Research Article

Multiple imputation based on restricted mean model for censored data

Lyrica Xiaohong Liu, Susan Murray*† and Alex Tsodikov

Most multiple imputation (MI) methods for censored survival data either ignore patient characteristics when imputing a likely event time, or place quite restrictive modeling assumptions on the survival distributions used for imputation. In this research, we propose a robust MI approach that directly imputes restricted lifetimes over the study period based on a model of the mean restricted life as a linear function of covariates. This method has the advantages of retaining patient characteristics when making imputation choices through the restricted mean parameters and does not make assumptions on the shapes of hazards or survival functions. Simulation results show that our method outperforms its closest competitor for modeling restricted mean lifetimes in terms of bias and efficiency in both independent censoring and dependent censoring scenarios. Survival estimates of restricted lifetime model parameters and marginal survival estimates regain much of the precision lost due to censoring. The proposed method is also much less subject to dependent censoring bias captured by covariates in the restricted mean model. This particular feature is observed in a full statistical analysis conducted in the context of the International Breast Cancer Study Group Ludwig Trial V using the proposed methodology. Copyright © 2011 John Wiley & Sons, Ltd.

Keywords: multiple imputation; restricted mean lifetime; survival; censoring

1. Introduction

In survival analysis, estimation of expected life over a fixed time window is often of interest, either non-parametrically or as a function of covariates. In addition, it is common to desire estimates of survival probabilities within particular subgroups. For example, in the International Breast Cancer Study Group (IBCSG) Ludwig Trial V, investigators would like to estimate a long-duration treatment effect on patient lifetimes over the 9-year study period, adjusting for tumor size, estrogen receptor (ER) status, number of positive nodes and age. They also want to compare marginal survival curves for the two treatment groups. The presence of right censoring makes standard analysis methods for fully observed data inappropriate, although they would be much simpler to implement if available.

We propose that since restricted means are of interest and may be modeled already as part of a thorough analysis of the IBCSG study, that we take advantage of the restricted mean model structure to augment censored outcomes via multiple imputation (MI). The resulting final analyses (regression parameters, estimated restricted means and non-parametric quantities) are based on more standard analytical tools using multiply imputed data sets and are hypothesized to be more efficient since imputes better utilize covariate information.

Several researchers have given attention to modeling censored restricted lifetimes as a function of patient characteristics. Karrison [1] used a generalized Cox model approach [2] with piecewise constant baseline hazards, and made appropriate transformations to the restricted mean scale that indirectly linked covariates to the restricted means. Extensions of this approach, also centered around a Cox model, were given by Chen and Tsiatis [3] and Zucker [4]. Andersen *et al.* [5] link covariate effects with the mean restricted lifetimes by using pseudo observations (POs) *in lieu* of the original outcomes

Department of Biostatistics, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109, U.S.A.
Correspondence to: *Dr. Susan Murray, Department of Biostatistics, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109, U.S.A.*
†*E-mail: skmurray@umich.edu*

and applying generalized linear models (GLM) [6]. None of these authors suggested the restricted mean as a tool for imputation of censored lifetimes.

Meanwhile, increasingly researchers have come to view censored data in the more traditional role of missing data, where MI is a popular strategy for appropriately addressing missing information in an analysis. For example, Faucett *et al.* [7] multiply impute survival outcomes via joint modeling of a change-point model and a time-dependent Cox proportional hazards model. Taylor *et al.* [8] develop non-parametric MI methods that reproduce Kaplan–Meier [9] estimates when no covariate information is available. Hsu *et al.* [10] use Cox models to more selectively build risk sets of individuals with similar hazards for MI, utilizing a non-parametric imputation procedure within this risk set. The advantages of imputation are longstanding, because many different analyses may be conducted using the multiply imputed data sets once they are obtained. An overview of several effective imputation strategies based upon observed data is given in Rubin [11] and Little and Rubin [12]. Most existing MI methods either assume parametric models acting on (and linking) the hazards of interest or are non-parametric in nature.

Our goal in this research is to produce multiply imputed data sets that directly model the missing outcomes of interest via a restricted mean structure. The resulting multiply imputed data sets incorporate individual information to gain efficiency in restricted mean model parameter estimation as well as in other analyses of interest, such as survival curve estimation and two-sample testing. The remainder of the manuscript is structured as follows: in Section 2, we describe the mean structure for restricted lifetimes given covariates. Section 3 introduces the restricted mean lifetime-based MI algorithm with some technical details of implementation included in the Appendix. In Section 4, we summarize several commonly used standard analyses applied to multiply imputed data sets. Section 5 presents finite sample simulation results. We return to the IBCSG study in Section 6 and report various analyses of interest. Discussion follows in Section 7.

2. Structure for restricted mean lifetimes used in multiple imputation

Suppose lifetime, T, has survival function, $S_T(t)$, with mean life $E(T) = \int_0^\infty S_T(t)\,dt$. With right-censored data, the data tend to support only estimated lifetimes restricted to the study period, or restricted means, $E\{\min(\tau, T)\} = \int_0^\tau S_T(t)\,dt$, where τ is within the range of the observed data [13].

To examine the restricted mean as function of covariates, Z, regression models have been developed. For instance, one approach is to assume a transformation model taking the form, $g(T) = \beta^T Z + \varepsilon$, where $\beta = (\beta_{[0]}, \beta_{[1]}, \ldots, \beta_{[p]})$ is a $(p+1)$-dimensional vector, ε is the residual vector with mean zero and g is some link function (e.g. [14–16]). Fully parametric models can be implemented if the residual distribution is known. Andersen *et al.* [5] use POs to model this mean structure, with the added advantage that few assumptions are required on the distribution of ε for their model to hold.

In particular, for each individual, PO i $(i = 1, \ldots, n)$ is defined as:

$$n \int_0^\tau \widehat{S}^{\mathrm{KM}}(t)\,dt - (n-1) \int_0^\tau \widehat{S}^{\mathrm{KM}(-i)}(t)\,dt,$$

where $\widehat{S}^{\mathrm{KM}}(t)$ is the Kaplan–Meier (KM) estimate for survival and $\widehat{S}^{\mathrm{KM}(-i)}(t)$ is the KM estimate excluding patient i. These POs are comparable in expectation to the distribution of the original restricted failure times, similar to the jackknife. Hence, this modeling approach addresses the censoring issue through transformation to uncensored values with identical restricted mean regression parameters. Andersen *et al.* recommend GLM analysis on log-transformed POs using an identity link.

We assume a similar mean structure with the idea of imputing for $\log\min(\tau, T)$ rather than using log-transformed POs. That is:

$$E[\log\{\min(\tau, T)\}|Z] = \beta^T Z. \tag{1}$$

The log transformation of $\min(\tau, T)$ continues to ensure that regression parameters apply to the real line rather than merely to positive values. Also, transforming $\min(\tau, T)$ before model fitting seems to produce better estimates of the intercept than if a log link were applied, which is useful in the context of imputation. We suspect that this is the case due to Jensen's Inequality since $\log E\{\min(\tau, T)|Z\} \geqslant E[\log\{\min(\tau, T)\}|Z]$ and we impute on the scale of $\log\{\min(\tau, T)\}$. Standard linear models can be used to fit (1) once multiply imputed data sets are created.

3. Multiple imputation algorithm

With the mean structure in (1), we have a natural way to fill in missing event times during the study window. We achieve this goal by developing restricted mean MI algorithm. The algorithm has two parts: first, we obtain desired parameter estimates as in (1); second, we append appropriate residuals to the estimated means to form an impute that better approximate variability of the original data.

The proposed algorithm is summarized in Steps 1–4, with further details of implementation following:

Step 1: By fitting GLM model (1) treating censored data as failures, we obtain initial parameter estimates $\widehat{\beta}^{(0)}$. Next, we use a pseudo EM algorithm described in the Appendix to obtain a converged and improved $\widehat{\beta}$. This algorithm takes into account the current estimate of $\widehat{\beta}$, its variability and the observed censoring time C_i for each value requiring imputation.

Step 2: We form imputes for censored patients by adding error terms to the estimated means $\widehat{\beta}^T Z$, where $\widehat{\beta}$ is obtained in Step 1. For patients with similar $\widehat{\beta}^T Z$, observed residuals are sampled; the detailed sampling procedure is described in Appendix B. Sampled error terms are required to yield an impute larger than the original censored value.

Step 3: Repeat Step 2 until we have M imputes for each censored value.

Step 4: Combine analysis from M imputed data sets to get the final parameter estimates and the associated variances.

Next, we describe the details of the algorithm. Suppose $T_1, \ldots, T_n$ come from a non-negative random variable, T, with survival function, S_T, and $C_1, \ldots, C_n$ come from a random variable, C, that may or may not depend on covariates in model (1), but are otherwise independent of T. Let $X_i = \min(T_i, C_i)$, $i = 1, \ldots, n$ be the observed times to event. Let $Y = \log\{\min(\tau, T)\}$ with τ a fixed positive constant.

In Step 1, we fit (1) to obtain initial values $\widehat{\beta}^{(0)}$ treating all observed data as failures. In practice, we have not found the initial value to have much influence on final parameter estimates, although several alternative choices for obtaining $\widehat{\beta}^{(0)}$ were explored. Naturally the more censoring in the data, the further away $\widehat{\beta}^{(0)}$ is from the true β when censored values are treated as failures. The next part of the algorithm is an iterative procedure to obtain $\widehat{\beta}^{(0)}, \widehat{\beta}^{(1)}, \widehat{\beta}^{(2)}, \ldots, \widehat{\beta}^{(k)}$, where the procedure is said to converge when $\max_j(|\hat{\beta}_{[j]}^{(k)} - \hat{\beta}_{[j]}^{(k-1)}|) < a$ for some small tolerance a, with $\hat{\beta}_{[j]}^{(k)}$ being the jth element of the vector $\widehat{\beta}^{(k)}$. Steps of the iterative procedure are located in Appendix A.

In Step 2, with the converged parameter $\widehat{\beta}$, $\widehat{E}(Y_i|C_i, Z_i)$ is calculated as $\widehat{\beta}^T Z_i$ for a censored patient i, then residual errors are added to $\widehat{E}(Y_i|C_i, Z_i)$ to reproduce appropriate variability. Details of sampling residuals are in Appendix B. Our assumption for constructing residuals is that patients with similar $\widehat{E}(Y_i|Z_i)$ will have a similar distribution of residuals and can be used to create an appropriate pool for selection. Residuals that result in imputes of $\log\{\min(\tau, T)\} < \log C_i$ are removed from further consideration. In the case of models with discrete covariates only, this residual pool reduces to patients with covariates identical to Z_i, and the impute ($\widehat{\beta}^T Z_i +$ residual) will essentially select one of the failed patients' death times from the pool as the imputed value. For continuous covariates, we sample residuals from patients whose $\widehat{E}(Y|Z)$ fall within some small b-margin of the censored patient's $\widehat{E}(Y_i|Z_i)$, so that the impute does not necessarily match any observed failure time from the original data set, but is shifted higher or lower depending on $\widehat{\beta}^T Z_i$. In either case, when a patient from the residual pool is selected with event times $> \tau$, we use $\log\tau$ for the imputed value since a reasonable shift from $\widehat{\beta}^T Z_i$ is not available in this case.

Finally in Step 4, for each censored value, we sample M of those residuals and add them to $\widehat{E}(Y_i|Z_i)$ as described above, resulting in a total of M imputed data sets to be analyzed.

4. Analyze multiply imputed data sets

Since we fill in the missing outcomes for censored people, many research problems become complete data problems and we can apply standard procedures to analyze M imputed data sets. In practice, $M = 10$ multiply imputed data sets are usually sufficient. Next, we summarize some most commonly desired analyses.

4.1. Estimating regression parameters based on multiply imputed data sets

Using a standard GLM modeling approach, each of the M imputed data set yields estimates $\widehat{\beta}_m^{\text{MI}} = (\hat{\beta}_{m,[0]}^{\text{MI}}, \hat{\beta}_{m,[1]}^{\text{MI}}, \ldots, \hat{\beta}_{m,[p]}^{\text{MI}})$, $m = 1, \ldots, M$ under model (1). The final estimates based on the multiply imputed data sets are $\widehat{\beta}^{\text{MI}} = \sum_{m=1}^{M} \widehat{\beta}_m^{\text{MI}} / M$. The associated variances are composed of within imputation variances W and between imputation variances B, respectively [11, 12].

The variances become

$$\text{Var}(\widehat{\beta}^{\text{MI}}) = W + (1 + M^{-1})B,$$

where $W = \sum_{m=1}^{M} \text{Var}(\widehat{\beta}_m^{\text{MI}})/M$ and $B = \sum_{m=1}^{M} (\widehat{\beta}_m^{\text{MI}} - \widehat{\beta}^{\text{MI}})^2/(M-1)$.

Similarly, covariances between the jth and kth elements of $\widehat{\beta}^{\text{MI}} = (\hat{\beta}_{[0]}^{\text{MI}}, \hat{\beta}_{[1]}^{\text{MI}}, \ldots, \hat{\beta}_{[p]}^{\text{MI}})$ are calculated as in [11, 12]:

$$\text{Cov}(\hat{\beta}_{[j]}^{\text{MI}}, \hat{\beta}_{[k]}^{\text{MI}}) = \sum_{m=1}^{M} \text{Cov}\{\hat{\beta}_{m,[j]}^{\text{MI}}, \hat{\beta}_{m,[k]}^{\text{MI}}\}/M + (1 + M^{-1}) \sum_{m=1}^{M} [\{\hat{\beta}_{m,[j]}^{\text{MI}} - \hat{\beta}_{[j]}^{\text{MI}}\}\{\hat{\beta}_{m,[k]}^{\text{MI}} - \hat{\beta}_{[k]}^{\text{MI}}\}]/(M-1),$$

where $\hat{\beta}_{m,[j]}^{\text{MI}}$, $j = 0, 1, \ldots, p$, $m = 1, \ldots, M$ is the jth element of estimate $\widehat{\beta}_m^{\text{MI}}$ from mth data set.

The hypothesis test for $\widehat{\beta}^{\text{MI}}$ and significance level is determined by the t distribution: $(\widehat{\beta}^{\text{MI}} - \beta)\text{Var}(\widehat{\beta}^{\text{MI}})^{-1/2} \sim t_v$, where $v = \{1 + (M+1)^{-1}W/B\}^2(M-1)$ based on the Satterthwaite approximation [11, 12].

4.2. Survival curve estimates based on multiply imputed data sets

Using the M imputed data sets, we calculate the KM survival estimates $\widehat{S}_1^{\text{KM}}(t), \ldots, \widehat{S}_M^{\text{KM}}(t)$, and obtain associated variances $\widehat{V}_1(t), \ldots, \widehat{V}_M(t)$ based on Greenwood's formula. For complete data sets, the KM survival estimates reduce to simple sample proportions of those alive at times t. The combined survival estimate $\widehat{S}^{\text{MI}}(t)$ and $\widehat{V}^{\text{MI}}(t)$ are calculated to be

$$\widehat{S}^{\text{MI}}(t) = M^{-1} \sum_{m=1}^{M} \widehat{S}_m^{\text{KM}}(t),$$

$$\widehat{V}^{\text{MI}}(t) = M^{-1} \sum_{m=1}^{M} \widehat{V}_m(t) + (1 + M^{-1}) \sum_{m=1}^{M} \{\widehat{S}_m^{\text{KM}}(t) - \widehat{S}^{\text{MI}}(t)\}/(M-1).$$

4.3. Log-rank test based on multiply imputed data sets

For the mth imputed data set ($m = 1, \ldots, M$), let $T_1^m < \cdots < T_L^m$ denote the ordered failure times, D_{ik}^m and Y_{ik}^m denote the number of failures and number at risk for group i at time T_k^m, $i = 1, 2$, $k = 1, \ldots, L$. Furthermore, let D_k^m and Y_k^m denote the corresponding values in whole sample, $E_{1k}^m = D_k^m Y_{1k}^m / Y_k^m$ be expected failures in group 1 and $V_{1k}^m = D_k^m Y_{1k}^m Y_{2k}^m (Y_k^m - D_k^m)/(Y_k^m)^2(Y_k^m - 1)$. The log-rank statistic for the mth imputed data set is given by:

$$Q_m^{\text{MI}} = \left(\sum_{k=1}^{L} V_{1k}^m\right)^{-1/2} \sum_{k=1}^{L} (D_{1k}^m - E_{1k}^m).$$

The combined log-rank statistic and corresponding variance then become:

$$Q^{\text{MI}} = \sum_{m=1}^{M} Q_m^{\text{MI}}/M,$$

$$V^{\text{MI}} = 1 + (1 + M^{-1}) \sum_{m=1}^{M} (Q_m^{\text{MI}} - Q^{\text{MI}})/(M-1),$$

since Q_m^{MI} is asymptotic standard normal for large samples.

The hypothesis test for treatment difference and significance level is determined by the t distribution: $Q^{\text{MI}}/\sqrt{V^{\text{MI}}} \sim t_v = \{1 + (M+1)^{-1}/\sum_{m=1}^{M}(Q_m^{\text{MI}} - Q^{\text{MI}})\}^2(M-1)$ based on the Satterthwaite approximation [11, 12].

5. Simulation study

In this section, we study finite sample properties of selected analyses based on restricted mean MI data sets, including restricted mean regression parameter estimation and marginal survival curve estimation.

GLM parameter estimates for the restricted mean model are produced using: (a) model (1) with an identity link, where $\log\{\min(\tau, T)\}$ is multiply imputed for censored observations using the restricted mean MI approach, (b) model (1) with an identity link applied to log-transformed POs as in Andersen *et al.* [5] and (c), the model in (a) applied to the uncensored (fully observed) data.

We first study the independent censoring case. In each of 1000 simulations, we perform the following procedure with a sample of size $n = 100$ and τ fixed at 1.5:

Step 1: For covariates, we generate bivariate normal $(0, 1)$ pairs $(Z_{1i} = z_{1i}, Z_{2i} = z_{2i})$, $i = 1, \ldots, n$ with correlation 0.3. We then transform one of these into a Uniform$(0, 1)$ distributed covariate by applying the inverse transform method, that is, $U_i = P(Z_{1i} \leqslant z_{1i})$. The second normal is transformed into a Bernoulli(0.5) covariate, $B_i = I(Z_{2i} \geqslant 0)$.

Step 2: We obtain the outcome of interest, $\min(\tau, T_i)$, $i = 1, \ldots, n$. Each failure time T_i is simulated from an exponential distribution with hazard rate, λ_i, that satisfies the mean structure (1) for a pre-specified $\beta = (\beta_0, \beta_1, \beta_2)$, B_i, U_i and τ. That is, for this simulation setup

$$E(Y_i | C_i, B_i, U_i) = \int_{-\infty}^{+\infty} y \, \mathrm{d}F_{Y_i}(y)$$

$$= \int_{-\infty}^{\log \tau^-} y f_{T_i}(\mathrm{e}^y) \, \mathrm{d}y + P(Y_i \geqslant \log \tau) \times \log \tau$$

$$= \int_{-\infty}^{\log \tau^-} y \lambda_i \mathrm{e}^y \mathrm{e}^{-\lambda_i \mathrm{e}^y} \, \mathrm{d}y + \mathrm{e}^{-\lambda_i \tau} \times \log \tau.$$

Therefore, λ_i is a numerical solution to

$$\int_{-\infty}^{\log \tau^-} y \lambda_i \mathrm{e}^y \mathrm{e}^{-\lambda_i \mathrm{e}^y} \, \mathrm{d}y + \mathrm{e}^{-\lambda_i \tau} \times \log \tau = \beta_0 + \beta_1 \times B_i + \beta_2 \times U_i$$

in terms of β_0, β_1, β_2, B_i, U_i and τ. This step gives us an uncensored data set for analyses using method (c).

Step 3: We generate independent censoring from an exponential distribution with rate chosen to yield approximately 30 per cent censoring prior to τ. Censored data analyses are based on $\min(\tau, X_i)$, where $X_i = \min(T_i, C_i)$.

Step 4: Using the censored data set generated from Step 3, we apply the MI algorithm described in Section 3 and obtain multiply imputed data sets. Then, we estimate model (1) regression parameters as in Section 4.1, and survival percentages as in Section 4.2.

Step 5: Using the censored data set generated from Step 3, we apply the PO approach to estimate model (1) regression parameters. We estimate survival percentages using the KM method. These analyses will be compared with those in Step 4.

Step 6: Using the uncensored data set generated from Step 2, we estimate model (1) regression parameters and survival percentages. These analyses represent the upper bound of available efficiency that is attainable for this setting.

The simulation for the dependent censoring case uses $n = 1000$ and $\tau = 2$. C_i is simulated using the exponential distribution $Exp(\lambda \times U_i)$, where λ is chosen to give approximately 30 per cent censoring. Otherwise, the procedure for simulation is similar to the independent case.

Table I displays the simulation results in the independent censoring case, where the true values of β are $(-1, 0.5, 0.5)$. For each approach (a), (b) and (c), we present bias as $\widehat{\beta} - \beta$, and the corresponding average estimated standard errors (SEs). We also calculate the empirical standard deviation (ESD) of the 1000 estimates and the proportion of simulations that cover the true values (empirical coverage probability, CP). To assess gains in efficiency, we give average 95 per cent confidence interval (95 per cent CI) widths over the 1000 simulations. The asymptotic relative efficiency (ARE) is defined as $\mathrm{Var}(\widehat{\beta}_{\mathrm{PO}})/\mathrm{Var}(\widehat{\beta} \text{ of interest})$, where $\mathrm{Var}(\widehat{\beta}_{\mathrm{PO}})$, the variance of estimates using the PO approach, is used as the reference variance.

Table I. Independent censoring: Comparison of estimates based on model (1) using method (a) restricted mean MI (MI) approach, method (b) pseudo observation (PO) approach and method (c) uncensored observations (Uncensored) ($n = 100$).

Parameter	Method	Bias*	SE†	ESD‡	CP§ (per cent)	Width of CI¶	ARE$^\parallel$
$\beta_0 = -1$	Uncensored	−0.005	0.286	0.314	91.7	1.12	1.36
	PO	−0.193	0.333	0.383	89.1	1.31	1.00
	MI	0.002	0.295	0.316	92.8	1.15	1.27
$\beta_1 = 0.5$	Uncensored	0.005	0.220	0.233	93.8	0.86	1.35
	PO	0.055	0.256	0.273	92.9	1.01	1.00
	MI	0.011	0.227	0.240	93.5	0.89	1.27
$\beta_2 = 0.5$	Uncensored	0.005	0.437	0.456	94.4	1.71	1.36
	PO	0.059	0.510	0.537	93.8	2.00	1.00
	MI	−0.042	0.451	0.452	95.1	1.77	1.28

*Bias is the average of $\widehat{\beta} - \beta$ over the simulations, i.e. the average estimated parameter is the shown bias plus the true parameter in column 1.

†SE is the average estimated standard errors over the simulations.

‡ESD is empirical standard deviation of the 1000 estimates.

§CP is the empirical coverage probability, i.e. the proportion of simulations that cover the true values.

¶Width of CI is the average 95 per cent confidence interval (95 per cent CI) widths over the 1000 simulations.

$^\parallel$ARE is asymptotic relative efficiency, defined as $\mathrm{Var}(\widehat{\beta}_{PO})/\mathrm{Var}(\widehat{\beta})$, where $\mathrm{Var}(\widehat{\beta}_{PO})$, the variance of estimates using the PO approach, is used as the reference variance.

The results show that the MI approach and uncensored data analysis yield approximately unbiased estimates. Aside from the intercept term, the PO method also appears unbiased. Because of difficulty in estimating the intercept term, the PO approach tends to have a lower coverage rate for the true $E[\log\{\min(\tau, T)\}]$ (83 per cent for the PO approach as opposed to 93 per cent for the MI method and 94 per cent for the uncensored data case).

The 95 per cent CI widths for regression parameters are around 12 per cent narrower using the MI method compared with the PO approach. Furthermore, they are very close to the 95 per cent CI widths based on uncensored data. The MI parameter estimates are 27–28 per cent more efficient in terms of the ARE than the PO approach. Hence, by assuming the mean structure as in (1) and including minimal assumptions on the variance, we are able to recover much of the efficiency lost due to censoring. Additional simulations under different parameter settings show similar patterns of results (data not shown).

The results for dependent censoring case are presented in Table II, where the true values of β are $(-1, 1, 0.5)$. The sample size used (1000) is comparable to the sample size in the IBCSG example in Section 6. Parameter estimates using the restricted mean MI approach are essentially unbiased, whereas the PO method was subject to bias as large as one standard deviation.

As mentioned in Section 1, various analyses can be conducted based on multiply imputed data sets. One example is to produce marginal survival estimates. Simulation results in the independent censoring case are shown in Table III at survival quantiles 60 per cent through 40 per cent $\approx S(\tau)$. Survival estimates using the MI method approximate true quantiles well, and as expected, the level of efficiency gain increases with increased censoring. Survival quantiles 100 per cent through 60 per cent, where censoring was minimal, showed only negligible differences in efficiency (data not shown). Simulations conducted in the dependent censoring case gave unbiased results for the MI estimated quantiles, whereas the KM method overestimated survival by approximately 3 per cent after the 60th quantile. Although gains in efficiency for survival estimates were seen using MI method in the dependent censoring case, gains were not nearly as attractive as in the independent censoring case (data not shown).

All simulations were repeated using an additional bootstrap step to provide a further level of variability in the selection of imputed values. That is, each of the M imputed data sets was produced from a different bootstrap sample of the original observed data, which further varied the distribution of parameter estimates in Section 3 Step 1 as well as the observed residual distribution in Section 3 Step 2. Although some authors (Taylor *et al.* [8], Rubin and Schenker [17], Heitjan and Little [18]) have found improved coverage using this approach, coverage probabilities in the simulations did not appreciably change (data not shown). A conservative recommendation would be to perform the analysis both with and without the bootstrap step, particularly for smaller sample sizes than the simulations shown here ($n < 100$).

Table II. Dependent censoring: Comparison of estimates based on model (1) using method (a) restricted mean MI (MI) approach, method (b) pseudo observation (PO) approach and method (c) uncensored observations (Uncensored) ($n = 1000$).

Parameter	Method	Bias*	SE[†]	ESD[‡]	CP[§] (per cent)	Width of CI[¶]	ARE[‖]
$\beta_0 = -1$	Uncensored	0.001	0.078	0.091	90.6	0.31	1.49
	PO	−0.163	0.095	0.119	57.3	0.37	1.00
	MI	0.002	0.082	0.099	90.0	0.32	1.33
$\beta_1 = 1$	Uncensored	0.001	0.066	0.071	93.5	0.26	1.49
	PO	0.131	0.081	0.093	63.7	0.32	1.00
	MI	0.004	0.069	0.074	93.5	0.27	1.38
$\beta_2 = 0.5$	Uncensored	−0.004	0.117	0.121	92.9	0.46	1.49
	PO	−0.083	0.142	0.148	89.2	0.56	1.00
	MI	−0.008	0.122	0.132	91.7	0.48	1.36

*Bias is the average of $\widehat{\beta} - \beta$ over the simulations, i.e. the average estimated parameter is the shown bias plus the true parameter in column 1.

[†]SE is the average estimated standard errors over the simulations.

[‡]ESD is empirical standard deviation of the 1000 estimates.

[§]CP is the empirical coverage probability, i.e. the proportion of simulations that cover the true values.

[¶]Width of CI is the average 95 per cent confidence interval (95 per cent CI) widths over the 1000 simulations.

[‖]ARE is asymptotic relative efficiency, defined as $\mathrm{Var}(\widehat{\beta}_{\mathrm{PO}})/\mathrm{Var}(\widehat{\beta})$, where $\mathrm{Var}(\widehat{\beta}_{\mathrm{PO}})$, the variance of estimates using the PO approach, is used as the reference variance.

Table III. Independent censoring: Comparison of survival estimates using restricted mean MI (MI) approach, KM with censored observations (Censored) and KM with uncensored observations (Uncensored) ($n = 100$).

Quantile (per cent)	Censoring per cent*	Method	Bias[†]	SE[‡]	ARE[§]
60	21.3	Uncensored	0.006	0.049	1.18
		Censored	0.006	0.053	1.00
		MI	−0.005	0.052	1.01
55	23.7	Uncensored	0.005	0.049	1.22
		Censored	0.005	0.055	1.00
		MI	−0.008	0.054	1.03
50	26.0	Uncensored	0.007	0.050	1.27
		Censored	0.007	0.056	1.00
		MI	−0.007	0.054	1.06
45	28.1	Uncensored	0.006	0.050	1.33
		Censored	0.006	0.057	1.00
		MI	−0.009	0.054	1.09
40	30.1	Uncensored	0.006	0.049	1.40
		Censored	0.004	0.058	1.00
		MI	−0.008	0.054	1.13

*Censoring per cent is the average censoring percentage up to corresponding quantile over the simulations.

[†]Bias is the average of true survival percentages minus estimated survival percentages over the simulations.

[‡]SE is the average estimated standard errors over the simulations.

[§]ARE is asymptotic relative efficiency, defined as $\mathrm{Var}(\widehat{S}_{\mathrm{PO}}^{\mathrm{KM}})/\mathrm{Var}(\widehat{S}^{\mathrm{KM}})$, where $\mathrm{Var}(\widehat{S}_{\mathrm{PO}}^{\mathrm{KM}})$, the variance of estimates using the PO approach, is used as reference variance.

6. IBCSG Ludwig Trial V example

We now apply standard analyses to restricted mean-based multiply imputed data sets from the IBCSG Ludwig Trial V study. The data consist of 1229 patients, where 59 patients are still at risk at 108 months, 551 have died and 669 are censored prior to 108 months of followup. Observed covariates include long-duration (LD) or short-duration (SD) treatment assignment, ER status (positive vs negative/unknown), tumor size (greater or less than 2 cm), number of nodes (0–3, 4–9 or 10+) and age (in decades).

The primary interest of the study is to examine the treatment effect over the study period. Just as in a traditional medical journal results section, we first describe marginal treatment effects via plots and
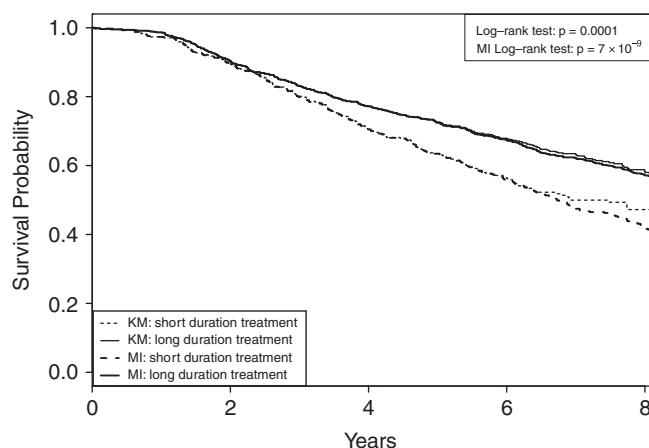
Figure 1. IBCSG survival by treatment and method of estimation.

point estimates using the KM method and our restricted mean MI method. Then, we test for treatment differences using the traditional log-rank test and the MI augmented log-rank test. Multivariate analysis results then assess adjusted treatment effects and other useful predictors.

As part of performing the MI procedure we estimate the restricted mean in terms of available data as follows:

$$E[\log\{\min(108, T)\}] = \beta_0 + \beta_1 \times I(\text{LD treatment}) + \beta_2 \times I(\text{ER positive})$$
$$+ \beta_3 \times I(\text{Tumor} \geqslant 2\,\text{cm}) + \beta_4 \times I(\text{positive nodes 4–9})$$
$$+ \beta_5 \times I(\text{positive nodes 10+}) + \beta_6 \times (\text{Age in decades}).$$

In other words, using the methods described in Section 3, for censored patients we multiply imputed failure times incorporating information on treatment assignment, ER status, node group categories, tumor size and age.

Marginal survival curves based on the KM method and the MI method are shown in Figure 1, with confidence intervals at year marks shown in Table IV. The rightmost column of Table IV summarizes the differences in point estimates of $\widehat{S}(t)$, according to method. Estimates between the two methods are similar during the first couple of years, but as censoring increases, so do differences between survival estimates across time. Particularly in the short-duration therapy, the difference in $\widehat{S}(8\,\text{years})$ approaches 6 per cent. This pattern was similar both with and without a bootstrap step included in the analysis. When investigating possible reasons for differences in tail survival estimates, we identified dependent censoring captured by age (Hazard Ratio for censoring: 0.81 per decade age increase, 95 per cent CI: (0.70, 0.93), p-value$=0.004$). That is, older patients who entered the trial were both less likely to be censored and had longer restricted lifetimes (to be discussed shortly in Table V as part of the multivariate analyses). The MI procedure accounts for this setting, giving lower survival estimates over time when compared with the KM method. The treatment differences are much larger once the dependent censoring bias related to age is accounted for. This is reflected in the much higher significance of the log-rank analysis on the restricted mean MI data sets (without bootstrap: $p=7 \times 10^{-9}$; with bootstrap: $p=2 \times 10^{-9}$) compared with the traditional log-rank test ($p=0.0001$). The marginal survival plots also indicate some non-proportionality early in the study duration, perhaps arguing the merits of a multivariate model not dependent on proportional hazard shapes to hold.

The significant treatment difference is maintained once we adjust for other risk factors (full results in Table V). The first three columns of Table V give the GLM model fit using the PO approach as in [5]. The remaining columns give the analysis based on our restricted mean MI procedure. We report $e^{\widehat{\beta}}$, where $\widehat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5, \hat{\beta}_6)$ associated with 95 per cent CIs and p-values. The parameters $e^{\widehat{\beta}}$ are interpreted as multiplicative effects, so that parameter estimates higher than one give longer estimated restricted lifetimes and estimates smaller than one give shorter estimated restricted lifetimes. For example, using the MI approach, long-duration treatment tends to prolong restricted lifetime by

Table IV. IBCSG Ludwig Trial V results II: Survival estimates using Kaplan–Meier (KM) approach and the restricted mean multiple imputation (MI) approach.

Treatment	T (year)	Censoring per cent	$\widehat{S}^{(KM)}(t)$ (95 per cent CI)*	$\widehat{S}^{(MI)}(t)$ (95 per cent CI)*	$\Delta^{\dagger}$
SD treatment					
	1	0.0	0.97 (0.96, 0.99)	0.97 (0.96, 0.99)	0.000
	2	0.2	0.90 (0.87, 0.93)	0.89 (0.87, 0.92)	−0.001
	3	0.5	0.80 (0.76, 0.84)	0.80 (0.76, 0.84)	−0.001
	4	0.7	0.71 (0.66, 0.75)	0.71 (0.66, 0.75)	−0.001
	5	2.7	0.63 (0.59, 0.68)	0.63 (0.59, 0.68)	−0.001
	6	14.8	0.57 (0.52, 0.61)	0.55 (0.51, 0.60)	−0.010
	7	28.8	0.50 (0.45, 0.55)	0.46 (0.41, 0.51)	−0.036
	8	38.5	0.47 (0.42, 0.53)	0.41 (0.31, 0.41)	−0.063
LD treatment					
	1	0.0	0.99 (0.98, 0.99)	0.99 (0.98, 0.99)	0.000
	2	0.1	0.90 (0.88, 0.92)	0.90 (0.88, 0.92)	0.000
	3	0.3	0.83 (0.80, 0.86)	0.83 (0.81, 0.86)	0.000
	4	0.8	0.77 (0.74, 0.80)	0.77 (0.74, 0.80)	0.000
	5	2.8	0.73 (0.70, 0.76)	0.73 (0.70, 0.76)	−0.001
	6	14.2	0.68 (0.65, 0.71)	0.68 (0.65, 0.71)	−0.003
	7	31.7	0.63 (0.60, 0.67)	0.62 (0.60, 0.66)	−0.008
	8	45.6	0.58 (0.54, 0.62)	0.57 (0.54, 0.61)	−0.008

*Statistical significance is defined when 95 per cent CI does not cover the true values with Type I error rate 5 per cent.
$^{\dagger}\Delta$ is defined as $\widehat{S}^{(KM)}(t) - \widehat{S}^{(MI)}(t)$, at different t.

Table V. IBCSG Ludwig Trial V results I: Estimated restricted mean model parameters using a pseudo observation (PO) approach and the restricted mean multiple imputation (MI) approach.

Parameter	Mean	PO method			MI method			ARE‡
		Estimate*	95 per cent CI†	*p*-Value	Estimate*	95 per cent CI†	*p*-Value	
Intercept	—	59.3	47.7, 73.7	<0.0001	58.3	46.9, 72.3	<0.0001	1.01
Long-duration treatment (LD)	0.66	1.12	1.04, 1.21	0.00300	1.14	1.06, 1.23	<0.0001	1.01
Positive nodes (4–9)	0.27	0.80	0.73, 0.87	<0.0001	0.79	0.72, 0.85	<0.0001	1.01
Positive nodes (10+)	0.16	0.56	0.50, 0.62	<0.0001	0.55	0.50, 0.61	<0.0001	1.01
Tumor ($\geqslant 2$ cm)	0.77	0.88	0.81, 0.96	0.00400	0.88	0.81, 0.96	0.00400	1.01
ER positive	0.54	1.23	1.14, 1.32	<0.0001	1.23	1.14, 1.32	<0.0001	1.01
Age (in decades)	5.00	1.04	0.99, 1.08	0.06000	1.04	1.00, 1.08	0.04000	1.01

*Parameter estimates from mean structure (1) are presented on the scale of $e^{\widehat{\beta}}$ so that estimates higher than one give longer mean restricted lifetimes and estimates smaller than one give shorter mean restricted lifetimes, i.e. multiplicative effects.
†Statistical significance is defined when 95 per cent CI does not cover the true values with Type I error rate 5 per cent.
‡ARE is asymptotic relative efficiency, defined as $\text{Var}(\widehat{\beta}_{\text{PO}})/\text{Var}(\widehat{\beta}^{\text{MI}})$, where $\text{Var}(\widehat{\beta}_{\text{PO}})$, the variance of estimates using the PO approach, is used as reference variance.

a factor of $e^{\widehat{\beta}_1} = 1.14$ (95 per cent CI: 1.06, 1.23), compared with the restricted lifetime on the SD arm, adjusted for other factors. Across covariate effects, we observe similar or slightly narrower 95 per cent CI widths and 1 per cent efficiency gains for the MI method versus the PO approach. These minor efficiency gains in the variability of the parameter estimates were also seen when including a bootstrap step. The multivariate analysis confirms the role of age as a factor causing some dependent censoring since age is both related to the event time as well as the censoring time as noted above. The significance of age straddles the p-value$=0.05$ depending on methodology used. According to all methods used, older age is associated with longer restricted lifetimes. The MI method without a bootstrap step indicates that for each decade increase in age, the estimated restricted lifetime increases

by 4.1 per cent $\{e^{\hat{\beta}_6^{MI}} = 1.041$, 95 per cent CI: $(1.002, 1.082)$, p-value $= 0.040\}$. This seems to be capturing the known risk of more aggressive breast cancer tumors diagnosed in younger patients. Incorporating a bootstrap step into the MI procedure gave corresponding results for age as $\{1.036$, 95 per cent CI: $(0.996, 1.076)$, p-value $= 0.076\}$, whereas the PO method gave $\{e^{\hat{\beta}_6^{PO}} = 1.038$, 95 per cent CI: $(0.998, 1.079)$, p-value $= 0.060\}$.

The MI procedure seems to be accounting for dependent censoring through slightly better estimation of the intercept, which affects estimation of restricted lifetimes, and therefore the values imputed in the MI algorithm. For instance, we may estimate restricted lifetimes for a typical patient in the SD treatment arm, using average patient profile values for other risk factors, i.e. $(58.3)(0.79)^{0.27}(0.55)^{0.16}(0.88)^{0.77}(1.23)^{0.54}(1.04)^{5.00} = 61.6$ months for the MI method (95 per cent CI: 57.9, 65.5), and similarly 62.1 months for the PO method (95 per cent CI: 58.3, 66.0). That is, a 50-year-old patient with a 27 per cent chance of having 4–9 positive nodes, a 16 per cent of having 10+ positive nodes, a 54 per cent chance of being ER positive and a 77 per cent chance of having tumor greater than 2 cm is expected to live 61.6 months out of a possible 108 months on study based on the MI method. The PO approach estimates a slightly longer restricted mean for this type of patient, likely connected to dependent censoring biases associated with age. Since younger (sicker) patients are being censored more often, the PO approach seems to be slightly overestimating the expected number of months lived during the 108-month duration.

7. Discussion

Using the restricted mean formulation, the shapes of the survival curves in relation to one another are not specified. Our method merely requires the mean structure to be correctly specified, i.e. the area under the survival curve to τ follows (1). It is also possible, of course, to recover restricted mean estimates from a Cox modeling framework using $\hat{S}(t|Z) = \hat{S}_0(t)^{e^{\hat{\beta}^T Z}}$, and in cases where hazards are truly proportional these estimates should be fully efficient. In two-sample testing literature when hazards are not proportional, Pepe and Fleming [19] indicated a substantial improvement in detecting treatment effects using differences in restricted means as opposed to the log-rank test (score test for Cox model). Previous authors advocating restricted mean models have not clearly laid out inference performance issues in relation to hazard-based models beyond the intuition gleaned from the two-sample testing setting.

Andersen *et al.* [5] use POs to create a modified data set and apply a similar mean structure as (1) for analysis. In studying the PO method and how it might be modified to provide imputes larger than the observed censoring times, we discovered a potential loss of statistical information available from Z. We considered imputes for C_i that add C_i plus a conditional PO created from the patients at risk at C_i. Suppose $\widehat{S}^{KM(-i)}(t|T \geqslant C_i)$ is the KM estimate with the person censored at C_i left out among those otherwise at risk at C_i. Then a conditional PO defined as $n \int_{C_i}^{\tau} \widehat{S}^{KM}(t|T \geqslant C_i)\,\mathrm{d}t - (n-1)\int_{C_i}^{\tau} \widehat{S}^{KM(-i)}(t|T \geqslant C_i)\,\mathrm{d}t$ would reduce to $\int_{C_i}^{\tau} \widehat{S}^{KM}(t|T \geqslant C_i)\,\mathrm{d}t$. The non-parametric estimate $\widehat{S}^{KM}(t|T \geqslant C_i)$ does not fully utilize covariate information from Z. The PO calculation for this special case may indicate why our restricted mean MI approach outperforms the traditional PO method in simulation with respect to efficiency. Looking at this special case of PO creation may also suggest why dependent censoring might influence the PO method in terms of bias. That is, a conditional PO $\widehat{S}^{KM}(t|T \geqslant C_i)$ may still be biased if censoring depends on Z, and the traditional PO method seems also subject to this same source of bias.

More recently, Andersen and Perme [20] suggested the use of POs based upon weighted KM estimates [21], $n^{-1}\sum_{g=1}^{G} n_g \widehat{S}_g^{KM}(t)$, where $\widehat{S}_g^{KM}(t)$ is a KM estimate in subgroup g of patients with categorical covariate $Z = g$ and n_g is the number of patients in subgroup g ($g = 1, \ldots, G$). When only a single categorical covariate is associated with the survival and censoring distributions, corresponding calculations of conditional pseudovalues created from those at risk at C_i reduce to $\int_{C_i}^{\tau} \widehat{S}_g^{KM}(t|T \geqslant C_i)\,\mathrm{d}t$, which is a maximum likelihood estimate of the restricted mean life for someone from group g surviving past C_i. Since all available covariate information is utilized in this scenario, we suspect that the PO approach will be fully efficient and unbiased for estimation of parameters in (1). Simulations using categorical covariates yield very similar results for parameter estimates based on weighted KM-based POs and restricted mean MI method (data not shown).

The approach of creating POs with KM estimates averaged across categorical covariate strata may be impractical in regression settings with many covariates. As covariate strata become more finely partitioned, technical difficulties of estimating survival consistently in the tails of the distribution arise since each stratum specific KM curve is only guaranteed consistency in the range of the observed outcomes for that stratum. When only a single continuous covariate is related to survival and censoring distributions, averaging KM estimates across categorical strata of the continuous covariate may still be viable since simulations by Murray and Tsiatis demonstrated that most efficiency gain and bias correction may be captured using roughly 3–5 strata.

With a single categorical covariate our restricted mean MI approach reduces to that of Hsu *et al.* [10] since risk set groupings based on either similar restricted means, as in our work, or similar hazards, as in Hsu *et al.*'s work, will result in groups with the same categorical covariate to impute from. As indicated in Hsu *et al.*, this special case produces marginal survival estimates similar in expectation to the weighted KM estimate described by Murray and Tsiatis as well as the survival estimates proposed by Malani [22]. Similarly, rank-based tests based on MI analyses with categorical covariates would be expected to perform similarly to those proposed by Mackenzie and Abrahamowicz [23].

To our knowledge, this is the first instance when the restricted mean lifetime has been used to impute censored survival data based on risk factors, increasing efficiency. Efficiency gains when using covariate information in marginal survival curve estimation have been seen in many other contexts by authors including Finkelstein and Schoenfeld [24], Gray [25] and Robins and Rotnitzky [26]. The MI method retains essential characteristics of the observed data and approximates the original distribution well. Final parameter estimates have good operating characteristics and have improved finite sample properties. Particularly appealing to clinicians is that the interpretation of the parameter estimates apply directly to days, months or years of life saved for different risk profiles. In addition, our method preserves the traditional benefits of MI such as transparency of variance calculation and availability of standard statistical software to analyze the augmented data sets.

Appendix A: Pseudo EM algorithm in restricted mean MI procedure

We assume the mean structure as in (1), but keep the distribution of Y otherwise unspecified. The goal is to predict each censored individual's mean restricted lifetime based on the converged $\widehat{\beta}^{(k)T} Z_i$ values.

We calculate $\widehat{\beta}^{(0)T} Z_i$ as the initial expected value of Y_i based on covariate profile Z_i. We use $\widehat{\beta}^{(0)}$ to estimate the initial expected value of Y_i based on $\widehat{\beta}^{(0)T} Z_i$ for each censored individual i. To get the next iterated estimate, $\widehat{\beta}^{(1)}$, we sample from $N\{\widehat{\beta}^{(0)T} Z_i, \text{Var}(\widehat{\beta}^{(0)T} Z_i)\}$ for each patient censored at $X_i = C_i$. We retain sample values that are greater than $\log C_i$, and then use the sample average to estimate $E(Y_i | C_i, Z_i)$, our E-step of the pseudo EM algorithm. After the 'E' step is completed for each censored individual, this completed data set is used to calculate $\widehat{\beta}^{(1)}$ under (1), our M-step. We repeat 'E' and 'M' steps until the convergence of $\widehat{\beta}^{(k)}$.

Appendix B: Residual sampling procedure in restricted mean MI algorithm

In order to generate an augmented data set, we append an appropriate residual to $\widehat{E}(Y_i | C_i, Z_i)$, for each censored patient i. We achieve this goal by creating an imputing pool with patients whose categorical covariates match Z_i, and whose $\widehat{E}(Y | Z)$ fall within some small b-margin of $\widehat{E}(Y_i | Z_i)$ (to account for similarity in continuous covariates). Once we form the imputing pool, we estimate the pool's survival curve using the KM method. Then we randomly select a probability, s_i, from a Uniform(a,1) distribution, where a is the minimum survival probability in this pool. Next, we select a failure time where the estimated survival probability from the KM curve equals to s_i. The corresponding residual is taken from the patient with the observed failure time that was selected. This is then the error term to be appended to $\widehat{E}(Y_i | C_i, Z_i)$. This approach is similar to the nearest neighborhood approach in [10] modified for the restricted mean setting.

Acknowledgements

References

1. Karrison K. Restricted mean life with adjustment for covariates. *Journal of American Statistics Association* 1987; **82**:1169–1176.
2. Cox DR. Regression models and life tables (with Discussion). *Journal of the Royal Statistical Society*, *Series B* 1972; **34**:187–200.
3. Chen PY, Tsiatis A. Causal inference on the difference of the restricted mean lifetime between two groups. *Biometrics* 2001; **57**:1030–1038.
4. Zucker DM. Restricted mean life with covariates: modification and extension of a useful survival analysis method. *Journal of American Statistics Association* 1998; **93**:702–709.
5. Andersen PK, Hansen M, Klein JP. Regression analysis of restricted mean survival time based on pseudo-observations. *Life Data Analysis* 2004; **10**:335–350.
6. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**:13–22.
7. Faucett CL, Schenker N, Taylor JM. Survival analysis using auxiliary variables via multiple imputation, with application to AIDS clinical trial data. *Biometrics* 2002; **58**:37–47.
8. Taylor JM, Murray S, Hsu CH. Survival estimation and testing via multiple imputation. *Statistics and Probability Letters* 2002; **58**:221–232.
9. Kaplan EL, Meier P. Non-parametric estimation from incomplete observations. *Journal of American Statistics Association* 1958; **53**:457–481, 562–563.
10. Hsu CH, Taylor JM, Murray S, Commenges D. Survival analysis using auxiliary variables via non-parametric multiple imputation. *Statistics in Medicine* 2006; **25**:3503–3517.
11. Rubin D. *Multiple Imputation for Nonresponse in Surveys*. Wiley: New York, 1987.
12. Little RJA, Rubin D. *Statistical Analysis with Missing Data* (2nd edn). Wiley: New York, 2002.
13. Irwin JO. The standard error of an estimate of expectational life. *Journal of Hygiene* 1949; **47**:188–189.
14. Buckley JD, James IR. Linear regression with censored data. *Biometrika* 1979; **66**:429–436.
15. Dabrowska DM, Doksum KA. Partial likelihood in transformation models with censored data. *Scandinavian Journal of Statistics* 1988; **10**:1–24.
16. Fine JP, Ying Z, Wei LJ. On a linear transformation model for censored data. *Biometrika* 1998; **85**:980–986.
17. Rubin DB, Schenker N. Multiple imputation in health-care databases: an overview and some applications. *Statistics in Medicine* 1991; **10**:585–598.
18. Heitjan DF, Little RJA. Multiple imputation for the fatal accident reporting system. *Applied Statistics* 1991; **40**:13–29.
19. Pepe M, Fleming T. Weighted Kaplan–Meier statistics—a class of distance tests for censored survival data. *Biometrics* 1989; **45**:497–507.
20. Andersen PK, Perme MP. Pseudo-observations in survival analysis. *Statistical Methods in Medical Research* 2010; **19**:71–99.
21. Murray S, Tsiatis AA. Nonparametric survival estimation using prognostic longitudinal covariates. *Biometrics* 1996; **52**:137–151.
22. Malani HM. A modification of the redistribution to the right algorithm using disease markers. *Biometrika* 1995; **82**:515–526.
23. Mackenzie T, Abrahamowicz M. Using categorical markers as auxiliary variables in log-rank tests and hazard ratio estimation. *The Canadian Journal of Statistics* 2005; **33**:201–219.
24. Finkelstein DM, Schoenfeld DA. Analyzing survival in the presence of an auxiliary variable. *Statistics in Medicine* 1994; **13**:1747–1754.
25. Gray RJ. A Kernel method for incorporating information on disease progression in the analysis of survival. *Biometrika* 1994; **81**:527–539.
26. Robins JM, Rotnitzky A. Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology—Methodological Issues*, Jewell N, Dietz K, Farewell V (eds). Birkhauser, 1992; 297–331. ISBN: 0817636323.