.963

In summary, when random effects are present in the treatment (i.e., diet) effects, then the sandwich estimators of variance proposed in this paper perform better than the classical estimator of variance. The modifications to the standard Wald test with the sandwich estimator appear to work well. When the treatment effects are constant across all clusters ($\sigma_a^2 = 0$, $a = 1, 2$), then the classical conditional model is correctly specified and it performs well. In these cases, the sandwich estimator of variance does not perform too badly.

4. Example: Meta-Analysis

We now analyze the motivating data mentioned in the Introduction. The results for the Sprague-Dawley rats are presented in the upper section of Table 6. The unconditional model is the model reported in Freedman et al. (1990) and described in the Introduction. The conclusion of that model is that all calories increase the risk of tumorigenesis, with fat calories increasing this risk even more than other calories. The values are almost identical to the conditional logistic analysis with the classical error variance. This is as expected since most of the groups have a large number of rats. The sandwich standard errors calculated using the standardized residuals are quite different. The standard error for the fat parameter is larger than the classical estimate, while the standard error for the total calories is smaller than the classical estimate. A possible reason for this low estimate is the high variability of the sandwich estimator: The standard errors do not capture all the difference between the two conditional methods because the classical model uses the normal distribution while the sandwich model uses the t -distribution with the Satterthwaite degrees of freedom. We list the estimated 95% confidence intervals to capture a fuller picture of the differences between the two methods: for the classical model, $0.087 \leq \beta_1 \leq 0.158$ and $0.068 \leq \beta_2 \leq 0.091$, while for the sandwich model, $0.109 \leq \beta_1 \leq 0.136$ and $0.059 \leq \beta_2 \leq 0.101$. Even using the t -distribution, the intervals for β_1 (TCAL) are larger for the classical model. Because so few experiments are used to estimate the TCAL effect and the variance of the sandwich estimates is large, we recommend using the classical confidence intervals for β_1 .

The second data set comprises mice bred for spontaneous tumors. In this data set, there are 57 groups of animals from 17 experiments. The average number of mice in each group was 38.7. Thirteen of the 17 experiments varied the total caloric intake (TCAL) between different groups. In most of the groups, the TCAL was measured, although there were a few in which we imputed the average of the TCAL for groups fed nonrestricted diets. In one of the experiments, the TCAL changed between groups but was not measured (White et al., 1944). The experiment reported a 46% calorie restriction in one of the groups but did not list the TCAL for the nonrestricted group. The results did not change significantly by changing the imputed value. The results are presented in the lower section of Table 6. Again we see very similar results between the unconditional model and the conditional one. The standard errors from the sandwich estimates are considerably larger than those of the conditional model. Again we compare the 95% confidence intervals: for the classical

Data
 Sprague-Dawley
 fed corn oil w
 kcal unknown
 Mice bred for
 spontaneous
 with kcal un

model, 0.521 \leq
 1.155 and 0.088
 than with the f
 groups, while t
 The advanta
 to the previous
 across experim
 unconditional l
 are significantly
 present analysis

5. Discussion

In this paper,
 particularly as
 to estimate a c
 tests that use t
 estimator and a

To relate the
 binary treatme
 Haenszel test
 meta-analysis,
 in the meta-ana
 Berlin et al., 19
 of variance is c
 The estimator

The use of th
 such as the tr
 estimating equ
 Further study
 prefer the stan
 estimators. The
 study and may

To our know
 has not previo
 estimators to
 1992). This ma
 Another reason
 recent methods
 therein) and th

Although ex
 misspecificati
 not modeled, t
 sample sizes at
 will give simila

The extended Mantel-Haenszel estimating procedure (EMH) forms the estimating equation for the i th cluster as the sum over a set of pairs of two sample conditional logistic regression score functions. The set of pairs is defined as all pairs of Bernoulli observations (i.e., animals) with one case (i.e., with tumor) and one control (i.e., no tumor). The estimating function for β is a linear combination of the estimating equations for each cluster. Two variance estimators have been proposed. Using a sandwich variance estimator, the EMH allows for dependency within cluster (i.e., experiment) and is valid as the number of clusters goes to infinity, just like the conditional logistic regression with sandwich estimators. The advantage of EMH is that, by using the other variance estimator, it may be used when the number of clusters is constant and the number of responses per cluster (i.e., animals per experiment) goes to infinity. The disadvantage of EMH as compared to the conditional logistic regression is that it is less efficient under constant treatment effects. An area of further work would be to compare the two methods under random treatment effects.

Finally, we caution against the indiscriminate use of the sandwich variance estimators with conditional logistic regression. Recall these estimators are valid as the number of clusters with information about the parameter grows large. Although the use of the t -distribution accounts for this small cluster sample in some way, we stress that the sandwich variance estimators are quite variable for these small cluster sizes. Thus, if the classical variance estimator is larger than the sandwich estimator (as in the first data set of Section 4), we recommend using the largest of the two estimates to give more robust inferences.

ACKNOWLEDGEMENT

We thank the referees for careful reading of the manuscript.

RÉSUMÉ

Motivés par une méta-analyse sur des expérimentations animales concernant l'effet de la consommation de graisse et de l'apport calorique sur la carcinogenèse mammaire, nous avons exploré l'utilisation des estimateurs "sandwich" de la variance dans les modèles logistiques conditionnels. La régression logistique conditionnelle classique suppose que les paramètres sont à effets fixes sur l'ensemble des clusters, alors que l'approche avec les estimateurs "sandwich" permet de faire des inférences avec des effets fixes aussi bien qu'avec des effets aléatoires. Cependant l'utilisation du test de Wald avec un estimateur "sandwich" suppose que chaque paramètre est estimé sur un grand nombre de clusters. Comme notre exemple viole cette condition, nous introduisons 2 modifications dans le test standard de Wald. D'abord, nous réduisons le biais de l'estimateur empirique de la variance (le milieu du "sandwich") en utilisant des résidus standardisés. Ensuite, nous réglons le problème des variances de ces estimateurs en utilisant une distribution de Student, plutôt qu'une distribution normale, où les degrés de libertés sont estimés par l'approximation de Satterthwaite. Au travers de simulations, nous montrons que ces estimateurs "sandwich" sont aussi bons que les estimateurs classiques lorsque les vrais effets sont fixes, et sont bien meilleurs que les estimateurs classiques lorsque les effets vrais sont aléatoires. Nous avons élargi le champ des simulations pour ces estimateurs "sandwich" à des cas où certains paramètres sont estimés à partir d'un petit nombre de clusters.

REFERENCES

- Berlin, J. A., Laird, N. M., Sacks, H. S., and Chalmers, T. C. (1989). A comparison of statistical methods for combining event rates from clinical trials. *Statistics in Medicine* **8**, 141-151.
- Binder, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistics Review* **51**, 279-292.
- Binder, D. A. (1992). Fitting Cox's proportional hazards models from survey data. *Biometrika* **79**, 139-147.
- Butler, S. M. and Louis, T. A. (1992). Random effects models with non-parametric priors. *Statistics in Medicine* **11**, 1981-2000.
- Carroll, K. K. and Khor, H. T. (1970). Effects of dietary fat and dose level of 7,12-Dimethylbenz-(α)anthracene on mammary tumor incidence in rats. *Cancer Research* **30**, 2260-2264.
- Carroll, R. J. and Ruppert, D. (1988) *Transformation and Weighting in Regression*. London: Chapman and Hall.
- Cochran, W. G. (1977). *Sampling Techniques*, 3rd edition. New York: John Wiley.
- Cox, D. R. and Snell, E. J. (1989). *Analysis of Binary Data*. London: Chapman and Hall.
- Davidian, M. and Carroll, R. J. (1987). Variance function estimation. *Journal of the American Statistical Association* **82**, 1079-1091.

DerSimonian, R.
7, 177-188.

Diggle, P. J., I.
Clarendon

Freedman, L. S.
tumors. *Sta*

Freedman, L. S.,
and the de
5710-5719.

Gail, M. H., Lut
control stu

Graubard, B. I.
Medicine St

Laird, N. M. and
963-974.

Liang, K. Y. (I
regression

Liang, K.-Y. and
Biometrika

Lin, D. Y. and
Journal of

McCulloch, C.
Journal of

McCullagh, P. a
and Hall.

Mehta, C. R. an
Medicine I

Neuhaus, J. M.
data. *Biom*

O'Neill, T. J. ar

Royall, R. M. (
Internation

Stratelli, R., Le
binary resp

White, F. R., W
of caloric r
of strain D
5, 43-48.

White, H. (1982

Zeger, S. L. an
outcomes.

Zeger, S. L., Li
estimating

See Section 2.1
the i th cluster

where $\Omega_i = \{u$
 $c(u)$ is the num
and $Y_j^* \leq m_{ij}$

- DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**, 177-188.
- Diggle, P. J., Liang, K.-Y., and Zeger, S. L. (1994). *Analysis of Longitudinal Data*. Oxford: Clarendon Press.
- Freedman, L. S. (1994). Meta-analysis of animal experiments on dietary fat intake and mammary tumors. *Statistics in Medicine* **13**, 709-718.
- Freedman, L. S., Clifford, C., and Messina, M. (1990). Analysis of dietary fat, calories, body weight, and the development of mammary tumors in rats and mice: A review. *Cancer Research* **50**, 5710-5719.
- Gail, M. H., Lubin, J. H., and Rubinstein, L. V. (1981). Likelihood calculations for matched case-control studies and survival studies with tied death times. *Biometrika* **68**, 703-707.
- Graubard, B. I. and Korn, E. L. (1994). Regression analysis with clustered data. *Statistics in Medicine* **13**, 509-522.
- Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963-974.
- Liang, K. Y. (1987). Extended Mantel-Haenszel estimating procedure for multivariate logistic regression models. *Biometrics* **43**, 289-299.
- Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13-22.
- Lin, D. Y. and Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association* **84**, 1074-1078.
- McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association* **92**, 162-170.
- McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*, 2nd edition. London: Chapman and Hall.
- Mehta, C. R. and Patel, N. R. (1995). Exact logistic regression: Theory and examples. *Statistics in Medicine* **14**, 2143-2160.
- Neuhaus, J. M. (1993). Estimation efficiency and tests of covariate effects with clustered binary data. *Biometrics* **49**, 989-996.
- O'Neill, T. J. and Barry, S. C. (1995). Truncated logistic regression. *Biometrics* **51**, 533-541.
- Royall, R. M. (1986). Model robust confidence intervals using maximum likelihood estimators. *International Statistical Review* **54**, 221-226.
- Stratelli, R., Laird, N., and Ware, J. H. (1984). Random-effects models for serial observations with binary response. *Biometrics* **40**, 961-971.
- White, F. R., White, J., Mider, G. B., Kelly, M. G., Heston, W. E., and David, P. W. (1944). Effect of caloric restriction on mammary-tumor formation in strain C3H mice and on the response of strain DBA to painting with methylcholanthrene. *Journal of the National Cancer Institute* **5**, 43-48.
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrics* **50**, 1-25.
- Zeger, S. L. and Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42**, 121-130.
- Zeger, S. L., Liang, K.-Y., and Albert, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics* **44**, 1049-1060.

Received April 1996; revised March and June 1997; accepted July 1997.

APPENDIX

Likelihood for Conditional Logistic Regression

See Section 2.1 for notation. Conditioning on $a' = [a_1 \dots a_K]$, the (conditional) likelihood for the i th cluster is proportional to

$$L_i = \frac{c(t_i) \exp(t_i' \beta)}{\sum_{u \in \Omega_i} c(u) \exp(u' \beta)}$$

where $\Omega_i = \{u : u = \sum_{j=1}^{n_i} Y_j^* z_{ij}, \sum_{j=1}^{n_i} Y_j^* = a_i, \text{ and } Y_j^* \text{ are integers with } Y_j^* \leq m_{ij} \text{ for all } j\}$ and $c(u)$ is the number of values of Y^* that give $u = \sum_{j=1}^{n_i} Y_j^* z_{ij}$ under the conditions $\sum_{j=1}^{n_i} Y_j^* = a_i$ and $Y_j^* \leq m_{ij}$ (see Cox and Snell, 1989; Mehta and Patel, 1995). In other words, the set Ω_i is the

forms the estimating equation for conditional logistic regression score observations (i.e., animals) with estimating function for β is a variance estimators have been dependency within cluster (i.e., just like the conditional logistic estimator, by using the other variance estimator and the number of responses advantage of EMH as compared to constant treatment effects. An random treatment effects.

which variance estimators with the number of clusters with the t -distribution accounts for variance estimators are quite the estimator is larger than the recommend using the largest of the

concernant l'effet de la mammaire, nous avons exploré les modèles logistiques conditionnels. Les modèles sont à effets fixes sur "sandwich" permet de faire des simulations. Cependant l'utilisation du modèle est estimé sur un grand nombre de simulations. Nous introduisons 2 modifications à l'estimateur empirique de la variance. Ensuite, nous réglons les degrés de liberté de Student, plutôt qu'une approximation de Satterthwaite. Les modèles "sandwich" sont aussi bons que les modèles "Student". Les meilleurs que les estimateurs de variance. Les simulations pour les degrés de liberté à partir d'un petit nombre

A comparison of statistical methods in *Statistics in Medicine* **8**, 141-151.

estimators from complex surveys.

survey data. *Biometrika* **79**,

parametric priors. *Statistics*

level of 7,12-Dimethylbenzanthracene. *Biometrika* **30**, 2260-2264.

ing in *Regression*. London:

John Wiley.

Chapman and Hall.

Journal of the American

set of all the different possible values of t_i given different possible values of Y_i with a_i total positive responses, and $c(t_i)$ is the number of ways to obtain t_i .

We take the derivative of the full (conditional) log-likelihood with respect to β to get the score equations $U(\beta)$,

$$\begin{aligned} U(\beta) &= \sum_{i=1}^K \frac{\partial}{\partial \beta} \log \left(\frac{c(t_i) e^{t_i' \beta}}{\sum_{u \in \Omega_i} c(u) e^{u' \beta}} \right) \\ &= \sum_{i=1}^K \frac{\partial}{\partial \beta} \left(\log[c(t_i)] + t_i' \beta - \log \left[\sum_{u \in \Omega_i} c(u) e^{u' \beta} \right] \right) \\ &= \sum_{i=1}^K (t_i - E(t_i | \beta)), \end{aligned}$$

where

$$E(t_i | \beta) = E(t_i | Z_i; a_i; \beta) = \frac{\sum_{u \in \Omega_i} u c(u) e^{u' \beta}}{\sum_{u \in \Omega_i} c(u) e^{u' \beta}}. \quad (11)$$

Also, $D_i = (\partial U_i(\beta)) / (\partial \beta')$ is a $p \times p$ matrix with the r, s th element equal to

$$\frac{\partial L_i^2(\beta)}{\partial \beta_r \partial \beta_s} = \frac{-\sum_{u_i \in \Omega_i} u_{ir} u_{is} c(u_i) e^{\beta u_i}}{\sum_{u_i \in \Omega_i} c(u_i) e^{\beta u_i}} + \left(\frac{\sum_{u_i \in \Omega_i} u_{ir} c(u_i) e^{\beta u_i}}{\sum_{u_i \in \Omega_i} c(u_i) e^{\beta u_i}} \right) \left(\frac{\sum_{u_i \in \Omega_i} u_{is} c(u_i) e^{\beta u_i}}{\sum_{u_i \in \Omega_i} c(u_i) e^{\beta u_i}} \right), \quad (12)$$

where u_{ir} is the r th row of u_i .

In s
with
or f
Thi
the
the
with
the

1. I
Sur
sore
trial
Oak
the
diag
EC
myc
exp
rece

D
For
infa
conf
prep
and
and
the
pote
end
evid
diffe
lost
to a
shou
the
L
viva

Em
Key