A Multiple Imputation Approach to Regression Analysis for Doubly Censored Data with Application to AIDS Studies

Wei Pan

Division of Biostatistics
School of Public Health
University of Minnesota
A460 Mayo Building (MMC 303)
Minneapolis, MN 55455-0378, USA

Http: //www.biostat.umn.edu/~weip
Email: weip@biostat.umn.edu
Phone: (612)626-2705, Fax: (612)626-0660

November 2000; Revised August 2001
A Multiple Imputation Approach to Regression Analysis for Doubly Censored Data with Application to AIDS Studies

Summary

Sun, Liao and Pagano (1999, Biometrics 55, 909-914) proposed an interesting estimating equation approach to Cox regression with doubly censored data. Here we point out that a modification of their proposal leads to a multiple imputation approach, where the double censoring is reduced to single censoring by imputing for the censored initiating times. For each imputed data set one can take advantage of many existing techniques and software for singly censored data. Under the general framework of multiple imputation, the proposed method is simple to implement, and can accommodate modeling issues such as model checking, which has not been adequately discussed previously in the literature for doubly censored data. Here we illustrate our method with an application to a formal goodness-of-fit test and a graphical check for the proportional hazards model for doubly censored data. We reanalyze a well-known AIDS data set.

Key words: AIDS; Data augmentation; Goodness-of-fit test; HIV; Interval censoring; NPMLE; Proportional hazards model.
1. Introduction

Several authors (e.g., Goggins, Finkelstein and Zaslavsky 1999; Sun, Liao and Pagano 1999, and references therein) have considered Cox regression with doubly censored data. As an extensively discussed example in the literature, doubly censored data can arise from acquired immune deficiency syndrome (AIDS) studies. In these studies, it is of interest to investigate the distribution of the AIDS incubation time, the time between the infection of human immunodeficiency virus (HIV) and the onset of AIDS. The problem is complicated by double censoring: both the HIV infection time and AIDS onset time can be censored. Due to the nature of the periodic screening the HIV infection time is often interval censored: the exact time of infection is unknown; it is only known to have happened between two examination times, the last negative test and the first positive test. Similarly the AIDS onset time may also be interval censored. As in Sun et al and Goggins et al, we only consider the situation where the AIDS onset time is possibly right censored.

2. Estimating Equations for Doubly Censored Data

Suppose that there are $n$ patients in the study. For the $i$th patient, denote his/her HIV infection time and AIDS onset time as $X_i$ and $T_i^a$ respectively. The study goal is to investigate the relation between the incubation time $S_i = T_i^a - X_i$ and some covariates $Z_i$, a $p$-vector. $T_i^a$ may be right-censored: there is a censoring variable $C_i$ such that we only observe $T_i = \min(T_i^a, C_i)$ and $\delta_i = I(T_i = T_i^a)$. First consider the simplified situation where $X_i$'s are observed. We adopt the Cox proportional hazards model (PHM), which specifies that the hazard function for patient $i$ with a given covariate vector $Z_i$ is

$$\lambda(t|Z_i) = \lambda_0(t) \exp(\beta'Z_i),$$

where $\lambda_0$ is the unknown baseline hazard function and $\beta = (\beta_1, \ldots, \beta_p)'$ is the vector of regression coefficients to be estimated. Using counting process notation, we define
Let \[ Y_i(t | X_i) = I(T_i - X_i \geq t) \text{ and } N_i(t | X_i) = I(T_i - X_i \leq t, \delta_i = 1). \] Let \( X = (X_1, ..., X_n) \) and

\[
S^{(j)}(\beta, t | X) = n^{-1} \sum_{i=1}^{n} Y_i(t | X_i) \exp(\beta^t Z_i) Z_i^{(j)}
\]

for \( j = 0, 1, 2 \), where for a column vector \( a \), \( a^{(2)} = ad^t \), \( a^{(1)} = a \) and \( a^{(0)} \) is the scalar 1. \( S^{(j)}(\beta, t | X) \) are related to the first and second order derivatives of Cox’s partial likelihood and their use is only to simplify notation. Under the PHM, we can estimate \( \beta \) by maximizing the partial likelihood, which is equivalent to solving the following score equation (i.e. the first derivative of the partial likelihood):

\[
U(\beta | X) = \int_0^\tau \frac{\sum_{i=1}^{n} \left\{ Z_i - \frac{S^{(1)}(\beta, t | X)}{S^{(0)}(\beta, t | X)} \right\} dN_i(t | X_i)}{= 0,}
\]

(1)

where \( \tau \) is the longest possible follow-up time. The covariance matrix of the resulting maximum partial likelihood estimate \( \hat{\beta} = \hat{\beta}(X) \) can be consistently estimated by the inverse of the Fisher information matrix \( I = I(\hat{\beta} | X) = -\partial U(\beta | X)/\partial \beta |_{\beta = \hat{\beta}} \).

Of course, due to interval censoring, we do not observe \( X_i \) directly, but only know that \( X_i \) lies in a time interval \([L_i, R_i]\). Supposing that the distribution of \( X_i \) is \( H \), Sun et al proposed to estimate \( \beta \) by solving the following estimating equation (see their equation (3)):

\[
U(\beta | D) := E_{X|D} U(\beta | X) = 0,
\]

(2)

where the expectation \( E_{X|D} \) is taken with respect to the conditional distribution of \( X \) given the observed data \( D = \{ [L_i, R_i], i = 1, ..., n \} \). For non-trivial data sets, Monte Carlo methods can be applied to solve (2). Suppose that \( X_{(1)}, ..., X_{(M)} \) are \( M \) sets of independent realizations from the conditional distribution of \( X \) given \( D \), then (2) can be approximated by

\[
\frac{1}{M} \sum_{m=1}^{M} U(\beta | X_{(m)}) = 0
\]

(3)

In practice, \( H \) is unknown and we replace it with an estimate \( \hat{H} \), such as the nonparametric maximum likelihood estimate (NPMLE) (Turnbull 1976). The NPMLE \( \hat{H} \) maximizes
the nonparametric likelihood \( l(H|D) = \prod_{i=1}^{n}[H(R_i) - H(L_i)] \) with respect to \( H \) under the constraint that \( H \) is a proper cumulative distribution function. Turnbull (1976) described an EM-type algorithm to compute the NPMLE iteratively. We use the NPMLE \( \hat{H} \) throughout the paper. Then for each \( m = 1, \ldots, M \), the imputed HIV infection time for patient \( i \), \( X_{(m)i} \), is drawn independently from \( \hat{H} \) conditional on \( L_i \leq X_{(m)i} \leq R_i \). In the following, we propose a modified approach using \( X_{(m)i} \)'s and show its advantages.

3. A Multiple Imputation Approach

3.1 Estimating regression coefficients

One needs to write special computer programs to solve (3) to estimate \( \beta \). Rather than that, we propose to solve \( U(\beta|X_{(m)}) = 0 \) to obtain \( \hat{\beta}_{(m)} \) for each \( m = 1, \ldots, M \). Then the final estimate is \( \bar{\beta} = \sum_{m=1}^{M} \hat{\beta}_{(m)}/M \). It can be shown that under suitable conditions \( \bar{\beta} \) is asymptotically equivalent to the estimator obtained by solving (3) (see also Lemma 1 of Wang and Robins, 1998).

The idea behind using the estimator \( \bar{\beta} \) is exactly that of Rubin’s multiple imputation (MI) (Rubin 1987). One attractive property of this approach is its ability to take advantage of many existing techniques and software for right censored data. For instance, each \( U(\beta|X_{(m)}) = 0 \), a standard partial likelihood score equation, can be solved directly in many statistical packages. In addition, MI also provides a simple formula to estimate the covariance matrix of \( \bar{\beta} \) (Rubin 1987). Denote the covariance estimate of \( \hat{\beta}_{(m)} \) as \( \mathcal{I}_{-1}^{(m)} = [-\partial U(\beta|X_{(m)})/\partial \beta^t|_{\beta=\hat{\beta}_{(m)}}]^{-1} \), then

\[
\tilde{\text{Cov}}(\bar{\beta}) = \frac{\sum_{m=1}^{M} \mathcal{I}_{-1}^{(m)}}{M} + \left(1 + \frac{1}{M}\right) \frac{\sum_{m=1}^{M}(\hat{\beta}_{(m)} - \bar{\beta})(\hat{\beta}_{(m)} - \bar{\beta})'}{M-1},
\] (4)

which is a sum of the within-imputation covariance (the first term) and the between-imputation covariance (the second term). The second term is inflated by a factor \( 1/M \) to take account of the finite number of imputations. Statistical inferences on the true \( \beta \)
are accomplished by using the approximately normal distribution of $\hat{\beta}$, $N(\beta, \text{Cov}(\hat{\beta}))$. (For small $M$, $t$-based inference may be preferred; see Rubin 1987.) Note that $M = 1$ leads to single imputation, and the formula (4) cannot be used to estimate the between-imputation covariance matrix. It is well known that single imputation usually leads to the underestimation of the variability (e.g. Rubin 1987).

3.2 Model checking for PHM

An attractive point of our proposed MI approach is its ability to take advantage of existing techniques for right-censored data. Here we demonstrate this point by applying it to check the proportional hazards assumption with doubly censored data, which has never been discussed in the literature. Both a formal goodness-of-fit (GOF) test and a graphical method will be described.

First we consider the simplified situation where $X$ is observed. In this situation we have only right censored data, for which there is a large literature on GOF tests for the PHM (e.g. Pettitt and Daud 1990; Grambsch and Therneau 1994 and references therein). Although in principle our method can be applied to many other GOF tests, here we restrict our discussion to a general class of GOF tests proposed by Grambsch and Therneau (1994), which includes many existing tests as special cases and can be adapted as a graphical method (see also Pettitt and Daud 1990). Its implementation in Splus and SAS facilitates its routine use in practice. The method is based on using time-varying regression coefficients as a general alternative to the PHM:

$$\beta_j(t) = \beta_j + \theta_j g_j(t),$$

where $g_j(t)$ is a specified non-zero predictable process, $j = 1, \ldots, p$. Testing the PHM is equivalent to testing $H_0$: $\beta_j(t) = \beta_j$ for all $j$. The PHM is violated if $\theta_j \neq 0$ for some $j$. To construct the test statistic, we need some notation. Let $Z_{(k)}$ be the covariate vector of the
subject with an event at time \( t_k = T_k^* - X_k \). Define

\[
M(\beta, t|X) = \frac{S^{(1)}(\beta, t|X)}{S^{(0)}(\beta, t|X)}, \quad V(\beta, t|X) = \frac{S^{(2)}(\beta, t|X)}{S^{(0)}(\beta, t|X)} - \left\{ \frac{S^{(1)}(\beta, t|X)}{S^{(0)}(\beta, t|X)} \right\}^2.
\]

The Schoenfeld (1982) residuals are

\[
r_k(\beta|X) = Z_{(k)} - M(\beta, t_k|X).
\]

Let \( G_k = G(t_k) \) be a diagonal matrix with the \((j, j)\)th element \( g_j(t_k) \). Denote \( \hat{r}_k = r_k(\hat{\beta}|X) \), \( \hat{V}_k = V(\hat{\beta}, t_k|X) \), where \( \hat{\beta} = \hat{\beta}(X) \) is the maximum partial likelihood estimator solving (1). Then Grambsch and Therneau proposed to use the test statistic (see their equations (10)–(11))

\[
R(X) = Q(X)' D^{-1} Q(X)
\]

(5)

with

\[
Q(X) = \sum G_k \hat{r}_k, \quad D(X) = \sum G_k \hat{V}_k G_k' - (\sum G_k \hat{V}_k)(\sum \hat{V}_k)^{-1}(\sum G_k \hat{V}_k)'.
\]

Under \( H_0 \), \( R(X) \) has an asymptotic chi-squared distribution with degrees of freedom \( p \).

Note that \( n^{-1} D(X) \) is a consistent covariance estimator of \( n^{-1/2} Q(X) \).

In the current context, based on each imputed \( X_{[m]} \), we can calculate the corresponding \( Q(X_{[m]}) \) and its covariance estimate \( D(X_{[m]}) \) for any \( m = 1, ..., M \). Then using Rubin’s formula we have

\[
\bar{Q} = \frac{\sum_{m=1}^{M} Q(X_{[m]})}{M},
\]

\[
\bar{Cov}(\bar{Q}) = \frac{\sum_{m=1}^{M} D(X_{[m]})}{M} + \left( 1 + \frac{1}{M} \right) \frac{\sum_{m=1}^{M} [Q(X_{[m]}) - \bar{Q}][Q(X_{[m]}) - \bar{Q}]'}{M - 1}.
\]

(6)

Hence the goodness-of-fit test statistic

\[
R(D) = \bar{Q}' \bar{Cov}(\bar{Q})^{-1} \bar{Q}
\]

(7)
will have an approximately chi-squared distribution with degrees of freedom $p$ under $H_0$.

In addition to the above formal test, visual inspection of plots for some nonparametric estimates of the time-varying regression coefficients $\beta_j(t)$ has proved to be useful as a graphical tool to check the PHM for right-censored data. Under the PHM, each $\beta_j(t)$ should be a constant. The functional forms of $\beta_j(t)$’s in these plots provide evidence for or against the proportional hazards assumption. Furthermore, once a violation of the PHM is suspected, the functional forms of $\beta_j(t)$’s are also suggestive to alternative models. In the same article, Grambsch and Therneau (1994) also described how to obtain nonparametric estimates of $\beta_j(t)$ for right-censored data, which are based on smoothing appropriately rescaled Schoenfeld residuals. In our setting, for each imputed data set, this graphical method can be directly applied. Examination of these plots based on multiply imputed data sets serves as a diagnostic tool to check the PHM. We will show its application in the example.

In S-Plus, for right censored data, PHMs can be fitted using the function `coxph()`, and Grambsch and Therneau’s GOF test and plots of nonparametric estimates of $\beta_j(t)$’s are all implemented in the function `cox.zph()`. Using some components of the returning values from the above two S-Plus functions, it is easy to construct our proposed test statistic $R(D)$. Several transformation functions $g_j(t)$ are available in `cox.zph()`. To take full advantage of existing S-Plus functions and to minimize the programming effort, we implemented our proposal in S-Plus.

4. Remarks

Sun et al proved that their estimate of $\beta$ is consistent and asymptotically normal, and our estimate is asymptotically equivalent to theirs. The relative simplicity of Sun et al’s approach is largely facilitated by a simplified imputation scheme: the HIV infection time $X$
is imputed based only on the estimated conditional distribution of $X$ given $D$. This approach can be summarized by two separate estimation steps. First, the distribution function $H$ of $X$ is consistently estimated by the NPMLE $\hat{H}$, using only the HIV infection times $D$ but ignoring any relevant information contained in the AIDS onset times. Second, use $\hat{H}$ to impute $X$ and thus estimate the regression coefficients $\beta$. Our approach bears this same spirit. It is reminiscent of pseudo maximum likelihood estimation (Gong and Samaniego 1981). In contrast, other previous approaches estimate $H$ and $\beta$ jointly, using the joint distribution of the HIV infection time and AIDS onset time, which is complex and does not have a closed form. An interesting example is Goggins et al’s Monte Carlo EM algorithm approach. Their estimating equation has the same form as (3). However, in their approach, $X$ is imputed based on the joint likelihood of the HIV infection time and AIDS onset time, which depends on the unknown $H$ and $\beta$. Hence, the step of imputing $X$ and the step of estimating the parameters based on the estimating equation need to be iterated. Although more complicated, Goggins et al’s approach is likely to be more efficient asymptotically than Sun et al’s approach. The reason is that Goggins et al’s approach yields the joint MLEs of both $H$ and $\beta$, which is not the case for the latter approach. Of course, the amount of asymptotic efficiency gain depends on many factors (e.g., the joint distribution of the HIV infection time and AIDS onset time) and needs to be further investigated. In the AIDS example to be discussed, we find that the results from two approaches are very close.

Another potential issue in Sun et al and our modification here is that $\hat{H}$, the estimated distribution of the HIV infection time $X$, is treated as if known when we impute $X$. Thus this imputation method is an example of the poor man’s data augmentation (PMDA) scheme (Wei and Tanner 1990). Intuitively, due to its failure to take account of the uncertainty in estimating $H$, the PMDA may underestimate the true variability of the resulting parameter estimates. An alternative is to use the Approximate Bayesian Bootstrap (ABB) (Rubin
It works as follows: during each imputation \( m \), we first take a bootstrap sample \( D_{(m)} \) from \( D \), then the NPMLE (or any other reasonable estimate) \( \hat{H}_{(m)} \) of \( H \) is obtained based on \( D_{(m)} \). Finally the imputed \( X_{(m)} \) is taken from \( \hat{H}_{(m)} \) (conditional on \( D \)). The remaining steps are the same as before. In our simulations, the performance of using ABB is close to that of PMDA, possibly due to the relatively light censoring. In practice, if it is suspected that the variation of \( \hat{H} \) may heavily influence the conclusion, or just to be conservative, we recommend using the ABB (see also Pan 2000a).

Note that some simple imputation methods of using the left-, right- or mid-point of the censoring interval to impute the HIV infection time have been used in the literature (e.g., Munoz and Xu 1996). Here we use a more sophisticated data augmentation method based on \( \hat{H} \). The reason is that the simple methods may lead to biased estimates (e.g., Goggins et al 1999).

5. Simulation

A small simulation study was conducted to evaluate the finite-sample performance of our proposed method. Our simulation set-up is similar to that in Goggins et al. A binary covariate was used to indicate group membership. The sample size for each group was 100. For both groups, the HIV infection time was simulated from a Lognormal distribution \( LN(3.8, 0.3) \). To mimic screening studies, we simulated a patient’s first visit as a random number from a uniform distribution \( U(0, 5) \). After the first visit, each patient is scheduled to have annual follow-ups. Whether one completes each annual follow-up is an independent Bernoulli variable. A patient’s HIV infection time is accordingly censored between two consecutive visits. Two probabilities of making an annual visit, 0.5 and 0.3, were used to result in moderate and heavy interval censoring with average censoring interval widths of 2.9 and 5.5 years respectively. The incubation time was simulated from two scaled
Weibull distributions with the same shape parameter for the two groups, \( W(2.5, 70.1) \) and \( W(2.5, 60) \), resulting a PHM with \( \beta = 0.389 \). The resulting AIDS onset time was subject to about 10% random right-censoring.

The results are presented in Table 1. To facilitate comparison, we also include the results based on the usual Cox regression with right censored data (i.e. when the HIV infection time is exactly known). Both PMDA and ABB were used to impute. It can be seen that their performances are close and satisfactory. It seems that all the estimates of \( \beta \) and their standard errors are essentially unbiased, leading to the coverage percentage of the 95% confidence interval (CI) close to the nominal level. The power for the 5% level normal-based test for \( H_0: \beta = 0 \) vs \( H_1: \beta \neq 0 \) is also reasonable. The size of the 5%-level GOF test for the PHM is within the nominal level.

It has been noted that the distribution of AIDS incubation time can be better modeled using a Lognormal than using a Weibull distribution (Munoz and Xu 1996). Although our simulation study was conducted under the Weibull distribution for the AIDS incubation time, we do not expect that our result will change much if a Lognormal distribution is used. The reason is that our method is semi-parametric without any fully distributional assumption on the AIDS incubation time.

6. An Example

Many authors have considered a cohort study of hemophiliacs who were treated in two hospitals in France (Kim et al 1993; Sun et al 1999; Goggins et al 1999, and references therein). The patients of the study were at risk for HIV infection through contaminated blood factor. At the end of study, there were 188 patients found to be HIV-infected, but the infection times were interval censored. Among them 41 progressed to AIDS (or related
symptoms). The data were presented in Kim et al and Sun et al. The patients were classified into either the heavily treated group or lightly treated group according to the amount of blood received (when treated for hemophilia). The goal here is to investigate the possible association between the treatment and the AIDS incubation time. We code the covariate $Z_i = 0$ or 1 if the $i$th patient was lightly or heavily treated. We first consider using our proposed PMDA to impute for interval-censored HIV infection times. With $M = 1000$, we obtained $\hat{\beta} = 0.712$ with estimated standard error 0.331, close to those (0.69 and 0.34 respectively) obtained by Kim et al and Goggins et al, and those (0.7039 and 0.2833) by Sun et al. In summary, it is confirmed that the heavily treated group had a significantly higher risk of the onset of AIDS after HIV infection.

To investigate how $\hat{\beta}$ depends on the number of imputations $M$, we tried various values of $M$ (Table 2). It can be seen that our $\beta$ estimate is relatively stable even with an $M$ as small as 5. This is probably related to the small between- and within-imputation variance ratio, which is only around 1%.

For the goodness-of-fit test, we used the S-Plus default transformation $g_j(t)$ based on the Kaplan-Meier estimator. The test statistics and $p$-values based on various numbers of imputations are presented in Table 2. Here due to the larger between- and within-imputation variance ratio (20%), there is some small variation in the obtained $p$-values with different $M$. But all $p$-values are larger than 0.2, indicating no strong evidence against the adequacy of the Cox PHM. This point can be also confirmed from Figure 1, where the estimated time-varying regression coefficients $\tilde{\beta}(t)$ based on six imputed data sets are plotted. Recall that the proportional hazards assumption is equivalent to a constant (i.e. non-time-varying) $\beta(t) \equiv \beta$. Hence, if $\tilde{\beta}(t)$ changes largely with the time $t$, it will cast some doubt on the adequacy of the PHM. Since in each panel of Figure 1, the 95% point-
wise confidence interval (almost) covers a constant horizontal zone, it provides no strong
evidence against the proportional hazards assumption. The points in each plot are the
rescaled Schoenfeld residuals. By smoothing these residuals one obtains $\hat{\beta}(t)$. Since each
panel is based on an imputed data set, a random realization consistent with the observed,
an overall interpretation of the result should be based on summarizing the information
present in all the panels. Hence, it would be helpful to combine the estimated time-varying
regression coefficients and their standard errors using the MI formula given in (4), and then
present an overall estimate and its confidence interval. Unfortunately, the current S-Plus
function \texttt{cox.zph()} does not provide $\hat{\beta}(t)$ and its standard error function directly, though
the functions can be plotted out. Of course, in principle, this proposal can be implemented.

The above analysis is based on using the PMDA. The results based on using the ABB
are similar. For instance, in comparison with Table 2, the regression coefficient estimate $\tilde{\beta}$
(SE) and the GOF statistic $R$ ($p$-value) are respectively $.711 (.331)$ and $1.4161 (.2341)$ for
$M = 5$, and $.709 (.331)$ and $1.2086 (.2716)$ for $M = 500$.

Finally we comment on the choice of the imputation number $M$. As shown in previous
studies (e.g. Rubin 1987), a small $M$ often suffices. In general, the choice of $M$ depends
on the trade-off between the computational cost and the loss of estimation efficiency. It
also depends on the severeness of the missingness (or censoring). With current computing
power, it is not really an issue to use a large $M$. It is also informative to look at the between-
and within-imputation variance ratio. If the ratio is large, then a large $M$ may be needed.
In general, based on our and others' experience, it seems natural to try several values of $M$
and see whether the results are close or not. If not, then a large $M$ should be used.
7. Discussion

We have shown that a modification of Sun et al’s method leads to the MI approach to doubly censored data. The advantage of the MI approach is its potential to take full use of existing methods and software for right or interval censored data. With doubly censored data, after imputing for the censored initiating times (i.e. HIV infection times in the AIDS example), we only have singly censored data, to which many existing methods can be applied. For instance, we have illustrated how to apply the goodness-of-fit test for right censored data to doubly censored data. Furthermore, if the AIDS onset time is also interval censored, after imputing for the interval-censored HIV infection time, we can apply any of the existing methods (e.g. Finkelstein 1986; Satten 1996; Kooperberg and Clarkson 1997) to the resulting singly interval-censored data. As demonstrated in other relevant applications (e.g. Taylor et al 1990; Betensky and Finkelstein 1999; Bebchuk and Betensky 2000; Pan 2000a, 2000b), we believe that the MI provides a simple and general approach for doubly censored data and other complex incomplete data (e.g. Alioum and Commenges 1996).

ACKNOWLEDGEMENT

The author is grateful to Drs Chap Le and Patricia Grambsch for helpful discussions on the subject, to Drs Bill Goggins and Dianne Finkelstein for help with the simulation set-up, and to Dr Lynn Eberly for reading an earlier draft. The author thanks a referee, an AE and the editor for constructive comments.
References


*Biometrics* **56**, 199-203.


Estimating the distribution of times from HIV seroconversion to AIDS using multiple imputation.  

Censored and Truncated Data.  

*Biometrika* **85**, 935-948.

Table 1: Simulation results with $n = 200$ and $\beta = 0.389$ based on 500 independent replications. We used $M = 5$ imputations. $r$ is the mean percentage of the between-imputation SE in the total SE of $\hat{\beta}$.

<table>
<thead>
<tr>
<th>HIV infection times $X$</th>
<th>uncensored</th>
<th>moderately censored</th>
<th>heavily censored</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PMDA</td>
<td>ABB</td>
<td>PMDA</td>
</tr>
<tr>
<td>Mean of $\hat{\beta}$</td>
<td>.3842</td>
<td>.3794</td>
<td>.3790</td>
</tr>
<tr>
<td>$SD(\hat{\beta})$</td>
<td>.1556</td>
<td>.1559</td>
<td>.1557</td>
</tr>
<tr>
<td>Mean $SE(\hat{\beta})$</td>
<td>.1525</td>
<td>.1527</td>
<td>.1528</td>
</tr>
<tr>
<td>$r$</td>
<td>-</td>
<td>.048</td>
<td>.054</td>
</tr>
<tr>
<td>Coverage of 95% CI</td>
<td>.942</td>
<td>.938</td>
<td>.940</td>
</tr>
<tr>
<td>Estimated power</td>
<td>.706</td>
<td>.702</td>
<td>.700</td>
</tr>
<tr>
<td>Size of GOF test</td>
<td>.040</td>
<td>.038</td>
<td>.036</td>
</tr>
</tbody>
</table>

Table 2: The estimate $\tilde{\beta}$ and its standard error $SE$, and the goodness-of-fit test statistic $R$ and $p$-value, with various numbers of imputations $M$.

<table>
<thead>
<tr>
<th>PMDA $M$</th>
<th>Estimate $\tilde{\beta}$</th>
<th>SE $SE$</th>
<th>GOF test $R$ Statistic $R$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>.716</td>
<td>.331</td>
<td>1.2285</td>
<td>.2677</td>
</tr>
<tr>
<td>10</td>
<td>.710</td>
<td>.330</td>
<td>0.9662</td>
<td>.3256</td>
</tr>
<tr>
<td>20</td>
<td>.716</td>
<td>.331</td>
<td>1.3171</td>
<td>.2511</td>
</tr>
<tr>
<td>100</td>
<td>.719</td>
<td>.331</td>
<td>1.5439</td>
<td>.2140</td>
</tr>
<tr>
<td>500</td>
<td>.711</td>
<td>.331</td>
<td>1.3732</td>
<td>.2413</td>
</tr>
<tr>
<td>1000</td>
<td>.712</td>
<td>.331</td>
<td>1.4529</td>
<td>.2281</td>
</tr>
</tbody>
</table>
Figure 1: Estimates (solid lines) of the time-varying regression coefficients and their 95% point-wise confidence intervals (dotted lines) based on six imputed data sets for the AIDS example.