Correlated Weibull Regression Model for Multivariate Binary Data

Reginald N. O. Odai[1], Joachim Kötting[1], John Kwagyan[2], George E. Bonney[3] and Wolfgang Urfer[1]

[1]Department of Statistics, University of Dortmund, D-44221 Dortmund, Germany
[2]Howard University, Washington, D.C., U.S.A.
[3]National Human Genome Center, Howard University, Washington, D. C., U.S.A.

Summary

The correlated Weibull regression model for the analysis of correlated binary data is presented. This regression model is based on Bonney's disposition model for the regression analysis of correlated binary outcomes. Parameter estimation was done through the maximum likelihood method. The correlated Weibull regression model was contrasted with the correlated logistic regression model. The results showed that both regression models were useful in explaining the familial aggregation of oesophageal cancer. The correlated logistic regression model fitted the oesophageal cancer data better than the correlated Weibull regression model for both the non-nested and nested cases. Furthermore, the correlated logistic regression model was computationally more attractive than the correlated Weibull regression model.

Key words: Correlated binary data; Non-nested disposition model; Nested disposition model; Weibull distribution.

1.   Introduction

The occurrence of correlated binary data in the study of familial diseases necessitates the characterisation of the dependence structure and the response probabilities associated with it. This calls for the introduction of the disposition model as a basis for analysing correlated

binary data. The development of the disposition model involves the derivation and parameterisation of the joint distribution on which the likelihood function is based. Here, the experimental unit is the nuclear family and the response is the disease status. In such studies, the methods of estimating the parameters of the models are of particular importance. Here, the maximum likelihood method will be used to analyse the models. Since closed-form solutions are not possible, the Newton-Raphson iteration method is applied to obtain maximum likelihood estimates of the parameter vector. With this study, potential risk factors for disease such as smoking and age can be identified. Also, it can be assessed whether the disease tends to aggregate in families as a result of common shared risks. Such knowledge is decisive for counselling in the aetiology of familial disease.

Liang and Zeger (1986) introduced the use of 'generalised estimating equations' (GEE), an extension of generalised linear models, for estimating regression parameters in situations when the vector of association parameters is a nuisance parameter. The approach is to use a working generalised linear model for the marginal distribution of the outcome variable. The method gives efficient estimates of regression coefficients, although estimates of the association among the binary outcomes can be inefficient. Liang, Zeger and Qaqish (1992) discussed the use of 'generalised estimating equations' (GEE1 and GEE2) for regression analysis of multivariate binary data, focusing on the regression and association parameters. They recommended the use of GEE1, introduced by Liang and Zeger (1986), when the association parameter is considered as a nuisance and the number of clusters is large relative to the size of each cluster. On the other hand, GEE2, introduced by Zhao and Prentice (1990), is preferable to GEE1 when there are few clusters and/or the association parameter is of primary interest. In order to accommodate the many complicating features associated with real data, Bonney (1998) derived joint distributions for constructing likelihood functions. The central aspects of his work concern the notion of disposition to an outcome. He used a moment series representation to derive the joint distributions. Bonney (1998) and Kwagyan (2000) developed estimation procedures for the non-nested and nested cases of the disposition model. An application of the disposition model for the analysis of ordinal tree damage in forest ecosystem was treated by Kötting, Bonney and Urfer (1998).

In this paper, computationally attractive models with readily interpretable dependence structure for the regression analysis of correlated binary data will be presented. Estimation is based on the log likelihood function, whose solutions can be solved by the Newton-Raphson

iteration. In Section 2, the standard Weibull distribution and its parameters will be discussed. Section 3 introduces the disposition model (Bonney, 1998) and its associated likelihood function. The first level extension of the disposition model will be considered in Section 4. Parameter estimation for the models will be treated in Section 5. Section 6 illustrates the methods with oesophageal cancer data. The last section contains a discussion of the methods and the experiences gained.

2.   The standard Weibull distribution

The purpose of this section is to review some basic concepts of survival theory of the standard Weibull distribution. This is necessary since there is a link between the constructions of the likelihood functions of the standard Weibull distribution and the correlated Weibull regression model. This link will be discussed at the end of Section 3.

Consider the two-parameter Weibull distribution denoted by $T \sim W(\phi, \rho)$ $(\phi > 0, \rho > 0)$, where T is the lifetime of a living organism or a product, or the time until the occurrence of some specified event, ϕ is the shape parameter and ρ is the scale parameter, and let $T_1, T_2, ..., T_n$ be a random sample of size n from T.

The probability density function (PDF), which is sometimes also called the unconditional failure rate, is given by

$$f_T(t; \phi, \rho) = \begin{cases} \phi \rho t^{\phi-1} \exp(-\rho t^\phi), t > 0, \\ 0, \qquad\qquad\qquad \text{otherwise} \end{cases} \qquad (2.1)$$

where $\phi > 0$, $\rho > 0$ are real parameters (Gross and Clark, 1975).

The cumulative distribution function (CDF)

$$F_T(t; \phi, \rho) = P(T \le t) = \begin{cases} 0, \qquad\qquad\qquad t \le 0 \\ 1 - \exp(-\rho t^\phi), t > 0 \end{cases} \qquad (2.2)$$

is called the lifetime distribution or failure distribution. If T represents time at death of an individual, $F_T(t; \phi, \rho)$ is the probability that an individual dies before time t. On the other hand, if T represents age of first occurrence of a certain event (e.g., chronic disease), then $F_T(t; \phi, \rho)$ represents age of onset distribution of the event (disease) (Gross and Clark, 1975; Elandt-Johnson and Johnson, 1980).

The survival function (SF), which is defined as the probability of an individual surviving beyond time t, is given by

$$S_T(t) = \Pr(T > t) = 1 - F_T(t; \phi, \rho) = \exp(-\rho t^{\phi}) \tag{2.3}$$

(Gross and Clark, 1975; Elandt-Johnson and Johnson, 1980). In survival analysis, $S_T(t)$ is more commonly used, instead of its complementary function, $F_T(t; \phi, \rho)$.

The hazard function (HF), which characterises the instantaneous failure rate when T = t, conditional on survival to time t, is defined mathematically as

$$h_T(t) = \lim_{\Delta t \to 0} \left[\frac{\Pr(t < T < t + \Delta t \mid T \geq t)}{\Delta t} \right] \tag{2.4}$$

(Gross and Clark, 1975). The hazard function, also termed the failure rate, may also be defined as a measure of proneness to failure. This can also be expressed as

$$h_T(t) = -\frac{\frac{d}{dy} S_T(t)}{S_T(t)} = -\frac{d}{dt} \log_e S_T(t) = \phi \rho t^{\phi - 1} \tag{2.5}$$

(Gross and Clark, 1975; Nelson, 1972). For values of the shape parameter, ϕ, less than 1, the hazard function is a decreasing function, for $\phi = 1$, the Weibull distribution is an exponential distribution and has a constant failure rate, and for $\phi > 1$, it is an increasing function of t (Nelson, 1972). An increasing hazard rate indicates that a unit of age t is more likely to fail in a given increment of time than it would be in the same increment of time at an earlier age. For example, the probability that an individual survives to age 71, given that he has lived to age 70, is greater than the probability that an individual survives to age 72, given that he has lived to age 71. Similarly, a decreasing hazard rate means that the unit is improving with age. For example, children who have undergone an operative procedure to correct a congenital

condition such as a heart defect represent a population exhibiting a decreasing hazard rate. This is because the principal risk of death is the surgery or complications immediately thereafter (Gross and Clark, 1975). A constant hazard rate results due to chance failures (e.g., accidents). Such random occurrences are often independent of age.

The failure rate function of a discrete distribution $\{p_k\}_{k=0}^{\infty}$ (e.g., geometric, binomial, poisson, etc.) is

$$h(k) = \frac{p_k}{\sum_{j=k}^{\infty} p_j} \tag{2.6}$$

(Barlow and Proschan, 1965). We note that in this case $h(k) \le 1$.

From (2.1), (2.3) and (2.5), it follows that

$$f_T(t) = h_T(t) S_T(t) . \tag{2.7}$$

Any distribution of survival times can be characterised by the three equivalent functions $f_T(t)$, $h_T(t)$ and $S_T(t)$.

In observational studies of the time to failure of units (e.g., breakdown of a machine, death of an individual), a group of data may be incomplete in the sense that some units may not have failed by the end of the study, or may have been withdrawn before the end of the study. Such data are said to be censored (Daintith and Nelson, 1989).

Censoring is said to be on the right when the item or subject is observed prior to failure or death. Since the event time is larger than the time of observation, such an observation provides information on the survival function, $S_T(t)$, evaluated at the time of observation (Klein and Moeschberger, 1997).

On the other hand, censoring is said to be on the left when failure or death occurs prior to some designated censoring time. Since the event time has already occurred, such an observation provides information on the cumulative distribution function, $F_T(t)$, evaluated at the time of observation (Klein and Moeschberger, 1997).

An observation corresponding to an exact event time provides information on the density function of T at this time, that is, $f_T(t)$ (Klein and Moeschberger, 1997).

The likelihood function may take the following form:

$$L \propto \prod_{j \in D} f_T(t_j) \prod_{j \in R} S_T(t_j) \prod_{j \in L} F_T(t_j), \qquad (2.8)$$

where, D is the set of death times, R the set of right-censored observations and L is the set of left-censored observations (Klein and Moeschberger, 1997). If the data set comprises only right-censored and left-censored observations, the above likelihood function reduces to

$$L \propto \prod_{j \in R} S_T(t_j) \prod_{j \in L} F_T(t_j). \qquad (2.9)$$

The following are some examples on censored data.

Ex. 1: In a particular clinical trial, suppose that all n patients are followed until death. Their recorded survival times are $t_1,...,t_n$, and it is assumed that the death density function for the jth patient is given by the Weibull density function. The likelihood function $L(t; \phi, \rho)$ is given by

$$L(t; \phi, \rho) = \prod_{j=1}^{n} f(t_j; \phi, \rho) = \prod_{j=1}^{n} \phi \rho t_j^{\phi-1} \exp(-\rho t_j^{\phi}) \qquad (2.10)$$

(Gross and Clark, 1975).

Ex. 2: Suppose that we only know that out of n individuals starting at time zero, r died before time t', and $(n - r)$ survived beyond t' (i.e., censored data). The statistical model for this set of data is binomial, so that the likelihood function is

$$L(t; \phi, \rho) = \binom{n}{r} F_T(t'; \theta)]^r [S_T(t';)\theta]^{n-r} \qquad (2.11)$$

(Elandt-Johnson and Johnson, 1980).

3.   The non-nested disposition model

Disposition, as defined by Bonney, is the tendency of an individual or group to manifest an outcome (e.g., to be affected by a disease). The central aspect of the development of the disposition model is to derive joint distributions that directly capture aggregation, if there should be any. In this section, there will be a brief presentation of the disposition model (Bonney, 1998) and its associated joint distribution function.

Consider a binary outcome $Y = 1$ or 0, with q_0 group-specific covariates, $Z_0^T = (Z_{01},...,Z_{0q_0})$, and p individual-specific covariates, $X_j^T = (X_{j1},...,X_{jp})$, $j = 1,...,n$, measured on several groups of individuals. We consider two types of dispositions here: the group disposition, δ_0, which is determined by the group-specific covariates, Z_0, and the individual disposition, δ_j (for individual j), which is determined by the group-specific covariates, Z_0, and the individual-specific covariates, X_j, $j = 1,...,n$.

Define the group or overall disposition, δ_0, by

$$\delta_0 = \frac{\mu_0}{\alpha_0},$$
(3.1)

where μ_0 is the baseline (i.e., $X_j = 0$, $j = 1,...,n$) disposition under no aggregation and α_0 is the relative disposition. Then, $\alpha_0 < 1$ corresponds to positive aggregation, $\alpha_0 = 1$ corresponds to no aggregation, and $\alpha_0 > 1$ corresponds to negative aggregation.

The logit of the group disposition can be written as

$$\log \frac{\delta_0}{1-\delta_0} = M_0(Z_0) + D_0(Z_0),$$
(3.2)

where

$$M_0(Z_0) = \log \frac{\mu_0}{1 - \mu_0} \tag{3.3}$$

and

$$D_0(Z_0) = \log \frac{\delta_0}{1 - \delta_0} - \log \frac{\mu_0}{1 - \mu_0}. \tag{3.4}$$

We term $M_0(Z_0)$ the logit of group disposition assuming no aggregation or the cluster logit mean risk and $D_0(Z_0)$ the excess disposition due to aggregation or the excess cluster logit disposition due to dependence among members of a group.

From (3.3) and (3.4), it follows that

$$\mu_0 = \frac{1}{1 + \exp\{-[M_0(Z_0)]\}}, \ \delta_0 = \frac{1}{1 + \exp\{-[M_0(Z_0) + D_0(Z_0)]\}} \tag{3.5}$$

and therefore

$$\alpha_0 = \frac{\mu_0}{\delta_0} = \frac{1 + \exp\{-[M_0(Z_0) + D_0(Z_0)]\}}{1 + \exp\{-[M_0(Z_0)]\}}. \tag{3.6}$$

Now, we decompose the logit of the individual disposition as

$$\log \frac{\delta_j}{1 - \delta_j} = M_0(Z_0) + D_0(Z_0) + W_j(X_j) =: \theta_j, \tag{3.7}$$

$j = 1,\ldots,n$, where $M_0(Z_0)$ and $D_0(Z_0)$ are as defined above, and $W_j(X_j)$ is a function of the individual-specific covariates. It follows that

$$\delta_j = \frac{1}{1 + \exp(-\theta_j)} = \frac{1}{1 + \exp\{-[M_0(Z_0) + D_0(Z_0) + W_j(X_j)]\}}, \tag{3.8}$$

$j = 1,\ldots,n.$

The joint probability for a group or cluster becomes

$$P(Y_1 = y_1,..., Y_n = y_n) = (1 - \alpha_0)\prod_{j=1}^{n}(1 - y_j) + \alpha_0\prod_{j=1}^{n}\delta_j^{y_j}(1 - \delta_j)^{1-y_j} , \qquad (3.9)$$

with α_0 and δ_j as defined in (3.6) and (3.8). Explicit derivation of the joint distribution can be found in Bonney (1998). If $\alpha_0 = 1$ or $D_0(Z_0) = 0$, equation (3.9) reduces to

$$P(Y_1 = y_1,..., Y_n = y_n) = \prod_{j=1}^{n}\delta_j^{y_j}(1 - \delta_j)^{1-y_j} , \qquad (3.10)$$

that is, the independence case. Explicit parameterisations for $M_0(Z_0)$ and $D_0(Z_0)$ are obtained by the linear models

$$M_0(Z_0) = \xi_{00} + \xi_{01}Z_{01} + ... + \xi_{0q_0}Z_{0q_0} \qquad (3.11)$$

and

$$D_0(Z_0) = \gamma_{00} + \gamma_{01}Z_{01} + ... + \gamma_{0q_0}Z_{0q_0} . \qquad (3.12)$$

The set of parameters to be determined in the model is

$$\lambda = (\xi_0, \gamma_0, \beta) = (\xi_{00},...,\xi_{0q_0}, \gamma_{00},...,\gamma_{0q_0}, \beta_1,...,\beta_p) .$$

It is now convenient to compare and contrast the standard Weibull distribution with the correlated Weibull regression model. We denote the likelihood function of the joint distribution in Equation (3.9) by $L_k(\lambda \mid y)$, $k = 1,...,K$:

$$L_k(\lambda \mid y) = (1 - \alpha_0)\prod_{j=1}^{n}(1 - y_j) + \alpha_0\prod_{j=1}^{n}\delta_j^{y_j}(1 - \delta_j)^{1-y_j} ,$$

$$\delta_j = \frac{1}{1 + \exp\{-[M_0(Z_0) + D_0(Z_0) + (1 - \exp(\beta_1 X_{j1} + ... + \beta_p X_{jp}))]\}} , \quad j = 1,...,n, \text{ and recall that}$$

the likelihood function for the standard Weibull distribution based on (2.9) is

9

$$L \propto \prod_{j \in R} S_T(t_j) \prod_{j \in L} F_T(t_j).$$

The following differences are observed. (1) In the case of the standard Weibull distribution, the response variable is a variable of time (continuous or discrete), whereas the response variable in Bonney's disposition model presented in this paper is the disease status, and therefore binary. (2) As opposed to the standard Weibull distribution whose most applied characterisation revolves around its role in extreme value theory (e.g., daily maximum or minimum temperatures, precipitation, etc.), Bonney's model is fitted with parameters like δ_j and α_0 to model the effect of influential factors and to capture aggregation in families, if there should be any. Here, variables of time (e.g., age) are regarded as covariates in the model. Our concern, however, is to determine the link between the standard Weibull distribution and the correlated Weibull regression model. Suppose T is the length of time until the occurrence of a certain disease, and consider a group of size n with survival times $T_1,...,T_n$, where T_j is censored or not at time t_j with the censoring indicator $y_j = 0$ if censored, and $y_j = 1$ if uncensored. Then, in the above likelihood functions, $y_j = 0$ in the correlated Weibull regression model corresponds to the survival function in the standard Weibull distribution, and $y_j = 1$ in the correlated Weibull regression model corresponds to the cumulative distribution function in the standard Weibull distribution. In other words,

$$L_k(\lambda \mid y) = (1-\alpha_0) + \alpha_0 \prod_{j=1}^{n}(1-\delta_j)$$

$$= (1-\alpha_0) + \alpha_0 \prod_{j=1}^{n} \frac{\exp\{-[M_0(Z_0) + D_0(Z_0) + (1-\exp(\beta_1 X_{j1} + ... + \beta_p X_{jp}))]\}}{1 + \exp\{-[M_0(Z_0) + D_0(Z_0) + (1-\exp(\beta_1 X_{j1} + ... + \beta_p X_{jp}))]\}}$$

corresponds to $L \propto \prod_{j \in R} S_T(t_j) = \prod_{j \in R} \exp(-\rho t_j^{\phi})$

and

$$L_k(\lambda \mid y) = \alpha_0 \prod_{j=1}^{n} \frac{1}{1 + \exp\{-[M_0(Z_0) + D_0(Z_0) + (1-\exp(\beta_1 X_{j1} + ... + \beta_p X_{jp}))]\}}$$

corresponds to $L \propto \prod_{j \in L} F_T(t_j) = \prod_{j \in L} \{1 - \exp(-\rho t_j^{\phi})\}$, with the above parameters as previously defined. Thus, in this sense, the two likelihood functions are equivalent.

4.   First level nesting

Consider a binary outcome $Y = 1$ or 0, with q_0 group-specific covariates, $Z_0^T = (Z_{01},...,Z_{0q_0})$, q subgroup-specific covariates, $Z_i^T = (Z_{i1},...,Z_{iq})$, $i = 1,...,m$, and p individual-specific covariates, $X_{ij}^T = (X_{ij1},...,X_{ijp})$, $i = 1,...,m$, $j = 1,...,n_i$, measured on several individuals. Bonney (1998) considered three types of dispositions here: the group (cluster) disposition, δ_0, which is determined by the group-specific covariates, Z_0, the subgroup disposition, δ_i, $i = 1,...,m$, which is determined by the group-specific covariates, Z_0, and the subgroup-specific covariates, Z_i, $i = 1,...,m$, and the individual disposition, δ_{ij}, $i = 1,...,m$, $j = 1,...,n_i$, which is determined by the group-specific covariates, Z_0, the subgroup-specific covariates, Z_i, $i = 1,...,m$, and the individual-specific covariates, X_{ij}, $i = 1,...,m$, $j = 1,...,n_i$.

Then, δ_0 and δ_i are given by

$$\delta_0 = \frac{\mu_0}{\alpha_0} \qquad (4.1)$$

and

$$\delta_i = \frac{\mu_i}{\alpha_i}, \qquad (4.2)$$

$i = 1,...,m$, where μ_0 is the group baseline disposition under no aggregation, μ_i is the subgroup baseline disposition under no aggregation, α_0 is the relative disposition with respect to the group and α_i is the relative disposition with respect to subgroup i, $i = 1,...,m$.

The logit of the individual disposition is then

$$\log \frac{\delta_{ij}}{1 - \delta_{ij}} = M_0(Z_0) + D_0(Z_0) + M_i(Z_i) + D_i(Z_i) + W_{ij}(X_{ij}) \; =: \; \theta_{ij}, \qquad (4.3)$$

$i = 1, \ldots, m, j = 1, \ldots, n_i$, where

$$M_0(Z_0) = \log \frac{\mu_0}{1 - \mu_0} \qquad (4.4)$$

is the cluster logit mean risk,

$$D_0(Z_0) = \log \frac{\delta_0}{1 - \delta_0} - \log \frac{\mu_0}{1 - \mu_0} \qquad (4.5)$$

is the excess cluster logit disposition due to dependence among members of the group,

$$M_i(Z_i) = \log \frac{\mu_i}{1 - \mu_i} - \log \frac{\delta_0}{1 - \delta_0}, \qquad (4.6)$$

$i = 1, \ldots, m$, is the excess on the logit scale of the mean risk in subgroup i above that due to the cluster disposition,

$$D_i(Z_i) = \log \frac{\delta_i}{1 - \delta_i} - \log \frac{\mu_i}{1 - \mu_i}, \qquad (4.7)$$

$i = 1, \ldots, m$, is the excess on the logit scale of the disposition in subgroup i that cannot be explained by the overall cluster disposition and the differences in μ_i, $i = 1, \ldots, m$, and

$$W_{ij}(X_{ij}), \qquad (4.8)$$

$i = 1, \ldots, m, j = 1, \ldots, n_i$, is a function of the individual-specific covariates.

From (4.4)-(4.7), it follows that

$$\mu_0 = \frac{1}{1 + \exp\{-[M_0(Z_0)]\}}, \quad \delta_0 = \frac{1}{1 + \exp\{-[M_0(Z_0) + D_0(Z_0)]\}},$$

$$\mu_i = \frac{1}{1 + \exp\{-[M_0(Z_0) + D_0(Z_0) + M_i(Z_i)]\}}, \; i = 1, \ldots, m,$$

$$\delta_i = \frac{1}{1 + \exp\{-[M_0(Z_0) + D_0(Z_0) + M_i(Z_i) + D_i(Z_i)]\}}, \quad i = 1,\ldots,m, \qquad (4.9)$$

and therefore

$$\alpha_0 = \frac{\mu_0}{\delta_0} = \frac{1 + \exp\{-[M_0(Z_0) + D_0(Z_0)]\}}{1 + \exp\{-[M_0(Z_0)]\}}, \qquad (4.10)$$

$$\alpha_i = \frac{\mu_i}{\delta_i} = \frac{1 + \exp\{-[M_0(Z_0) + D_0(Z_0) + M_i(Z_i) + D_i(Z_i)]\}}{1 + \exp\{-[M_0(Z_0) + D_0(Z_0) + M_i(Z_i)]\}}, \qquad (4.11)$$

$i = 1,\ldots,m$, and

$$\delta_{ij} = \frac{1}{1 + \exp(-\theta_{ij})} = \frac{1}{1 + \exp\{-[M_0(Z_0) + D_0(Z_0) + M_i(Z_i) + D_i(Z_i) + W_{ij}(X_{ij})]\}},$$
$$(4.12)$$

$i = 1,\ldots,m, \; j = 1,\ldots,n_i$.

With these, the joint probability for the first level nesting becomes

$$P(Y_{11} = y_{11},\ldots,Y_{mn_i} = y_{mn_i}) = (1 - \alpha_0)\prod_{i=1}^{m}\prod_{j=1}^{n_i}(1 - y_{ij})$$

$$+ \alpha_0\prod_{i=1}^{m}\left\{(1 - \alpha_i)\prod_{j=1}^{n_i}(1 - y_{ij}) + \alpha_i\prod_{j=1}^{n_i}\delta_{ij}^{y_{ij}}(1 - \delta_{ij})^{1-y_{ij}}\right\}. \qquad (4.13)$$

The derivation of the joint distribution can be found in Bonney (1998). Explicit parameterisations for $M_0(Z_0)$, $D_0(Z_0)$, $M_i(Z_i)$ and $D_i(Z_i)$ are obtained by the linear models

$$M_0(Z_0) = \xi_{00} + \xi_{01}Z_{01} + \ldots + \xi_{0q_0}Z_{0q_0}, \qquad (4.14)$$

$$D_0(Z_0) = \gamma_{00} + \gamma_{01}Z_{01} + \ldots + \gamma_{0q_0}Z_{0q_0}, \qquad (4.15)$$

13

$$M_i(Z_i) = \xi_1 Z_{i1} + \ldots + \xi_q Z_{iq},\qquad\qquad(4.16)$$

$i = 1,\ldots,m$, and

$$D_i(Z_i) = \gamma_1 Z_{i1} + \ldots + \gamma_q Z_{iq},\qquad\qquad(4.17)$$

$i = 1,\ldots,m$.

The set of parameters to be determined in the model is

$$\lambda = (\xi, \gamma, \beta) = (\xi_{00}, \ldots, \xi_{0q_0}, \xi_1, \ldots, \xi_q, \gamma_{00}, \ldots, \gamma_{0q_0}, \gamma_1, \ldots, \gamma_q, \beta_1, \ldots, \beta_p).$$

If $\alpha_i = 1$ or $D_i(Z_i) = 0$, $i = 1,\ldots,m$, equation (4.13) reduces to the non-nested case. Also, if $\alpha_0 = 1$ and $\alpha_i = 1$, or equivalently, if $D_0(Z_0) = 0$ and $D_i(Z_i) = 0$, equation (4.13) reduces to the independence case.

5.   Estimation

The method of maximum likelihood is used to determine estimates of the unknown model parameters, $\lambda = (\xi, \gamma, \beta)$. Since closed-form solutions are not possible here, the Newton-Raphson iteration method is applied to obtain estimates of the parameter vector. The Newton-Raphson method requires the first and second derivatives of the log likelihood functions. To estimate the parameters in the model, the joint function of all the clusters is required, but there is no loss of generality if the joint function of a cluster is considered.

The maximum likelihood estimations for the correlated logistic regression model have been done by Bonney (1998) and Kwagyan (2000). The presentation in this paper is based on their work.

5.1　Parameter estimation for the non-nested case

Denote the likelihood function of the joint probability in Equation (3.9) by $L_k(\lambda \mid y)$, $k = 1,...,K$:

$$L_k(\lambda \mid y) = (1 - \alpha_0)\prod_{j=1}^{n}(1 - y_j) + \alpha_0 \prod_{j=1}^{n}\delta_j^{y_j}(1 - \delta_j)^{1-y_j}$$

$$= (1 - \alpha_0)\prod_{j=1}^{n}(1 - y_j) + \alpha_0 L_{\pi j}, \tag{5.1.1}$$

where $L_{\pi j} = \prod_{j=1}^{n} L_j$, $L_j = \delta_j^{y_j}(1 - \delta_j)^{1-y_j}$ and

$$\delta_j = \frac{1}{1 + \exp\{-[M_0(Z_0) + D_0(Z_0) + (1 - \exp(\beta_1 x_{j1} + ... + \beta_p x_{jp}))]\}}, j = 1,...,n.$$

This gives the score function

$$U_k(\lambda) = A_k(\lambda)\alpha_0^{*} + B_k(\lambda)\left[\sum_{j=1}^{n} U_j\right], \tag{5.1.2}$$

$k = 1,...,K$, where $\alpha_0^{*}(\lambda) = -(1 - \delta_0)\frac{\delta}{\delta\lambda}D_0(Z_0) + \delta_0(1 - \alpha_0)\frac{\delta}{\delta\lambda}M_0(Z_0)$,

$$A_k(\lambda) = \frac{\alpha_0\left[L_{\pi j} - \prod_{j=1}^{n}(1 - y_j)\right]}{L_k}, k = 1,...,K, \; B_k(\lambda) = \frac{\alpha_0}{L_k}L_{\pi j}, \; k = 1,...,K,$$

$$U_j(\lambda \mid y) = (y_j - \delta_j)\theta_j^{(1)} = (y_j - \delta_j)\frac{\delta}{\delta\lambda}[M_0(Z_0) + D_0(Z_0) + W_j(X_j)]$$

$$= (y_j - \delta_j)\begin{pmatrix} \dfrac{\delta}{\delta\xi_0}M_0(Z_0) \\ \dfrac{\delta}{\delta\gamma_0}D_0(Z_0) \\ \dfrac{\delta}{\delta\beta}W_j(X_j) \end{pmatrix} = (y_j - \delta_j)\begin{pmatrix} Z_0' \\ Z_0' \\ -X_j \exp(\beta^T X_j) \end{pmatrix},$$

$Z_0^T = (1, Z_{01}, Z_{02},...,Z_{0q_0})$, $\beta^T = (\beta_1,...,\beta_p)$ and $X_j^T = (X_{j1},...,X_{jp})$, $j = 1,...,n$.

The Hessian matrix is given by

$$H_k(\lambda) = \frac{\prod_{j=1}^{n}(1-y_j)}{L_k} A_k \alpha_0^* \alpha_0^{*T} + \frac{\prod_{j=1}^{n}(1-y_j)}{L_k} B_k \left[\alpha_0^* \left[\sum_{j=1}^{n} U_j \right]^T + \left[\sum_{j=1}^{n} U_j \right] \alpha_0^{*T} \right]$$

$$+ \frac{(1-\alpha_0)\prod_{j=1}^{n}(1-y_j)}{L_k} B_k \left[\sum_{j=1}^{n} U_j \right] \left[\sum_{j=1}^{n} U_j \right]^T + B_k \left[\sum_{j=1}^{n} H_j \right] + A_k \frac{\delta}{\delta\lambda} \alpha_0^*, \qquad (5.1.3)$$

$k = 1,...,K$, where

$$H_j(\lambda) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -(y_j - \delta_j)X_j X_j^T \exp(\beta^T X_j) \end{pmatrix}$$

$$- \delta_j(1-\delta_j) \begin{pmatrix} Z_0' Z_0^T & Z_0' Z_0^T & -Z_0'[X_j \exp(\beta^T X_j)]^T \\ Z_0' Z_0^T & Z_0' Z_0^T & -Z_0'[X_j \exp(\beta^T X_j)]^T \\ -[X_j \exp(\beta^T X_j)]Z_0^T & -[X_j \exp(\beta^T X_j)]Z_0^T & X_j X_j^T \exp(2\beta^T X_j) \end{pmatrix},$$

$j = 1,...,n$.

For the correlated logistic regression model, the following are the corresponding expressions for δ_j, $U_j(\lambda)$ and $H_j(\lambda)$:

$$\delta_j = \frac{1}{1 + \exp\{-[M_0(Z_0) + D_0(Z_0) + (\beta_1 x_{j1} + ... + \beta_p x_{jp})]\}}, \quad j = 1,...,n,$$

$$U_j(\lambda \mid y) = (y_j - \delta_j) \begin{pmatrix} Z_0' \\ Z_0' \\ X_j \end{pmatrix}, \quad j = 1,...,n,$$

and

$$H_j(\lambda) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} - \delta_j(1-\delta_j) \begin{pmatrix} Z_0' Z_0^T & Z_0' Z_0^T & Z_0' X_j^T \\ Z_0' Z_0^T & Z_0' Z_0^T & Z_0' X_j^T \\ X_j Z_0^T & X_j Z_0^T & X_j X_j^T \end{pmatrix}$$

$$= - \delta_j(1-\delta_j) \begin{pmatrix} Z_0' Z_0^T & Z_0' Z_0^T & Z_0' X_j^T \\ Z_0' Z_0^T & Z_0' Z_0^T & Z_0' X_j^T \\ X_j Z_0^T & X_j Z_0^T & X_j X_j^T \end{pmatrix}, \quad j = 1,...,n.$$

The Fisher Information matrix is

$$I_k(\lambda) = \alpha_0 \sum_{j=1}^{n} I_j(\lambda) \ - \ A_k^* \alpha_0^* \alpha_0^{*T} \ - \ B_k^* \left[\alpha_0^* \left[\sum_{j=1}^{n} U_j^* \right]^T + \left[\sum_{j=1}^{n} U_j^* \right] \alpha_0^{*T} \right]$$

$$- \ B_k^*(1-\alpha_0) \left[\sum_{j=1}^{n} U_j^* \right] \left[\sum_{j=1}^{n} U_j^* \right]^T , \qquad (5.1.4)$$

$k = 1,\dots,K$, where

$$I_j(\lambda) = \delta_j(1-\delta_j)\theta_j^{(1)'}\theta_j^{(1)T}$$

$$= \delta_j(1-\delta_j) \begin{pmatrix} Z_0'Z_0^T & Z_0'Z_0^T & -Z_0'[X_j \exp(\beta^T X_j)]^T \\ Z_0'Z_0^T & Z_0'Z_0^T & -Z_0'[X_j \exp(\beta^T X_j)]^T \\ -[X_j \exp(\beta^T X_j)]Z_0^T & -[X_j \exp(\beta^T X_j)]Z_0^T & X_j X_j^T \exp(2\beta^T X_j) \end{pmatrix},$$

$j = 1,\dots,n$, and A_k^*, B_k^*, U_j^*, and U_j^{*T} are the resulting values of A_k, B_k, U_j, and U_j^T evaluated at $y = 0$ (see also Bonney (1998) for the logistic version).

5.2 Parameter estimation for the first level nesting

Denote the likelihood function of the joint probability in Equation (4.13) by $L_k(\lambda \mid y)$, $k = 1,\dots,K$:

$$L_k(\lambda|y) \ = \ (1-\alpha_0)\prod_{i=1}^{m}\prod_{j=1}^{n_i}(1-y_{ij}) \ + \ \alpha_0 L_{\pi i}, \qquad (5.2.1)$$

where $L_{\pi i} = \prod_{i=1}^{m} L_i$, $L_i = (1-\alpha_i)\prod_{j=1}^{n_i}(1-y_{ij}) + \alpha_i L_{\pi j}$, $L_{\pi j} = \prod_{j=1}^{n_j} L_j$, $L_j = \delta_{ij}^{y_{ij}}(1-\delta_{ij})^{1-y_{ij}}$ and

$$\delta_{ij} = \frac{1}{1+\exp\{-[M_0(Z_0)+D_0(Z_0)+M_i(Z_i)+D_i(Z_i)+(1-\exp(\beta_1 x_{ij1}+\dots+\beta_p x_{ijp}))]\}},$$

$i = 1,\dots,m$, $j = 1,\dots,n_i$.

The corresponding score function is

$$U_k(\lambda \mid y) = A_k(\lambda)\alpha_0^* + B_k(\lambda)\left[\sum_{i=1}^m U_i\right], \tag{5.2.2}$$

$k = 1,\ldots,K$, where $\alpha_0^*(\lambda) = -(1-\delta_0)\dfrac{\delta}{\delta\lambda}D_0(Z_0) + \delta_0(1-\alpha_0)\dfrac{\delta}{\delta\lambda}M_0(Z_0)$,

$$A_k(\lambda) = \frac{\alpha_0\left[L_{\pi i} - \prod_i^m \prod_j^{n_i}(1-y_{ij})\right]}{L_k}, \ k = 1,\ldots,K, \ B_k(\lambda) = \frac{\alpha_0 L_{\pi i}}{L_k}, \ k = 1,\ldots,K,$$

$$L_{\pi i}(\lambda \mid y) = \prod_{i=1}^m L_i, \ L_i = (1-\alpha_i)\prod_{j=1}^{n_i}(1-y_{ij}) + \alpha_i \prod_{j=1}^{n_i}\delta_{ij}^{y_{ij}}(1-\delta_{ij})^{1-y_{ij}}, \ i = 1,\ldots,m,$$

$$U_i(\lambda \mid y) = A_i(\lambda)\alpha_i^* + B_i(\lambda)\left[\sum_{j=1}^{n_i}U_j\right],$$

$$\alpha_i^* = -(1-\delta_i)\frac{\delta}{\delta\lambda}D_i(Z_i) + \delta_i(1-\alpha_i)\frac{\delta}{\delta\lambda}[M_0(Z_0) + D_0(Z_0) + M_i(Z_i)],$$

$$A_i(\lambda) = \frac{\alpha_i\left[L_{\pi j} - \prod_{j=1}^{n_i}(1-y_{ij})\right]}{L_i}, \ B_i(\lambda) = \frac{\alpha_i}{L_i}L_{\pi j}, \ i = 1,\ldots,m,$$

$$U_j(\lambda \mid y) = (y_{ij} - \delta_{ij})\theta_{ij}^{(1)} = (y_{ij} - \delta_{ij})\frac{\delta}{\delta\lambda}[M_0(Z_0) + D_0(Z_0) + M_i(Z_i) + D_i(Z_i) + W_{ij}(X_{ij})]$$

$$= (y_{ij} - \delta_{ij})\begin{pmatrix} \dfrac{\delta}{\delta\xi_0}M_0(Z_0) \\ \dfrac{\delta}{\delta\gamma_0}D_0(Z_0) \\ \dfrac{\delta}{\delta\xi_i}M_i(Z_i) \\ \dfrac{\delta}{\delta\gamma_i}D_i(Z_i) \\ \dfrac{\delta}{\delta\beta}W_{ij}(X_{ij}) \end{pmatrix} = (y_{ij} - \delta_{ij})\begin{pmatrix} Z_0^{'} \\ Z_0^{'} \\ Z_i \\ Z_i \\ -X_{ij}\exp(\beta^T X_{ij}) \end{pmatrix},$$

$Z_0^T = (1, Z_{01}, Z_{02},\ldots,Z_{0q_0})$, $Z_i^T = (Z_{i1},\ldots,Z_{iq})$, $\beta^T = (\beta_1,\ldots,\beta_p)$ and $X_{ij}^T = (X_{ij1},\ldots,X_{ijp})$,

$i = 1,\ldots,m$, $j = 1,\ldots,n_i$.

The Hessian matrix is given by

$$H_k(\lambda) = B_k\left[\sum_{i=1}^{m}H_i(\lambda)\right] + \frac{\prod_{i=1}^{m}\prod_{j=1}^{n_i}(1-y_{ij})}{L_k}A_k\alpha_0^*\alpha_0^{*T} + \frac{\prod_{i=1}^{m}\prod_{j=1}^{n_i}(1-y_{ij})}{L_k}B_k(1-\alpha_0)\left[\sum_{i=1}^{m}U_i\right]\left[\sum_{i=1}^{m}U_i\right]^T$$

$$+ \frac{\prod_{i=1}^{m}\prod_{j=1}^{n_i}(1-y_{ij})}{L_k}B_k\left\{\left[\sum_{i=1}^{m}U_i\right]\alpha_0^{*T} + \alpha_0^*\left[\sum_{i=1}^{m}U_i\right]^T\right\} + A_k\frac{\delta}{\delta\lambda^T}\alpha_0^*, \qquad (5.2.3)$$

$k = 1,\ldots K$, where

$$H_i = \frac{\prod_{j=1}^{n_i}(1-y_{ij})}{L_i}A_i\alpha_i^*\alpha_i^{*T} + \frac{\prod_{j=1}^{n_i}(1-y_{ij})}{L_i}B_i\left[\alpha_i^*\left[\sum_{j=1}^{n_i}U_j\right]^T + \left[\sum_{j=1}^{n_i}U_j\right]\alpha_i^{*T}\right]$$

$$+ \frac{(1-\alpha_i)\prod_{j=1}^{n_i}(1-y_{ij})}{L_i}B_i\left[\sum_{j=1}^{n_i}U_j\right]\left[\sum_{j=1}^{n_i}U_j\right]^T + B_i\left[\sum_{j=1}^{n_i}H_j\right] + A_i\frac{\delta}{\delta\lambda}\alpha_i^*$$

and

$$H_j(\lambda) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -(y_{ij}-\delta_{ij})X_{ij}X_{ij'}^T\exp(\beta^TX_{ij}) \end{pmatrix}$$

$$- \delta_{ij}(1-\delta_{ij})\begin{pmatrix} Z_0'Z_0^T & Z_0'Z_0^T & Z_0'Z_i^T & Z_0'Z_i^T & -Z_0'w^T \\ Z_0'Z_0^T & Z_0'Z_0^T & Z_0'Z_i^T & Z_0'Z_i^T & -Z_0'w^T \\ Z_iZ_0^T & Z_iZ_0^T & Z_iZ_i^T & Z_iZ_i^T & -Z_iw^T \\ Z_iZ_0^T & Z_iZ_0^T & Z_iZ_i^T & Z_iZ_i^T & -Z_iw^T \\ -wZ_0^T & -wZ_0^T & -wZ_i^T & -wZ_i^T & X_{ij}X_{ij}^T\exp(2\beta^TX_{ij}) \end{pmatrix},$$

$w = [X_{ij}\exp(\beta^TX_{ij})]$, $i = 1,\ldots,m$, $j = 1,\ldots,n$.

For the correlated logistic regression model, we have the following corresponding expressions for δ_{ij}, $U_j(\lambda \mid y)$ and $H_j(\lambda)$:

$$\delta_{ij} = \frac{1}{1 + \exp\{-[M_0(Z_0) + D_0(Z_0) + M_i(Z_i) + D_i(Z_i) + (\beta_1 x_{ij1} + ... + \beta_p x_{ijp})]\}},$$

$i = 1,...,m$, $j = 1,...,n_i$,

$$U_j(\lambda \mid y) = (y_{ij} - \delta_{ij}) \begin{pmatrix} Z_0^{'} \\ Z_0^{'} \\ Z_i \\ Z_i \\ X_{ij} \end{pmatrix}, \quad i = 1,...,m, \ j = 1,...,n_i,$$

and

$$H_j(\lambda) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} - \delta_{ij}(1 - \delta_{ij}) \begin{pmatrix} Z_0^{'}Z_0^{'T} & Z_0^{'}Z_0^{'T} & Z_0^{'}Z_i^{T} & Z_0^{'}Z_i^{T} & Z_0^{'}X_{ij}^{T} \\ Z_0^{'}Z_0^{'T} & Z_0^{'}Z_0^{'T} & Z_0^{'}Z_i^{T} & Z_0^{'}Z_i^{T} & Z_0^{'}X_{ij}^{T} \\ Z_iZ_0^{'T} & Z_iZ_0^{'T} & Z_iZ_i^{T} & Z_iZ_i^{T} & Z_iX_{ij}^{T} \\ Z_iZ_0^{'T} & Z_iZ_0^{'T} & Z_iZ_i^{T} & Z_iZ_i^{T} & Z_iX_{ij}^{T} \\ X_{ij}Z_0^{'T} & X_{ij}Z_0^{'T} & X_{ij}Z_i^{T} & X_{ij}Z_i^{T} & X_{ij}X_{ij}^{T} \end{pmatrix}$$

$$= \delta_{ij}(1 - \delta_{ij}) \begin{pmatrix} Z_0^{'}Z_0^{'T} & Z_0^{'}Z_0^{'T} & Z_0^{'}Z_i^{T} & Z_0^{'}Z_i^{T} & Z_0^{'}X_{ij}^{T} \\ Z_0^{'}Z_0^{'T} & Z_0^{'}Z_0^{'T} & Z_0^{'}Z_i^{T} & Z_0^{'}Z_i^{T} & Z_0^{'}X_{ij}^{T} \\ Z_iZ_0^{'T} & Z_iZ_0^{'T} & Z_iZ_i^{T} & Z_iZ_i^{T} & Z_iX_{ij}^{T} \\ Z_iZ_0^{'T} & Z_iZ_0^{'T} & Z_iZ_i^{T} & Z_iZ_i^{T} & Z_iX_{ij}^{T} \\ X_{ij}Z_0^{'T} & X_{ij}Z_0^{'T} & X_{ij}Z_i^{T} & X_{ij}Z_i^{T} & X_{ij}X_{ij}^{T} \end{pmatrix}, \quad i = 1,...,m, \ j = 1,...,n.$$

The Fisher information matrix for the first level nesting is

$$I_k(\lambda) = \alpha_0 \sum_{i=1}^{m} I_i(\lambda) - A_k^{*}\alpha_0^{*}\alpha_0^{*T} - B_k^{*}(1 - \alpha_0)\left[\sum_{i=1}^{m} U_i^{*}\right]\left[\sum_{i=1}^{m} U_i^{*}\right]^{T}$$

$$- B_k^{*}\left\{\alpha_0^{*}\left[\sum_{i=1}^{m} U_i^{*}\right]^{T} + \left[\sum_{i=1}^{m} U_i^{*}\right]\alpha_0^{*T}\right\}, \quad (5.2.4)$$

$k = 1,...,K$, where

$$I_i(\lambda) = \alpha_i \sum_{j=1}^{n_i} I_j - A_i^* \alpha_i^* \alpha_i^{*T} - B_i^*(1-\alpha_i)\left[\sum_{j=1}^{n_i} U_j^*\right]\left[\sum_{j=1}^{n_i} U_j^*\right]^T - B_i^*\left\{\alpha_i^*\left[\sum_{j=1}^{n_i} U_j^*\right]^T + \left[\sum_{j=1}^{n_i} U_j^*\right]\alpha_i^{*T}\right\},$$

$i = 1,\ldots,m,$

$$I_j(\lambda) = \delta_{ij}(1-\delta_{ij})\theta_{ij}^{(1)}\theta_{ij}^{(1)T}$$

$$= \delta_{ij}(1-\delta_{ij})\begin{pmatrix}
Z_0' Z_0^{'T} & Z_0' Z_0^{'T} & Z_0' Z_i^{T} & Z_0' Z_i^{T} & -Z_0' w^T \\
Z_0' Z_0^{'T} & Z_0' Z_0^{'T} & Z_0' Z_i^{T} & Z_0' Z_i^{T} & -Z_0' w^T \\
Z_i Z_0^{'T} & Z_i Z_0^{'T} & Z_i Z_i^{T} & Z_i Z_i^{T} & -Z_i w^T \\
Z_i Z_0^{'T} & Z_i Z_0^{'T} & Z_i Z_i^{T} & Z_i Z_i^{T} & -Z_i w^T \\
-w Z_0^{'T} & -w Z_0^{'T} & -w Z_i^{T} & -w Z_i^{T} & X_{ij} X_{ij}^T \exp(2\beta^T X_{ij})
\end{pmatrix},$$

$w = [X_{ij}\exp(\beta^T X_{ij})]$, $i=1,\ldots,m$, $j = 1,\ldots,n_i$, and A_k^*, B_k^*, A_i^*, B_i^*, U_i^*, and U_j^*, are the resulting values of A_k, B_k, A_i, B_i, U_i, and U_j evaluated at $y = 0$ (see also Kwagyan (2000) for the logistic version).

6.   Illustrations

Data were collected in the Yangcheng County, Shanxi Province, the Peoples Republic of China, designed to assess the presence of familial aggregation of oesophageal cancer. There were 2951 clusters (families), parents and siblings forming two subgroups of individuals. Cluster sizes were distributed as follows:

Cluster size	3	4	5	6	7	8	9	10	11	12	13
Number of clusters	623	819	659	412	232	129	43	23	8	2	1

The independent variables were smoking status, alcohol, age, sib size (sibsize) and mean sib age (meansibage). There were no group-specific covariates. The subgroup-specific covariates consisted of sibsize and meansibage, and the individual-specific covariates consisted of smoking status, alcohol and age. Smoking status was coded as 0 for non-smokers and 1 for smokers, alcohol was coded as 0 for non-drinkers and 1 for drinkers, and age was measured in years. The response variable Y was coded as 0 for unaffected and 1 for affected.

6.1 Model for the non-nested case

In this subsection, we shall determine the correlated logistic and the correlated Weibull regression models. We shall also compare the model fit of the two regression models.

We note that there are no group-specific covariates in the data set. Therefore, the cluster logit mean risk, $M_0(Z_0)$, and the excess cluster logit disposition due to dependence among members of a group, $D_0(Z_0)$, become $M_0(Z_0) = \xi_{00}$ and $D_0(Z_0) = \gamma_{00}$, respectively (see Equations (3.11) and (3.12)). We also note that in the non-nested case aggregations in subgroups are not considered. The only variables in the model are therefore the individual-specific covariates: smoking status (X_1), alcohol (X_2) and age (X_3). Thus, the function that describes the effects of the individual-specific covariates becomes $W_j(X_j) = 1 - \exp(\beta_1 X_{j1} + \beta_2 X_{j2} + \beta_3 X_{j3})$, $j = 1,\ldots,n$, for the correlated Weibull regression model and $W_j(X_j) = \beta_1 X_{j1} + \beta_2 X_{j2} + \beta_3 X_{j3}$, $j = 1,\ldots,n$, for the correlated logistic regression model, for the jth individual. The set of parameters to be determined is therefore $\lambda = (\xi, \gamma, \beta) = (\xi_{00}, \gamma_{00}, \beta_1, \beta_2, \beta_3)$.

Table 6.1.1 presents the results of the correlated Weibull regression model (left panel) and of the correlated logistic regression model (right panel). The table shows regression parameter estimates, standard deviations of the parameter estimates and Wald statistics for determining whether the parameters in the model are needed.

We note that as opposed to the correlated logistic regression model, where a positive value of the coefficient of the individual-specific covariate indicates increased probability for a disease, a negative value of the coefficient of the individual-specific covariate is indicative of increased probability for a disease for the correlated Weibull regression model. For both models, a positive value of the coefficient of the group-specific covariate increases the probability for a disease. For example, the negative coefficient of age in the correlated Weibull regression model indicates that age increases the probability for oesophageal cancer. All the coefficients in Table 6.1.1 are statistically significant in both the correlated Weibull regression model and the correlated logistic regression model, when compared to the $(1-\frac{\alpha}{2})$th quantile of the standard normal distribution at an α-level of 0.05 (i.e., $u_{0.975} = 1.96$).

Table 6.1.1: Parameter estimates, standard deviations and Wald statistics using the correlated Weibull and the correlated logistic regression models

Variable	Parameter	Correlated Weibull regression model			Correlated logistic regression model		
		Parameter estimate	Standard deviation	Wald statistic	Parameter estimate	Standard deviation	Wald statistic
constant	ξ_{00}	-2.4630	0.0387	63.6434	-3.7617	0.0934	40.2752
constant	γ_{00}	0.1272	0.0319	3.9875	0.0510	0.0250	2.0400
smoking	β_1	-0.6657	0.2673	2.4905	0.5006	0.0597	8.3853
alcohol	β_2	2.1720	0.2581	8.4153	-1.1208	0.1701	6.5891
age	β_3	-0.0262	0.0027	9.7037	0.0364	0.0016	22.7500

Critical value for the rejection of the null hypothesis: $u_{0.975} = 1.96$.

To test the hypothesis of 'no aggregation of oesophageal cancer in a cluster', we test the hypothesis that $D_0(Z_0) = 0$, or more specifically, $\gamma_{00} = 0$. We do this by performing the likelihood ratio test and the Wald's test.

For the correlated Weibull regression model, the log likelihood under the null hypothesis is $\log L_0 = -5673.0479$ and the log likelihood based on the full data is $\log L_1 = -5665.1874$. The likelihood ratio test statistic is therefore LR_W = -2[-5673.0479-(-5665.1874)] = 15.7210, which is significant when compared to a chi-square distribution with one degree of freedom (i.e., $\chi_1^2 = 3.8415$). For the correlated logistic regression model, the corresponding values are $\log L_0 = -5494.8614$ and $\log L_1 = -5492.7594$. The likelihood ratio test statistic is therefore LR_L = -2[-5494.8614 – (-5492.7594)] = 4.2040, which is also significant.

We now perform the Wald's tests. In Table 6.1.1, the value of γ_{00} is 0.1272 for the correlated Weibull regression model. The value of the Wald statistic is $Z_W = 3.9875$, and the critical value is $u_{0.975} = 1.96$. Because $Z_W > u_{0.975}$, the null hypothesis will be rejected (see, for example, Garthwaite et al., 1995). The conclusion is that there is evidence of familial aggregation of oesophageal cancer. For the correlated logistic regression model, the Wald

statistic is $Z_L = 2.0400$. Since the Wald statistic is large, the null hypothesis will be rejected, indicating that there is significant aggregation of oesophageal cancer in the families.

We finally compare the model fit of the correlated Weibull regression model with that of the correlated logistic regression model using the Akaike's Information Criterion (AIC) (Akaike, 1974). The AIC of the correlated Weibull regression model is

$AIC_W = -2\log L_1 + 2(\text{number of estimated parameters}) = 11330.3748 + 10 = 11340.3748,$

and that of the correlated logistic regression model is

$AIC_L = -2\log L_1 + 2(\text{number of estimated parameters}) = 10985.5187 + 10 = 10995.5187.$

The correlated logistic regression model has minimum AIC, and therefore fits the data better.

6.2 Model for the first level nesting

Since there are no group-specific covariates in the data set, the cluster logit mean risk, $M_0(Z_0)$, and the excess cluster logit disposition due to dependence among members of a group, $D_0(Z_0)$, become $M_0(Z_0) = \xi_{00}$ and $D_0(Z_0) = \gamma_{00}$, respectively (see Equations (4.14) and (4.15)). Two subgroups are nested within each family: parents form the first subgroup (i.e., $i = 1$) and siblings the second (i.e., $i = 2$). No variables are available for subgroup 1. The variables for subgroup 2 are sibsize and meansibage. Therefore, the excess on the logit scale of the mean risk in group 2 above that due to the cluster disposition, $M_2(Z_2)$, and the excess on the logit scale of the disposition within group 2 that cannot be explained by the overall cluster disposition and differences in baseline disposition under no aggregation in the group, $D_2(Z_2)$, become $M_2(Z_2) = \xi_1 Z_{21} + \xi_2 Z_{22}$ and $D_2(Z_2) = \gamma_1 Z_{21} + \gamma_2 Z_{22}$, respectively (see Equations (4.16) and (4.17)).

The individual-specific covariates are smoking status (X_1), alcohol (X_2) and age (X_3). Thus, the function that describes the effects of the individual-specific covariates becomes $W_{ij}(X_{ij}) = 1 - \exp(\beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ij3})$, $i = 1,\ldots,m$, $j = 1,\ldots,n_i$, for the correlated Weibull regression model and $W_{ij}(X_{ij}) = \beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ij3}$, $i = 1,\ldots,m$, $j = 1,\ldots,n_i$, for the

correlated logistic regression model, for the jth individual in group i. The set of parameters to be estimated is therefore $\lambda = (\xi, \gamma, \beta) = (\xi_{00}, \xi_1, \xi_2, \gamma_{00}, \gamma_1, \gamma_2, \beta_1, \beta_2, \beta_3)$.

Table 6.2.1 provides analysis of the oesophageal cancer data. The table gives maximum likelihood estimates, standard deviations and Wald statistics for the correlated Weibull regression model (left panel) and the correlated logistic regression model (right panel).

Table 6.2.1: Parameter estimates, standard deviations and Wald statistics using the correlated Weibull and the correlated logistic regression models

Variable	Parameter	Correlated Weibull regression model			Correlated logistic regression model		
		Parameter estimate	Standard deviation	Wald statistic	Parameter estimate	Standard deviation	Wald statistic
constant	ξ_{00}	-4.4426	0.1154	38.4974	-4.6036	0.1243	37.0362
sibsize	ξ_1	0.0172	0.0146	1.1781	0.0183	0.0152	1.2039
meansibage	ξ_2	0.0412	0.0019	21.6842	0.0365	0.0021	17.3810
constant	γ_{00}	-0.0965	0.0342	2.8216	-0.1042	0.0342	3.0468
sibsize	γ_1	-0.0117	0.0149	0.7852	-0.0179	0.0123	1.4553
meansibage	γ_2	0.0077	0.0015	5.1333	0.0081	0.0013	6.2308
smoking	β_1	-1.2751	0.3082	4.1372	0.5812	0.0654	8.8869
alcohol	β_2	2.2346	0.3157	7.0782	-0.9633	0.1768	5.4485
age	β_3	-0.0247	0.0046	5.3696	0.0191	0.0020	9.5500

Critical value for the rejection of the null hypothesis: $u_{0.975} = 1.96$.

The negative coefficient of age in the correlated Weibull regression model indicates that age increases the probability for oesophageal cancer. With the exception of ξ_1 and γ_1, all the coefficients of both regression models are statistically significant.

The hypotheses to be tested are $H_0 : \gamma = 0$ and $H_1 : \gamma \neq 0$. The following critical values will be used in this subsection for the rejection of H_0: $u_{0.975} = 1.96$ for the 1-parameter Wald's test and $\chi^2_{3,0.95} = 7.8147$ for the likelihood ratio test.

The Wald's test rejects the null hypotheses $'\gamma_{00} = 0'$ and $'\gamma_2 = 0'$ of both the correlated Weibull regression model and the correlated logistic regression model, since the test statistics are large. The conclusion is that there is significant aggregation of oesophageal cancer in families and in siblings. It follows that the meansibage affects the familial aggregation of oesophageal cancer. On the other hand, the null hypothesis $'\gamma_1 = 0'$ of both disposition models cannot be rejected, since the test statistics are small. Hence, the sibsize does not affect the familial aggregation of oesophageal cancer.

For the correlated Weibull regression model, the maximised log likelihood from which γ is omitted is $\log L_0 = -5361.6679$, and the full log likelihood is $\log L_1 = -5323.3685$. The likelihood ratio statistic is therefore $\mathrm{LR}_W = -2[-5361.6679 - (-5323.3685)] = 76.5988$.

For the correlated logistic regression model, the corresponding values are $\log L_0 = -5353.7628$ and $\log L_1 = -5309.0410$. The likelihood ratio statistic is therefore $\mathrm{LR}_L = -2[-5353.7628 - (-5309.0410)] = 89.4436$. Thus, for both disposition models, significant familial aggregation is observed (see, for example, Wilks, 1938).

The AIC of the correlated Weibull regression model is $\mathrm{AIC}_W = 10664.7370$ and that of the correlated logistic regression model is $\mathrm{AIC}_L = 10636.0820$. The correlated logistic regression model minimises the AIC, and is therefore considered to be the more appropriate model.

7. Discussion

The correlated Weibull regression models for correlated binary data have been presented. The objective of the analyses has been to assess familial aggregation of diseases. In Section 6, the model fit of the correlated Weibull regression model was compared to that of the correlated logistic regression model using the Akaike Information Criterion (AIC). The model that minimised the AIC was considered to give a better fit to the oesophageal cancer data. The correlated logistic regression model fitted the data better than the correlated Weibull regression model for both the non-nested and nested cases. On the whole, the correlated

logistic regression model was computationally more feasible than the correlated Weibull regression model.

The data processing was done using the statistical program package SAS, and computations were made in the C programming language. Further research has to be done to study the performance of the correlated Weibull regression model as the level of nesting gets deeper.

Acknowledgements

References

[1]   Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Acoustics 19, 716-723.

[2]   Barlow, R. E. and Proschan, F. (1965). Mathematical Theory of Reliability. John Wiley and Sons, New York.

[3]   Bonney, G. E. (1998). Regression analysis of disposition to correlated binary outcomes. Department of Biostatistics Technical Report No. 1998-001, Fox Chase Cancer Center, Philadelphia, PA, U.S.A.

[4]   Daintith, J. and Nelson, R. D. (1989). The Penguin Dictionary of Mathematics. Cox and Wyman Limited, Great Britain.

[5]   Deutsch, R. (1965). Estimation theory. Prentice-Hall, New Jersey.

[6]   Elandt-Johnson, R. C. and Johnson, N. L. (1980). Survival models and data analysis. John Wiley and Sons, U.S.A.

[7]   Garthwaite, P. H., Jolliffe, I. T. and Jones, B. (1995). Statistical inference. Prentice Hall, London, New York.

[8]   Gross, A. J. and Clark, V. A. (1975). Survival distributions: Reliability Applications in the Biomedical Sciences. John Wiley and Sons, New York.

[9]   Klein, J. P. and Moeschberger, M. L. (1997). Survival Analysis: Techniques for Censored and Truncated Data. Springer-Verlag, New York.

[10]   Kötting, J., Bonney, G. E. and Urfer, W. (1998). Disposition models for the analysis of dynamic changes in Forest-Ecosystems. Department of statistics Technical Report No. 24/1998, University of Dortmund, Germany.

[11]   Kwagyan, J. (2000). Maximum likelihood inference in the models of disposition for correlated binary outcomes. Ph.D. Thesis, The Temple University Graduate School.

[12]   Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalised linear models. Biometrika 73, 1, 13-22.

[13]   Liang, K.-Y., Zeger, S. L. and Qaqish, B. (1992). Multivariate regression analysis for categorical data. Journal of the Royal Statistical Society, Series B 54, 3-40.

[14]   Nelson, W. B. (1972). Theory and Applications of Hazard Plotting for Censored Failure Data. Technometrics 14, 945-965.

[15]   Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. Annals of mathematical statistics, 9, 60-62.

[16]   Zhao, L. P. and Prentice, R. L. (1990). Correlated binary regression using a quadratic exponential model. Biometrika 77, 642-648.