Published in final edited form as:

Comput Stat Data Anal. 2010 July 1; 54(7): 1817–1823. doi:10.1016/j.csda.2010.01.035.

Regression analysis of clustered interval-censored failure time data with informative cluster size

Xinyan Zhang* and Jianguo Sun

Department of Statistics, University of Missouri, Columbia, MO 65211, USA

Abstract

Correlated or clustered failure time data often occur in medical studies, among other fields (Cai and Prentice, 1995; Kalbfleisch and Prentice, 2002), and sometimes such data arise together with interval censoring (Wang et al., 2006). Furthermore, the failure time of interest may be related to the cluster size. For example, Williamson et al. (2008) discussed such an example arising from a lymphatic filariasis study. A simple and common approach to the analysis of these data is to simplify or convert interval-censored data to right-censored data due to the lack of proper inference procedures for direct analysis of these data. In this paper, two procedures are presented for regression analysis of clustered failure time data that allow both interval censoring and informative cluster size. Simulation studies are conducted to evaluate the presented approaches and they are applied to a motivating example.

Keywords

Interval censoring; Informative cluster size; Weibull model; Within-cluster resampling

1. Introduction

Correlated or clustered failure time data occur in many fields such as medical studies (Cai and Prentice, 1995; Kalbfleisch and Prentice, 2002). In many cases the failure times of interest may not be observed exactly but are known only to belong to certain intervals. Such data are usually referred to as interval-censored failure time data, and they could arise naturally in, for example, periodic follow-up studies where each study subject is observed only at discrete time points (Finkelstein and Wolfe, 1986; Sun, 2006; Wang et al., 2006). In this paper, we will discuss regression analysis of clustered interval-censored data where the failure times of interest may be related to the cluster size. In other words, the cluster size may be informative for the failure times of interest and thus obviously needs to be taken into account in the analysis.

An example of clustered interval-censored failure time data with informative cluster size, which motivated this study, can be found in a lymphatic filariasis study discussed by Williamson et al. (2008) among others. The goal of the study was to compare the effect of

co-administration of diethylcarbamazine and albendazole (DEC/ALB) (new treatment) versus DEC alone (standard treatment) for the treatment of lymphatic filariasis (LF). In men, the problem is often characterized by one or more nests of adult filarial worms in the scrotum and the characteristic movement of the living adult worms can be detected by ultrasound. The effective treatment is expected to kill the worms in all of the nests. This study followed 47 men for a year since their treatment and they were periodically examined by ultrasound to see if the worms were still alive. For the times to the clearance of the worms in each nest, the variable of interest, only clustered interval-censored data are available with each study subject as a cluster and the cluster size being the number of nests of adult filarial worms in each subject. To give an idea about the correlation between the time to the clearance and the number of nests, Table 1, reproduced from Williamson et al. (2008), gives the average percentages of nests cleared during the one year follow-up for all study subjects. One can easily see that it took a longer time to clear the worms in men with more nests, suggesting that the time to clearance and the number of nests are related. Studies on all teeth of the same person give another common example of clustered failure time data which can easily become clustered interval-censored data with informative cluster size. In the dental study for a patient with chronic periodontitis, the effects of cigarette smoking and hygiene status may predict tooth survival. However, people found that the patients with more teeth have higher molar survival probability. As a result, cluster size is informative to tooth survival.

A number of methods have been proposed for the analysis of clustered right-censored failure time data assuming that the cluster size is noninformative (Cai and Prentice, 1995; Clayton and Cuzick, 1985; Hougaard, 1987, 2000). In the presence of informative cluster size, among others, Dunson et al. (2003) gave a Bayesian procedure that models the relationship between the failure times of interest and the cluster size through a latent variable. Williamson et al. (2008) and Cong et al. (2007) also considered the same problem and investigated a weighted score function approach and a within-cluster resampling (WCR) procedure. The former is essentially a weighted estimating equation approach, while the latter, following Hoffman et al. (2001), samples a single subject from each cluster and transforms the data to usual univariate failure time data. While it is commonly used in clustered data analysis and avoids the correlation problem between subjects within the same cluster, the WCR approach is usually computationally intensive. In the following, we will generalize these two approaches to clustered interval-censored data under the parametric model framework.

The rest of the article is organized as follows. In Section 2, we will first introduce the notation and model and then present two procedures for regression analysis of clustered interval-censored data. The first one bases the estimation on a weighted estimating equation and the other makes use of the WCR approach. Following Williamson et al. (2008), we will assume that the failure times of interest have Weibull distributions. It is straightforward to generalize the methodology to other parametric models. Section 3 gives some results from a simulation study conducted for evaluating the performance of the presented methods, which indicate that they perform well for practical situations. The methodology is applied to the motivating lymphatic filariasis data discussed above in Section 4, and Section 5 contains some discussion and concluding remarks.

2. Inference procedures

Consider a survival study that includes N clusters of subjects. Let T_{ij} denote the failure time of interest for subject j in the i th cluster with an associated vector of covariates X_{ij} , $i = 1, \dots, N, j = 1, \dots, n_i$, where n_i is the cluster size for the i th cluster. In general, the covariates may include two parts, one being cluster-specific and the other being subject-specific. It will be assumed that the T_{ij} 's may be dependent for the subjects within the same cluster but are independent for subjects from different clusters. In this paper, we will assume that T_{ij} follows the Weibull distribution with the survival function given by

$$S(t; X_{ij}) = \exp\left(-e^{\beta' X_{ij}} t^\gamma\right), \quad (1)$$

where β denotes the vector of regression parameters and γ the shape parameter. It is straightforward to generalize the inference procedures given below to other parametric models. In the following, we will assume that the failure times of interest, the T_{ij} 's, may be related to the cluster sizes, n_i 's, and the main goal is to estimate β and γ . Nonignorable cluster size can be defined as any violation of the property that $E(T_{ij}|n_i, X_{ij}) = E(T_{ij}|X_{ij})$. Nonignorability occurs when the cluster size is related to the risk for the outcome of interest.

For estimation of β and γ , suppose that only interval-censored data are available. Specifically, let $(L_{ij}, R_{ij}]$ denote the observed interval for T_{ij} . That is, $T_{ij} \in (L_{ij}, R_{ij}]$ and the observed data are $\{(L_{ij}, R_{ij}], X_{ij}; i = 1, \dots, N, j = 1, \dots, n_i\}$. It will be assumed that T_{ij} , L_{ij} and R_{ij} are independent given X_{ij} following Sun (2006). First we will present an estimating equation-based estimation procedure and then describe a WCR-based procedure.

2.1. An estimating equation-based procedure

To present the procedure, note that under model (1), the likelihood contribution from the j th subject in cluster i has the form

$$\exp\left(-e^{\beta' X_{ij}} L_{ij}^\gamma\right) - \exp\left(-e^{\beta' X_{ij}} R_{ij}^\gamma\right).$$

If the data is right-censored, this contribution would be $\exp\left(-e^{\beta' X_{ij}} C_{ij}^\gamma\right)$, where C_{ij} denotes the observed censoring time.

If assuming that the T_{ij} 's are independent for the subjects within the same cluster, one can obtain the pseudo-log-likelihood function as

$$l(\beta, \gamma) = \sum_{i=1}^N \sum_{j=1}^{n_i} \log \left\{ \exp\left(-e^{\beta' X_{ij}} L_{ij}^\gamma\right) - \exp\left(-e^{\beta' X_{ij}} R_{ij}^\gamma\right) \right\}. \quad (2)$$

Then it is natural to estimate β and γ by their values that maximize $l(\beta, \gamma)$ or solving the resulting score functions. The same idea has been used for the analysis of multivariate

failure time data and is often referred to as the working independence assumption (Wang et al., 2006).

Let $\Phi_{ij} = \exp(\beta' X_{ij})$ and ω_i be the weight for cluster i . Motivated by the pseudo-log-likelihood function given in (2), we propose to estimate β and γ by $\hat{\beta}_e$ and $\hat{\gamma}_e$ defined as the solutions to the equations

$$U_{\beta}(\hat{\beta}, \hat{\gamma}) = \sum_{i=1}^N \sum_{j=1}^{n_i} U_{ij}(\hat{\beta}; \hat{\gamma}) = 0, \quad U_{\gamma}(\hat{\beta}, \hat{\gamma}) = \sum_{i=1}^N \sum_{j=1}^{n_i} U_{ij}(\hat{\gamma}; \hat{\beta}) = 0,$$

where

$$U_{ij}(\beta; \gamma) = \omega_i \frac{\exp(-\Phi_{ij} R_{ij}^{\gamma}) (\Phi_{ij} R_{ij}^{\gamma} X_{ij}) - \exp(-\Phi_{ij} L_{ij}^{\gamma}) (\Phi_{ij} L_{ij}^{\gamma} X_{ij})}{\exp(-\Phi_{ij} L_{ij}^{\gamma}) - \exp(-\Phi_{ij} R_{ij}^{\gamma})}$$

and

$$U_{ij}(\gamma; \beta) = \omega_i \frac{\exp(-\Phi_{ij} R_{ij}^{\gamma}) (\Phi_{ij} R_{ij}^{\gamma} \log R_{ij}) - \exp(-\Phi_{ij} L_{ij}^{\gamma}) (\Phi_{ij} L_{ij}^{\gamma} \log L_{ij})}{\exp(-\Phi_{ij} L_{ij}^{\gamma}) - \exp(-\Phi_{ij} R_{ij}^{\gamma})}.$$

It is easy to see that $U_{\beta}(\beta, \gamma)$ and $U_{\gamma}(\beta, \gamma)$ are the weighted score functions from the pseudo-log-likelihood function $l(\beta, \gamma)$. With respect to the selection of the weights ω_i 's, the simplest and most commonly used one is $\omega_i = n_i^{-1}$. Some other choices are discussed in the simulation study.

Let $U(\beta, \gamma) = (U'_{\beta}(\beta, \gamma), U'_{\gamma}(\beta, \gamma))'$. To solve the equations $U(\beta, \gamma) = 0$, one can use the Newton–Raphson algorithm. By applying the estimating equation theory, one can show that $\hat{\beta}_e$ and $\hat{\gamma}_e$ are consistent and their joint distribution can be approximated by the multivariate normal distribution with the covariance matrix estimated by

$$\Sigma_e(\hat{\beta}_e, \hat{\gamma}_e) = H^{-1}(\hat{\beta}_e, \hat{\gamma}_e) V(\hat{\beta}_e, \hat{\gamma}_e) H^{-1}(\hat{\beta}_e, \hat{\gamma}_e),$$

where $H(\beta, \gamma) = U(\beta, \gamma) / (\beta, \gamma)$ and

$$V_e(\beta, \gamma) = \sum_{i=1}^N \begin{bmatrix} \sum_{j=1}^{n_i} U_{ij}(\beta; \gamma) & \sum_{j=1}^{n_i} U'_{ij}(\beta; \gamma) & \sum_{j=1}^{n_i} U_{ij}(\beta; \gamma) & \sum_{j=1}^{n_i} U_{ij}(\gamma; \beta) \\ \sum_{j=1}^{n_i} U_{ij}(\gamma; \beta) & \sum_{j=1}^{n_i} U'_{ij}(\beta; \gamma) & \left\{ \sum_{j=1}^{n_i} U_{ij}(\gamma; \beta) \right\}^2 \end{bmatrix}.$$

2.2. A WCR-based procedure

One observation is randomly sampled with replacement from each of the N clusters. The resampled dataset of size N can be analyzed by a generalized linear model since the observations in the resampled data are independent. If this process is repeated a large number of times, Q , say, where each of the Q analyses provides a consistent estimator of parameter of interest, we can estimate the within-cluster resampling estimates by taking the average of the Q resample-based estimates.

To present the WCR-based estimation procedure, first note that if $n_i = 1$ for all i , $l(\beta, \gamma)$ becomes the full likelihood and the estimates $\hat{\beta}_e$ and $\hat{\gamma}_e$ are the maximum likelihood estimates. To make use of this feature, one can apply the idea behind WCR to estimate β and γ as follows.

Let Q be an integer. For each $q = 1, \dots, Q$, let $\{(L_i^q, R_i^q), X_i^q\}$ denote a random sample from $\{(L_{ij}, R_{ij}), X_{ij}; i = 1, \dots, N, j = 1, \dots, n_i\}$, $\hat{\beta}_e^q$ and $\hat{\gamma}_e^q$ the estimates $\hat{\beta}_e$ and $\hat{\gamma}_e$ defined above or the maximum likelihood estimates of β and γ based on the data $\{(L_i^q, R_i^q), X_i^q; i = 1, \dots, N\}$. Then we can define the WCR-based estimates of β and γ as

$$\hat{\beta}_w = \frac{1}{Q} \sum_{q=1}^Q \hat{\beta}_e^q, \quad \hat{\gamma}_w = \frac{1}{Q} \sum_{q=1}^Q \hat{\gamma}_e^q.$$

Although the determination of the estimates may require more computation time depending on N and Q , the implementation of the procedure is straightforward.

A major feature of the estimation procedure given above is that it avoids the correlation issue among the subjects within the same cluster. Also, as mentioned above, all individual estimates $\hat{\beta}_e^q$ and $\hat{\gamma}_e^q$ are efficient. It can easily be shown that $\hat{\beta}_w$ and $\hat{\gamma}_w$ are consistent and one can approximate their joint distribution by the multivariate normal distribution with the covariance matrix estimated by

$$\frac{1}{Q} \sum_{q=1}^Q \Sigma_e(\hat{\beta}_e^q, \hat{\gamma}_e^q) - \frac{1}{Q-1} \sum_{q=1}^Q \begin{pmatrix} \hat{\beta}_e^q - \hat{\beta}_w \\ \hat{\gamma}_e^q - \hat{\gamma}_w \end{pmatrix} \begin{pmatrix} \hat{\beta}_e^{q'} - \hat{\beta}_w' & \hat{\gamma}_e^{q'} - \hat{\gamma}_w' \end{pmatrix}. \quad (3)$$

For estimation of the covariance matrix of $\hat{\beta}_w$ and $\hat{\gamma}_w$, as an alternative, one can replace $\Sigma_e(\hat{\beta}_e^q, \hat{\gamma}_e^q)$ in (3) by the inverse of the observed Fisher information matrix from $l(\beta, \gamma)$ calculated at $\beta = \hat{\beta}_e^q$ and $\gamma = \hat{\gamma}_e^q$.

3. A simulation study

To assess the performance of the inference procedures presented in the previous section, a simulation study was conducted with the focus on the estimation of regression parameters. In the study, we assumed that there existed two covariates X_{1i} and X_{2ij} with X_{1i} being the cluster-specific covariate taking value 0 or 1 with probability 0.5 and X_{2ij} being the subject-

specific covariate generated from the uniform distribution over $(-0.5, 0.5)$. To generate the clustered failure times (the T_{ij} 's), we first generated the latent variables (the W_i 's) from the positive stable distribution (Wei et al., 1989) by $W_i = (\xi_i / E_i)^{\frac{1-\alpha}{\alpha}}$, $i = 1, \dots, N$. Here E_i is a random number from the exponential distribution with mean one and

$$\xi_i = \left[\frac{\sin(\alpha U_i)}{\sin(U_i)} \right]^{\frac{1}{1-\alpha}} \frac{\sin[(1-\alpha)U_i]}{\sin(\alpha U_i)}$$

with α being an unknown parameter and U_i being equal to π times a random number from the uniform distribution over $(0, 1)$. The parameter α represents the correlation among the failure times within a cluster with $\alpha \rightarrow 1$ corresponding to the independent case and $\alpha \rightarrow 0$ meaning completely determined by each other. Given the W_i 's, the failure times T_{ij} 's were assumed to follow the conditional Weibull distribution with the survival function

$$S(t_{ij}; X_{ij}, W_i, \beta^*, \gamma^*) = \exp\left(-W_i e^{\beta_1^* X_{1i} + \beta_2^* X_{2ij}} t_{ij}^{\gamma^*}\right), \quad (4)$$

where $\beta^* = (\beta_1^*, \beta_2^*)'$ and γ^* are some unknown parameters.

Specifically, to generate the T_{ij} 's, we first generated the informative cluster size n_i according to W_i . If W_i was less than or equal to the median of the positive stable distribution, n_i was assumed to follow the binomial distribution $B(7, 0.75)$ and otherwise to follow the binomial distribution $B(7, 0.25)$. In both cases, the sizes 0 and 7 were discarded. Given W_i and n_i , $T_{i1}, \dots, T_{in_i}$ were assumed to be independent and generated from (4). The parameter values were specified as $(\beta_1^*, \beta_2^*, \gamma^*) = (1.0, 1.0, 2.0)$ in the conditional Weibull distribution. Then one can easily show that the marginal survival function of the T_{ij} 's is given by (1) with $\beta_1 = \alpha\beta_1^*$, $\beta_2 = \alpha\beta_2^*$ and $\gamma = \alpha\gamma^*$. For the generation of censoring intervals $(L_{ij}, R_{ij}]$, we mimic the medical follow-up studies and assumed that each subject was supposed to be examined eight times (Pan, 2000). The initial time was randomly generated and the length between the examination times was assumed to be constant. Furthermore, it was assumed that a subject may miss the scheduled examination times with probability 0.3 for the first four time points and 0.5 for the last four time points. Then the observed interval for T_{ij} was given by the two examination times right before and after T_{ij} . Here, we use the convention that $L_{ij} = R_{ij}$ means that we have an exact observation on the j th failure time of the i th subject and $R_{ij} = \infty$ means that the observation on T_{ij} is right-censored, and $L_{ij} = 0$ means the observation on T_{ij} is left-censored.

Tables 2 and 3 present the simulation results on the estimation of β_1 , β_2 and γ with their true values being $(0.2, 0.2, 0.4)$, $(0.5, 0.5, 1.0)$ and $(0.8, 0.8, 1.6)$, respectively. Table 2 is for the case of $N = 50$ to mimic the lymphatic filariasis study and Table 3 is for the case of $N = 500$. For all results, 2500 replications were used and the tables include the means of the estimates given in the previous section (MEAN), the averages of the estimated standard errors of the estimates (ESE), and the sample standard deviations of the estimates based on the simulated data (STD). In addition to the two estimates of β_1 , β_2 and γ proposed in Section 2, for comparison, we also considered the estimates based on the estimating equations $U(\beta, \gamma) = 0$,

letting $\omega_i = 1$. The resulting estimates are referred to as the unweighted score estimates (UWSE) in the tables, while the two estimates proposed in Section 2 are referred to as the weighted score estimates (WSE) and the WCR estimates (WCRE), respectively. For the WCRE, the results given here are based on $Q = 1000$.

It can be seen from the tables that both the WSE and the WCRE of the regression parameters seem to be reasonably unbiased and the bias becomes smaller as the number of clusters increases, as expected. Also the variance estimates given for the two estimates seem to perform reasonably well too as they are close to the sample variances. In contrast, the simulation results clearly showed that the UWSE is biased. That is, in the presence of informative cluster sizes, one would get biased results if not taking into account the informative cluster size in the analysis.

In addition to letting $\omega_i = n_i^{-1}$, we also investigated some other weights such as $\frac{1}{\sqrt{n_i}}$ and $\text{var}(n_i)$, and the simulation results showed that n_i^{-1} is optimal. Choosing the inverse of the cluster size for the weight results in unbiased estimation if the goal of the analysis is to estimate the marginal effects weighted at the cluster level. However, this might not be true if the goal is the estimation of the marginal effects weighted at the unit level. However, these approaches may not necessarily work for a full likelihood model of clustered survival data, and the inverse of the cluster size may not be optimal in that case.

4. Analysis of the lymphatic filariasis study

In this section, we apply the inference procedures proposed in Section 2 to the LF study discussed before on examining the effect of the co-administration of DEC and ALB against DEC alone for the LF treatment. ALB is an anti-parasitic drug, manufactured by GlaxoSmithKline, and it is commonly used to treat interstima worm infections. When co-administered with DEC, it helps break to the cycle of LF transmission between mosquitoes and humans, and by using ultrasound, the doctor can detect the movement of the living adult worms. As mentioned above, the study consists of 47 men followed for one year and periodically examined by ultrasound to determine whether any of the worms were still alive in the nests. The variable of interest is the nest-specific time from the treatment administration to the clearance of the worms: it took much longer to clear a nest in men with multiple nests than in men with fewer nests. This indicates that the cluster size may be informative.

Among 47 study subjects, 22 received the co-administration of DEC and ALB, while the others were given DEC alone. In total, 78 adult worm nests were detected by ultrasound, and the cluster size n_i ranges from 1 to 5, as seen in Table 1. Two covariates were of particular interest to the investigators. One is of course the treatment indicator and the other is the age of the subject in years, ranging from 16 to 66.

For subject i , define X_{1i} to be 0 if subject i was given the co-administration of DEC and ALB and 1 otherwise, and let X_{2i} be the age, $i = 1, \dots, 47$. Note that here we only have cluster-specific covariates. The application of the two estimation procedures proposed above gave the estimated effects in Table 4 of both treatment and age on the time to the clearance

of the worms along with the p -values for testing the effects equal to zero. The results suggest that there is essentially no difference on adding ALB to DEC for the LF treatment, and it seems that it took longer time for old patients to have their worms cleared than for young patients. By treating the observed interval-censored data as right-censored data and using the same model, a similar result is obtained in Williamson et al. (2008) for the treatment effect, but their result indicated that the age had no significant effect on the clearance time. To give a graphical idea about the age effect, Fig. 1 presents the predicted survival curves for two persons receiving both DEC and ALB treatments together at ages of 20 and 25, approximately 25 and 75 percentiles of the observed ages, respectively.

Note that, in Table 4, we have presented the estimated treatment and age effects given by WCRE using several different values for Q . As can be seen, the estimates are quite robust for Q greater than 100. We did the same in the simulation study and got similar results.

5. Discussion and concluding remarks

This paper has discussed the regression analysis of interval-censored clustered failure time data in the presence of informative cluster sizes for which no appropriate approach seems to have been available previously. To take into account the informative cluster sizes, two estimation procedures were presented under the parametric framework with one being based on a weighted pseudo score function and the other making use of the WCR idea. The simulation study showed that both procedures work reasonably well and one would obtain biased results if ignoring the informative cluster size. The approaches presented were applied to a motivating lymphatic filariasis study and they indicated that the worm clearance time may be related to the patient age.

It is worth noting that both estimation procedures presented here are marginal procedures as they rely only on the marginal model (1). As an alternative, sometimes one may want to specify a full model for the correlated or clustered failure times, and in this case, the estimation can be performed by using the maximum likelihood approach. An advantage of the full model approach is that it would allow estimation of the relationship among the clustered failure times, which would not be possible by using the marginal model approaches presented in the previous sections. Of course, a difficult issue related to the full model approach is the specification and verification of the model.

In the discussion so far, we have assumed that the cluster size may be related to the failure times of interest but censoring intervals are independent of the failure times of interest given covariates. Sometimes one may face situations where the censoring intervals may also be related to the failure times of interest even conditional on the covariates, and this can happen in, for example, medical follow-up studies. Some methods have been developed for the analysis of such failure time data for the case where the cluster size is equal to one or usual interval-censored failure time data (Sun, 2006). However, established procedures do not seem to exist for usual clustered interval-censored data.

References

- Cai J, Prentice RL. Estimating equations for hazard ratio parameters based on correlated failure time data. *Biometrics*. 1995; 82:151–164.
- Clayton DG, Cuzick J. Multivariate generations of the proportional hazards model. *Journal of the Royal Statistical Society, Series A*. 1985; 148:82–117.
- Cong X, Yin G, Shen Y. Marginal analysis of correlated failure time data with informative cluster size. *Biometrics*. 2007; 63:663–672. [PubMed: 17825000]
- Dunson DB, Chen Z, Harry J. Bayesian joint models of cluster size and subunit-specific outcomes. *Biometrics*. 2003; 59:521–530. [PubMed: 14601753]
- Finkelstein DM, Wolfe RA. Isotonic regression for interval censored survival data using an E-M algorithm. *Communications in Statistics: Theory and Methods*. 1986; 15:2493–2505.
- Hoffman EB, Sen PK, Weinberg CR. Within cluster resampling. *Biometrika*. 2001; 88:1121–1134.
- Hougaard P. Modelling multivariate survival. *Scandinavian Journal of Statistics*. 1987; 14:291–304.
- Hougaard, P. *Analysis of Multivariate Survival Data*. Springer-Verlag; New York: 2000.
- Kalbfleisch, JD.; Prentice, RL. *The Statistical Analysis of Failure Time Data*. second ed.. John Wiley; New York: 2002.
- Pan W. A multiple imputation approach to Cox regression with interval-censored data. *Biometrics*. 2000; 56:199–203. [PubMed: 10783796]
- Sun, J. *The Statistical Analysis of Interval-censored Failure Time Data*. Springer; New York: 2006.
- Wang L, Sun L, Sun J. A Goodness-of-fit test for the marginal Cox model for correlated interval-censored failure time data. *Biometrical Journal*. 2006; 5:1–9.
- Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariable incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association*. 1989; 84:1065–1073.
- Williamson J, Kim HY, Manathuga A, Addiss DG. Modeling survival data with informative cluster size. *Statistics in Medicine*. 2008; 27:543–555. [PubMed: 17640035]

Fig. 1.
Predicted survival functions for persons given both DEC and ALB together.

Table 1

Percentages of nests cleared during 360 days in the lymphatic filariasis study.

Number of nests	Percentage cleared
1	81.8
2	62.5
3	50
4 or 5	33.3

Table 2Estimation of parameters with size $N = 50$.

Parameter		WSE	UWSE	WCRE
$\beta_1 = 0.2$	MEAN	0.218	-0.148	0.204
	ESE	0.224	0.233	0.190
	STD	0.231	0.230	0.222
$\beta_2 = 0.2$	MEAN	0.210	0.252	0.213
	ESE	0.347	0.271	0.315
	STD	0.333	0.315	0.399
$\gamma = 0.4$	MEAN	0.426	0.588	0.404
	ESE	0.118	0.099	0.112
	STD	0.114	0.150	0.123
$\beta_1 = 0.5$	MEAN	0.515	0.284	0.500
	ESE	0.173	0.181	0.167
	STD	0.183	0.212	0.188
$\beta_2 = 0.5$	MEAN	0.513	0.583	0.531
	ESE	0.319	0.287	0.351
	STD	0.320	0.324	0.362
$\gamma = 1.0$	MEAN	1.048	1.231	1.002
	ESE	0.196	0.171	0.144
	STD	0.209	0.214	0.172
$\beta_1 = 0.8$	MEAN	0.816	0.704	0.816
	ESE	0.150	0.155	0.175
	STD	0.154	0.164	0.205
$\beta_2 = 0.8$	MEAN	0.813	0.892	0.804
	ESE	0.360	0.330	0.303
	STD	0.361	0.360	0.354
$\gamma = 1.6$	MEAN	1.616	1.803	1.674
	ESE	0.276	0.240	0.303
	STD	0.280	0.288	0.380

Table 3

Estimation of parameters with size $N = 500$.

Parameter		WSE	UWSE	WCRE
$\beta_1 = 0.2$	MEAN	0.202	-0.147	0.216
	ESE	0.069	0.072	0.071
	STD	0.069	0.088	0.084
$\beta_2 = 0.2$	MEAN	0.199	0.244	0.197
	ESE	0.103	0.087	0.092
	STD	0.101	0.099	0.104
$\gamma = 0.4$	MEAN	0.414	0.556	0.406
	ESE	0.032	0.030	0.043
	STD	0.032	0.048	0.040
$\beta_1 = 0.5$	MEAN	0.501	0.265	0.515
	ESE	0.055	0.058	0.052
	STD	0.055	0.062	0.051
$\beta_2 = 0.5$	MEAN	0.501	0.566	0.486
	ESE	0.099	0.090	0.094
	STD	0.101	0.094	0.113
$\gamma = 1.0$	MEAN	1.003	1.189	1.056
	ESE	0.062	0.053	0.045
	STD	0.063	0.062	0.060
$\beta_1 = 0.8$	MEAN	0.802	0.773	0.806
	ESE	0.047	0.049	0.051
	STD	0.047	0.047	0.062
$\beta_2 = 0.8$	MEAN	0.801	0.855	0.816
	ESE	0.107	0.100	0.110
	STD	0.113	0.101	0.074
$\gamma = 1.6$	MEAN	1.602	1.721	1.599
	ESE	0.084	0.071	0.088
	STD	0.085	0.079	0.094

Table 4

Estimates of covariate effects for the lymphatic filariasis study.

Covariate		WSE	WCRE ($Q = 100$)	WCRE ($Q = 1000$)	WCRE ($Q = 2000$)
Group (DEC/ALB = 0, DEC = 1)	Estimate	-0.267	-0.267	-0.265	-0.264
	<i>p</i> -value	0.277	0.294	0.301	0.303
Age (years)	Estimate	-0.088	-0.088	-0.088	-0.088
	<i>p</i> -value	<0.001	<0.001	<0.001	<0.001