Semiparametric regression based on quadratic inference function for multivariate failure time data with auxiliary information

Feifei Yan1,4 · Lin Zhu4 · Yanyan Liu2 · Jianwen Cai3 · Haibo Zhou3

Received: 8 October 2019 / Accepted: 23 December 2020 / Published online: 8 January 2021
© The Author(s) 2021

Abstract
This paper deals with statistical inference procedure of multivariate failure time data when the primary covariate can be measured only on a subset of the full cohort but the auxiliary information is available. To improve efficiency of statistical inference, we use quadratic inference function approach to incorporate the intra-cluster correlation and use kernel smoothing technique to further utilize the auxiliary information. The proposed method is shown to be more efficient than those ignoring the intra-cluster correlation and auxiliary information and is easy to implement. In addition, we develop a chi-squared test for hypothesis testing of hazard ratio parameters. We evaluate the finite-sample performance of the proposed procedure via extensive simulation studies. The proposed approach is illustrated by analysis of a real data set from the study of left ventricular dysfunction.

Keywords Multivariate failure time data · Validation sample · Quadratic inference function · Chi-squared test

Yanyan Liu
liuyy@whu.edu.cn

1 School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan 430074, Hubei, People’s Republic of China
2 School of Mathematics and Statistics, Wuhan University, Wuhan 430072, Hubei, People’s Republic of China
3 Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7420, USA
4 School of Science, East China University of Technology, Nanchang 330013, Jiangxi, People’s Republic of China
1 Introduction

This paper is aimed at developing improved inference procedure for multivariate failure time data with auxiliary information. Large cohort studies often involve thousands or more subjects and the studies, especially when involving failure time outcomes, could last for many years. It is often that the measurement of the primary covariate can only be obtained for a random subset of the study cohort due to technical difficulties or financial limitations. On the other hand, some auxiliary information that is less precise but highly correlated to the primary exposure can be cheaply collected for all cohort members. The auxiliary information could be a mismeasured surrogate to the true covariate, or any covariate that is informative about the true covariate. An example is from the left ventricular dysfunction (SOLVD 1991) prevention study, which aims to assess the effects of risk factors on the time (possibly censored) to heart failure and the first myocardial infarction. One of the most important risk factors is patient’s ejection fraction (EF), which can be precisely measured by using a standardized radionucleotide technique, but the cost is very high. Therefore, EF is only measured on a randomly chosen subset of all cohort, while a less precise but cheaper measurement of EF was ascertained for all the patients using a nonstandardized technique. Because each patient could experience both heart failure and the first myocardial infarction, statistical methods for handling multivariate failure time data with covariate measurement error would be required.

Proper use of auxiliary information has been proved to lead to improved efficiency of survival estimates in multivariate failure time data with auxiliary information. For example, Hu and Lin (2004) proposed a corrected estimation function under the assumption that the error is symmetrically distributed. Liu et al. (2009) and Liu et al. (2010) developed estimated pseudo-partial likelihood method for multivariate failure time data with discrete and continuous auxiliary variable, respectively. Liu et al. (2012) and Fan and Wang (2009) studied this problem under the assumption that the intra-cluster subjects have common baseline hazard. The above studies are based on the marginal hazards model, the intra-cluster correlation, however, is ignored in the estimation procedures and only adjusted in the inference step by applying a robust sandwich variance estimate. The practice of ignoring the intra-cluster correlation would result in some loss of efficiency.

Some authors have proposed to incorporate correlation explicitly into the estimating equations to improve the efficiency of estimate in dealing with multivariate failure time data with covariates being fully observed. For example, Cai and Prentice (1995, 1997) added a weight matrix based on the inverse of correlation matrix of marginal martingales into the partial likelihood score equation. Simulation studies have shown that their approach is more efficient than that using independent structure when cluster size is small. However, their method is computation intensive when the cluster size is large because the computation involves an estimation of very high dimensional weighting correlation matrix. To overcome this shortcoming, Xue et al. (2010) developed a different approach by applying the method of quadratic inference function (QIF). Their method avoids to explicitly estimate the correlation parameters and is easy to implement especially when cluster size is large. We note that both these two
methods assume that the covariates could be observed completely and therefore cannot be applied directly to SOLVD data.

Motivated by the advantages of the QIF method provided, we extend this method to the analysis of multivariate failure time data with auxiliary information. Here, we assume that the auxiliary covariate is continuous. We propose an estimated QIF method and study the asymptotic properties of the proposed estimator. The proposed method inherits the merit of QIF method which avoids the estimation of nuisance correlation parameters and is computationally easy to implement. Under certain regularity conditions, we establish the asymptotic normality of resulting estimator. Simulation studies show that our proposed method can improve the estimation efficiency compared with that ignoring dependent structure, such as the method by Liu et al. (2010). In addition, we study the problem of hypothesis testing, propose a proper test statistic which have a chi-squared limiting distribution under the null hypothesis.

The rest of the article is organized as follows. In Sect. 2, we introduce the model and describe the proposed estimation procedure. In Sect. 3, the large-sample properties of the proposed estimator are presented. In Sect. 4, a chi-squared test is developed for hypothesis testing. In Sect. 5, the finite-sample performance of the proposed procedures is assessed through extensive simulation studies. We illustrate the proposed method through analysis of a real data set from SOLVD study in Sect. 6. Some concluding remarks are given in Sect. 7 and the technical proofs are provided in “Appendix”.

2 Model and estimation

2.1 Preliminaries

Suppose that the whole cohort consists of \( n \) independent clusters, and each cluster contains \( K \) correlated failure types. Let \((i, k)\) denote the \( k \)th \((k = 1, \ldots, K)\) subject in the \( i \)th \((i = 1, \ldots, n)\) cluster. Let \( \tilde{T}_{ik} \) and \( C_{ik} \) be potential failure time and censoring time for subject \((i, k)\). With censoring, one observes \( T_{ik} = \min(\tilde{T}_{ik}, C_{ik}) \) and \( \Delta_{ik} = I(\tilde{T}_{ik} \leq C_{ik}) \), where \( I(\cdot) \) is the indicator function. Let \( Z_{ik}(t) \) be a \( p \)-vector of possibly time-dependent covariates.

For subject \((i, k)\), the hazard function \( \lambda_{ik}(t; \tilde{Z}_{ik}(t)) \) takes the following form:

\[
\lambda_{ik}(t; \tilde{Z}_{ik}(t)) = \lambda_{0k}(t) \exp\{\beta^T \tilde{Z}_{ik}(t)\}, \tag{2.1}
\]

where \( \beta \) is a \( p \)-vector of unknown regression parameters and \( \lambda_{0k}(t) \) is an unspecified marginal baseline hazard function pertaining to the \( k \)th failure type.

Note that model (2.1) includes as a special case the failure-type-specific model (Wei et al. 1989; Greene and Cai 2004) \( \lambda_{ik}(t; \tilde{Z}_{ik}(t)) = \lambda_{0k}(t) \exp\{\beta^T \tilde{Z}_{ik}(t)\} \), which allows for different covariate effect for different \( k \). This can be seen by defining \( \beta = (\beta_1^T, \ldots, \beta_k^T, \ldots, \beta_K^T)^T \) and \( \tilde{Z}_{ik}^*(t) = (\tilde{Z}_{ik}(t), \ldots, \tilde{Z}_{ik}(t), \ldots, \tilde{Z}_{ik}(t))^T \) in the model \( \lambda_{ik}(t; \tilde{Z}_{ik}^*(t)) = \lambda_{0k}(t) \exp\{\beta^T \tilde{Z}_{ik}^*(t)\} \). For simplicity, we write \( \lambda_{ik}(t; \tilde{Z}_{ik}(t)) \) as \( \lambda_{ik}(t) \) in the following.

Let \( \Lambda_{0k}(t) = \int_0^t \lambda_{0k}(u) \, du \) be the marginal cumulative baseline hazard function for the \( k \)th failure type. Let \( N_{ik}(t) = \Delta_{ik} I(T_{ik} \leq t) \) and \( Y_{ik}(t) = I(T_{ik} \geq t) \) be
the observed counting process and the at-risk indicator process. For convenience, write the relative risk function as \( r_{ik}(t; \beta) = \exp(\beta^T \tilde{Z}_{ik}(t)) \). Let \( M_{ik}(t) = N_{ik}(t) - \int_0^t Y_{ik}(u) r_{ik}(u; \beta_0) \lambda_{0k}(u) du \) be the marginal martingale process, where \( \beta_0 \) is the true parameter. Given \( \beta, \Lambda_{0k}(t) \) can be estimated consistently by the following Breslow type estimator (Breslow 1972):

\[
\tilde{\Lambda}_{0k}(t; \beta) = \int_0^t \frac{\sum_{i=1}^n dN_{ik}(u)}{\sum_{i=1}^n Y_{ik}(u) r_{ik}(u; \beta)}.
\]

Given \( \beta_0 \), it follows that \( M_{ik}(t) \) could be estimated as follows:

\[
\tilde{M}_{ik}(t; \beta_0) = N_{ik}(t) - \int_0^t Y_{ik}(u) r_{ik}(u; \beta_0) \tilde{\Lambda}_{0k}(dt; \beta_0).
\]

Let \( \tau \) be the end time of study. Write \( \tilde{M}_i(t; \beta) = (\tilde{M}_{i1}(t; \beta), \ldots, \tilde{M}_{iK}(t; \beta))^T \). To improve the estimation efficiency, Cai and Prentice (1995) added proper weight matrix to the pseudo-partial likelihood equation, and proposed to obtain estimate of \( \beta \) through solving the following equation:

\[
\sum_{i=1}^n \int_0^\tau R_i^T(t; \beta) W_i(t; \beta) \tilde{M}_i(dt; \beta) = 0,
\]

where \( W_i(t; \beta) = \text{corr}^{-1}(M_i(T_i; \beta)) \) is the weight matrix with

\[
M_i(T_i; \beta) = (M_{i1}(T_{i1}; \beta), \ldots, M_{iK}(T_{iK}; \beta))^T,
\]

and

\[
R_i^T(t; \beta) = \left( \frac{r_{i1}^{(1)}(t; \beta)}{r_{i1}(t; \beta)}, \ldots, \frac{r_{iK}^{(1)}(t; \beta)}{r_{iK}(t; \beta)} \right),
\]

with \( r^{(j)}(t; \beta) \) denotes the \( j \)th derivative of \( r(t; \beta) \) with respect to \( \beta \). The weight matrix measures the intra-cluster correlation and is important to improve estimation efficiency. However, when the cluster size \( K \) is large, the estimation of weight matrix is computationally expensive. To overcome this shortcoming, Xue et al. (2010) proposed a QIF method which is based on the following generalized estimating equation:

\[
\sum_{i=1}^n \int_0^\tau R_i^T(t; \beta) \Xi_i^{1/2}(t; \beta) \Sigma_i^{-1}(\alpha) \Xi_i^{-1/2}(t; \beta) \tilde{M}_i(dt; \beta) = 0,
\]

where \( \Xi_i(t; \beta) = \text{diag}(\lambda_{i1}(t), \ldots, \lambda_{iK}(t)) \) and \( \Sigma_i(\alpha) \) is the working correlation matrix whose common structure is specified by a vector of nuisance correlation parameters \( \alpha \). The inverse of the working correlation is approximated by a linear combination

\[\Xi\]
of several pre-specified symmetric basis matrices, namely,
\[ \Sigma^{-1} \approx a_1 B_1 + \ldots + a_L B_L, \quad (2.5) \]
where \( B_1, \ldots, B_L \) are known basis matrices and \( a_1, \ldots, a_L \) are unknown coefficients.

Substituting (2.5) in (2.4) leads to a linear combination of the elements of the following vector
\[ G_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} g_i(\beta) \]
\[ = \frac{1}{n} \sum_{i=1}^{n} \left( \int_0^T R_i^T(t; \beta) \Xi_i^{1/2}(t; \beta) B_1 \Xi_i^{-1/2}(t; \beta) \hat{M}_i(dt; \beta) \right) \]
\[ \vdots \]
\[ \int_0^T R_i^T(t; \beta) \Xi_i^{1/2}(t; \beta) B_L \Xi_i^{-1/2}(t; \beta) \hat{M}_i(dt; \beta) \right) \]

As there are more equations than unknown parameters, Xue et al. (2010) proposed to estimate \( \beta \) by minimizing the following QIF:
\[ Q_n(\beta) = G_n^T(\beta) W_n^{-1}(\beta) G_n(\beta), \quad (2.6) \]
where \( W_n(\beta) = \frac{1}{n^2} \sum_{i=1}^{n} g_i(\beta) g_i^T(\beta) \). We denote the solution as \( \tilde{\beta}_Q \) in the following.

In the implementation of QIF, there being an additional issue that the diagonal matrix \( \Xi_i(t; \beta) \) involves the unknown baseline hazard function \( \lambda_0(t; \beta) \). Xue et al. (2010) suggested a kernel smoothed estimator \( \tilde{\lambda}_0(t; \beta) \) as follows,
\[ \tilde{\lambda}_0(t; \beta) = \frac{1}{v_k} \sum_{i=1}^{n} \kappa \left( t - T_{ik} \right) \Delta \tilde{\lambda}_0(T_{ik}; \beta), \quad (2.7) \]
where \( \kappa(\cdot) \) is the Epanechnikov kernel function with \( v_k \) being the rule-of-thumb bandwidth, and \( \Delta \tilde{\lambda}_0(t; \beta) = \lambda_0(t; \beta) - \lambda_0(t^-; \beta) \) with \( \lambda_0(t; \beta) \) being the Breslow estimator given in (2.2).

2.2 Estimated QIF for marginal hazards model

Consider the situation that the primary covariate can only be ascertained in validation set. Let \( \tilde{Z}_{ik}(t) \) consist of two parts, \( X_{ik}(t) \) and \( Z_{ik}(t) \), where \( X_{ik}(t) \) is the primary variable which can only be observed in the validation set and \( Z_{ik}(t) \) is the vector of the remaining covariates that are measured precisely for the full cohort. Accordingly, write the true parameter as \( \beta_0 = (\beta_{10}, \beta_{20}) \) with \( \beta_{10} \) and \( \beta_{20} \) pertaining to \( X_{ik}(t) \) and \( Z_{ik}(t) \), respectively. Denote \( A(t) \) as a time-dependent auxiliary variable for the primary covariate \( X(t) \). \( A(\cdot) \) can be measured for all cohort members. Suppose that \( A \) provides no additional information to model given \( X \), i.e.,
\[ \lambda(t; X(t), Z(t), A(t)) \equiv \lambda(t; X(t), Z(t)). \]
Use $\eta_{ik} = 1$ or 0 to indicate whether the subject $(i, k)$ is in the validation set or not. Let $V_k = \{ i : \eta_{ik} = 1 \}$ and $\overline{V}_k = \{ i : \eta_{ik} = 0 \}$ denote the $k$th marginal validation set and non-validation set, respectively. Then the observed data are:

$$
\begin{align*}
&\{T_{ik}, \Delta_{ik}, \eta_{ik}, Y_{ik}(t), X_{ik}(t), Z_{ik}(t), A_{ik}(t)\} \quad \text{if } i \in V_k \\
&\{T_{ik}, \Delta_{ik}, \eta_{ik}, Y_{ik}(t), Z_{ik}(t), A_{ik}(t)\} \quad \text{if } i \in \overline{V}_k.
\end{align*}
$$

According to Liu et al. (2009), when subject $(i, k)$ is in non-validation set, the hazard function given observed data can be written as:

$$
\lambda_{ik}(t; Z_{ik}(t), A_{ik}(t)) = \lambda_{0k}(t)e^{\beta_1^T Z_{ik}(t)}E\{e^{\beta_1^T X_{ik}(t)}| Y_{ik}(t) = 1, A_{ik}(t), Z_{ik}(t)\} = \lambda_{0k}(t)e^{\beta_1^T Z_{ik}(t)}E\{e^{\beta_1^T X_{ik}(t)}| Y_{ik}(t) = 1, A_{ik}(t)\},
$$

where $A_{ik}(t)$ denotes all the possible auxiliary information, which may include the auxiliary covariate $A_{ik}(t)$ and the part from $Z_{ik}(t)$. Therefore, the induced relative risk function is

$$
r_{ik}(t; \beta) = \varphi_{ik}(t; \beta)\eta_{ik} + \psi_{ik}(t; \beta)(1 - \eta_{ik}),$$

where $\varphi_{ik}(t; \beta) = \exp[\beta_1^T X_{ik}(t) + \beta_2^T Z_{ik}(t)]$, and

$$
\psi_{ik}(t; \beta) = \exp[\beta_2^T Z_{ik}(t)]E\{e^{\beta_1^T X_{ik}(t)}| Y_{ik}(t) = 1, A_{ik}(t)\}.
$$

If the conditional density of $X_{ik}(t)$, written as $f(x|T_{ik} \geq t, A_{ik}^*)$, is a known function up to a parameter $\theta$, then $(\beta, \theta)$ can be estimated by using the induced risk function to replace risk function in equations (2.3) or (2.4). However, misspecification of such parameterization may lead to biased estimates. We use empirical method to estimate $\psi_{ik}(t; \beta)$ and then replace it with the corresponding estimate.

In this paper, we consider the often encountered case that both the primary covariate $X_{ik}(t)$ and the auxiliary variable $A_{ik}^*(t)$ are one-dimensional. The unknown part of induced relative risk function in non-validation set is estimated by kernel smoothing method

$$
\hat{\psi}_{ik}(t; \beta) = \frac{\sum_{j \in V_k} Y_{jk}(t)\Psi(\mu_k^{-1}(A_{jk}(t) - A_{ik}^*(t)))e^{\beta_1^T X_{jk}(t)}e^{\beta_2^T Z_{jk}(t)}}{\sum_{j \in V_k} Y_{jk}(t)\Psi(\mu_k^{-1}(A_{jk}(t) - A_{ik}^*(t)))}, \quad (2.8)
$$

where $\Psi(\cdot)$ is a kernel function, $\mu_k$ is the bandwidth. Imputation of the relative risk by interpolation would be used when the denominator is 0. Therefore, the estimate of the relative risk is

$$
\hat{r}_{ik}(t; \beta) = \varphi_{ik}(t; \beta)\eta_{ik} + \hat{\psi}_{ik}(t; \beta)(1 - \eta_{ik}).
$$

Replacing $r_{ik}(t; \beta)$ by $\hat{r}_{ik}(t; \beta)$ in the notations in Sect. 2.1, we obtain an estimated version of $\hat{R}_i$, $\hat{\Lambda}_{0k}$, $\lambda_{0k}$, $\Xi_i$, $\hat{M}_i$, $g_i$ and $G_n$. To differentiate, write as $\hat{R}_i$, $\hat{\Lambda}_{0k}$, $\hat{\lambda}_{0k}$, $\hat{\Xi}_i$, $\hat{M}_i$, $\hat{g}_i$ and $\hat{G}_n$. © Springer
\( \hat{\mathbf{M}}_i, \hat{g}_i \) and \( \hat{G}_n \). It yields an estimated QIF as
\[
\hat{Q}_n(\beta) = \hat{G}_n^T(\beta) \hat{W}_n^{-1}(\beta) \hat{G}_n(\beta),
\]
where \( \hat{W}_n(\beta) = \frac{1}{n^2} \sum_{i=1}^{n} \hat{g}_i(\beta) \hat{g}_i^T(\beta) \). \( \beta_0 \) can be estimated by minimizing \( \hat{Q}_n(\beta) \), i.e.,
\[
\hat{\beta}_Q = \arg\min_{\beta} \hat{Q}_n(\beta).
\]
(2.9)

To reduce the computation burden, we approximate the first and the second order derivatives of \( \hat{Q}_n(\beta) \) as in Qu et al. (2000) as follows.
\[
\begin{align*}
\hat{Q}_n^{(1)}(\beta) &= 2(\hat{G}_n^{(1)}(\beta)) ^T \hat{W}_n^{-1}(\beta) \hat{G}_n(\beta) + o_p(1), \\
\hat{Q}_n^{(2)}(\beta) &= 2(\hat{G}_n^{(1)}(\beta)) ^T \hat{W}_n^{-1}(\beta) \hat{G}_n^{(1)}(\beta) + o_p(1).
\end{align*}
\]

Then, Newton-Raphson algorithm can be applied by using the approximation.

### 3 Asymptotic properties

In this section, we present the asymptotic properties of the proposed estimated QIF estimator \( \hat{\beta}_Q \), and provide standard error formula for it.

Let \( n_k \) denote the number of subjects in \( V_k \) and assume \( n_k/n \to \rho_k \) as \( n \to \infty \), where \( \rho_k \) represents the probability of subject \((i, k)\) being sampled into the \( k \)th marginal validation set. Under the conditions listed in “Appendix”, we demonstrate the asymptotic behavior of \( \hat{\beta}_Q \) in the following theorems.

**Theorem 1** Under conditions (C1)–(C9) in “Appendix”, the following results hold:

1. The proposed estimator \( \hat{\beta}_Q \) is a consistent estimator of \( \beta_0 \).
2. \( \sqrt{n}(\hat{\beta}_Q - \beta_0) \) is asymptotically normally distributed with mean zero and variance matrix \( \Sigma_Q(\beta_0) = (J_0^T W_0^{-1} J_0)^{-1} \), where \( J_0 = J(\beta_0) \), with
\[
J(\beta_0) = \begin{pmatrix} 
\Gamma(\beta_0, B_1) \\
\vdots \\
\Gamma(\beta_0, B_L)
\end{pmatrix},
\]
and \( \Gamma(\beta_0, B) \) and \( \omega_0 \) are defined in (A.1) and (A.2) in “Appendix”, respectively.

The asymptotic covariance \( \Sigma_Q(\beta_0) \) can be consistently estimated by
\[
\hat{\Sigma}_Q(\beta_Q) = \{(\hat{J}^T(\beta_Q) \hat{W}^{-1}(\beta_Q) \hat{J}(\beta_Q))^{-1},
\]
where
\[
\begin{pmatrix} 
\hat{\Gamma}(\beta, B_1) \\
\vdots \\
\hat{\Gamma}(\beta, B_L)
\end{pmatrix}.
\]
with
\[
\hat{\Gamma}(\beta, B) = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \Delta_{ik} \left( \sum_{m=1}^{K} \frac{\partial \hat{R}_{im}(T_{ik}; \beta) \hat{\phi}_{imk}(T_{ik}; \beta, B)}{\partial \beta} - \frac{\hat{S}_{k}^3(T_{ik}; \beta, B)}{\hat{S}_{k}^0(T_{ik}; \beta)} \right)
\]
\[
- \frac{\hat{S}_{k}^3(T_{ik}; \beta, B) + \hat{S}_{k}^2(T_{ik}; \beta, B)}{\hat{S}_{k}^0(T_{ik}; \beta)} + \frac{\hat{S}_{k}^2(T_{ik}; \beta, B)(\hat{S}_{k}^1(T_{ik}; \beta))^T}{(\hat{S}_{k}^0(T_{ik}; \beta))^2},
\]
and
\[
\hat{W} = (\hat{\omega}(\hat{\beta}_Q, B_j, B_{j'}))_{j, j'=1}^L,
\]
where
\[
\hat{\omega}(\beta, B_j, B_{j'}) = n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{m=1}^{K} \hat{U}_{ik}(\beta, B_j) \hat{U}_{im}^T(\beta, B_{j'}),
\]
with
\[
\hat{U}_{ik}(\beta, B) = \Delta_{ik} \tilde{\xi}_{ik}(T_{ik}; \beta, B) - \sum_{l=1}^{n} \Delta_{lk} \tilde{\xi}_{ik}(T_{lk}; \beta, B) \frac{Y_{ik}(T_{lk}) \hat{\gamma}_{ik}(T_{lk}; \beta)}{n \hat{S}_{k}^0(T_{lk}; \beta)}
\]
\[
- \frac{n - n_k}{n_k} \sum_{l=1}^{n} \Delta_{lk} \tilde{\xi}_{ik}(T_{lk}; \beta, B) \frac{Y_{ik}(T_{lk}) (\hat{\gamma}_{lk}(T_{lk}; \beta) - \hat{\psi}_{lk}(T_{lk}; \beta))}{n \hat{S}_{k}^0(T_{lk}; \beta)},
\]
where \(\hat{S}_{k}^d (d = 0, \ldots, 5)\) and \(\hat{\xi}_{ik}(t; \beta, B)\) are defined in “Appendix”.

Remark 1 As a special case of the estimated QIF estimator, the estimator using the independent working correlation is denoted as \(\hat{\beta}_I\), which is the same as the EPPL estimator of Liu et al. (2010), but different expressions of the variance matrix \(\Sigma_I(\beta_0)\) of the asymptotic distribution of \(\sqrt{n}(\hat{\beta}_I - \beta_0)\) and its estimator \(\hat{\Sigma}_I(\beta_I)\) are provided under the conditions (C1)–(C8) in “Appendix”. From the corresponding expressions of \(\Sigma_Q(\beta_0)\) and \(\hat{\Sigma}_Q(\beta_Q)\), we can obtain that
\[
\Sigma_I(\beta_0) = \Gamma^{-1}(\beta_0, I) \omega(\beta_0, I, I) \Gamma^{-1}(\beta_0, I),
\]
and
\[
\hat{\Sigma}_I(\beta_I) = \hat{\Gamma}^{-1}(\beta_I, I) \hat{\omega}(\hat{\beta}_I, I, I) \hat{\Gamma}^{-1}(\beta_I, I).
\]

4 Inference on hazard ratio parameters

The QIF is built on an objective function, which provides a natural way to make inference about the hazard ratio parameter \(\beta\). Suppose that \(\beta\) is partitioned into \(\gamma\) and \(\delta\), where \(\gamma\) is vector of hazard ratio parameters of interest with dimension \(p_0\), and \(\delta\) is a vector of nuisance parameters with dimension \(p - p_0\). As a special case, we also allow \(p_0 = p\), with \(\beta = \gamma\) and \(\delta\) being absent.
To test
\[ H_0 : \gamma = \gamma_0 \quad \text{versus} \quad H_1 : \gamma \neq \gamma_0, \]
we propose a test statistic
\[ T = \hat{Q}_n(\gamma_0, \tilde{\delta}) - \hat{Q}_n(\tilde{\gamma}, \tilde{\delta}), \tag{4.1} \]
where
\[ \tilde{\delta} = \arg\min_\delta \hat{Q}_n(\gamma_0, \delta), \quad (\tilde{\gamma}, \tilde{\delta}) = \arg\min_{(\gamma, \delta)} \hat{Q}_n(\gamma, \delta). \]

The values of \( \hat{Q}_n(\gamma_0, \tilde{\delta}) \) and \( \hat{Q}_n(\tilde{\gamma}, \tilde{\delta}) \) measure how well the model fits the data under \( H_0 \) and \( H_1 \), respectively. Under \( H_0 \), the difference between \( \hat{Q}_n(\gamma_0, \tilde{\delta}) \) and \( \hat{Q}_n(\hat{\gamma}, \hat{\delta}) \) should be very small. However, under \( H_1 \), \( \hat{Q}_n(\gamma_0, \tilde{\delta}) \) should be systematically larger than \( \hat{Q}_n(\tilde{\gamma}, \tilde{\delta}) \).

**Theorem 2** Suppose conditions (C1)–(C9) in “Appendix” are satisfied, under \( H_0 \), the test statistic \( T \) asymptotically follows chi-squared distribution with \( p_0 \) degrees of freedom.

**Comment** To prove Theorem 2, we rewrite that
\[
\hat{Q}_n(\gamma_0, \tilde{\delta}) - \hat{Q}_n(\tilde{\gamma}, \tilde{\delta}) = (\hat{Q}_n(\gamma_0, \tilde{\delta}) - Q_n(\gamma_0, \tilde{\delta})) - (\hat{Q}_n(\tilde{\gamma}, \tilde{\delta}) - Q_n(\tilde{\gamma}, \tilde{\delta})) \\
+ (Q_n(\gamma_0, \tilde{\delta}) - Q_n(\gamma_0, \tilde{\delta}^*)) - (Q_n(\tilde{\gamma}, \tilde{\delta}) - Q_n(\tilde{\gamma}^*, \tilde{\delta}^*)) \\
+ (Q_n(\gamma_0, \tilde{\delta}^*) - Q_n(\tilde{\gamma}^*, \tilde{\delta}^*)),
\]
where \( Q_n \) is defined as in (2.6), \( \tilde{\delta}^* = \arg\min_\delta Q_n(\gamma_0, \delta) \), and \( (\tilde{\gamma}^*, \tilde{\delta}^*) = \arg\min_{(\gamma, \delta)} Q_n(\gamma, \delta) \). From the proof of Theorem 1, we can obtain that the first two brackets equal to \( o_p(1) \). In addition, from the conclusions of the previous Theorem 1 and the Theorem 1 in Xue et al. (2010), both \( \hat{\beta}_Q \) and \( \tilde{\beta}_Q \) are consistent estimators of \( \beta_0 \), then we can have that the third and the fourth brackets also equal to \( o_p(1) \). Furthermore, the last bracket asymptotically approaches to a random variable which follows chi-squared distribution with \( p_0 \) degrees of freedom, the proof of the last one is similar to the proof of Theorem 1 in Qu et al. (2000).

### 5 Simulation studies

In this section, we conduct simulation studies to evaluate the finite-sample behavior of the proposed method. We first evaluate the performance of proposed estimator in Sect. 5.1 and then the performance of inference method in Sect. 5.2.
5.1 Performance of estimated QIF estimator

We compare the proposed estimator with the QIF estimator proposed by Xue et al. (2010) based only on the validation set and the EPPL estimator ($\hat{\beta}_I$) of Liu et al. (2010), which utilizes the auxiliary information but does not consider the intra-cluster correlation in the estimate of $\beta$. The proposed estimator takes both the intra-cluster correlation and the auxiliary information into account.

The covariates $(X_{i1}, X_{i2}, \ldots, X_{iK})$, which are only observed in the validation sets in the real studies, are generated independently from uniform distribution $U(0, 1)$. The covariates $(Z_{i1}, Z_{i2}, \ldots, Z_{iK})$ are independent binary covariates taking value one with probability 0.5. The multivariate failure times $\tilde{T}_{i1}, \tilde{T}_{i2}, \ldots, \tilde{T}_{iK}$ are generated from multivariate (Clayton and Cuzick 1985) model with the joint survival function

\[
S(t_1, \ldots, t_K; D_1, \ldots, D_K) = \left( \sum_{k=1}^{K} \exp(\theta^{-1}\lambda_{0k}t_k e^{\beta_k^T D_k}) - (K - 1) \right)^{-\theta},
\]

where $D_k = (X_k^T, Z_k^T)^T$, and $\beta_k$, which may vary with the failure type, is the corresponding parameter of $D_k$, and $\theta$ is the dependence parameter, a larger value of which represents a weaker dependence between the failure times. We set $\theta = 0.25, 0.5$ or $2$, which presents a varying degree of correlation between the generated failure times, and the baseline hazard function $\lambda_{0k} = 1$. The simulated failure times $(t_1, \ldots, t_K)$ are generated by using the algorithm described in Cai and Shen (2000) through

\[
t_1 = -e^{-\beta_1^T D_1} \ln(1 - u_1),
\]

\[
t_k = \theta e^{-\beta_k^T D_k} \ln \left( (k - 1) - \sum_{i=1}^{k-1} a_i + \left( \sum_{i=1}^{k-1} a_i - (k - 2) \right) (1 - u_k)^{-((\theta + k - 1)^{-1})} \right),
\]

for $k = 2, \ldots, K$, where $a_i = \exp(\theta^{-1}t_i e^{\beta_i^T D_i})$ for $i = 1, \ldots, k - 1$, and $(u_1, \ldots, u_K)$ are generated from uniform distribution over interval $(0, 1)$. Censoring times are generated from $U(0, c)$, where $c$ is a selected constant to achieve a specified censoring rate.

Notice that the true correlation structure of the Clayton model is exchangeable, the working correlation in (2.9) is taken to be exchangeable, and the corresponding estimated QIF estimator is denoted as $\hat{\beta}_{QE}$. We calculated another estimated QIF estimator $\hat{\beta}_{QA}$ using the misspecified AR(1) working correlation. The corresponding resulting estimators of the QIF method based only on the validation set are denoted as $\tilde{\hat{\beta}}_{QEV}$ and $\tilde{\hat{\beta}}_{QAV}$, respectively.

To estimate the induced relative risk function and the baseline hazard function, we apply the Epanechnikov kernel function in (2.8) and (2.7) with bandwidths $\mu_k = 2\hat{\sigma}(A_{V_k}) n_k^{-1/3}$ and $\nu_k = 1.06\hat{\sigma}(T_k) n_k^{-1/5}$, respectively, where $\hat{\sigma}(\cdot)$ is the sample standard deviation function, $A_{V_k}$ is the part of auxiliary covariate $A_k$ in the $k$th marginal validation set. We choose the nearest neighbor interpolation to estimate the induced relative function when the denominator in (2.8) is 0. In addition, it is worth noting...
that $\hat{\lambda}_{0k}(t)$ may be 0 at some locations because the Epanechnikov kernel function is of bounded support, which could make the diagonal matrix $\hat{\Xi}_i(t; \beta)$ not invertible. If this happens, we replace $\hat{\lambda}_{0k}(t)$ with the average of values at the non-zero locations.

We consider two types of simulations:

1. $\beta_k$s are the same for different failure type, i.e. $\beta_1 = \ldots = \beta_K = \beta$.
2. $\beta_k$ varies across failure type.

The auxiliary covariate $A_k$ is generated from $X_k$ via

$$A_k = X_k + \epsilon_k,$$

where $\epsilon_k$ follows a normal distribution $N(0, \sigma^2)$, the positive parameter $\sigma$ controls the strength of association between $A_k$ and $X_k$. Each simulation is repeated 1000 times.

**Simulation study (1)**

In the first simulation, we set the true parameter $\beta = (\beta^{(1)}, \beta^{(2)})^T = (0.693, -0.2)^T$, validation proportion $\rho_k = 0.5$, association parameter $\sigma = 0.1$. The number of independent clusters is $n = 200$, with $K = 4$ or 8 failure types in each cluster.

Tables 1, 2 and 3 demonstrate the simulation results for estimates of parameter $\beta$ for each method under different censoring rates 10%, 40% and 80%. The sample mean and sample standard deviation of the 1000 estimates, the average of estimated standard errors and the coverage rate of the 95% confidence intervals for the true parameter are listed in the Est, SD, SE and CR columns, respectively. RE, the ratio of the empirical variance of $\hat{\beta}_I$ to that of $\hat{\beta}_{QE}$ or $\hat{\beta}_{QA}$, is the estimated relative efficiency of estimated QIF estimators relative to $\hat{\beta}_I$. We summarize the results as follows: (i) The estimates of all the methods are all approximately unbiased. Moreover, the estimators of the asymptotic standard errors are approximately equal to the empirical standard deviations. The corresponding 95% confidence intervals calculated by the estimated standard errors provide reasonable coverage rates. This suggests that the estimates of asymptotic standard errors for all methods work well. (ii) For each considered scenario, the estimator $\hat{\beta}_I$ using auxiliary information is more efficient than the estimators $\tilde{\beta}_{QEV}$ and $\tilde{\beta}_{QAV}$ using validation set only. However, $\hat{\beta}_I$ loses efficiency when the degree of correlation within a cluster becomes stronger. (iii) As $K$ increases, the empirical standard deviations (SD) of all the estimators decrease. That is naturally because of the increase in the total amount of data. (iv) As $\theta$ decreases, the efficiency gain of estimated QIF estimators relative to $\hat{\beta}_I$ increases. From Table 1, estimated QIF estimators are more efficient than the other estimators for all combinations of $\theta$ and $K$. We also observe the same trend from Table 2. From Table 3, however, the estimated QIF estimators are less efficient than $\hat{\beta}_I$, although REs are very close to 1, in several cases due to the reduction of correlation when censoring rate is 80%. Furthermore, as expected, $\hat{\beta}_{QE}$ with correct working correlation is always more efficient than $\hat{\beta}_{QA}$ with misspecified working correlation. (v) The validation proportion of the incomplete covariate has effect on the values of RE, especially for the first parameter. For example, when $K = 4$ and 10% censoring, the REs of $\hat{\beta}_I$ relative to $\hat{\beta}_I$ for $\hat{\beta}^{(1)}$ decrease from (3.34, 2.26, 1.21) to (2.85, 2.03, 1.18), when validation proportion decreases from
0.5 to 0.3 (results not shown). However, when we increased \( n \), not only the CRs but also the REs increased.

**Simulation study (2)**

In practical studies, one may be interested in the failure-type-specific model

\[
\lambda_{ik}(t; \tilde{Z}_{ik}(t)) = \lambda_{0k}(t) \exp\{\beta_k^T \tilde{Z}_{ik}(t)\},
\]

which allows the regression parameters varying with the failure type. We simulate \( K = 2 \) failure types in each cluster, the true parameter \( \beta_1 = (\beta_1^{(1)}, \beta_1^{(2)})^T = (0.693, -0.2)^T \) and \( \beta_2 = (\beta_2^{(1)}, \beta_2^{(2)})^T = (0.5, -0.262)^T \). Since the cluster size is 2, we only need to consider the exchangeable working correlation structure. Consider three settings of \( n \) and censoring rate (CE), \((n, CE) = (300, 10\%), (300, 40\%), (700, 80\%)\). The simulation results are shown in Table 4. From this table, we can observe similar results as in Simulation Study (1).

**5.2 Performance of inference method**

We also conduct simulation studies to assess the performance of the proposed chi-squared test method. The data are generated from the same model as in Simulation Study (1) with \( \beta = (\beta^{(1)}, \beta^{(2)})^T = (0.693, -0.2)^T \), \( n = 200 \), \( K = 4 \), \( \theta = 0.25 \), \( \rho_k = 0.5 \), \( \sigma = 0.1 \), and censoring rate is 10%. First, we are interested in testing \( H_0 : \beta^{(1)} = 0.693 \) versus \( H_1 : \beta^{(1)} \neq 0.693 \). Since the dimension of \( \beta^{(1)} \) is 1, the test statistic \( T \) in (4.1) asymptotically follows \( \chi_1^2 \), where \( \tilde{\beta}^{(2)} \) is calculated by minimizing \( \hat{Q}_n(0.693, \beta^{(2)}) \) with exchangeable or AR(1) working correlation. Figure 1 shows Q–Q plot based on 1000 replications. It is clear that the plots indicate proximity to the \( \chi_1^2 \) distribution for both exchangeable and AR(1) working correlation. We also examine the power of the proposed test under \( H_1 : \beta^{(1)} = \beta^{(1)}_* \). The powers with significance level \( \alpha = 0.05 \) are calculated when \( \beta^{(1)}_* \) takes different values in \([0.3, 0.693]\). According to the simulation results, when \( \beta^{(1)}_* = 0.693 \), i.e., the alternative hypothesis collapses into the null hypothesis, powers are 0.051 and 0.061 for exchangeable and AR(1) working correlation, respectively. It shows that the proposed chi-squared test gives the right level for testing. Figure 2 plots the power functions of the chi-squared test for the estimated QIF method and the QIF method with two different working correlations, and the EPPL method. We can observe that the power functions decrease rapidly as \( \beta^{(1)}_* \) gets closer to the true value (0.693), but the power function for exchangeable working correlation is always larger than that for AR(1) working correlation, thus the test with correct working correlation is more powerful than the one with misspecified working correlation. Nonetheless, powers of the chi-squared test for either of the estimated QIF methods are larger than those of the other two methods. In addition, the power for the EPPL method that utilizes the auxiliary information is larger than those for QIF method based on the validation set only. Similarly, we also consider the hypothesis test that \( H_0 : \beta^{(2)} = -0.2 \), and
| $K$ | $\theta$ | Method  | $\hat{\beta}^{(1)}$ Est | $\hat{\beta}^{(1)}$ SD | $\hat{\beta}^{(1)}$ SE | $\hat{\beta}^{(1)}$ CR | $\hat{\beta}^{(1)}$ RE | $\hat{\beta}^{(2)}$ Est | $\hat{\beta}^{(2)}$ SD | $\hat{\beta}^{(2)}$ SE | $\hat{\beta}^{(2)}$ CR | $\hat{\beta}^{(2)}$ RE |
|-----|---------|---------|-----------------|-----------------|----------------|---------|-------|-----------------|-----------------|----------------|---------|-------|
| 4   | 0.25    | $\tilde{\beta}_{QEV}$ | 0.686 | 0.164 | 0.163 | 0.952 | – | – | 0.202 | 0.098 | 0.091 | 0.926 | – |
|     |         | $\tilde{\beta}_{QAV}$ | 0.682 | 0.171 | 0.169 | 0.948 | – | – | 0.205 | 0.104 | 0.095 | 0.917 | – |
|     |         | $\hat{\beta}_I$ | 0.675 | 0.143 | 0.141 | 0.945 | – | – | 0.203 | 0.077 | 0.075 | 0.938 | – |
|     |         | $\hat{\beta}_E$ | 0.684 | 0.078 | 0.078 | 0.942 | 3.34 | – | 0.200 | 0.039 | 0.038 | 0.936 | 3.93 |
|     |         | $\tilde{\beta}_{QAV}$ | 0.680 | 0.089 | 0.088 | 0.933 | 2.62 | – | 0.201 | 0.047 | 0.045 | 0.941 | 2.65 |
| 0.5 |         | $\hat{\beta}_{QEV}$ | 0.683 | 0.166 | 0.167 | 0.955 | – | – | 0.202 | 0.103 | 0.094 | 0.925 | – |
|     |         | $\hat{\beta}_{QAV}$ | 0.679 | 0.174 | 0.173 | 0.946 | – | – | 0.205 | 0.107 | 0.097 | 0.926 | – |
|     |         | $\hat{\beta}_I$ | 0.673 | 0.142 | 0.139 | 0.943 | – | – | 0.202 | 0.077 | 0.075 | 0.941 | – |
|     |         | $\hat{\beta}_E$ | 0.681 | 0.094 | 0.091 | 0.935 | 2.26 | – | 0.199 | 0.048 | 0.047 | 0.940 | 2.55 |
|     |         | $\tilde{\beta}_{QAV}$ | 0.676 | 0.103 | 0.101 | 0.930 | 1.91 | – | 0.201 | 0.055 | 0.053 | 0.945 | 1.97 |
| 2   |         | $\hat{\beta}_{QEV}$ | 0.678 | 0.184 | 0.179 | 0.939 | – | – | 0.203 | 0.111 | 0.102 | 0.932 | – |
|     |         | $\hat{\beta}_{QAV}$ | 0.676 | 0.188 | 0.181 | 0.924 | – | – | 0.204 | 0.112 | 0.103 | 0.927 | – |
|     |         | $\hat{\beta}_I$ | 0.673 | 0.139 | 0.137 | 0.946 | – | – | 0.201 | 0.078 | 0.075 | 0.938 | – |
|     |         | $\hat{\beta}_E$ | 0.675 | 0.127 | 0.123 | 0.938 | 1.21 | – | 0.200 | 0.069 | 0.067 | 0.941 | 1.24 |
|     |         | $\tilde{\beta}_{QAV}$ | 0.672 | 0.133 | 0.127 | 0.928 | 1.11 | – | 0.201 | 0.072 | 0.069 | 0.935 | 1.15 |
| 8   | 0.25    | $\hat{\beta}_{QEV}$ | 0.699 | 0.122 | 0.113 | 0.939 | – | – | 0.202 | 0.066 | 0.062 | 0.933 | – |
|     |         | $\hat{\beta}_{QAV}$ | 0.697 | 0.130 | 0.121 | 0.929 | – | – | 0.201 | 0.070 | 0.067 | 0.936 | – |
|     |         | $\hat{\beta}_I$ | 0.683 | 0.106 | 0.103 | 0.941 | – | – | 0.201 | 0.054 | 0.054 | 0.949 | – |
|     |         | $\hat{\beta}_E$ | 0.686 | 0.061 | 0.059 | 0.938 | 3.05 | – | 0.200 | 0.027 | 0.026 | 0.942 | 4.12 |
|     |         | $\tilde{\beta}_{QAV}$ | 0.687 | 0.068 | 0.066 | 0.940 | 2.38 | – | 0.200 | 0.031 | 0.031 | 0.949 | 2.96 |
| 0.5 |         | $\hat{\beta}_{QEV}$ | 0.698 | 0.126 | 0.116 | 0.927 | – | – | 0.200 | 0.069 | 0.064 | 0.923 | – |
| $K$ | $\theta$ | Method       | $\hat{b}^{(1)}$ | $\hat{b}^{(2)}$ |
|-----|----------|--------------|----------------|----------------|
|     |          | Est  | SD  | SE  | CR  | RE  | Est  | SD  | SE  | CR  | RE  |
| 2   | 0.123    | $\tilde{\beta}_{QAV}$ | 0.695 | 0.133 | 0.123 | 0.935 | $\tilde{\beta}_{QAV}$ | -0.200 | 0.072 | 0.068 | 0.931 | - |
|     |          |      |     |    |     |     | 0.135 | 0.125 | 0.922 | - |     |
|     | 0.231    | $\tilde{\beta}_{QE}$ | 0.688 | 0.068 | 0.065 | 0.931 | $\tilde{\beta}_{QE}$ | -0.200 | 0.032 | 0.032 | 0.946 | 2.89|
|     |          |      |     |    |     |     | 0.077 | 0.073 | 0.929 | 1.80 | -     |
|     | 0.200    | $\tilde{\beta}_{QA}$ | 0.687 | 0.077 | 0.073 | 0.929 | $\tilde{\beta}_{QA}$ | -0.200 | 0.037 | 0.037 | 0.939 | 2.17|
|     |          |      |     |    |     |     | 0.077 | 0.073 | 0.929 | 1.80 | -     |
| 2   | 0.128    | $\tilde{\beta}_{QEV}$ | 0.698 | 0.135 | 0.125 | 0.922 | $\tilde{\beta}_{QEV}$ | -0.199 | 0.074 | 0.071 | 0.941 | - |
|     |          |      |     |    |     |     | 0.139 | 0.128 | 0.918 | - |     |
|     | 0.200    | $\tilde{\beta}_{QAV}$ | 0.694 | 0.099 | 0.098 | 0.935 | $\tilde{\beta}_{QAV}$ | -0.200 | 0.075 | 0.073 | 0.945 | - |
|     |          |      |     |    |     |     | 0.099 | 0.098 | 0.935 | - |     |
|     | 0.199    | $\tilde{\beta}_{VA}$ | 0.690 | 0.087 | 0.084 | 0.938 | $\tilde{\beta}_{VA}$ | -0.199 | 0.044 | 0.045 | 0.948 | 1.37|
|     |          |      |     |    |     |     | 0.087 | 0.084 | 0.938 | 1.29 | -     |
|     | 0.048    | $\tilde{\beta}_{QA}$ | 0.687 | 0.094 | 0.090 | 0.931 | $\tilde{\beta}_{QA}$ | -0.200 | 0.048 | 0.048 | 0.944 | 1.16|
|     |          |      |     |    |     |     | 0.094 | 0.090 | 0.931 | 1.13 | -     |

$\hat{\beta}_{QEV}$ is the estimator of the QIF method with exchangeable working correlation based only on the validation set, while $\tilde{\beta}_{QAV}$ is the one with AR(1) working correlation. $\hat{\beta}_I$ is the EPPL estimator using the independent structure. $\hat{\beta}_{QE}$ and $\hat{\beta}_{QA}$ are the estimators of the proposed estimated QIF method with exchangeable and AR(1) working correlation, respectively.
Table 2  Simulation results for common effect size across failure type: $\beta = (\beta^{(1)}, \beta^{(2)})^T = (0.693, -0.2)^T$ under the censoring rate 40%

| $K$ | $\theta$ | Method | $\hat{\beta}^{(1)}$ | $\hat{\beta}^{(2)}$ |
|-----|----------|--------|---------------------|---------------------|
|     |          |        | Est     | SD    | SE    | CR | RE | Est     | SD    | SE    | CR | RE |
| 4   | 0.25     | $\tilde{\beta}_{QEV}$ | 0.679   | 0.213 | 0.206 | 0.936 | – | – | 0.205   | 0.121 | 0.117 | 0.946 | – |
|     |          | $\tilde{\beta}_{QAV}$ | 0.676   | 0.222 | 0.213 | 0.932 | – | – | 0.206   | 0.124 | 0.121 | 0.950 | – |
|     |          | $\hat{\beta}_I$      | 0.681   | 0.177 | 0.169 | 0.936 | – | – | 0.205   | 0.093 | 0.092 | 0.950 | – |
|     |          | $\hat{\beta}_{QE}$    | 0.687   | 0.120 | 0.116 | 0.941 | 2.18 | – | – | 0.201   | 0.062 | 0.061 | 0.946 | 2.20 |
|     |          | $\hat{\beta}_{QA}$    | 0.682   | 0.131 | 0.127 | 0.934 | 1.82 | – | – | 0.202   | 0.069 | 0.068 | 0.949 | 1.79 |
| 0.5 |          | $\tilde{\beta}_{QEV}$ | 0.673   | 0.218 | 0.212 | 0.935 | – | – | 0.206   | 0.127 | 0.120 | 0.941 | – |
|     |          | $\tilde{\beta}_{QAV}$ | 0.670   | 0.224 | 0.217 | 0.934 | – | – | 0.207   | 0.129 | 0.123 | 0.940 | – |
|     |          | $\hat{\beta}_I$      | 0.680   | 0.174 | 0.168 | 0.939 | – | – | 0.204   | 0.091 | 0.092 | 0.954 | – |
|     |          | $\hat{\beta}_{QE}$    | 0.685   | 0.132 | 0.131 | 0.944 | 1.73 | – | – | 0.202   | 0.070 | 0.070 | 0.952 | 1.66 |
|     |          | $\hat{\beta}_{QA}$    | 0.679   | 0.143 | 0.140 | 0.935 | 1.48 | – | – | 0.203   | 0.076 | 0.076 | 0.954 | 1.45 |
| 2   |          | $\tilde{\beta}_{QEV}$ | 0.668   | 0.225 | 0.221 | 0.939 | – | – | 0.204   | 0.134 | 0.127 | 0.944 | – |
|     |          | $\tilde{\beta}_{QAV}$ | 0.666   | 0.226 | 0.222 | 0.934 | – | – | 0.205   | 0.134 | 0.127 | 0.937 | – |
|     |          | $\hat{\beta}_I$      | 0.679   | 0.166 | 0.166 | 0.942 | – | – | 0.202   | 0.092 | 0.092 | 0.954 | – |
|     |          | $\hat{\beta}_{QE}$    | 0.678   | 0.156 | 0.157 | 0.940 | 1.14 | – | – | 0.202   | 0.088 | 0.086 | 0.950 | 1.11 |
|     |          | $\hat{\beta}_{QA}$    | 0.675   | 0.161 | 0.160 | 0.940 | 1.07 | – | – | 0.203   | 0.090 | 0.088 | 0.947 | 1.05 |
| 8   | 0.25     | $\tilde{\beta}_{QEV}$ | 0.702   | 0.152 | 0.143 | 0.929 | – | – | 0.201   | 0.084 | 0.080 | 0.944 | – |
|     |          | $\tilde{\beta}_{QAV}$ | 0.701   | 0.159 | 0.152 | 0.940 | – | – | 0.201   | 0.088 | 0.085 | 0.944 | – |
|     |          | $\hat{\beta}_I$      | 0.695   | 0.122 | 0.123 | 0.948 | – | – | 0.203   | 0.066 | 0.066 | 0.947 | – |
|     |          | $\hat{\beta}_{QE}$    | 0.696   | 0.085 | 0.082 | 0.941 | 2.08 | – | – | 0.201   | 0.042 | 0.041 | 0.944 | 2.39 |
|     |          | $\hat{\beta}_{QA}$    | 0.697   | 0.094 | 0.092 | 0.942 | 1.68 | – | – | 0.201   | 0.048 | 0.048 | 0.944 | 1.85 |
| 0.5 |          | $\tilde{\beta}_{QEV}$ | 0.706   | 0.157 | 0.147 | 0.934 | – | – | 0.200   | 0.087 | 0.083 | 0.945 | – |
|     |          | $\tilde{\beta}_{QAV}$ | 0.702   | 0.163 | 0.154 | 0.948 | – | – | 0.200   | 0.090 | 0.087 | 0.950 | – |
|     |          | $\hat{\beta}_I$      | 0.697   | 0.120 | 0.121 | 0.946 | – | – | 0.202   | 0.066 | 0.065 | 0.951 | – |

Seminparametric regression based on quadratic inference...
| $K$ | $\theta$ | Method | $\hat{\beta}^{(1)}$ | $\hat{\beta}^{(2)}$ |
|-----|--------|--------|------------------|------------------|
|     |        |        | Est   | SD  | SE  | CR   | RE   | Est   | SD  | SE  | CR   | RE   |
|     |        | $\hat{\beta}_{QE}$ | 0.700 | 0.092 | 0.090 | 0.945 | 1.69 | – | 0.200 | 0.048 | 0.047 | 0.955 | 1.85 |
|     |        | $\hat{\beta}_{QA}$ | 0.699 | 0.102 | 0.099 | 0.938 | 1.39 | – | 0.200 | 0.054 | 0.053 | 0.950 | 1.48 |
| 2   |        | $\tilde{\beta}_{QEV}$ | 0.700 | 0.165 | 0.155 | 0.939 | – | – | 0.199 | 0.089 | 0.088 | 0.947 | – |
|     |        | $\tilde{\beta}_{QAV}$ | 0.697 | 0.168 | 0.158 | 0.934 | – | – | 0.200 | 0.091 | 0.090 | 0.945 | – |
|     |        | $\tilde{\beta}_I$ | 0.695 | 0.120 | 0.118 | 0.949 | – | – | 0.200 | 0.063 | 0.065 | 0.950 | – |
|     |        | $\hat{\beta}_{QE}$ | 0.698 | 0.111 | 0.108 | 0.943 | 1.16 | – | 0.199 | 0.058 | 0.059 | 0.946 | 1.19 |
|     |        | $\hat{\beta}_{QA}$ | 0.695 | 0.116 | 0.113 | 0.950 | 1.07 | – | 0.201 | 0.062 | 0.062 | 0.950 | 1.04 |

See Table 1
Table 3  Simulation results for common effect size across failure type: $\beta = (\beta^{(1)}, \beta^{(2)})^T = (0.693, -0.2)^T$ under the censoring rate 80% 

| $K$ | $\theta$ | Method | $\hat{\beta}^{(1)}$ | $\hat{\beta}^{(2)}$ |
|-----|----------|--------|----------------------|----------------------|
|     |          |        | Est  | SD  | SE  | CR  | RE  | Est  | SD  | SE  | CR  | RE  |
| 4   | 0.25     | $\tilde{\beta}_{QEV}$ | 0.655 | 0.412 | 0.374 | 0.924 | – | – | 0.216 | 0.250 | 0.215 | 0.921 | – |
|     |          | $\tilde{\beta}_{QAV}$ | 0.649 | 0.414 | 0.379 | 0.928 | – | – | 0.216 | 0.244 | 0.219 | 0.923 | – |
|     |          | $\tilde{\beta}_{I}$ | 0.684 | 0.308 | 0.288 | 0.939 | – | – | 0.198 | 0.164 | 0.159 | 0.941 | – |
|     |          | $\tilde{\beta}_{QE}$ | 0.695 | 0.264 | 0.248 | 0.936 | 1.36 | – | 0.199 | 0.147 | 0.136 | 0.935 | 1.24 |
|     |          | $\tilde{\beta}_{QA}$ | 0.685 | 0.277 | 0.259 | 0.943 | 1.23 | – | 0.200 | 0.153 | 0.143 | 0.936 | 1.15 |
| 0.5 |          | $\tilde{\beta}_{QEV}$ | 0.652 | 0.435 | 0.381 | 0.929 | – | – | 0.211 | 0.241 | 0.221 | 0.938 | – |
|     |          | $\tilde{\beta}_{QAV}$ | 0.651 | 0.412 | 0.384 | 0.933 | – | – | 0.209 | 0.238 | 0.222 | 0.939 | – |
|     |          | $\tilde{\beta}_{I}$ | 0.689 | 0.304 | 0.287 | 0.939 | – | – | 0.196 | 0.160 | 0.159 | 0.946 | – |
|     |          | $\tilde{\beta}_{QE}$ | 0.691 | 0.282 | 0.266 | 0.938 | 1.17 | – | 0.199 | 0.155 | 0.147 | 0.940 | 1.07 |
|     |          | $\tilde{\beta}_{QA}$ | 0.682 | 0.290 | 0.273 | 0.938 | 1.10 | – | 0.200 | 0.156 | 0.151 | 0.945 | 1.06 |
| 2   |          | $\tilde{\beta}_{QEV}$ | 0.658 | 0.432 | 0.385 | 0.918 | – | – | 0.210 | 0.248 | 0.224 | 0.932 | – |
|     |          | $\tilde{\beta}_{QAV}$ | 0.660 | 0.417 | 0.386 | 0.915 | – | – | 0.210 | 0.250 | 0.224 | 0.936 | – |
|     |          | $\tilde{\beta}_{I}$ | 0.698 | 0.296 | 0.287 | 0.942 | – | – | 0.195 | 0.165 | 0.159 | 0.945 | – |
|     |          | $\tilde{\beta}_{QE}$ | 0.685 | 0.298 | 0.282 | 0.938 | 0.99 | – | 0.202 | 0.167 | 0.157 | 0.936 | 0.98 |
|     |          | $\tilde{\beta}_{QA}$ | 0.683 | 0.299 | 0.283 | 0.934 | 0.98 | – | 0.202 | 0.167 | 0.157 | 0.936 | 0.97 |
| 8   | 0.25     | $\tilde{\beta}_{QEV}$ | 0.705 | 0.269 | 0.260 | 0.937 | – | – | 0.203 | 0.147 | 0.148 | 0.953 | – |
|     |          | $\tilde{\beta}_{QAV}$ | 0.696 | 0.284 | 0.269 | 0.949 | – | – | 0.205 | 0.149 | 0.154 | 0.961 | – |
|     |          | $\tilde{\beta}_{I}$ | 0.706 | 0.214 | 0.206 | 0.937 | – | – | 0.205 | 0.111 | 0.113 | 0.953 | – |
|     |          | $\tilde{\beta}_{QE}$ | 0.710 | 0.178 | 0.169 | 0.939 | 1.44 | – | 0.204 | 0.093 | 0.092 | 0.942 | 1.43 |
|     |          | $\tilde{\beta}_{QA}$ | 0.706 | 0.195 | 0.183 | 0.928 | 1.20 | – | 0.206 | 0.100 | 0.100 | 0.956 | 1.23 |
| $K$ | $\theta$ | Method | $\hat{\beta}^{(1)}$ | $\hat{\beta}^{(2)}$ |
|-----|---------|--------|--------------------|----------------|
|     |         |        | Est | SD  | SE  | CR | RE | Est | SD  | SE  | CR | RE |
| 0.5 | 0.699   | 0.281  | 0.266 | 0.934 | – | – | 0.202 | 0.151 | 0.153 | 0.949 | – |
| 2   | 0.692   | 0.288  | 0.271 | 0.930 | – | – | 0.204 | 0.153 | 0.156 | 0.947 | – |
| 0.5 | 0.707   | 0.204  | 0.204 | 0.937 | – | – | 0.202 | 0.111 | 0.112 | 0.954 | – |
| 2   | 0.709   | 0.188  | 0.183 | 0.938 | 1.19 | – | 0.202 | 0.100 | 0.101 | 0.954 | 1.22 |
| 0.5 | 0.705   | 0.199  | 0.193 | 0.929 | 1.06 | – | 0.204 | 0.106 | 0.106 | 0.949 | 1.10 |
| 2   | 0.684   | 0.291  | 0.273 | 0.932 | – | – | 0.205 | 0.164 | 0.157 | 0.937 | – |
| 0.5 | 0.681   | 0.290  | 0.273 | 0.929 | – | – | 0.204 | 0.161 | 0.157 | 0.942 | – |
| 2   | 0.705   | 0.207  | 0.203 | 0.943 | – | – | 0.203 | 0.112 | 0.112 | 0.954 | – |
| 0.5 | 0.702   | 0.207  | 0.198 | 0.942 | 1.00 | – | 0.205 | 0.111 | 0.109 | 0.947 | 1.02 |
| 2   | 0.700   | 0.209  | 0.200 | 0.944 | 0.98 | – | 0.206 | 0.112 | 0.110 | 0.947 | 0.99 |

See Table 1
Table 4  Simulation results for varying effect size across failure type: $\beta_1 = (\beta_1^{(1)}, \beta_1^{(2)})^T = (0.693, -0.2)^T$, $\beta_2 = (\beta_2^{(1)}, \beta_2^{(2)})^T = (0.5, -0.262)^T$

| $\theta$ | Method | $\hat{\beta}_1^{(1)}$ | $\hat{\beta}_2^{(1)}$ | $\hat{\beta}_1^{(2)}$ | $\hat{\beta}_2^{(2)}$ |
|----------|--------|------------------|------------------|------------------|------------------|
| 0.25     | $\tilde{\beta}_{QEV}$ | 0.696 0.294 0.272 0.941 | -0.191 0.170 0.155 0.931 | -0.477 0.296 0.270 0.929 | -0.263 0.172 0.156 0.928 |
|          | $\tilde{\beta}_I$      | 0.693 0.229 0.222 0.940 | -0.194 0.121 0.123 0.948 | -0.495 0.226 0.222 0.943 | -0.257 0.127 0.124 0.946 |
|          | $\tilde{\beta}_{QE}$   | 0.681 0.139 0.130 0.934 | 2.69 -0.200 0.073 0.069 0.930 2.75 | 0.493 0.139 0.128 0.915 2.62 -0.263 0.070 0.069 0.949 3.31 |
| 0.5      | $\tilde{\beta}_{QEV}$ | 0.694 0.301 0.278 0.934 | -0.190 0.172 0.158 0.926 | -0.484 0.302 0.277 0.929 | -0.263 0.176 0.160 0.925 |
|          | $\tilde{\beta}_I$      | 0.693 0.229 0.222 0.941 | -0.194 0.121 0.123 0.948 | -0.498 0.226 0.222 0.936 | -0.258 0.127 0.124 0.946 |
|          | $\tilde{\beta}_{QE}$   | 0.682 0.170 0.158 0.937 1.81 -0.198 0.090 0.085 0.932 1.82 | 0.495 0.170 0.157 0.930 1.76 -0.263 0.087 0.085 0.947 2.13 |
| 2        | $\tilde{\beta}_{QEV}$ | 0.688 0.315 0.291 0.938 | -0.189 0.179 0.166 0.932 | -0.482 0.315 0.291 0.921 | -0.267 0.184 0.168 0.926 |
|          | $\tilde{\beta}_I$      | 0.693 0.229 0.222 0.941 | -0.194 0.121 0.123 0.948 | -0.502 0.225 0.222 0.950 | -0.262 0.123 0.124 0.961 |
|          | $\tilde{\beta}_{QE}$   | 0.677 0.223 0.206 0.932 1.05 -0.195 0.117 0.113 0.937 1.08 | 0.490 0.220 0.206 0.938 1.05 -0.266 0.117 0.114 0.943 1.10 |

$n = 300$, censoring rate 10%
| $\theta$ | Method | $\hat{\beta}_1^{(1)}$ | $\hat{\beta}_1^{(2)}$ | $\hat{\beta}_2^{(1)}$ | $\hat{\beta}_2^{(2)}$ |
|---|---|---|---|---|---|
| 2 | $\tilde{\beta}_{QE}$ | 0.680 | 0.387 | 0.357 | 0.933 | -0.200 | 0.214 | 0.205 | 0.935 | - | 0.481 | 0.411 | 0.366 | 0.913 | - | -0.276 | 0.225 | 0.212 | 0.932 | - |
| | $\hat{\beta}_I$ | 0.702 | 0.273 | 0.269 | 0.953 | -0.195 | 0.142 | 0.148 | 0.956 | - | 0.508 | 0.282 | 0.275 | 0.939 | - | -0.260 | 0.152 | 0.153 | 0.950 | - |
| | $\tilde{\beta}_{QE}$ | 0.683 | 0.271 | 0.259 | 0.942 | 1.01 | -0.200 | 0.142 | 0.142 | 0.950 | 1.01 | 0.491 | 0.282 | 0.264 | 0.932 | 1.00 | -0.268 | 0.148 | 0.147 | 0.946 | 1.05 |
| $n = 700$, censoring rate 80% | | | | | |
| 0.25 | $\tilde{\beta}_{QE}$ | 0.672 | 0.404 | 0.397 | 0.948 | -0.213 | 0.241 | 0.229 | 0.947 | - | 0.477 | 0.471 | 0.418 | 0.912 | - | -0.283 | 0.259 | 0.243 | 0.938 | - |
| | $\hat{\beta}_I$ | 0.701 | 0.311 | 0.299 | 0.940 | -0.200 | 0.164 | 0.166 | 0.955 | - | 0.506 | 0.323 | 0.314 | 0.942 | - | -0.262 | 0.169 | 0.176 | 0.958 | - |
| | $\tilde{\beta}_{QE}$ | 0.686 | 0.283 | 0.272 | 0.936 | 1.20 | -0.206 | 0.150 | 0.151 | 0.948 | 1.18 | 0.494 | 0.305 | 0.286 | 0.938 | 1.12 | -0.270 | 0.162 | 0.160 | 0.948 | 1.09 |
| 0.5 | $\tilde{\beta}_{QE}$ | 0.670 | 0.412 | 0.401 | 0.938 | -0.213 | 0.245 | 0.232 | 0.952 | - | 0.473 | 0.476 | 0.422 | 0.922 | - | -0.283 | 0.267 | 0.246 | 0.934 | - |
| | $\hat{\beta}_I$ | 0.701 | 0.311 | 0.299 | 0.940 | -0.200 | 0.164 | 0.166 | 0.955 | - | 0.510 | 0.328 | 0.314 | 0.943 | - | -0.267 | 0.172 | 0.176 | 0.957 | - |
| | $\tilde{\beta}_{QE}$ | 0.683 | 0.299 | 0.286 | 0.942 | 1.08 | -0.207 | 0.160 | 0.159 | 0.950 | 1.04 | 0.494 | 0.327 | 0.301 | 0.931 | 1.01 | -0.276 | 0.172 | 0.168 | 0.947 | 1.00 |
| 2 | $\tilde{\beta}_{QE}$ | 0.672 | 0.420 | 0.405 | 0.933 | -0.215 | 0.247 | 0.234 | 0.947 | - | 0.474 | 0.458 | 0.425 | 0.935 | - | -0.278 | 0.274 | 0.248 | 0.934 | - |
| | $\hat{\beta}_I$ | 0.701 | 0.311 | 0.299 | 0.940 | -0.200 | 0.164 | 0.166 | 0.955 | - | 0.521 | 0.316 | 0.314 | 0.948 | - | -0.263 | 0.183 | 0.176 | 0.948 | - |
| | $\tilde{\beta}_{QE}$ | 0.687 | 0.313 | 0.295 | 0.930 | 0.99 | -0.207 | 0.165 | 0.165 | 0.947 | 0.99 | 0.499 | 0.324 | 0.310 | 0.941 | 0.96 | -0.271 | 0.187 | 0.174 | 0.940 | 0.96 |

The cluster size $K = 2$, validation proportion $\rho_k = 0.5$, association parameter $\sigma = 0.1$
\( H_1 : \beta^{(2)} \neq -0.2 \), and compute the powers under \( H_1 : \beta^{(2)} = \beta^{(2)}_n \) when \( \beta^{(2)}_n \) varies in \([-0.2, 0]\). Similar results are obtained but not presented in this paper due to space limitation.

6 Analysis of SOLVD data

We apply the proposed method to Left Ventricular Dysfunction (SOLVD 1991) study in this section. The SOLVD study was a randomized, double-masked, placebo-controlled trial between 1986 and 1991. The trial had a three-year recruitment and a two-year follow-up. The basic inclusion criteria for the prevention trial were: age between 21 and 80 years, inclusive, no overt symptoms of congestive heart failure, and left ventricular EF less than 35%. EF is a number between 0 and 100 that measures the efficiency of the heart in ejecting blood. A total of 4228 patients with asymptomatic left ventricular dysfunction were randomly assigned to receive either enalapril or placebo at one of the 83 hospitals linked to 23 centers in the United States, Canada, and Belgium. Liu et al. (2009) and Liu et al. (2010) have analyzed this data without considering the intra-cluster correlation.

The primary clinical issues of interest are the effects of covariates on the risk of heart failure and on the first nonfatal myocardial infarction (MI) after adjusting for the confounding variables. The covariates of interest are ejection fraction, patient’s gender (SEX), which is coded 1 for male and 0 for female, treatment (TRT), which is coded 1 for enalapril and 0 for placebo, and patient’s age (AGE), which is measured in years. In the SOLVD study, the covariates SEX, TRT and AGE were recorded for almost all of the patients, but only 108 among the total of 4228 patients have their ejection fraction recorded.
fraction accurately measured using a standardized radionucleotide technique (LVEF). A related nonstandardized measure (EF) was ascertained for all the patients. Therefore, the nonstandardized measure (EF) is a surrogate measure for the standardized measure (LVEF) in this case.

In terms of the notation in the previous sections, we set \( X_{ik} = LVEF_{ik} \), \( A_{ik} = EF_{ik} \), \( Z_{ik} = (TRT_{ik}, SEX_{ik}, AGE_{ik})^T \), where \( k \) denotes failure type with \( k = 1 \) for heart failure and \( k = 2 \) for nonfatal MI and \( i \) denotes the patient with \( i = 1, \ldots, 4228 \). Let \( \beta_k = (\beta_{1k}, \beta_{2k})^T \) be the unknown regression coefficients, we fit the following marginal hazards model to the SOLVD data:

\[
\lambda_{ik}(t; X_{ik}, Z_{ik}) = \lambda_{0k}(t) \exp(\beta_{1k} X_{ik} + \beta_{2k}^T Z_{ik}).
\]

Since the primary covariate LVEF is continuous and severely incomplete, we need to estimate the induced relative risk function using the validation set. Furthermore, Liu...
et al. (2009) found that, given EF, the LVEF is conditionally independent of the rest of the covariates, thus $\psi_{ik}(t; \beta)$ can be estimated through (2.8) with $A^*_ik = A_{ik}$.

Table 5 presents the data analysis results for two methods, the proposed estimated QIF method, and the EPPL method which ignores the intra-cluster correlation between the failure times. It can be seen that the parameter estimates from two methods are close but the proposed estimated QIF method have smaller standard errors. To test whether or not the covariates have significant effects on the times of heart failure and nonfatal MI, we calculate p-values from both the two-sided Z-test and the chi-squared test for the two methods. The results indicate that, at 0.05 significance level, all of the covariates are statistically significant for heart failure by the proposed method, while SEX is not significant from the EPPL method. From both methods, only TRT is significant for the risk of nonfatal MI. By the proposed method, the risk of heart failure decreases by 3.92% (95% CI [1.83%, 5.97%]) with 1% increase in LVEF, the risk increases by 2.63% (95% CI [1.43%, 3.85%]) per year increase in age, males have about 29.46% (95% CI [7.19%, 46.39%]) lower risk for heart failure than females, and enalapril reduces the risk by 33.77% (95% CI [20.52%, 44.80%]).

To illustrate the prediction of the survival probability for a subject, Fig. 3 shows the estimated survival curves of heart failure and nonfatal MI for a 69-year-old male patient with LVEF of 28% (the median of LVEF), receiving enalapril. The survival curves by the two methods are very close, but the pointwise confidence intervals from the proposed method are narrower than those from the EPPL method.

7 Concluding remarks

We proposed an estimated QIF approach for multivariate failure time data when the primary covariate is ascertained on a subset of full cohort but auxiliary information

| Covariate | Proposed method | EPPL method |
|-----------|----------------|-------------|
|           | $Coe$ | $SE$ | $P$-value | $Coe$ | $SE$ | $P$-value |
|           | Z-test | $\chi^2$-test | Z-test | $\chi^2$-test | |
| For heart failure | | | | | | |
| LVEF      | −0.040 | 0.011 | <0.001 | −0.045 | 0.012 | <0.001 | |
| TRT       | −0.412 | 0.093 | <0.001 | −0.454 | 0.106 | <0.001 | |
| SEX       | −0.349 | 0.140 | <0.001 | −0.318 | 0.165 | 0.054 | 0.072 |
| AGE       | 0.026 | 0.006 | <0.001 | 0.023 | 0.012 | 0.045 | 0.009 |
| For nonfatal MI | | | | | | |
| LVEF      | 0.006 | 0.011 | 0.566 | 0.534 | 0.023 | 0.015 | 0.111 | 0.102 |
| TRT       | −0.433 | 0.116 | <0.001 | <0.001 | −0.391 | 0.131 | 0.003 | 0.002 |
| SEX       | 0.084 | 0.196 | 0.669 | 0.294 | 0.048 | 0.214 | 0.822 | 0.815 |
| AGE       | 0.006 | 0.006 | 0.275 | 0.715 | 0.004 | 0.008 | 0.651 | 0.597 |

Proposed method is referred to as the estimated QIF method with exchangeable working correlation. EPPL method is the method which uses the independent structure.
Fig. 3 Survival curves by the proposed method (bold curve) and the EPPL method (thin curve) for heart failure and nonfatal myocardial infarction of a subject with covariates LVEF = 28%, TRT = 1, SEX = 1 and AGE = 69, along with the corresponding 95% pointwise confidence intervals (dotted curves)

is available on the full cohort. For our proposed approach, we allowed the censoring times for different failures for a subject to be different. It is worth noting that in practice the censoring times for different failures for a subject are usually the same. This can be treated as a special situation of our general set up and the proposed method is applicable.

In this article, we consider the situation that auxiliary variable is continuous. The method is based on the kernel smoothing technique and therefore is nonparametric with respect to the association between the missing covariate and corresponding auxiliary. QIF method has advantage of incorporating intra-cluster correlation in the
estimation procedure. Compared with other existing methods (e.g., Liu et al. 2010) where intra-cluster correlation is not considered, the proposed procedure can improve the estimation efficiency without requiring the specification of the correlation formula. Another advantage of proposed method is that it is easy to implement. In this work, we consider the situation when the dimension of continuous auxiliary covariates is low, further research is needed when the dimension is high.

Acknowledgements We are grateful to the associate editor and the reviewers for their detailed and constructive comments which led to the improvements in the paper. This research is supported in part by the National Key Research and Development Project of China No. 2018YFC1314603 (Liu), the National Science Foundation of China (NSFC) Grants 11771366 (Yan, Liu), the U.S. National Institute of Health Grants P01 CA142538 (Cai, Zhou), P42ES031007 (Zhou) and P30 ES010126 (Zhou).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Appendix

For \( i = 1, \ldots, n \), let \( \Phi_i(t; \beta, B) = \Xi_i^{1/2}(t; \beta)B\Xi_i^{-1/2}(t; \beta) \), and \( \Phi_i(t; \beta, B) = \{\phi_{lmk}\}_{m,k=1}^{K} \). For \( k = 1, \ldots, K \), some quantities are defined as follows:

\[
S_k^d(t; \beta) = n^{-1} \sum_{l=1}^{n} Y_{lk}(t) r_{lk}^{(d)}(t; \beta), \quad (d = 0, 1),
\]

\[
S_k^d(t; \beta, B) = n^{-1} \sum_{l=1}^{n} \sum_{m=1}^{K} Y_{lk}(t) R_{lm}(t; \beta)\phi_{lmk}(t; \beta, B)(r_{lk}^{(d-2)}(t; \beta))^T, \quad (d = 2, 3),
\]

\[
S_k^4(t; \beta, B) = n^{-1} \sum_{l=1}^{n} \sum_{m=1}^{K} Y_{lk}(t) R_{lm}(t; \beta)\{\partial \phi_{lmk}(t; \beta, B) / \partial \beta^T\} r_{lk}(t; \beta),
\]

\[
S_k^5(t; \beta, B) = n^{-1} \sum_{l=1}^{n} \sum_{m=1}^{K} Y_{lk}(t) \{\partial R_{lm}(t; \beta) / \partial \beta^T\} \phi_{lmk}(t; \beta, B) r_{lk}(t; \beta).
\]

For \( d = 0, \ldots, 5 \), define \( \tilde{S}_k^d \) by substituting \( \hat{r}_{lk}(t; \beta) \), \( \hat{R}_{lm}(t; \beta) \) and \( \hat{\phi}_{lmk}(t; \beta, B) = \hat{\lambda}_{lm}^{1/2}(t; \beta) \Xi_{lm}^{-1/2}(t; \beta) \) for \( r_{lk}(t; \beta) \), \( R_{lm}(t; \beta) \) and \( \phi_{lmk}(t; \beta, B) \) in \( S_k^d \), respectively. Furthermore, we also define

\[
E_k(t; \beta, B) = \frac{S_k^2(t; \beta, B)}{S_k^0(t; \beta)}, \quad V_k(t; \beta, B) = \frac{S_k^3(t; \beta, B)}{S_k^0(t; \beta)} - \frac{S_k^2(t; \beta, B)(S_k^1(t; \beta))}{(S_k^0(t; \beta))^2}.
\]
We introduce some conditions to ensure the consistency and asymptotic normality of $\hat{\beta}_Q$ and $\hat{\beta}_T$ as follows:

(C1) $\lambda_{ok}(\tau) = \int_0^\tau \lambda_{ok}(t)dt < \infty$, for $k = 1, \ldots, K$.

(C2) $P\{Y_{ik}(t) = 1\} > 0$, for all $t \in [0, \tau]$, $i = 1, \ldots, n$.

(C3) There exists a compact set $\mathcal{B}$, containing $\beta_0$ as its interior point.

(C4) Multivariate kernel function $\Psi(\cdot)$ is non-negative and uniformly bounded with finite support satisfying that $\int \Psi(u)du = 1$ and $\int \Psi^2(u)du < \infty$. Furthermore, $\Psi(\cdot)$ has order $\alpha_0$ in the sense that $\alpha_0 \equiv \inf\{|\alpha| > b; \int_{\mathbb{R}^b} u^\alpha \Psi(u)du \neq 0\}$, where $b$ is the dimension of $u$, $u = (u_1, \ldots, u_b)$, $\alpha = (\alpha_1, \ldots, \alpha_b)$ with $\alpha_i$s being non-negative integers, $u^\alpha = u_1^{\alpha_1} \cdots u_b^{\alpha_b}$.

For $k = 1, \ldots, K$, the bandwidth matrix $\mu_k$ satisfies that $\sqrt{n}\parallel\mu_k\parallel^{\alpha_0} \to 0$ and $\log n/(\sqrt{n}\parallel\mu_k\parallel^b) \to 0$.

(C5) Let $\alpha_0$ be as in (C4), $h_{k}(v, s) = \frac{\partial^2 H_k(1,v,s)}{\partial v \partial s}$ has the $\alpha_0$th continuous derivative with respect to $v$, where $H_k(u, v, s)$ is the joint distribution function of $(\eta_k Y_k(t), A_k^*(t), X_k(t))$ for given $t$.

(C6) There exist scalar, vector and matrix functions $s^d_k(t; \beta)(d = 0, 1), s^d_k(t; \beta, B)$ $(d = 2, \ldots, 5)$, such that for all $k = 1, \ldots, K$ and all constant matrix $B$,

$$\sup_{t \in [0, \tau], \beta \in \mathcal{B}} \|S^d_k - s^d_k\| \to 0,$$

in probability.

(C7) Let $s^d_k(t; \beta)(d = 0, 1), s^d_k(t; \beta, B)(d = 2, \ldots, 5)$ be as in Condition (C6), and set $e^1_k(t; \beta, B) = s^1_k(t; \beta, B)/s^0_k(t; \beta), e^2_k(t; \beta, B) = s^3_k(t; \beta, B)/s^0_k(t; \beta) - s^2_k(t; \beta, B)(s^1_k(t; \beta, B))^T/(s^0_k(t; \beta))^2$. Then for all $\beta \in \mathcal{B}, t \in [0, \tau]$, and $k = 1, \ldots, K$,

$$s^1_k(t; \beta) = \partial s^0_k(t; \beta)/\partial \beta,$$

$$s^3_k(t; \beta, B) = \partial s^2_k(t; \beta, B)/\partial \beta - s^4_k(t; \beta, B) - s^5_k(t; \beta, B),$$

and $s^0_k(t; \beta)$ is bounded away from 0 on $[0, \tau] \times \mathcal{B}$. For all basis matrix $B$, the matrix

$$\Gamma(\beta_0, B) = -\sum_{k=1}^K \int_0^\tau v^k(t; \beta_0, B)s^0_k(t; \beta_0)\lambda_{ok}(t)dt \quad (A.1)$$

is negative definite.

(C8) There exists a matrix function $\omega(\cdot, \cdot, \cdot)$, such that for any $K \times K$ constant matrices $B_1, B_2$,

$$n^{-1} \sum_{i=1}^n \sum_{k=1}^K \int_0^\tau \xi_{ik}(t; \beta, B_1)\xi_{ik}^T(t; \beta, B_2)s^0_k(t; \beta)\lambda_{ok}(t)dt \to \omega(\beta, B_1, B_2),$$

\(\copyright\) Springer
in probability uniformly for $\beta \in \mathcal{B}$, where

$$\xi_{ik}(t; \beta, B) = \sum_{m=1}^{K} R_{im}(t; \beta) \phi_{imk}(t; \beta, B) - e_k(t; \beta, B).$$

Furthermore, for any set of basis matrices $\{B_j, j = 1, \ldots, L\}$, the matrix

$$W_0 = W(\beta_0) = \{\omega(\beta_0, B_j, B_{j'})\}_{j, j' = 1}^{L}$$

is positive definite.

(C9) The baseline hazard rates $\{\lambda_{0k}(\cdot), k = 1, \ldots, K\}$ are twice continuously differentiable on $[0, \tau]$.

**Proof of Theorem 1** Consistency

Following the arguments of Xue et al. (2010), one can show that $\hat{\beta}_Q$ is consistent for $\beta_0$ provided that:

(i) $\partial \hat{G}_n(\beta) / \partial \beta^T$ exists and is continuous, and it converges in probability to a fixed function, say $J(\beta)$, uniformly for $\beta \in \mathcal{B}$;

(ii) $\hat{G}_n(\beta_0) \rightarrow 0$ in probability;

(iii) $n \hat{W}_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \hat{R}_i(\beta) \hat{R}_i^T(\beta)$ converges in probability to a constant matrix $W(\beta)$ uniformly for $\beta \in \mathcal{B}$;

(iv) $n \hat{W}_n(\beta_0)$ is positive definite with probability going to 1 as $n \rightarrow \infty$.

According to the first conclusion of Lemma 1 in Liu et al. (2010), under conditions (C2), (C4) and (C5), for $a = 0, 1, 2$, we can prove that in probability

$$\sup_{t \in [0, \tau], \beta \in \mathcal{B}} \| \hat{\psi}^{(a)}_{ik}(t; \beta) - \psi^{(a)}_{ik}(t; \beta) \| \rightarrow 0.$$

Since the kernel smoothed estimator $\hat{\lambda}_{0k}(t)$ in (2.7) is a consistent estimator for $\lambda_{0k}(t)$, then, by the definitions of $\hat{S}_k^d$ and $S_k^d$ and the remaining conditions, it follows that

$$\sup_{t \in [0, \tau], \beta \in \mathcal{B}} \| \hat{S}_k^d - S_k^d \| \rightarrow 0,$$

in probability for $d = 0, 1, \ldots, 5$.

Denote

$$\hat{U}(\beta, B) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} \hat{R}_i^T(t; \beta) \hat{Z}_{i1/2}(t; \beta) B \hat{Z}_{i-1/2}(t; \beta) \hat{M}_i(dt; \beta),$$

then $\hat{G}_n(\beta) = (\hat{U}^T(\beta, B_1), \ldots, \hat{U}^T(\beta, B_L))^T$.

After simple algebraic manipulations, we obtain that

$$\hat{U}(\beta, B) = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_{0}^{\tau} \hat{\xi}_{ik}(t; \beta, B) dN_{ik}(t)$$
\[
\begin{align*}
&= \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_{0}^{\tau} \tilde{\xi}_{ik}(t; \beta, B) dM_{ik}(t) \\
&\quad + \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_{0}^{\tau} \tilde{\xi}_{ik}(t; \beta, B) Y_{ik}(t) r_{ik}(t; \beta) \lambda_{0k}(t) dt,
\end{align*}
\]

where \( \tilde{\xi}_{ik}(t; \beta, B) = \sum_{m=1}^{K} \hat{R}_{im}(t; \beta) \hat{\phi}_{imk}(t; \beta, B) - \hat{S}_{k}(t; \beta)/\hat{S}_{k}(t; \beta) \), and

\[
\frac{\partial \hat{U}(\beta, B)}{\partial \beta^{T}} = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_{0}^{\tau} \frac{\partial \hat{\xi}_{ik}(t; \beta, B)}{\partial \beta^{T}} dM_{ik}(t) \\
&\quad + \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} \int_{0}^{\tau} \frac{\partial \hat{\xi}_{ik}(t; \beta, B)}{\partial \beta^{T}} Y_{ik}(t) r_{ik}(t; \beta) \lambda_{0k}(t) dt,
\] (7.3)

with

\[
\frac{\partial \hat{\xi}_{ik}(t; \beta, B)}{\partial \beta^{T}} = \sum_{m=1}^{K} \frac{\partial \hat{R}_{im}(t; \beta) \hat{\phi}_{imk}(t; \beta, B)}{\partial \beta^{T}} + \frac{\hat{S}_{k}(t; \beta, B)(\hat{S}_{k}(t; \beta))^{T}}{(\hat{S}_{k}(t; \beta))^{2}} \\
&\quad - \frac{\hat{S}_{k}(t; \beta, B) + \hat{S}_{k}(t; \beta, B) + \hat{S}_{k}(t; \beta, B)}{\hat{S}_{k}(t; \beta)}.
\]

Clearly, \( \frac{\partial \hat{U}(\beta, B)}{\partial \beta^{T}} \) is continuous. For any constant matrix \( B \), the first term on the right hand side of (7.3) is a local square integrable martingale, hence by Lenglart inequality, we can show that it converges to zero in probability, uniformly for \( \beta \in \mathcal{B} \). Let \( \Gamma(\beta, B) \) denote the uniform convergence limit of the second term, we can show that

\[
\frac{\partial \hat{U}(\beta, B)}{\partial \beta^{T}} \to \Gamma(\beta, B) = - \sum_{k=1}^{K} \int_{0}^{\tau} \psi_{k}(t; \beta, B) s_{k}^{0}(t; \beta) \lambda_{0k}(t) dt
\]

in probability, uniformly in \( \beta \in \mathcal{B} \). Thus, \( \partial G_{n}(\beta)/\partial \beta^{T} \) converges to

\[
\begin{bmatrix}
\Gamma(\beta, B_{1}) \\
\vdots \\
\Gamma(\beta, B_{L})
\end{bmatrix}
\]

in probability, uniformly for \( \beta \in \mathcal{B} \). Therefore, (i) is satisfied.

Denote

\[
\hat{U}_{1}(\beta, B) = \frac{1}{n} \sum_{k=1}^{K} \sum_{i=1}^{n} \int_{0}^{\tau} \hat{\xi}_{ik}(t; \beta, B) dM_{ik}(t),
\]
and

$$\hat{U}_2(\beta, B) = \frac{1}{n} \sum_{k=1}^{K} \sum_{i=1}^{n} \int_0^\tau \hat{\xi}_{ik}(t; \beta, B) Y_{ik}(t)r_{ik}(t; \beta)\lambda_0(t)dt.$$  

When $\beta = \beta_0$, by Lenglart inequality, we can show that

$$\sqrt{n} \hat{U}_1(\beta_0, B) = \frac{1}{\sqrt{n}} \sum_{k=1}^{K} \sum_{i=1}^{n} \int_0^\tau \xi_{ik}(t; \beta_0, B)dM_{ik}(t) + o_p(1),$$

where $\xi_{ik}(t; \beta, B) = \sum_{m=1}^{K} R_{im}(t; \beta)\phi_{im}(t; \beta, B) - e_k(t; \beta, B)$. From the arguments in Zhou and Wang (2000), it can be shown that

$$\sqrt{n} \hat{U}_2(\beta_0, B) = -\frac{1}{\sqrt{n}} \sum_{k=1}^{K} \sum_{i=1}^{n} \frac{n-n_k}{n_k} F_{ik}(\beta_0) + o_p(1),$$

where

$$F_{ik}(\beta) = \int_0^\tau \xi_{ik}(t; \beta, B)(r_{ik}(t; \beta) - \psi_{ik}(t; \beta))Y_{ik}(t)\lambda_0(t)dt.$$  

Note that $F_{ik}(\beta) = 0$ if $i \in \mathcal{V}_k$. Then we have

$$\sqrt{n} \hat{U}(\beta_0, B) = \frac{1}{\sqrt{n}} \sum_{k=1}^{K} \sum_{i=1}^{n} \int_0^\tau \xi_{ik}(t; \beta_0, B)dM_{ik}(t)$$

$$-\frac{1}{\sqrt{n}} \sum_{k=1}^{K} \sum_{i=1}^{n} \frac{n-n_k}{n_k} F_{ik}(\beta_0) + o_p(1),$$

where $M_{ik}(t)$ is a mean-zero martingale and $E(F_{ik}(\beta_0)) = 0$, then by strong law of large numbers, we have $\hat{U}(\beta_0, B)$ converges to zero with probability 1, thus (ii) is satisfied.

Denote

$$\hat{U}(\beta, B) = \sum_{i=1}^{n} \sum_{k=1}^{K} \hat{U}_{ik}(\beta, B),$$

then we can obtain

$$\hat{g}_i(\beta) = (n \sum_{k=1}^{K} \hat{U}_{ik}^T(\beta, B_1), n \sum_{k=1}^{K} \hat{U}_{ik}^T(\beta, B_2), \ldots, n \sum_{k=1}^{K} \hat{U}_{ik}^T(\beta, B_L))^T,$$
and

\[
\hat{W}_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \tilde{g}_i(\beta) \tilde{g}_i^T(\beta)
\]

\[
= \begin{pmatrix}
O(\beta, B_1, B_1) & \cdots & O(\beta, B_1, B_L) \\
\vdots & \ddots & \vdots \\
O(\beta, B_L, B_1) & \cdots & O(\beta, B_L, B_L)
\end{pmatrix},
\]

where

\[
O(\beta, B_j, B_{j'}) = \frac{1}{n} \sum_{i=1}^{n} K \sum_{k=1}^{K} \sum_{l=1}^{K} \tilde{U}_{ik}(\beta, B_j) \tilde{U}_{il}^T(\beta, B_{j'})
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} K \sum_{k=1}^{K} \sum_{l=1}^{K} \left\{ \int_{0}^{\tau} \xi_{ik}(t; \beta, B_j) dM_{ik}(t) \right\} \left\{ \int_{0}^{\tau} \xi_{il}(t; \beta, B_{j'}) dM_{il}(t) \right\}^T + o_p(1).
\]

Similarly as in Xue et al. (2010), by condition (C8), we can show that \( O(\beta, B_j, B_{j'}) \) converges in probability to \( \omega(\beta, B_j, B_{j'}) \), and \( n^\frac{1}{2} \hat{W}_n(\beta) \) converges in probability to \( W(\beta) = [\omega(\beta, B_j, B_{j'})]_{j, j'=1}^{L} \) uniformly for \( \beta \in \mathcal{B} \), hence (iii) and (iv) are satisfied.

Asymptotic normality

Define \( D_i(B) = \sum_{k=1}^{K} \int_{0}^{\tau} \xi_{ik}(t; \beta_0, B) dM_{ik}(t) \), from the previous statements, for any constant matrix \( B \), we can obtain that \( \sqrt{n} \tilde{U}(\beta_0, B) \) is asymptotically equivalent to \( n^{-1/2} \sum_{i=1}^{n} D_i(B) \), which is a sum of independent \( p \)-vector random variables with mean zero and variance \( \text{var}(D_i(B)) \). By condition (C8) and the multivariate central limit theorem, we can show that \( \sqrt{n} \tilde{U}(\beta_0, B) \) converges in distribution to a normal random vector with mean zero and covariance matrix \( \omega(\beta_0, B, B) \), and \( \sqrt{n} \tilde{G}_n(\beta_0) \) converges in distribution to a normal random vector, denoted as \( \mathcal{X} \), with mean zero and covariance matrix \( W_0 \) as defined in (A.2).

By Taylor expansion of \( \hat{Q}_n^{(1)}(\beta) \) around the true parameter \( \beta_0 \), we have

\[
\sqrt{n}(\hat{\beta}_Q - \beta_0) = -[n^{-1} \hat{Q}_n^{(2)}(\beta^*)]^{-1}[n^{-1/2} \hat{Q}_n^{(1)}(\beta_0)],
\]

where \( \beta^* \) is between \( \beta_0 \) and \( \hat{\beta}_Q \). From the argument in the proof of Theorem 1 and the consistency of \( \hat{\beta}_Q \), \( n^{-1} \hat{Q}_n^{(2)}(\beta^*) \) and \( n^{-1/2} \hat{Q}_n^{(1)}(\beta_0) \) converge to \( 2J_0^T W_0^{-1} J_0 \) and \( 2J_0^T W_0^{-1} \mathcal{X} \) in probability, respectively. Hence, we can obtain that in distribution

\[
\sqrt{n}(\hat{\beta}_Q - \beta_0) \rightarrow N(0, (J_0^T W_0^{-1} J_0)^{-1}).
\]

\( \Box \)
References

Breslow NE (1972) Discussion of the paper by D. R. Cox. J R Stat Soc Ser B 34:216–217
Cai JW, Prentice RL (1995) Estimating equations for hazard ratio parameters based on correlated failure time data. Biometrika 82:151–164
Cai JW, Prentice RL (1997) Regression estimation using multivariate failure time data and a common baseline hazard function model. Lifetime Data Anal 3:197–213
Cai JW, Shen Y (2000) Permutation tests for comparing marginal survival functions with clustered failure time data. Stat Med 19:2963–2973
Clayton D, Cuzick J (1985) Multivariate generalizations of the proportional hazard model. J R Stat Soc Ser A 148:82–117
Fan Z, Wang X (2009) Marginal hazards model for multivariate failure time data with auxiliary covariates. J Nonparametr Stat 21:771–786
Greene WF, Cai JW (2004) Measurement error in covariates in the marginal hazards model for multivariate failure time data. Biometrics 60:987–996
Hu C, Lin DY (2004) Semiparametric failure time regression with replicates of mismeasured covariates. J Am Stat Assoc 99:105–118
Liu Y, Zhou H, Cai JW (2009) Estimated pseudopartial-likelihood method for correlated failure time data with auxiliary covariates. Biometrics 65:1184–1193
Liu Y, Wu Y, Zhou H (2010) Multivariate failure times regression with a continuous auxiliary covariate. J Multivar Anal 101:679–691
Liu Y, Yuan Z, Cai JW, Zhou H (2012) Marginal hazard regression for correlated failure time data with auxiliary covariates. Lifetime Data Anal 18:116–138
Qu A, Lindsay BG, Li B (2000) Improving generalised estimating equations using quadratic inference functions. Biometrika 87:823–836
SOLVD Investigators (1991) Effect of enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure. N Engl J Med 325:293–302
Wei LJ, Lin DY, Weissfeld L (1989) Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. J Am Stat Assoc 84:1065–1073
Xue L, Wang L, Qu A (2010) Incorporating correlation for multivariate failure time data when cluster size is large. Biometrics 66:393–404
Zhou H, Wang C-Y (2000) Failure time regression with continuous covariates measured with error. J R Stat Soc Ser B 62:657–665

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.