Published in final edited form as:
Sci China Ser A Math. 2009 June 1; 52(6): 1113–1138. doi:10.1007/s11425-009-0076-9.

Gaining Efficiency via Weighted Estimators for Multivariate Failure Time Data*

Jianqing Fan[1,2], **Yong Zhou**[2,3], **Jianwen Cai**[4], and **Min Chen**[3]

[1] Department of Operations Research and Financial Engineering, Princeton, NJ08544

[2] Department of Statistics, Shanghai University of Finance and Economics, Shanghai 200433

[3] Academy of Mathematics and System Science, Chinese Academy of Science, Beijing 100190

[4] Department of Biostatistics, University of North Carolina at Chapel Hill, Chepel Hill, NC 27599-7420

Abstract

Multivariate failure time data arise frequently in survival analysis. A commonly used technique is the working independence estimator for marginal hazard models. Two natural questions are how to improve the efficiency of the working independence estimator and how to identify the situations under which such an estimator has high statistical efficiency. In this paper, three weighted estimators are proposed based on three different optimal criteria in terms of the asymptotic covariance of weighted estimators. Simplified close-form solutions are found, which always outperform the working independence estimator. We also prove that the working independence estimator has high statistical efficiency, when asymptotic covariance of derivatives of partial log-likelihood functions is nearly exchangeable or diagonal. Simulations are conducted to compare the performance of the weighted estimator and working independence estimator. A data set from Busselton population health surveys is analyzed using the proposed estimators.

Keywords

Marginal hazard model; pseudo-partial likelihood; working independence estimator; optimal weight

1 Introduction

Statistical estimation and inference for the marginal hazard models for multivariate failure time data are vital in survival analysis. Multivariate failure time data arise frequently in biomedical research and financial credit risk analysis. For example, such data could arise when related subjects in clusters are at risk of a failure time event or study subjects are at risk of recurrence of the same event. Another example is that default of a firm can have contagious effect on the default time of other firms, particularly for those in the same sector or industry. A key feature of this type of data is that the failure times may be dependent. When there is at most one event for each subject and these subjects are mutually independent, the Cox (1972) proportional hazards model has commonly been used to assess the effects of covariates on failure times.

The Cox model is a simple and mathematically convenient way to study and explain covariate effects. However, in many biomedical studies, the covariate effects from cluster or recurrent

Corresponding Author: Yong Zhou, Institute of Applied Mathematics, Academy of Mathematics and System Science, Chinese Academy of Science, Beijing 100080, P. R. China, yzhou@amss.ac.cn.
Dedicated to Professor Zhidong Bai in honor of his 65th birthday

data can be more complicated than the specified structure and new analytic challenges arise in assessing covariate effects in correlated multivariate survival data. Beyond the traditional independence model, there are infinitely many possibilities to model the dependence among the clusters of survival data. Depending on the background of studies, one often chooses a specific form that reasonably explains the objective of the study. For example, the effect of covariate variables and confounding correlation on the hazard risk may vary with the level of clusters. This leads naturally to consider the following Cox (1972) type marginal hazard model:

$$\lambda_{ij}(t)=\lambda_{j0}(t)\exp\{\beta^T \boldsymbol{Z}_{ij}(t)\}, \quad t \geq 0. \tag{1}$$

for the jth failure type of the ith subject($i = 1, 2, \cdots, n, j = 1, 2, \cdots, J_i$), and $\lambda_{j0}(t)$ is an unspecified "baseline" hazard function pertaining to the jth failure type, and $\boldsymbol{\beta}$ is an unknown parameter with dimension p. The model also accommodates the cluster data with different baseline hazard functions for different clusters by regarding j as the index of cluster and i as the subject within the j-th index. Considerable efforts have been made on the marginal hazard model (1). See, for example, Wei, Lin, and Weissfeld (1989), Cai and Prentice (1995, 1997), Spiekerman and Lin (1998), and Cai (1999) for parametric covariate effect studies and Cai *et al.* (2007) and Cai *et al.* (2007, 2008) for nonparametric covariate effect models.

The most commonly used analysis for model (1) is the working independence analysis in which the dependence of the data is ignored when estimating the unknown parameters. Applying a standard Cox procedure will produce consistent estimators but the standard error estimators need to be computed differently to reflect the possible dependence of the multivariate failure time. Lee and Kapadia (1992) showed how standard error estimates should be computed under the independence assumption. Wei *et al.* (1989),Cai and Prentice (1995,1997) and Spiekerman and Lin (1998) also proposed procedures for calculating the standard errors. In many applications, ignoring dependence structure presented in data cannot result in efficient estimators for the unknown parameters. Similarly to estimating equations considered by Liang and Zeger (1986), we can introduce weighted estimating equations to improve the efficiency of such estimators. Liang and Zeger (1986) pointed out that the use of working independence correlation structures may result in a notable loss of efficiency in the generalized linear model, when the correlation coefficient is large, or the non-homogeneity in covariate is large. One hopes naturally to identify the situations under which ignoring the dependence is statistically acceptable in estimating unknown parameters in the marginal hazards model. This will be the subject of Section 4.

Cai and Prentice (1995, 1997) introduced a weighted approach to estimate the regression parameter $\boldsymbol{\beta}$ in the distinguished baseline marginal hazard model (1). The weighting matrix is allowed to depend on survival time and unknown parameters. The asymptotic properties they established apply to general weight matrix safisfying their specified condition and they considered using the inverse of the estimated correlation matrix as the weight matrix in detail. However, they did not consider the problem of finding the optimal weight. Gray and Li (2002) developed an algorithm for choosing the optimal weights, but this method cannot be applied to multiple covariate. Recently, Cai et al (2007, CFZZ) proposed a weighted estimation method for marginal Cox models with varying-coefficients for multivariate failure time data. The weight was selected by minimizing the asymptotic variance of the estimator of varying-coefficients. Glidden (2007) suggested a copula method to capture dependence for clustered failure-time data. Yu and Lin (2008) studied marginal proportional hazards model with the effect of covariates modelled nonparametrically for correlated failure time data. Chen, Chen and Ying (2008, CCY) proposed an estimator for regression parameter in marginal Cox model for multivariate failure time data by linear combination of martingale residuals.

The procedure proposed in this paper is different from the above approaches. The proposed weighted estimators of parameters are to select "optimal" weight by combination of the score functions of the partial likelihood instead of linear combination of martingale residuals or minimizing (CCY) the asymptotic variance of estimator of varying-coefficients (CFZZ). Unlike Glidden (2007), we do not assume any particular structure for the dependence. Our aim is to give a simple weighting scheme that outperforms the procedure based on the working independence. Through mathematical simplifications, our optimal weighting schemes admit closed form. This makes practical implementations very simple and stable.

In a different setting, Heyde (1997) proposed a method for combining estimating equations to improve efficiency, which is in the same spirit of the general method of moments of Hensen (1982). The optimal weighting matrix can be explicitly found. This method is more general and more efficient. However, a serious drawback of such an approach is the need for calculation of the inverse of a large estimated covariance matrix of the score functions of the partial likelihood. Because of estimation in high-dimensionality, the implementation of such an approach is often unstable and inaccurate. The efficiency gain cannot always be materialized due to noise accumulation in the estimated large covariance matrix, which contains many elements. This motivated us to consider other weighting schemes, too. To overcome these difficulties, various working correlation were proposed, such as Zeger and Liang (1985) and Liang and Zeger (1986). On the other hand, when the dimensionality is small, the method can be viable. See related research in Andersen (2004), Schaubel and Cai (2005), Larocque et al (2007), and Kuk (2007). We will extend the method of Heyde (1997) to the correlated censored data and compare with other schemes.

The paper is organized as follows. Section 2 summarizes the results for marginal partial likelihood method and introduces a simple weighting scheme for combining correlated likelihood. In section 3, we introduce simple strategies for selecting weights that will improve the working independence estimator. The setting under which the working independence is nearly optimal will be discussed in Section 4. Simulation results and real data applications are given in Section 5. Concluding remarks is provided in Section 6.

2 Summary of marginal partial likelihood

2.1 Estimation with working independence

Suppose that there is a random sample of size n from an underlying population. Let i denote the individual and j denote the type of failure one might experience. For individual i and failure type j, let T_{ij} ($i = 1, 2, \cdots, n$, $j = 1, 2, \cdots, J_i$) denote the failure time, C_{ij} the censoring time, and $X_{ij} = min(T_{ij}, C_{ij})$ the observed event time with censoring indicator Δ_{ij}. We assume that the censoring times are independent of the failure times given the covariates, i.e. the censoring is noninformative. The observed data structure is

$$\{\mathbf{X}_{ij}, \Delta_{ij}, \mathbf{Z}_{ij}\} \quad \text{for } i=1, \cdots, n; j=1, 2, \cdots, J_i,$$

where $\mathbf{Z}_{ij} = (Z_{ij1}, \cdots, Z_{ijp})^T$ is a p×1 vector of covariates.

Let $N_{ij}(t) = I(X_{ij} \le t, \Delta_{ij} = 1)$ denote the counting process for failure and $Y_{ij}(t) = I(X_{ij} \ge t)$ denote the at risk process. A commonly used marginal model that links the covariate with failure time is postulated in (1).

For ease of presentation, we drop the dependence of covariates on time, with the understanding that the methods and proofs in this paper are applicable to external time dependent covariates

NIH-PA Author Manuscript
NIH-PA Author Manuscript

(Kalbfleisch and Prentice 2002). For a given failure type j, we define the following marginal partial likelihood,

$$L_j(\beta)=\prod_{i=1}^{n_j}\left\{\frac{\exp\{\beta^T\mathbf{Z}_{ij}\}}{\sum_{l\in\mathcal{R}_j(X_{ij})}\exp\{\beta^T\mathbf{Z}_{lj}\}}\right\}^{\Delta_{ij}},\quad j=1,2,\cdots,J \tag{2}$$

where n_j is the number of individuals for the jth failure type, $J = \max\{J_1, \cdots, J_n\}$, and $R_j(t) = \{i: X_{ij} \geq t\}$ denotes the set of the individuals at risk for failure type j just prior to time t. If all the failure times are independent, the partial likelihood function for inference of $\boldsymbol{\beta}$ in model (1) is

$$L(\beta)=\prod_{j=1}^{J}L_j(\beta). \tag{3}$$

To avoid the technicality of tail problems, frequently only the data up to certain time point τ are used. From the partial likelihood (2), by multiplying a constant n_j^{-1}, we have the following marginal partial log-likelihood:

$$\ell_j(\beta,\tau)=n_j^{-1}\sum_{i=1}^{n_j}\int_0^{\tau}\beta^T\mathbf{Z}_{ij}(u)dN_{ij}(u)-n_j^{-1}\sum_{i=1}^{n_j}\int_0^{\tau}\log\{\sum_{l=1}^{n_j}Y_{lj}(u)\exp(\beta^T\mathbf{Z}_{lj}(u))\}dN_{ij}(u), \tag{4}$$

for $j = 1, 2, \cdots, J$. For the sake of simplicity, write $\ell_j(\boldsymbol{\beta}) = \ell_j(\boldsymbol{\beta}, \tau)$.

As in Cai and Prentice (1995), by introducing $X_{ij} = T_{ij} = 0$ if necessary, we can drop the dependence of J_i on individual i, with the understanding that varying cluster size can be accommodated. For the sake of simplicity, we assume that $J_i = J$, $i = 1, 2, \cdots, n$.

Marginal pseudo-partial log-likelihood under working independence is defined by

$$\ell(\beta)=\sum_{j=1}^{J}\ell_j(\beta). \tag{5}$$

The working independence partial likelihood estimator $\hat{\boldsymbol{\beta}}_I$ is obtained by maximizing the same function as the partial likelihood for the Cox model with independent failure times. The asymptotic normality of $\hat{\boldsymbol{\beta}}_I$ can be demonstrated using a similar technique as in Andersen and Gill (1982).

2.2 Asymptotic normality

In marginal models, the asymptotic normality of the pseudo-partial likelihood estimator has been derived by Cai and Prentice (1995). For a vector $\mathbf{a}$, we denote $\mathbf{a}^{\otimes 0} = 1$, $\mathbf{a}^{\otimes 1} = \mathbf{a}$, and $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$. Let

$$S_{jk}(\beta,u)=\frac{1}{n}\sum_{i=1}^{n}Y_{ij}(u)\exp(\beta^T\mathbf{Z}_{ij}(u))(\mathbf{Z}_{ij}(u))^{\otimes k},\quad k=0,1,2$$

and $s_{jk}(\boldsymbol{\beta}, u)$ be its asymptotic limits. Write

$$E_j(\beta,t)=\frac{S_{j1}(\beta,t)}{S_{j0}(\beta,t)} \text{ and } V_j(\beta,t)=\frac{S_{j2}(\beta,t)}{S_{j0}(\beta,t)}-E_j(\beta,t)^{\otimes 2}.$$

Set $e_j = s_{j1}/s_{j0}$, $v_j=s_{j2}/s_{j0}-e_j^{\otimes 2}$ and

$$\sum\nolimits_j(\beta)=\int_0^\tau v_j(\beta,t)s_{j0}(\beta,t)\lambda_{j0}(t)dt.$$

It has been shown by Cai and Prentice (1995) and Clegg et al.(2000) that under some regularity assumptions, we have

$$\sqrt{n}(\widehat{\beta}_I-\beta)\xrightarrow{\mathcal{L}}N(0,\mathcal{I}(\beta))$$

where

$$\mathcal{I}(\beta)=\mathbf{A}^{-1}(\beta)\sum\nolimits_m(\beta)\mathbf{A}^{-1}(\beta)$$

in which $\mathbf{A}(\beta)=\sum_{j=1}^{J}\sum_j(\beta)$ and $\sum_m(\beta)=\sum_{l=1}^{J}\sum_{k=1}^{J}E\{\Pi_{1k}(\beta)(\Pi_{1l}(\beta))^T\}$, with

$$\Pi_{jk}(\beta)=\int_0^\tau\left[Z_{jk}-\frac{s_{k1}(\beta,u)}{s_{k0}(\beta,u)}\right]dM_{jk}(u). \tag{6}$$

The unknown quantities in (6) can be estimated by the substitution method as follows. A natural estimator to $\mathbf{D}_{kl}(\boldsymbol{\beta}) = E\{\Pi_{1k}(\boldsymbol{\beta})\Pi_{1l}(\boldsymbol{\beta})^T\}$ is

$$\widehat{\mathbf{D}}_{kl}(\widehat{\beta}_I)=n^{-1}\sum_{j=1}^{n}\widehat{\varphi}_{jk}\widehat{\varphi}_{jl}^T \tag{7}$$

where

$$\widehat{\varphi}_{jk}=\Delta_{jk}\left\{Z_{jk}(X_{jk})-\frac{S_{k1}(\widehat{\beta}_I,X_{jk})}{S_{k0}(\widehat{\beta}_I,X_{jk})}\right\}-\frac{1}{n}\sum_{m=1}^{n}\frac{\Delta_{mk}Y_{jk}(X_{mk})\exp\{\widehat{\beta}_I^T Z_{jk}(X_{mk})\}}{S_{k0}(\widehat{\beta}_I,X_{mk})}\left\{Z_{jk}(X_{mk})-\frac{S_{k1}(\widehat{\beta}_I,X_{mk})}{S_{k0}(\widehat{\beta}_I,X_{mk})}\right\}. \quad (8)$$

We then estimate the covariance matrix $\boldsymbol{\Sigma}_m(\boldsymbol{\beta})$ by

$$\widehat{\sum}_m(\widehat{\beta}_I)=\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{J}\sum_{k=1}^{J}\widehat{\varphi}_{ik}\varphi_{il}^T.$$

Similarly, $\mathbf{A}(\boldsymbol{\beta})$ can be estimated by $\widehat{\mathbf{A}}(\widehat{\beta}_I)=\sum_{j=1}^{J}\widehat{\sum}_j(\widehat{\beta}_I)$, where

$$\widehat{\sum}_j(\beta)=\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau\frac{S_{j2}(\beta,u)S_{j0}(\beta,u)-(S_{j1}(\beta,u))^{\otimes 2}}{(S_{j0}(\beta,u))^2}dN_{ij}(u). \quad (9)$$

Hence, the asymptotic covariance of the working independence estimator can be consistently estimated by

$$\widehat{\mathcal{I}}(\widehat{\beta}_I)=\widehat{\mathbf{A}}(\widehat{\beta}_I)^{-1}\widehat{\sum}_m(\widehat{\beta}_I)\widehat{\mathbf{A}}(\widehat{\beta}_I)^{-1}.$$

2.3 Weighted partial likelihood approach

To utilize the dependence among subjects within a cluster or failure time of different types, we consider the weighted partial likelihood. Specifically, instead of using $\ell(\boldsymbol{\beta})$ in (5), we consider the following weighted partial likelihood function:

$$\ell_W(\beta)=\sum_{j=1}^{J}w_j\ell_j(\beta), \quad (10)$$

where $\ell_j(\boldsymbol{\beta})$ is defined as in (4) and $\{w_j\, j = 1, \cdots, J\}$ are unknown weights. For simplicity, write $\mathbf{W} = (w_1, \cdots, w_J)^T$. The proposed estimator is denoted by $\hat{\boldsymbol{\beta}}_W$ that maximizes $\ell_W(\boldsymbol{\beta})$, which is convex when all weights are nonnegative. Note that when $w_j = 1/J$, the estimator is the working independence estimator.

By similar arguments as in Cai and Prentice(1995), it can be shown that

$$\mathrm{var}(\widehat{\beta}_W)=\mathbf{A}_W^{-1}(\beta)\sum_W(\beta)\mathbf{A}_W^{-1}(\beta),$$

which will be denoted by $\sum_W^\star(\beta)$, where $\mathbf{A}_W(\beta)=\sum_{j=1}^{J}w_j\sum_j(\beta)$ and

$$\sum\nolimits_{w}(\beta)=\sum_{j=1}^{J}w_j^2\sum\nolimits_{j}(\beta)+2\sum_{k<l}^{n}w_k w_l \mathbf{D}_{kl}(\beta). \quad (11)$$

The asymptotic covariance can be estimated by using the substitution method as in Section 2.2.

Theorem 1—Under the assumptions of Cai and Prentice (1995), if $\boldsymbol{\beta}_0$ is the true value of $\boldsymbol{\beta}$, we have

$$\sqrt{n}(\widehat{\beta}_w-\beta_0)\xrightarrow{\mathcal{L}}N(0,\sum\nolimits_{w}^{\star}(\beta_0)).$$

3 Selection of Weights

It is difficult to find the optimal weights that minimize the asymptotic covariance matrix $\sum_{w}^{\star}(\beta_0)$, particularly when J is large. Hence, our goal reduces to choose viable weights with close forms that outperform the working independence, which corresponds to the choice $w_j = 1/J$ for all failure types. For simplicity of the notation, we consider the population version. It is understood that the unknown quantities will be estimated by using the substitution method. We will consider the following three criteria.

3.1 Componentwise variance

This criterion attempts to choose the weight to minimize the variance of $\hat{\beta}_k$, where β_k is the kth component of $\boldsymbol{\beta}$. Such a choice depends on k. Let $\sum_{j}^{(k)}$ denote the kth diagonal entry of matrix $\boldsymbol{\Sigma}_j$ and $\mathbf{D}_{ij}^{(k)}$ be defined similarly. Direct minimization of the k-th diagonal element of $\sum_{w}^{\star}$ is not feasible or analytic. Hence, an approximation solution is sought.

Assume that $\boldsymbol{\Sigma}_j(\boldsymbol{\beta}) \approx b_j\,\boldsymbol{\Sigma}$ for a given $\boldsymbol{\Sigma}$ and scale b_j. For multivariate failure time data, this assumption implies that the information matrix for different failure types has the same structure and only differs by a multiple of a constant. This condition will be satisfied if the survival probability for different failure types are proportional over time and the covariates considered for different failure types follow the same distribution. Under assumption $\boldsymbol{\Sigma}_j(\boldsymbol{\beta}) \approx b_j\boldsymbol{\Sigma}$, we try to minimize the variance $var(\hat{\beta}_k)$ of the kth component of $\boldsymbol{\beta}$. When $\boldsymbol{\Sigma}_j(\boldsymbol{\beta}) \approx b_j\boldsymbol{\Sigma}$ is true, the variance $\sum_{w}^{\star}(\beta_0)$ of $\hat{\beta}$ becomes

$$\sum\nolimits^{-1}\left(\sum_{j=1}^{J}w_j^2\sum\nolimits_{j}+2\sum_{i<j}w_i w_j\mathbf{D}_{ij}\right)\sum\nolimits^{-1}\left(\sum_{j=1}^{J}w_j b_j\right)^{-2}. \quad (12)$$

Since $\boldsymbol{\Sigma}^{-1}$ does not depend on $\mathbf{W}$, one possibility is to minimize the matrix

$$\left(\sum_{j=1}^{J} w_j^2 \sum\nolimits_j + 2\sum_{i<j} w_i w_j \mathbf{D}_{ij}\right)\left(\sum_{j=1}^{J} w_j b_j\right)^{-2}.$$

Focusing only on the kth diagnonal element of

$$\left(\sum_{j=1}^{J} w_j^2 \sum\nolimits_j + 2\sum_{i<j} w_i w_j \mathbf{D}_{ij}\right)\left(\sum_{j=1}^{J} w_j b_j\right)^{-2},$$

we would like to find **W** to minimize

$$\left(\sum_{j=1}^{J} w_j^2 \sum\nolimits_j^{(k)} + 2\sum_{i<j} w_i w_j \mathbf{D}_{ij}^{(k)}\right)\left(\sum_{j=1}^{J} w_j b_j\right)^{-2}.$$

Thus our problem becomes to minimize

$$\left(\sum_{j=1}^{J} w_j^2 \sum\nolimits_j^{(k)} + 2\sum_{i<j} w_i w_j \mathbf{D}_{ij}^{(k)}\right) = \mathbf{w}^T \mathbf{D} \mathbf{w}, \tag{13}$$

subject to $\sum_{j=1}^{J} w_j b_j = 1$, where $\mathbf{D}$ is a $(J \times J)$ symmetric matrix with diagonal elements $\sum_j^{(k)}$ and off-diagonal elements $\mathbf{D}_{ij}^{(k)}$. By the Lagrange multiplier method, we can easily show that the solution is given by

$$\mathbf{w} = \mathbf{D}^{-1}\mathbf{b}/\mathbf{b}^T\mathbf{D}^{-1}\mathbf{b}, \tag{14}$$

where $\mathbf{b}$ is a vector with elements being b_j ($j = 1, 2, \cdots, J$).

Another possibility is to directly minimize the asymptotic variance of the kth component $\hat{\beta}_k$ of $\hat{\boldsymbol{\beta}}$. According to (12), this is given by

$$\left(\sum_{j=1}^{J} w_j^2 \tilde{\sum}_j^{(k)} + 2\sum_{i<j} w_i w_j \tilde{\mathbf{D}}_{ij}^{(k)}\right)\left(\sum_{j=1}^{J} w_j b_j\right)^{-2}, \tag{15}$$

where $\tilde{\sum}_j^{(k)}$ and $\tilde{\mathbf{D}}_{ij}^{(k)}$ are the kth diagonal element of $\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{\Sigma}^{-1}\mathbf{D}_{ij}\boldsymbol{\Sigma}^{-1}$, respectively. Estimating $\widehat{\sum} = \sum_{j=1}^{J} \widehat{\sum}_j / J$, the problem admits a similar solution to (14).

Different choices of the vector **b** give different suboptimal solutions and our strategy is to take the one with the smallest var($\hat{\beta}_k$). Here are several choices based on the assumption $\mathbf{\Sigma}_j(\boldsymbol{\beta}) \approx b_j\mathbf{\Sigma}$:

1. $b_j = \sum_j^{(k)}(\beta)$ $(j = 1,\cdots,J)$;
2. $b_j = \mathrm{tr}(\mathbf{\Sigma}_j(\boldsymbol{\beta}))$, $(j = 1, \cdots, J)$, where tr is the trace of a matrix;
3. $b_j = 1$ for all j.

With those three choices of **b'***s*, we obtain their corresponding weighting vectors as in (14). We now pick the best weighting scheme among the four weighting schemes: The three just constructed and the working independence weight. The resulting weighting scheme always improves the working independence estimator, in terms of the efficiency of estimating β_k. This method will be abbreviated as "CW". Note that we can also get thee additional weighting schemes from (15) using the same choice of the vector **b**, resulting in seven choices instead of three choices of **W**. This would improve our estimation method further. Unfortunately, this was not implemented in our numerical studies.

3.2 Total variance

The total variance method is to choose the weights to optimize the performance of all entries of $\boldsymbol{\beta}$ simultaneously. It intends to solve the following optimization problem:

$$\min_W \mathrm{tr}\left(\sum\nolimits_w^{\star}(\beta)\right). \tag{16}$$

Again, the close-form solution cannot be found and a simplified version is sought.

Assume again that $\mathbf{\Sigma}_j(\boldsymbol{\beta}) \approx b_j\mathbf{\Sigma}$. The optimization problem reduces to

$$\mathrm{tr}\left(\sum_{j=1}^{J} w_j^2 \sum^{-1} \sum\nolimits_j \sum^{-1} + 2\sum_{k<l} w_k w_l \sum^{-1} \mathbf{D}_{kl} \sum^{-1}\right) = \mathbf{w}^T\mathbf{H}\mathbf{w},$$

subject to $\sum_{j=1}^{J} w_j = 1$, where **H** is a symmetric matrix with diagonal elements $\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_j\mathbf{\Sigma}^{-1})$ and off-diagonal elements $\mathrm{tr}(\mathbf{\Sigma}^{-1}\,\mathbf{D}_{kl}\,\mathbf{\Sigma}^{-1})$. Extending the constraint to $\sum_{j=1}^{J} w_j b_j = 1$, by the Lagrange multiplier method, the solution is explicitly given by

$$\mathbf{w} = \mathbf{H}^{-1}\mathbf{b}/\mathbf{b}^T\mathbf{H}^{-1}\mathbf{b}, \tag{17}$$

where **b** is a vector with elements being b_j $(j = 1, 2, \cdots, J)$. The unknown parameters tr $(\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_j\mathbf{\Sigma}^{-1})$ and $\mathrm{tr}(\mathbf{\Sigma}^{-1}\,\mathbf{D}_{kl}\,\mathbf{\Sigma}^{-1})$ can be estimated respectively by $\mathrm{tr}(\hat{\mathbf{\Sigma}}^{-1}\,\hat{\mathbf{\Sigma}}_j\,\hat{\mathbf{\Sigma}}^{-1})$ and $\mathrm{tr}(\hat{\mathbf{\Sigma}}^{-1}\,\hat{\mathbf{D}}_{kl}\,\hat{\mathbf{\Sigma}}^{-1})$, where $\widehat{\sum} = \sum_{j=1}^{J}\widehat{\sum}_j/J$.

Similar to the componentwise case, the following are two possible choices of b_j.

1. $b_j = \mathrm{tr}(\hat{\mathbf{\Sigma}}_j)$, $(j = 1, \cdots, J)$, or

2. b_j's always are 1.

Let the corresponding solution (17) be respectively $\mathbf{W}^{(A)}$ and $\mathbf{W}^{(D)}$.

Another simple and workable procedure is to simply take $w_j = [\mathrm{tr}(\boldsymbol{\Sigma}_j(\hat{\boldsymbol{\beta}})^{-1})]^{-1}$ as weights. The intuition is that each failure type gives an estimate of $\boldsymbol{\beta}_0$, with covariance matrix $\sum_j^{-1}$, or overall accuracy $\mathrm{tr}(\sum_j^{-1})$. More informative (about $\boldsymbol{\beta}$) failure types are assigned with heavier weights. The method to estimate parameter $\boldsymbol{\beta}$ from this weight will be denoted by $\mathbf{W}^{(d)}$.

Counting the working independence weight $\mathbf{W}^{(I)}$, we have four weighting schemes to choose: $\mathbf{W}^{(A)}$, $\mathbf{W}^{(D)}$, $\mathbf{W}^{(d)}$ and $\mathbf{W}^{(I)}$. We will pick the one with the smallest total variance $\mathrm{tr}\{\sum_w^{\star}(\widehat{\beta})\}$. The method will be denoted by "WT".

3.3 General scoring method

We now consider a general weighting scheme as in Heyde (1997), which combines J groups of estimating equations for $\boldsymbol{\beta}$ from J failure types:

$$\mathbf{W}\mathcal{S}(\beta)=0, \tag{18}$$

where $\mathbf{W}$ is a $p \times Jp$ weight matrix and $\mathcal{S}(\beta)=((\ell_1'(\beta))^T, (\ell_2'(\beta))^T, \cdots, (\ell_J'(\beta))^T)^T$ is a vector of score functions, with the score of the marginal pseudo-partial likelihood function $\ell_j'(\beta)$ defined by the gradient vector of (4).

Suppose that our aim is to find the weighting matrix $\mathbf{W}$ such that the asymptotic covariance matrix is minimized. Following the argument in Heyde (1997), the optimal weight matrix is given by $\mathbf{W}_{opt} = E(\mathcal{S}')\boldsymbol{\Sigma}^{-1}$, where $\mathcal{S}' = \partial\mathcal{S}/\partial\boldsymbol{\beta}$ is the first order derivative of $\mathcal{S}(\boldsymbol{\beta})$, a $(Jp) \times p$ matrix, and $\boldsymbol{\Sigma}$ is the asymptotic variance matrix of the score function $\mathcal{S}(\boldsymbol{\beta})$, a $(Jp) \times (Jp)$ matrix. See also Hensen (1982).

The weight $\mathbf{W}_{opt}$ depends on unknown $\boldsymbol{\beta}$. This can be estimated by using a working independence estimator. With weight $\mathbf{W}_{opt}$ being estimated, we can now solve (18) to obtain a two-stage estimator. This method will be abbreviated as "WS".

A drawback of this method is that it estimates a $(Jp)\times(Jp)$ matrix and another $(Jp)\times p$ matrix. The elements in each of these two matrices are estimated with errors. When J or p is large, it is not clear how close the estimate of $\mathbf{W}_{opt}$ is. In other words, the vector that we get $\hat{\mathbf{W}}_{opt}$ and the vector we want $\mathbf{W}_{opt}$ can be different. Hence, the efficiency gain is not always materialized. This is why we introduce two simplified schemes in Sections 3.1 and 3.2. However, if $\mathbf{W}_{opt}$ can be estimated correctly, the general scoring method with the optimal weight will provide the lowest asymptotic variance of $\hat{\beta}$ compared to minimizing the trace of the variance or choosing a specific $\mathbf{b}$.

4 Optimality of the working independence estimator

We would like now to identify the situations under which the working independence gives a nearly optimal solution. For those situations, there is no need to attempt to improve the working independence estimator. This also helps us design simulation settings in which working independence is not optimal.

Let us consider the optimization (13) again. When covariance matrix of marginal score functions $\ell'_j(\beta)$ ($j = 1, 2, \cdots, J$) is dominated by diagonal block $\boldsymbol{\Sigma} j$ ($j = 1, 2 \cdots, J$), then matrix $\mathbf{D}$ is also dominated by its diagonal entries. If the matrix $\mathbf{D}$ is dominated by its diagonal elements, then the solution (14) is the uniform weight which corresponds to the working independence estimator. This can be seen by assuming ideally that $\mathbf{D}$ in (13) is a diagonal element and $b_j = \sum_j^{(k)}$ that makes the assumption $\boldsymbol{\Sigma}(\boldsymbol{\beta}) \approx b_j \boldsymbol{\Sigma}$ valid for the k-diagonal element of the matrices. In other words, when the covariance matrix of the score functions $\ell'_j(\beta)$ ($j = 1, 2, \cdots, J$) is dominated by diagonal blocks $\boldsymbol{\Sigma}_j$ ($j = 1, 2 \cdots, J$), it implies that dependencies between the failure types in the same cluster are ignorable. Consequently, samples from different failure types are approximately independent of each other and the working independence weight can be reasonably applied in such a situation.

Another situation that the working independence is optimal is that the information contained in failure types is exchangeable. By exchangeability, we mean that the partial likelihood from each failure type plays a symmetric role: the blocks $\mathbf{D}_{ij}$ ($i, j = 1, 2, \cdots, J, i \neq j$) of covariance matrix $\boldsymbol{\Sigma}_m$ of the score functions $\ell'_j(\beta)$ ($j = 1, 2, \cdots, J$) are the same and $\boldsymbol{\Sigma} j$ ($j = 1, 2, \cdots, J$) are approximately equal. In this case, the optimization problem (13) becomes

$$\sum_{j=1}^{J} w_j^2 \sum^{(k)} + 2\sum_{i<j} w_i w_j \mathbf{D}^{(k)},$$

subject to the constraint $\sum_{j=1}^{J} w_j = 1$, where $\boldsymbol{\Sigma}^{(k)}$ is the k-th diagonal element of the matrices $\boldsymbol{\Sigma}$ and $\mathbf{D}^{(k)}$ is the k-th diagonal element of the matrix $\mathbf{D}$. This optimization problem is symmetric in w_j. By using the Lagrange multiplier method, one can easily show that the uniform weight is the optimal choice. In other words, even if dependency between the clusters is very strong, as long as they are exchangeable, the working independence estimator is still an optimal choice.

Using the total variance as the criterion, the results are the same as those of the componentwise criterion. In fact, we have verified these properties in our simulations that the working independence weight is the best when the asymptotic covariance is exchangeable or diagonal-dominated matrix and data are balanced.

5 Numerical studies

5.1 Simulations

Simulation studies are used to evaluate the performance of the proposed estimation methods. Multivariate failure times are generated from a multivariate extension of the model of Clayton and Cuzick (1985) in which the joint survival function of ($T_1, \cdots, T_J$) given ($\mathbf{Z}_1, \cdots, \mathbf{Z}_J$) is:

$$F(t_1, \cdots, t_J; \mathbf{Z}_1, \cdots, \mathbf{Z}_J) = \left\{\sum_{j=1}^{J} S_j(t_j)^{-1/\theta} - (J-1)\right\}^{-\theta}, \tag{19}$$

where $S_j(t)$ is the marginal survival probability for the jth member, depending on covariates $\mathbf{Z}_j$, and θ is a parameter that controls the degree of dependence among survival times. The relationship between Kendall's tau and θ is $\tau = 1/(2\theta + 1)$. The marginal distribution of T_j, given the covariate $\mathbf{Z}_j$, is governed by the following hazard rate:

$$\lambda_j(t)=\lambda_{0j}(t)\exp\{\beta^T \mathbf{Z}_j\}. \tag{20}$$

The multivariate survival times can be generated from the above specifications via the following conditional relation:

$$F(t_h|t_1,\cdots,t_{h-1})=1-\left(\sum_{i=1}^{h} S_i(t_i)^{-1/\theta}-(h-1)\right)^{-\theta-h+1}\left(\sum_{i=1}^{h} S_i(t_i)^{-1/\theta}-(h-2)\right)^{\theta+h-1}. \tag{21}$$

That is, for each draw, we first generate T_1 from model (20), then generate T_2 using (21) after getting $T_1 = t_1$, simulate T_3 after getting $T_1 = t_1$ and $T_2 = t_2$, and so on.

In our simulation studies, we take $J = 3$, $\lambda_{0j}(t)=\lambda^*_{0j}$, corresponding to the exponential distribution, and $\theta = 0.05$, 0.5 and 2, which represent strong, moderate and weak positive dependence, respectively. Censoring times C_{ij} are generated independently from uniform distribution over $(0, c)$, where c is a constant which controls the censoring rate. In all simulation studies, $c = 1$ is taken, which results in different censoring rates for different settings. The number of simulations are 500.

In the first simulated example, we take $\beta = (-0.05, 0.2, -1.5)^T$ and $(\lambda^*_{01}, \lambda^*_{02}, \lambda^*_{03})=(1, 2, 8)$. We now describe how the covariate vector $\mathbf{Z}_j = (Z_{j1}, Z_{j2}, Z_{j3})^T$ $(j = 1, 2, 3)$ is simulated for each draw. Z_{11} is a binary random variable with probability 0.5 taking values 0 or 1, respectively, and Z_{12} is a normal random variable with mean 1 and standard deviation 1, $Z_{13} = 0.8(U_i - 0.5)$ for the standard uniform random variable, and these three random variables are independent. For $j = 2$ and 3, in order to generate dependent covariates, we take covariates by $Z_{j1} = Z_{11}$ and

$$Z_{jk}=\sqrt{j}\delta_k+0.5\varepsilon_{jk}, \text{ for } j=2,3, \text{ and } k=2,3,$$

where (δ_2, δ_3) is generated from the multivariate normal distribution with mean 0, standard deviation 1 and correlation coefficient 0.5, and ε_{jk} is generated from independent standard normal distribution. The censoring rate for $c = 1$ is approximately 53%.

Table 1 summarizes the simulation results for the pseudo-partial likelihood estimator of $\boldsymbol{\beta}$ with 3 event types for each individual, while Table 2 summarizes the results with 30% of event type 2 missing and 50% of event type 3 missing. The averages, among 500 simulations, of Kendall's τ and Pearson correlation coefficients ρ summarize the degree of dependence among three survival times (before they are censored). They are summarized as (a_1, a_2, a_3), representing the correlation between the failure types 1 and 2, 1 and 3, and 2 and 3, respectively. They give us an idea of the degree of dependence among the survival times. The columns $\hat{\beta}_j$ show the averages of the estimate among 500 simulations and the columns "SE" give the averages of the estimated standard errors. The empirical standard errors of the 500 estimates are given in the columns "SD", and the coverage rates of the 95% confidence intervals are summarized in the column "CR". The columns "Ratio" show the percentage of variance reduction by using the weighted estimator in comparison with the working independence estimator.

First of all, all estimators are approximately unbiased and columns SE are close to the columns SD for all estimators, which indicates good performance of the standard error formulas in the presence of strong or modest dependency. The coverage rates of the 95% confidence intervals are close to the nominal level for those cases. The variance reduction of the weighted method is clearly shown in the columns "Ratio". The stronger the dependence among the correlated events, the more the reduction of the variance by using the weighted estimator. In other words, the more efficiency gain.

In the second example, we take the coefficient paramter $\beta = (0, -\log(2)/4, 0.05)^T$, and $(\lambda^*_{01}, \lambda^*_{02}, \lambda^*_{03})=(1, 1, 1)$. The generation of covariate vector is as follows. We first generate the covariate vector $\mathbf{Z}_1 = (Z_{11}, Z_{12}, Z_{13})^T$ from a multivariate normal distribution with marginal mean of 0, standard deviation being 1, and the correlation between Z_{1l} and Z_{1k} of $2^{-|l-k|}$. Then, for $j = 2$ and 3, we take $\mathbf{Z}_j = j\mathbf{Z}_1 + 0.5\mathbf{e}_j$, where $\mathbf{e}_j$ is generated from the trivariate standard normal distribution. We conduct the simulation with the same combination of event types and obtain similar results. In fact, the efficiency improvement by the weighted estimator is even more dramatic. We omit the details.

We have also conducted various experiments to verify the claims in Section 4. If we generate the covariate $\mathbf{Z}_j$ independently with the same distribution, then the score functions are exchangeable. In this case, no matter how small the dependent parameter θ is (i.e. how strong the dependence of the failure times is), our simulation results show that both weighted methods in Sections 3.1 and 3.2 have approximately the same efficiency, as expected from Section 4. See Table 3.

In the third example, we consider the simplest example with $p = 1$ and $\boldsymbol{\beta} = 0$. This avoids the impact of the dependence of covariates on the dependence of the survival times. The baseline hazard is taken as $(\lambda^*_{01}, \lambda^*_{02}, \lambda^*_{03})=(1, 1, 1)$. The covariate (Z_1, Z_2, Z_3) for each type of failure are generated as follows. The covariate Z_1 in failure type 1 is generated from standard normal distribution, and $Z_j = jZ_1 + 0.5\,\varepsilon_j$ for $j = 2, 3$, in which ε_j is also generated from the standard normal distribution. The percent of censoring for $c = 1$ is about 63%. The results are summarized in Tables 4 and 5. The improvements of the weighted estimators are now even more pronounced.

5.2 Busselton Population Health Surveys

We illustrate the proposed method by analyzing a data set from the Busselton Population Health Surveys. The Busselton Population Health Surveys are a series of cross-sectional health surveys conducted in the town of Busselton in Western Australia. Every 3 years from 1966 to 1981, general health information for adult participants were collected by means of questionnaire and clinical visit. Details of the study are described in Cullen (1972) and Knuiman et al. (1994). Data for several cardiovascular risk factors are available for 2202 persons who make up 619 families. In this analysis we are interested in investigating the effect of cardiovascular risk factors on the risk of death due to cardivascular disease (CVD) based on these family data. Since the death times of the family members might be correlated due to genetic factors and cohabitation, we are dealing with multivariate failure time data.

The risk factors we considered here includes age, gender, SBP, DBP, body mass index (bmi), serum cholesterol level (chol), and smoking status. Participant's age was measured in years. Serum cholesterol was determined from a blood sample and the unit used in this analysis is mmol/L. Body mass index was derived as weight (kg) divided by the square of height (m). Smoking status is coded as 1 for current smoker and 0 otherwise.

If a person took part in more than one of the Busselton surveys, only one record from the survey at which that person's age was closest to 45 years is included. Forty-eight percent of the participants are males (gender=0 for male and 1 for female). The average age in the data analyzed is 41.7 years, ranging from 16.3 to 89.0 years old. The average cholesterol reading was 5.65 mmol/L. The average body mass index was 24.8 kg/m^2. The prevalence of the neversmokers, ex-smokers, and current smokers are 49 %, 17 %, and 34%, respectively. Of the 619 families, there are 154 families with one event, 28 families with two events, and 3 families with more than two events. There are 219 observed events in all. The censorship is very heavy, 93.73% of data are censored in this data set.

For this analysis, we are interested in investigating the effect of the risk factors on hazard rate of death. We consider the following model:

$$\lambda_{ij}(t)=\lambda_{0j}(t)\exp\{\beta_1*\mathrm{age}_{ij}+\beta_2*\mathrm{gender}_{ij}+\beta_3*SBP_{ij}+\beta_4*\mathrm{DBP}_{ij}+\beta_5*\mathrm{bmi}_{ij}+\beta_6*\mathrm{chol}_{ij}+\beta_7*\mathrm{smoke2}_{ij}\},$$

where j = 1 and 2 denote the parents and the children of the family, respectively.

For this data and this model, we perform the proposed estimating procedures to estimate the coefficients. The procedure includes the working independence (WI), componentwise (CW), trace (TW), and generalized score methods (WS). The estimated coefficients are summarized in Table 6. The results are very similar among those four methods, with the generalized score method having slightly smaller SEs. This is probably due to the heavy censoring of the data, which makes working independence more attractive.

6 Concluding Remarks

We have proposed several weighted estimators based on different criteria for estimating regression coefficients in Cox model with failure time data. Our aim is to come up with a simple weighting scheme that performs at least as good as the working independence estimator. We identify the situations under which the working independence estimators are in fact effective. Specifically, when the covariance matrix of the score functions of the partial likelihood function for each failure types or clusters are nearly block diagonal or exchangeable, the working independence is indeed nearly optimal. The generalized score method, which efficiently combines the estimating equations, is an efficient method. However, the efficiency cannot always be realized, particularly when J and p are both large.

Acknowledgments

This work was partially supported by a grant from National Natural Science Foundation of China (No. 10628104). Fan's research was also partially supported by NIH R01-GM072611 and NSF grant DMS-0714554. Zhou's research was supported by grants from the National Natural Science Foundation of China (NSFC) (No. 10731010 and No. 10628104), the National Basic Research Program (No. 2007CB814902) and Creative Research Groups of China (No. 10721101). Cai's research was partially supported by the National Institutes of Health grant R01-HL57444. The authors gratefully acknowledge helpful comments by Professor Haibo Zhou. We thank Professor Matthew Knuiman and the Busselton Population Medical Research Foundation for permission to use the data in the illustration.

References

Andersen LW. Composite likelihood and two-stage estimation in family studies. Biostatistics 2004;5:15–30. [PubMed: 14744825]

Andersen PK, Gill RD. Cox's regression model for counting process: A large sample study. The annals of statistics 1982;10:1100–1120.

Cai J, Fan J, Jiang J, Zhou H. Partially linear hazard regression for multivariate survival data. Jour Ameri Statist Assoc 2007;102:538–551.

Cai J, Fan J, Jiang J, Zhou H. Partially linear hazard regression with varying-coefficients for multivariate survival data. Journal Royal Statistical Soc B 2008;70:141–158.

Cai J, Fan J, Zhou H, Zhou Y. Marginal hazard models with varying-coefficients for multivariate failure time data. The Annals of Statistics 2007;35:324–354.

Cai J, Prentice RL. Estimating equations for hazard ratio parameters based on correlated failure time data. Biometrika 1995;82:151–164.

Cai J, Prentice RL. Regression estimation using multivariate failure time data and a common baseline hazard function model. Lifetime Data Analysis 1997;3:197–213. [PubMed: 9384652]

Cai J. Hypothesis testing of hazard ratio parameters in marginal models for multivariate failure time data. Lifetime Data Analysis 1999;5:39–53. [PubMed: 10214001]

Chen K, Chen Y, Ying Z. Analysis of multivariate failure time data via using marginal proportional hazard model. 2008 Manuscription.

Clayton D, Cuzick J. Multivariate generalizations of the proportional hazards model. J Roy Statist Soc Ser A 1985;148:82–117.

Clegg LX, Cai J, Sen PK, Kupper LL. Misspecification of marginal hazards models in multivariate failure time data. Biostatistics, Ser B 2000;62:25–42.

Cox DR. Regression models and life-tables (with discussion). J Roy Statist Soc, Ser B 1972;34:187–220.

Cullen KJ. Mass Health Examinations in the Busselton Population, 1966 to 1970. Australia Journal of Medicine 1972;2:714–718.

Glindden DV. Pairwise dependence diagnostics for clustered failure-time data. Biometrika 2007;94:371–385.

Gray RJ, Li Y. Optimal weight functions for marginal proportional hazards analysis of clustered failure time data. Lifetime Data Analysis 2002;8:5–19. [PubMed: 11878224]

He X, Tong XW, Sun J, Cook R. Regression analysis of multivariate panel count data. Biostatistics 2008;9:234–248. [PubMed: 17626224]

Hansen LP. Large sample properties of generalized method of moments estimators. Econometrica 1982;50:1029–1054.

Heyde, CC. A general approach to optimal parameter estimation. Springer Series in Statistics; New York: 1997. Quasi-likelihood and its application.

Knuiman MW, Cullen KJ, Bulsara MK, Welborn TA, Hobbs MST. Mortality Trends 1965 to 1989 in Busselton, the Site of Repeated Health Surveys and Interventions. Australia Journal of Public Health 1994;18:129–135.

Kuk YC. A hybrid pairwise likelihood method. Biometrika 2007;94:939–952.

Larocque D, Nevalainen J, Oja H. A weighted multivariate sign test for cluster-correlated data. Biometrika 2007;94:267–283.

Lee KR, Kapadia CH. Mean squared error efficient estimators of the variance components. Metrika 1992;39:21–26.

Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73:13–22.

Schaubel D, Cai J. Analysis of clustered recurrent event data with application to hospitalization rates among renal failure patients. Biostatistics 2005;9:404–419. [PubMed: 15831581]

Spiekerman CF, Lin DY. Marginal regression models for multivariate failure time data. J Amer Statist Assoc 1998;93:1164–1175.

Wei LJ, Lin DY, Weissfeld L. Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions. Journal of the American Statistical Association 1989;84:1065–1073.

Yu ZS, Lin XH. Nonparametric regression using local kernel estimating equations for correlated failure time data. Biometrika 2008;95:123–137.

Zeger SL, Liang KY, Self SG. The analysis of binary longitudinal data with time-independent covariates. Biometrika 1985;72:31–38.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Results with n = (100, 100, 100) for the first simulated example

	$\beta_1 = -0.05$				$\beta_2 = 0.2$				$\beta_3 = -1.5$				
Method	$\hat{\beta}_1$	**SE**	**SD**	**CR**	$\hat{\beta}_2$	**SE**	**SD**	**CR**	$\hat{\beta}_3$	**SE**	**SD**	**CR**	**Ratio**
$\theta = 0.05$, $\tau = (0.8259, 0.7990, 0.8042)$, $\rho = (0.9530, 0.9339, 0.9347)$													
WI	−0.074	0.265	0.278	0.94	0.215	0.137	0.135	0.96	−1.570	0.405	0.404	0.96	1.000
CW	−0.074	0.249	0.255	0.93	0.215	0.116	0.112	0.96	−1.560	0.346	0.348	0.95	0.772
WT	−0.075	0.251	0.260	0.93	0.216	0.120	0.117	0.97	−1.570	0.358	0.362	0.95	0.811
WS	−0.069	0.241	0.268	0.92	0.209	0.110	0.116	0.95	−1.537	0.330	0.356	0.93	0.708
$\theta = 0.5$, $\tau = (0.5044, 0.4972, 0.5097)$, $\rho = (0.8106, 0.7951, 0.8021)$													
WI	−0.058	0.257	0.244	0.96	0.212	0.133	0.121	0.97	−1.551	0.386	0.355	0.97	1.000
CW	−0.059	0.246	0.235	0.96	0.211	0.117	0.118	0.94	−1.532	0.350	0.349	0.94	0.846
WT	−0.057	0.248	0.239	0.95	0.212	0.121	0.116	0.96	−1.542	0.354	0.345	0.95	0.867
WS	−0.048	0.238	0.248	0.94	0.209	0.111	0.121	0.92	−1.536	0.333	0.352	0.93	0.774
$\theta = 2$, $\tau = (0.2523, 0.2577, 0.2810)$, $\rho = (0.4797, 0.4696, 0.4917)$													
WI	−0.054	0.247	0.210	0.98	0.208	0.129	0.107	0.99	−1.530	0.371	0.309	0.98	1.000
CW	−0.051	0.241	0.215	0.97	0.203	0.118	0.115	0.94	−1.507	0.347	0.324	0.96	0.894
WT	−0.049	0.244	0.224	0.97	0.203	0.120	0.109	0.97	−1.518	0.350	0.319	0.97	0.911
WS	−0.046	0.233	0.229	0.95	0.204	0.112	0.117	0.93	−1.513	0.330	0.339	0.94	0.817

WI, CW, WT and WS denote the estimators based on working independence, componentwise, trace method and generalized score method, respectively.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 2

Results based on n = (100, 70, 50) for first simulated example

Method	$\hat{\beta}_1$ ($\beta_1 = 0.05$)	SE	SD	CR	$\hat{\beta}_2$ ($\beta_2 = 0.2$)	SE	SD	CR	$\hat{\beta}_3$ ($\beta_3 = -1.5$)	SE	SD	CR	Ratio
$\theta = 0.05$, $\tau = (0.8211, 0.7914, 0.7972)$, $\rho = (0.7161, 0.5584, 0.7517)$													
WI	−0.057	0.313	0.325	0.94	0.204	0.179	0.193	0.94	−1.550	0.521	0.533	0.94	1.000
CW	−0.055	0.308	0.319	0.94	0.203	0.172	0.186	0.93	−1.528	0.505	0.517	0.94	0.945
WT	−0.056	0.317	0.331	0.93	0.202	0.174	0.186	0.94	−1.537	0.505	0.517	0.95	0.962
WS	−0.055	0.294	0.327	0.92	0.195	0.161	0.189	0.91	−1.480	0.474	0.525	0.92	0.840
$\theta = 0.1$, $\tau = (0.7679, 0.7449, 0.7497)$, $\rho = (0.7096, 0.5535, 0.7461)$													
WI	−0.055	0.311	0.326	0.94	0.202	0.179	0.190	0.95	−1.545	0.519	0.526	0.95	1.000
CW	−0.052	0.307	0.320	0.95	0.203	0.172	0.185	0.94	−1.527	0.503	0.519	0.95	0.947
WT	−0.055	0.316	0.333	0.93	0.200	0.174	0.184	0.94	−1.533	0.504	0.518	0.94	0.964
WS	−0.054	0.293	0.325	0.92	0.193	0.161	0.186	0.91	−1.479	0.472	0.526	0.92	0.841
$\theta = 2$, $\tau = (0.2456, 0.2538, 0.2801)$, $\rho = (0.3572, 0.2803, 0.4594)$													
WI	−0.049	0.291	0.251	0.98	0.197	0.167	0.142	0.98	−1.538	0.481	0.400	0.98	1.000
CW	−0.047	0.289	0.253	0.97	0.198	0.163	0.146	0.97	−1.530	0.473	0.399	0.98	0.969
WT	−0.047	0.295	0.260	0.97	0.195	0.165	0.144	0.97	−1.533	0.473	0.401	0.98	0.982
WS	−0.042	0.277	0.269	0.96	0.195	0.152	0.153	0.95	−1.515	0.443	0.434	0.95	0.860

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

Results with n = (100, 100, 100) for the second simulated example with c = 5

	$\beta_1 = 0$				$\beta_2 = -log(2)/4$				$\beta_3 = 0.05$				
est	$\hat{\beta}_1$	**SE**	**SD**	**CR**	$\hat{\beta}_2$	**SE**	**SD**	**CR**	$\hat{\beta}_3$	**SE**	**SD**	**CR**	**Ratio**
θ = 0.05, τ = (0.8279, 0.8287, 0.8281), ρ = (0.9522, 0.9523, 0.9513)													
WI	−0.001	0.078	0.084	0.93	−0.173	0.087	0.088	0.94	0.049	0.078	0.076	0.95	1.000
CW	−0.000	0.077	0.084	0.92	−0.173	0.087	0.088	0.94	0.049	0.077	0.077	0.94	0.990
WT	−0.001	0.078	0.084	0.92	−0.172	0.087	0.088	0.95	0.049	0.078	0.076	0.95	0.996
WS	0.001	0.075	0.087	0.90	−0.172	0.084	0.092	0.92	0.048	0.075	0.081	0.93	0.932
θ = 5.0, τ = (0.0887, 0.0901, 0.0868), *coeff* = (0.1929, 0.1817, 0.1824)													
WI	0.000	0.077	0.083	0.92	−0.173	0.087	0.093	0.92	0.047	0.077	0.079	0.94	1.000
CW	0.000	0.077	0.083	0.91	−0.173	0.086	0.093	0.92	0.047	0.077	0.079	0.94	0.990
WT	0.001	0.077	0.083	0.92	−0.174	0.087	0.092	0.91	0.046	0.077	0.079	0.94	0.997
WS	0.001	0.075	0.086	0.90	−0.173	0.084	0.096	0.92	0.046	0.075	0.082	0.92	0.936

Table 4

Results based on n = (100, 100, 100) for the third simulated example

	$\beta_1 = 0$				
Method	$\hat{\beta}_1$	**SE**	**SD**	**CR**	**Ratio**
$\theta = 0.05$, $\tau = (0.8998, 0.9005, 0.8998)$, $\rho = (0.9961, 0.9961, 0.9961)$					
WI	−0.001	0.063	0.060	0.94	1.000
CW	0.000	0.046	0.047	0.94	0.541
WT	−0.001	0.054	0.052	0.94	0.720
WS	−0.000	0.046	0.047	0.92	0.525
$\theta = 0.5$, $\tau = (0.4945, 0.4953, 0.4954)$, $\rho = (0.8267, 0.8257, 0.8267)$					
WI	0.000	0.032	0.032	0.95	1.000
CW	0.001	0.027	0.031	0.91	0.735
WT	0.000	0.029	0.031	0.93	0.819
WS	0.001	0.027	0.031	0.90	0.704
$\theta = 2$, $\tau = (0.1983, 0.2004, 0.1992)$ $\rho = (0.4208, 0.4155, 0.4175)$					
WI	0.000	0.056	0.048	0.98	1.000
CW	0.001	0.051	0.053	0.94	0.836
WT	0.001	0.053	0.050	0.95	0.876
WS	0.001	0.050	0.053	0.93	0.798

Table 5

Results based on n = (100, 70, 50) for the third simulated example

	$\beta_1 = 0$				
Method	$\hat{\beta}_1$	**SE**	**SD**	**CR**	**Ratio**
θ = 0.05, τ = (0.8954, 0.8902, 0.8904), ρ = (0.9958, 0.9958, 0.9958)					
WI	0.001	0.081	0.085	0.94	1.000
CW	0.001	0.074	0.080	0.93	0.857
WT	0.001	0.075	0.078	0.94	0.875
WS	0.001	0.072	0.084	0.91	0.796
θ = 0.5, τ = (0.4924, 0.4883, 0.4915), ρ = (0.8119, 0.8123, 0.8129)					
WI	0.002	0.076	0.077	0.95	1.000
CW	0.002	0.072	0.076	0.93	0.909
WT	0.003	0.073	0.076	0.94	0.916
WS	0.002	0.070	0.077	0.92	0.843
θ = 2, τ = (0.1957, 0.1967, 0.1980), ρ = (0.3963, 0.4018, 0.3999)					
WI	0.002	0.072	0.067	0.95	1.000
CW	0.001	0.070	0.068	0.94	0.936
WT	0.002	0.070	0.068	0.94	0.943
WS	0.002	0.067	0.070	0.93	0.875

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 6

Estimated parameters for the data from Busselton Population Health Surveys

		β_1	β_2	β_3	β_4	β_5	β_6	β_7
WI	Est	0.1261	−0.7240	0.0127	0.0180	0.0031	0.2155	0.5834
	SE	0.0080	0.1514	0.0036	0.0072	0.0202	0.0559	0.1388
CW	Est	0.1262	−0.7204	0.0127	0.0183	0.0031	0.2220	0.6030
	SE	0.0080	0.1513	0.0036	0.0072	0.0202	0.0554	0.1379
WT	Est	0.1261	−0.7240	0.0127	0.0180	0.0031	0.2155	0.5834
	SE	0.0080	0.1514	0.0036	0.0072	0.0202	0.0559	0.1388
WS	Est	0.1220	−0.8370	0.0142	0.0097	0.0130	0.3523	0.6595
	SE	0.0073	0.1480	0.0035	0.0069	0.0200	0.0448	0.1320