

UCLA

Department of Statistics Papers

Title

Multiple Imputation for Threshold-Crossing Data with Interval Censoring

Permalink

<https://escholarship.org/uc/item/1kq3n8zd>

Authors

Fredierick J. Dorey
Roderick J.A. Little
Nathaniel Schenker

Publication Date

2011-10-24

MULTIPLE IMPUTATION FOR THRESHOLD-CROSSING DATA WITH INTERVAL CENSORING

FREDERICK J. DOREY

Division of Orthopaedic Surgery, UCLA School of Medicine, Los Angeles, CA 90024, U.S.A.

RODERICK J. A. LITTLE

Department of Biomathematics, UCLA School of Medicine, Los Angeles, CA 90024, U.S.A.

AND

NATHANIEL SCHENKER

Department of Biostatistics, UCLA School of Public Health, Los Angeles, CA 90024-1772, U.S.A.

SUMMARY

Medical statistics often involve measurements of the time when a variable crosses a threshold value. The time to threshold crossing may be the outcome variable in a survival analysis, or a time-dependent covariate in the analysis of a subsequent event. This paper presents new methods for analysing threshold-crossing data that are interval censored in that the time of threshold crossing is known only within a specified interval. Such data typically arise in event-history studies when the threshold is crossed at some time between data-collection points, such as visits to a clinic. We propose methods based on multiple imputation of the threshold-crossing time with use of models that take into account values recorded at the times of visits. We apply the methods to two real data sets, one involving hip replacements and the other on the prostate specific antigen (PSA) assay for prostate cancer. In addition, we compare our methods with the common practice of imputing the threshold-crossing time as the right endpoint of the interval. The two examples require different imputation models, but both lead to simple analyses of the multiply imputed data that automatically take into account variability due to imputation.

1. INTRODUCTION

Frequently in medicine, patients are considered in a new risk category when they cross a threshold on some prognostic variable. Because patient examinations occur only periodically, the exact time of crossing the threshold is often unknown. In such cases, one knows only that the time falls within a specified interval; that is, the threshold crossing is 'interval censored'.

An example of interval-censored threshold-crossing data that we consider in this paper involves hip replacements. The routine follow-up of patients who have received a total hip arthroplasty includes an evaluation of radiographs for the presence of radiolucent lines between the prosthesis and the cement interface. Although some lines frequently appear soon after surgery, substantial lines of a specified magnitude indicate that the prosthesis has begun to loosen from the cement and that the patient has an increased risk of an eventual clinical condition, such as severe pain or fracture, that may necessitate revision surgery with insertion of a new prosthesis. The crossing of this threshold by the radiolucent lines typically results in a closer surveillance of

the patient as well as medical intervention, in the sense that the patient may be advised to reduce his activity level.

Another example that we consider involves prostate cancer. One treatment for this disease is radical prostatectomy (removal of the prostate). After surgery, patients are usually followed annually for clinical detection of disease recurrence. Since about 1985, a radioimmunoassay, the prostate specific antigen (PSA) assay, has been available; usually this assay can detect the presence of prostate cancer much earlier than can a clinical examination or bone scan. The level of the PSA assay reflects the magnitude of the cancer present. When a patient's assay exceeds a specified level, it is considered detectably positive, and indicates cancer recurrence. There have been only a negligible number of recurrences among assays that were not detectably positive. Thus a clinical issue is not the utility of the assay as a prognostic device, but rather how long it takes for the clinical examination also to detect disease recurrence. We can answer this question only with existing data sets because current practice is that, as soon as the PSA assay becomes detectably positive, the patient receives adjuvant therapy. Hence, the normal disease course can no longer be observed.

A common practice in analyses of medical data is to date events as having occurred when they were first detected. In particular, for orthopaedic data such as in the above example, one would date the threshold-crossing time as the first visit when one detected radiolucent lines of a certain magnitude. This practice corresponds to imputing the unknown crossing time at the right endpoint of the interval of possible times. The procedure clearly involves some bias, and may lead to inflated estimates of the impact of the threshold variable on the risk of a subsequent event. In a randomized prospective trial or a study of a serious disease, with little time between follow-up visits, we can probably ignore the bias since the time interval in which the threshold crossing occurred is short. For studies of chronic diseases, however, in which patients (or prostheses) may survive for many years after surgery, the bias could be substantial since the time between follow-up visits is typically much longer.

The examples of threshold crossings discussed in this paper illustrate a particular kind of interval-censored data, where one records an event history and knows only that the dates of some events lie within specified intervals. Maximum-likelihood methods for interval-censored outcomes have been considered by Peto,¹ Turnbull,² Finkelstein and Wolfe,³ Finkelstein,⁴ and Bacchetti and Jewell.⁵ Our problem differs from standard interval censoring in that the event is triggered by a continuous variable crossing a threshold, and measures of the variable are available at the interval endpoints. The methods that we propose make use of this information. We also apply the methods to situations in which the threshold-crossing time is a covariate in a Cox proportional hazards model, a problem which is not addressed in the current literature on interval censoring.

It would be of interest to extend maximum-likelihood approaches to the forms of interval censoring considered here. Existing maximum-likelihood algorithms for the (simpler) problems involving interval-censored outcomes are complex, however, involving iterative algorithms and (in the case of parametric models) numerical integration to evaluate the likelihood function.⁶

In this paper, we propose and apply simple, easily implemented approximate solutions for interval-censored crossing times based on multiply imputing the events within the intervals. Multiple-imputation methods for related censoring problems in AIDS data have been developed by Muñoz *et al.*⁷ and Taylor *et al.*⁸ The multiple imputations for our problems are comparatively simple to implement, and are based on approximate models for the crossing times that take into account important covariate information. As in previous approaches to the analysis of interval-censored data, we assume that the process of interval censoring is non-informative; more

technically, we assume that the data are ‘coarsened at random’ in the sense defined by Heitjan and Rubin.⁹

Section 2 presents a brief introduction to multiple imputation. Sections 3 and 4 develop specific methods that we apply to the two examples mentioned above, that is, the hip-replacement data and the radical-prostatectomy data. A concluding discussion appears in Section 5.

2. MULTIPLE IMPUTATION

We can view the presence of interval-censored threshold-crossing data as a problem of missing data in that the exact time of crossing the threshold is missing. A common method for handling missing data is to impute a value for each missing datum. A major attraction of imputation is the creation of a data set that one can analyse with standard statistical software developed for complete data. Imputation of just one value for each missing datum, however, fails to reflect the uncertainty about which value to fill in, and leads to underestimates of variability and invalid statistical inferences (see for example Rubin and Schenker^{10, 11}).

Multiple imputation^{12, 13} is a method that avoids this problem while retaining the advantages of imputation. We sample several, say M , draws from a distribution for the missing data estimated using the available data; this results in M completed data sets, one with the first set of imputes, one with the second set of imputes, and so on. We analyse each completed data set using standard complete-data methods, and then we combine the M analyses as follows.

Suppose that we base the standard complete-data inference for a parameter θ on a point estimate $\hat{\theta}$ and its associated variance U , so that multiple imputation and the subsequent complete-data analyses result in M sets of statistics, $\hat{\theta}_m$ and U_m , $m = 1, \dots, M$. The multiple-imputation estimate is then

$$\bar{\theta} = \Sigma \hat{\theta}_m / M,$$

with associated variance

$$T = \bar{U} + (1 + M^{-1})B,$$

where

$$\bar{U} = \Sigma U_m / M$$

estimates the within-imputation variance and

$$B = \Sigma (\hat{\theta}_m - \bar{\theta})^2 / (M - 1)$$

estimates the between-imputation variance (Reference 10 and Reference 13, Chapter 3). The ratio $(1 + M^{-1})B/\bar{U}$, which measures the proportional increase in variance due to imputation, also indicates the fraction of information that is lost due to missing data; larger values of the ratio are associated with larger fractions of missing information.

Multiple imputation is justified theoretically as simulating random draws of the missing values from their posterior predictive distribution given the observed data (Reference 13, Chapter 3). Thus a strict application of multiple imputation would require a full specification of a model for the data and the missing-data mechanism. In addition, drawing each set of multiple imputations should involve two stages. First, we draw a value of the parameter (vector) of the model from its posterior distribution, and second, we draw the set of imputations conditional upon the drawn value of the parameter. This two-stage procedure allows us to build into the imputations the uncertainty due to estimating the parameter of the model.

Although multiple imputation has a Bayesian justification, it also tends to produce inferences with good frequency properties, since the uncertainty due to imputation is properly reflected in the analysis (Reference 10 and Reference 13, Chapter 4). This is often true even when one uses approximations to the true posterior distribution for the missing data. Moreover, References 10, 14 and 13 (Chapter 4) show that for typical fractions of missing data, only a few imputations (say, $3 \leq M \leq 10$) are needed to produce point estimates that are nearly fully efficient and inferences that are approximately valid. Generally, the larger the fraction of missing information, the larger M should be.

In the applications presented in Sections 3 and 4, we apply the general principles of multiple imputation; we use, however, simple models for the missing values as well as approximate methods for drawing parameters from posterior distributions. This approach, suggested by Rubin,¹⁵ Little and Rubin,¹⁶ and Rubin and Schenker,¹¹ is based on the idea that if the amount of information that we lose due to missing data is only small or moderate, then we should find sensible but less than pristine methods of imputation adequate. Little⁶ and Heitjan and Rubin¹⁷ provide evidence for this contention in different but related problems.

3. THE HIP-REPLACEMENT PROBLEM

Here we consider imputation methods for the hip-replacement problem described in Section 1. In all applications of multiple imputation, we create $M = 10$ sets of imputations.

3.1. Data on radiographic evaluations

Three hundred and fifty-two orthopaedic patients at the UCLA Medical Center received cemented THARIES resurfacing prostheses and then had one or more clinical follow-ups that included radiographic evaluations. The common guideline for these evaluations is that when the radiolucent lines cover 100 per cent of the prosthesis with a maximum width of at least 2 mm, the patient has increased risk of eventual revision surgery; we refer to the crossing of this threshold as 'radiographic loosening'.

In practice, the determination of the time to radiographic loosening is subject to error from the measurement of the maximum width of translucent lines. Following standard practice in analysing data of this kind, we choose to ignore this measurement error and regard the variable of interest as the 'measured' rather than 'true' time to radiographic loosening. Thus, our focus is not on the measurement-error problem but on the missing-data problem that arises when the crossing times are interval censored.

Because of measurement error, the maximum width, say w , may not be monotonically increasing over time, so radiographic loosening may not be a unique event. We can avoid this problem conceptually by defining the threshold crossing to be the *first* observed occurrence of $w = 2$ mm. In our models for imputation that incorporate the values of w at the interval endpoints, we assume that w is monotonically increasing.

Of the 352 patients, there were 167 with radiographic loosening. For 94 of these 167 patients, the time of crossing of the 2 mm threshold is unknown; we know only that the crossing occurred between two visits. The first visit, at time t_L , had a measured maximum width of $w_L < 2$ mm, and the second visit, at time t_R , had a measured maximum width of $w_R > 2$ mm. Thus we need to impute the crossing time, t_C , within the interval (t_L, t_R) .

3.2. Imputation methods

We describe three deterministic methods of imputation and two random methods for multiple imputation.

Deterministic methods

We refer to the method that sets $t_C = t_R$ for purposes of statistical analyses as ‘right-point’ imputation. As mentioned in Section 1, this method is likely to result in a biased analysis. In the absence of any information regarding where in the interval (t_L, t_R) the threshold crossing lies, a more reasonable deterministic procedure is ‘mid-point’ imputation, that is, setting $t_C = (t_L + t_R)/2$. Since the maximum width measurements are available at the interval endpoints, however, we could incorporate them in the imputation procedure as well. For instance, if w_R is closer to 2 mm than is w_L , it is plausible that t_R is closer to t_C than is t_L . This suggests the use of ‘linear-interpolation’ imputation, that is, setting

$$t_C = t_L + s(t_R - t_L),$$

where

$$s = (2 - w_L)/(w_R - w_L). \tag{1}$$

Multiple-imputation methods

The technique of multiple imputation requires imputes that are random draws from a distribution, rather than deterministic values. Corresponding to mid-point and linear-interpolation imputation, we develop multiple-imputation methods that draw imputed values from distributions over the interval with means given by the deterministic methods. We find it convenient to formulate the multiple-imputation methods using distributions on the unit interval $(0, 1)$; therefore we describe the procedures in terms of imputing the ‘crossing fraction’

$$r = (t_C - t_L)/(t_R - t_L),$$

that is, the fraction of the interval (t_L, t_R) that we must travel to reach t_C .

A simple analogue to mid-point imputation in the absence of any information on where in the interval the impute should lie is ‘uniform’ imputation: impute the crossing fraction r as uniformly distributed over $(0, 1)$, or equivalently, impute t_C as uniformly distributed over (t_L, t_R) . We create multiple imputes by repeating this procedure independently M times for each missing value.

The uniform distribution is not appropriate for generating a stochastic version of the linear-interpolation method, since it is centred at one-half rather than at the ‘interpolation value’ s defined in equation (1). A flexible and convenient family of distributions over $(0, 1)$ is the beta family with mean μ and variance $\kappa\mu(1 - \mu)$, where μ and κ are parameters. The analogue to linear-interpolation imputation is to set the mean μ equal to the value of s for the case in question.

The task of estimating κ remains. For 22 of the patients with radiographic loosening, the times of the 2 mm threshold crossing are available along with smaller and larger maximum-width measurements at the two adjacent visits; thus we can compute values of both the crossing fraction r and the interpolation value s . We can estimate the parameter κ from these 22 ‘complete cases’ as follows. Let r_i and $s_i, i = 1, \dots, 22$, denote the observed values of r and s . Under the assumption that r_i has a beta distribution with mean s_i and variance $\kappa s_i(1 - s_i)$, the quantity $(r_i - s_i)^2$ estimates the variance of r_i and hence $(r_i - s_i)^2/[s_i(1 - s_i)]$ estimates κ . Averaging the 22 estimates of κ yields

$$\hat{\kappa} = (1/22)\Sigma(r_i - s_i)^2/[s_i(1 - s_i)].$$

To approximate drawing κ from its posterior distribution (Section 2), we use the bootstrap¹⁸ that is, for each set of imputations, we draw a random sample of size 22 (the number of complete cases) with replacement from the complete cases and we compute a new value of $\hat{\kappa}$ from this bootstrap

sample. The use of the bootstrap to reflect uncertainty in the parameter estimates has been suggested in other contexts by Rubin and Schenker¹⁰ and Heitjan and Little.¹⁹

We summarize the 'beta' imputation procedure as follows. For $m = 1, \dots, M$: (i) simulate a bootstrap replicate of $\hat{\kappa}$, say $\hat{\kappa}_m$; (ii) for each interval-censored case, draw the crossing fraction r from the beta distribution with mean equal to the interpolation value s and variance equal to $\hat{\kappa}_m s(1-s)$ and then set $t_C = t_L + r(t_R - t_L)$.

As mentioned in Section 2, multiple imputation should in principle draw the imputed values from the posterior predictive distribution of the missing data given the observed data. The beta imputation procedure introduced above is only an approximation to this ideal. For example, rather than drawing the parameter κ from its posterior, we use a bootstrap procedure. Perhaps more importantly, the imputation model does not incorporate all the observed data, such as patients' ages, whether they eventually had revision surgery, and calendar time of initial surgery (to capture cohort effects); to keep methods simple, our model uses only the most clearly relevant information for imputing the threshold-crossing time, namely the interval endpoints and width measurements. The approximate procedures used here and in Section 4 are based on the assumption that the data omitted from the imputation models do not contain much information about the missing values. Some simple analyses discussed in Section 5 lend support to this assumption for the hip-replacement data.

3.3. Kaplan–Meier analysis of survival

As a simple analysis of the risk of eventual revision surgery associated with radiographic loosening, Kaplan–Meier estimates of the survival distribution after radiographic loosening have been obtained, where 'survival' refers to not having revision surgery. To illustrate variability of estimates across multiple imputations under a particular model, Figure 1 presents the estimated survival functions for three sets of imputations created using the beta method. There is a tendency for the curves to become more variable later in time.

For all of the methods of Section 3.2, Table I gives the estimated probabilities of survival to three and five years after radiographic loosening. Right-point imputation results in substantially lower estimated probabilities of survival than do the other imputation methods. Thus, with the right-point method, we obtain higher apparent probabilities of revision surgery following radiographic loosening, corroborating the statement in Section 1 that the right-point method can result in inflated estimates of risk.

The mid-point and linear-interpolation methods yield similar point estimates at three and five years. The estimates are noticeably lower than those for their multiple-imputation analogues, however. This is most likely due to the fact that the Kaplan–Meier estimate is not linear in the crossing times: whereas the deterministic methods compute the Kaplan–Meier estimate at the mean of the distribution of missing crossing times, the multiple-imputation methods simulate the approximate mean of the Kaplan–Meier estimate over the distribution of the missing crossing times.

The uniform and beta multiple-imputation procedures also yield similar point estimates at both three and five years. This is not surprising, considering that for the 94 interval-censored cases, the average value of s (that is, the average interpolation value) is 0.48, indicating that the means of the distributions under both methods centre near one-half. Moreover, for the 22 complete cases, there is not a strong relation between s and the crossing time r (correlation = 0.04); thus, in this example, incorporation of the values of s into the imputation model may not add much information.

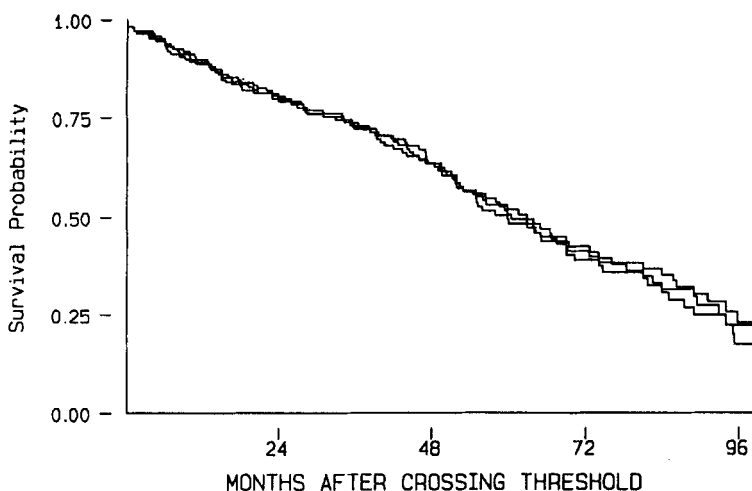


Figure 1. Kaplan-Meier estimates of the survival function for the hip-replacement example using three sets of imputations under the beta method

Table I. Kaplan-Meier estimates (standard errors in parentheses) of the probabilities of survival (no revision surgery) to three and five years after radiographic loosening

	Three years	Five years
<i>Deterministic imputation</i>		
Right-point	0.598 (0.043)	0.384 (0.050)
Mid-point	0.644 (0.040)	0.468 (0.047)
Linear interpolation	0.673 (0.040)	0.466 (0.048)
<i>Multiple imputation</i>		
Uniform		
Estimate	0.705 (0.039)	0.509 (0.048)
Ratio $((1 + M^{-1})B/\bar{U})$	7%	9%
Beta		
Estimate	0.726 (0.038)	0.513 (0.049)
Ratio $((1 + M^{-1})B/\bar{U})$	8%	11%

Standard errors for the deterministic methods are computed using Greenwood's formula. Standard errors for the multiple-imputation methods are given by $[\bar{U} + (1 + M^{-1})B]^{1/2}$ (see Section 2), with the U_m determined by Greenwood's formula.

The multiple-imputation analyses for beta imputation show that the fractional increase in variance due to imputation, $(1 + M^{-1})B/\bar{U}$, is in the moderate range of 8 to 11 per cent. This is perhaps surprising since 56 per cent (94/167) of the cases were interval censored. However, the fraction of missing information, as measured by the fractional increase in variance, depends on factors other than the simple rate of missing data, including the quantity being estimated and the variability of the imputations; we discuss this further in Section 5.

Table II. Estimates of the coefficient for radiographic loosening (standard errors in parentheses) in the Cox regression of time from initial surgery to revision surgery on radiographic loosening and age at initial surgery

<i>Deterministic imputation</i>	
Right-point	1.119 (0.186)
Mid-point	0.824 (0.183)
Linear interpolation	0.804 (0.183)
<i>Multiple imputation</i>	
Uniform	
Estimate	0.842 (0.184)
Ratio $((1 + M^{-1})B/\bar{U})$	1%
Beta	
Estimate	0.820 (0.183)
Ratio $((1 + M^{-1})B/\bar{U})$	1%

Standard errors for the deterministic methods are computed from the observed information matrix. Standard errors for the multiple-imputation methods are given by $[\bar{U} + (1 + M^{-1})B]^{1/2}$ (see Section 2), with the U_m determined from the observed information matrix.

3.4. Cox regression analysis of risk

To examine the change in the risk of eventual revision surgery associated with radiographic loosening, we estimated the Cox proportional-hazards regression model using time from initial surgery to revision surgery as the outcome variable, age at initial surgery as a fixed covariate, and the occurrence of radiographic loosening as a binary time-dependent covariate. The parameter of interest (θ) is the coefficient of the covariate for radiographic loosening. Results appear in Table II.

Right-point imputation yields a point estimate substantially larger than those for the other imputation methods, and thus inflates the apparent increase in the probability of revision surgery associated with radiographic loosening. Once again, there appears a positive bias associated with the right-point method. The mid-point, linear-interpolation, uniform, and beta imputation procedures yield similar point estimates.

The results for multiple imputation indicate that the fractional increases in variance due to imputation are equal to 1 per cent, much lower than the values seen for the Kaplan–Meier analysis. This is due in part to the lower missing-data rate of 27 per cent (94/352), but it also illustrates the dependence of the fractional increase in variance on the quantity estimated (see Section 5). The multiple-imputation point estimates are more than four times as large as their standard errors. Thus, radiographic loosening appears to increase the risk of revision surgery, and the effect is highly statistically significant.

4. THE RADICAL-PROSTATECTOMY PROBLEM

We require somewhat different imputation methods for the radical-prostatectomy problem introduced in Section 1, as we now discuss. Once again, we apply all multiple-imputation procedures with $M = 10$ sets of imputations.

4.1. Data on the PSA assay

One hundred and ninety (190) patients received radical retropubic prostatectomies for clinical localized prostate cancer at UCLA between 1977 and 1989 and had one or more PSA assays administered during follow-up examinations. The assay is considered detectably positive if its value crosses the 0.4 ng/ml threshold. As in the hip-replacement problem, our focus is on the ‘measured’ time of threshold crossing rather than the ‘true’ time that would be recorded in the absence of assay error. Fifty-nine (59) of the patients had one or more PSA assay values detectably positive, and 20 out of these 59 received a clinical diagnosis of recurrence before the study’s conclusion.

Of the patients with PSA assays that crossed the threshold, only 11 have known crossing times; the times are interval-censored for the other 48. In the absence of measurement error, we expect the assay values for a patient to remain constant at 0.2 ng/ml until a certain point in time at which they begin to increase. Thus, for each interval-censored case, we know that the threshold-crossing occurred between two visits. The first visit, at time t_L , has an assay value of 0.2 (plus error). The second visit, at time t_R , and all subsequent visits, have detectably positive and monotonically increasing values (plus error). We need to impute the crossing time, t_C , within the interval (t_L, t_R) based on an extrapolation from the detectably positive values.

For developing imputation models in the next section, it is useful to classify the incomplete cases according to the number of detectably positive PSA readings as follows. Seventeen (17) cases have only one positive reading, 21 cases have exactly two, and 10 cases have three or more. Some illustrative plots of detectably positive values for different patients appear in Figure 2.

4.2. Imputation methods

The right-point, mid-point, and uniform imputation methods are identical to those described in Section 3.2 for the hip-replacement problem. To incorporate observed data other than the interval endpoints, however, into an imputation method for the radical-prostatectomy problem, we must discuss models other than the simple linear-interpolation model used for the hip-replacement data, owing to differences in the PSA assay response curve.

Recall that the linear-interpolation method of Section 3.2 assumes in theory that the maximum widths are increasing linearly over time inside the interval (t_L, t_R) . As discussed in Section 4.1, the theoretical response curve for the PSA assay values remains at 0.2 ng/ml over time until a certain point at which it increases monotonically. Inspections of the existing literature as well as several patients with multiple detectably positive PSA assay values suggest that once the response curve becomes positive, its logarithm increases linearly over time; that is, the detectably positive values are consistent with an exponential growth model. Figure 3 illustrates that the PSA response curve is expected to lie below the interpolating line within the interval (t_L, t_R) ; equivalently, the PSA curve should predict a larger value for t_C than should linear interpolation.

Suppose that for each patient, the growth curve for the PSA assay values v has the form

$$\log(v) = \beta_0 + \beta_1 t + \varepsilon, \tag{2}$$

where t denotes the time from prostatectomy to the PSA measurement, β_0 and β_1 may vary across patients, and the error term ε has mean 0 and variance σ^2 that is constant across patients. Given estimates of the intercept β_0 and the slope β_1 in (2), it is possible to estimate the threshold-crossing time t_C by setting $\varepsilon = 0$ and $v = 0.4$ ng/ml and then solving for t . This is an example of the ‘linear calibration’ or ‘inverse regression’ problem (Reference 20, Chapter 9; Reference 21, Chapter 1). We now describe deterministic and random imputation procedures based on linear calibration.

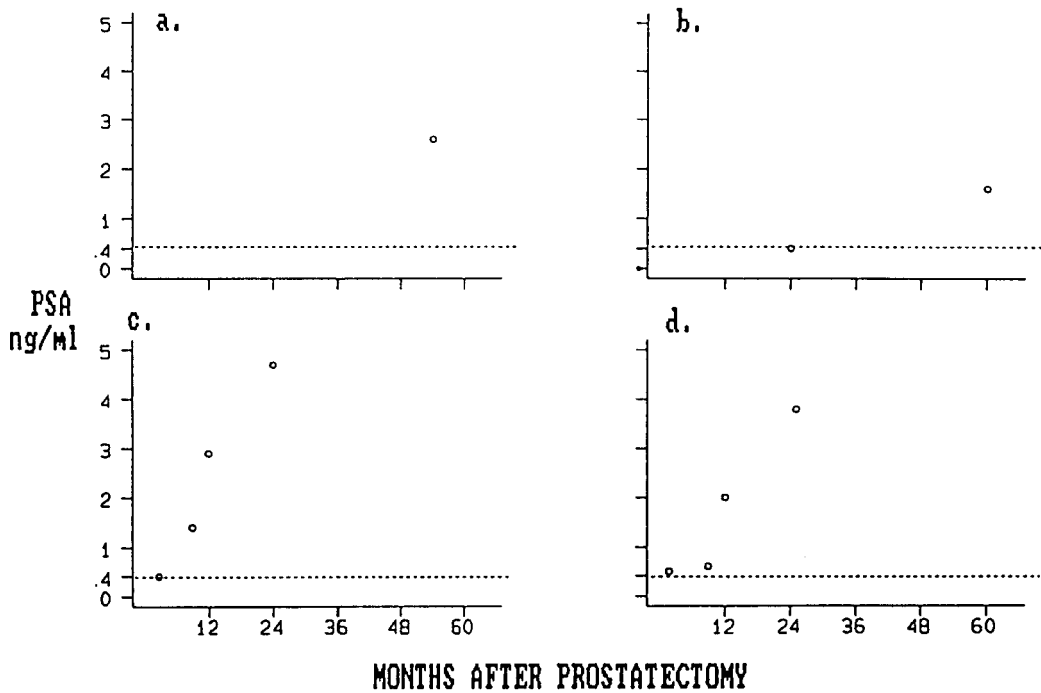


Figure 2. Detectably positive PSA readings for four patients. Patient (a) has only one reading at 54 months after surgery; patient (b) has two readings with the threshold-crossing time known; patient (c) has four readings with a known crossing time; patient (d) has four readings but the crossing time is unknown

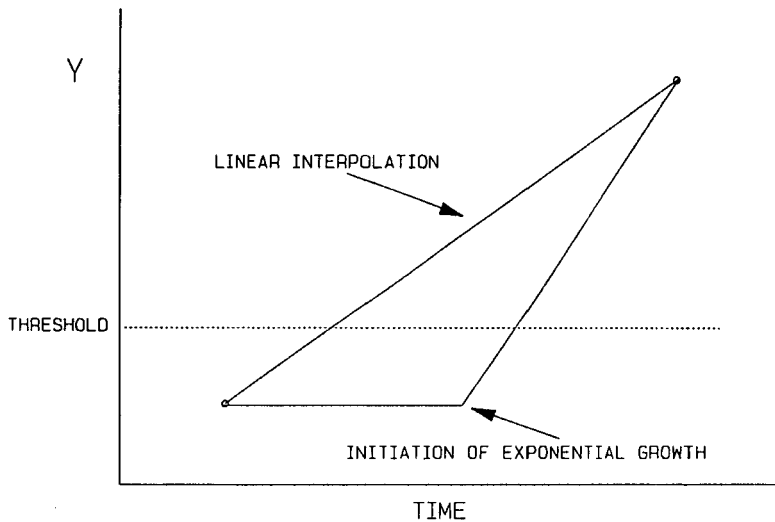


Figure 3. Two theoretical functions for the change in a variable (Y) within the censoring interval. The linear-interpolation and beta imputation methods for the hip-replacement example are based on the interpolating line, with Y denoting the maximum width of the radiolucent lines. The deterministic- and random-calibration methods for the radical-prostatectomy example are based on the piecewise linear function, with Y denoting the logarithm of the PSA value

The deterministic-calibration method

For each interval-censored case with two or more detectably positive PSA assay values, we obtain least-squares estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ of the intercept and slope in (2) and then impute the threshold-crossing time as

$$\hat{t}_C = [\log(0.4) - \hat{\beta}_0] / \hat{\beta}_1. \tag{3}$$

When only one detectably positive reading exists for an interval-censored case, say v_R at time t_R , we impute the crossing time t_C as follows. Let $\bar{\beta}_1$ be the average slope estimate over all cases (including complete ones) having two or more detectably positive values. Then we draw a line with slope $\bar{\beta}_1$ passing through the point (t_R, v_R) , and we impute the t -coordinate corresponding to $v = 0.4$ ng/ml on this line for t_C . Another way to describe this procedure is that we estimate β_1 from all of the cases having two or more detectably positive values, we estimate β_0 from the point (t_R, v_R) for the incomplete case in question, and then we impute t_C by substituting these estimates into (3).

Because the above procedures do not use the left endpoint t_L of the interval, they occasionally predict a value for t_C that lies to the left of the interval. When this occurs, we replace the predicted value for t_C by the linear-interpolation prediction, since as discussed above, linear interpolation provides an expected lower bound for t_C that lies inside the interval. We suspect that this fix-up is a minor source of bias, and we prefer the following random method.

The random-calibration method

Conditional on σ^2 , an approximate standard error for the predicted value (3) is

$$SE = (\sigma / \hat{\beta}_1) [1 + 1/q + (\hat{t}_C - \bar{t})^2 / \sum (t_i - \bar{t})^2]^{1/2}, \tag{4}$$

where q is the number of points used in estimating the intercept β_0 , t_1, t_2, \dots , are the times at which there are detectably positive readings for the interval-censored case in question, and \bar{t} is the average of these times (Reference 20, Chapter 9; Reference 21, Chapter 1). Of the three terms in square brackets in (4), the first reflects uncertainty due to predicting in individual value, the second reflects uncertainty due to estimating the intercept, and the third reflects uncertainty due to estimating the slope. The use of (4) is most appropriate when the estimate $\hat{\beta}_1$ differs significantly from zero, which was usually the case for the data set discussed here.

The random-calibration imputation method first obtains an estimate $\hat{\sigma}^2$ of σ^2 by pooling the error sums of squares over the least-squares regression fits for all cases (including complete ones) with three or more detectably positive PSA assay values and then dividing by the total degrees of freedom, say ν . For each set of imputations ($m = 1, \dots, M$), we then draw a random value of σ^2 as

$$\sigma^2 \sim \nu \hat{\sigma}^2 / \chi^2,$$

where χ^2 is a chi-square random variable with ν degrees of freedom; this step is equivalent to drawing σ^2 from its posterior distribution under the Jeffreys prior (Reference 22, Chapter 2) and is done to reflect uncertainty due to estimating σ^2 . Given the drawn value of σ^2 , we create imputations for t_C as follows. For interval-censored cases with two or more detectably positive readings, the imputed value is

$$\hat{t}_C + z SE, \tag{5}$$

where \hat{t}_C is the deterministic impute given by (3), z is a standard normal random variable, and SE is given by (4); note that q in (4) is simply the number of detectably positive readings. For

Table III. Kaplan–Meier estimates (standard errors in parentheses) of the probabilities of survival (no clinical detection) to two and four years after the first detectably positive PSA reading

	Two years	Four years
<i>Deterministic imputation</i>		
Right-point	0.455 (0.105)	0.216 (0.119)
Mid-point	0.729 (0.064)	0.589 (0.077)
Calibration	0.765 (0.062)	0.547 (0.084)
<i>Multiple imputation</i>		
Uniform		
Estimate	0.730 (0.066)	0.527 (0.092)
Ratio $((1 + M^{-1})B/\bar{U})$	9%	12%
Random calibration		
Estimate	0.749 (0.069)	0.487 (0.102)
Ratio $((1 + M^{-1})B/\bar{U})$	11%	22%

Standard errors for the deterministic methods are computed using Greenwood's formula. Standard errors for the multiple-imputation methods are given by $[\bar{U} + (1 + M^{-1})B]^{1/2}$ (see Section 2), with the U_m determined by Greenwood's formula.

interval-censored cases with only one detectably positive reading: (i) we draw a slope β_1 at random from the set of slope estimates for all cases (including complete ones) with two or more detectably positive readings; (ii) we compute t_c by the deterministic method with the drawn value of β_1 replacing $\hat{\beta}_1$; and (iii) we determine the imputed value by (4) and (5) with $q = 1$ and the term in (4) reflecting variability due to estimating the slope omitted (since we have already incorporated this variability by drawing a slope randomly).

As with the deterministic-calibration method, the random method occasionally imputes a value for t_c that lies outside the interval. When this occurs, we repeat the random procedure independently (except for the drawing of σ^2 , which is not repeated) until either we obtain a value inside the interval or we have made ten repetitions. If we do not obtain an acceptable value in ten repetitions, then we use the linearly interpolated value.

Note that the imputation method described here is quite complicated to explain, but is easy to implement, involving only linear regression fits and draws from normal and chi-square distributions. In contrast, a full maximum-likelihood solution that captures the main features of the interval-censoring process appears to us very hard to develop and implement.

4.3. Kaplan–Meier analysis of survival

We computed Kaplan–Meier estimates for the probabilities of survival to two and four years after the PSA assay first became detectably positive, with 'survival' referring to lack of clinical detection of cancer recurrence. This addresses the clinical issue discussed in Section 1. Results for the three deterministic and two random multiple-imputation methods appear in Table III.

As in the hip-replacement example, the right-point imputation method results in much lower estimated survival probabilities (higher estimated 'risk') than do the other methods, thus indicating substantial bias. The mid-point, uniform, deterministic-calibration, and random-calibration

procedures yield similar estimates, although the four-year estimates for the multiple-imputation methods are somewhat lower than their deterministic counterparts.

The assessment of variability due to missing data from the multiple imputes is useful for this example, with the estimated fractional increase in variance reaching a substantial 22 per cent for the four-year estimate.

5. DISCUSSION

This paper has proposed methods for multiply imputing interval-censored data on threshold crossings that seek to model the occurrence of the crossing within the interval in simple ways. For the two examples considered, these procedures resulted in point estimates that differed substantially from those obtained using the common fix-up of setting the threshold crossing time at the right endpoint of the interval. Typically, the right-point method yielded inflated estimates of risk compared with the other methods. We suspect that the methods proposed here are superior to the right-point method because they are based on more reasonable models for the threshold-crossing times given the observed data. We could formulate even more detailed models that incorporate, for example, subsequent outcomes of interest; this would result, however, in increasing complexities in estimating the models and resultant difficulties in implementing the imputation procedures.

As discussed in Section 3.2, the use of less detailed models assumes that the variables omitted do not contain much information about the missing values. If covariates were available that were thought to predict the threshold-crossing fraction r , then the models described could be modified to include them, although in our examples the small numbers of complete cases severely limit the number of covariates that can be incorporated satisfactorily. Variables involved in the main analyses should have relatively high priority for inclusion in the imputation model, since omitting them may lead to bias if they influence the predictive distribution of r . (See the theoretical development in Reference 13, Chapter 3.) As a limited evaluation of the effect of omitting covariates in the hip-replacement example, we examined plots for the complete cases described in Section 3.2 of the interpolation value r versus two variables that were used in the Cox regression analysis but were omitted from the imputation model; the two variables were age and time from radiographic loosening to revision surgery. The plots showed little relationship between the two omitted variables and r , implying that the variables contain little information for imputing r . Thus, for the hip-replacement example, we found no evidence that the use of less detailed models for imputation entails bias.

The fractional increase in variance due to missing data, as assessed using multiple imputation, ranged from 1 per cent to 22 per cent for the analyses presented here. These values were lower than one might have anticipated given that the missing-data rates ranged from 27 per cent (94/352, for the Cox regression analysis of Section 3) to 81 per cent (48/59, for the Kaplan–Meier analysis of Section 4). This illustrates, however, that the actual fraction of missing information generally tends to be lower than the missing-data rate for a variable if there is auxiliary information available (such as interval endpoints and other variables) that predicts the variable well. In addition, the fraction of missing information can vary greatly depending on the quantity estimated, as illustrated by the differences between the Cox regression results and the Kaplan–Meier results in Section 3.

To illustrate these ideas further, consider the simple problem of estimating the mean from a sample of values, some of which are interval censored, using the sample mean and its usual associated estimate of variance. Let m_i and v_i denote the mean and variance of the posited imputation model for the i th sample value; if the case is not interval censored then m_i is the

observed value and $v_i = 0$. Under suitable conditions, the expected value of the fractional increase in variance due to imputation over the distribution of the missing values is approximated by

$$f(1 + s_m^2/\bar{v})^{-1}, \quad (6)$$

where f is the fraction of missing values, s_m^2 is the sample variance of the m_i and \bar{v} is the sample mean of the v_i . Expression (6) shows that the expected fractional increase in variance depends not only on f , but on the ratio \bar{v}/s_m^2 ; this ratio measures the average variance of the imputed values relative to the variability in the data set. As the ratio increases from zero to infinity, approximation (6) increases from zero to f . Of course, approximation (6) does not apply directly to analyses other than estimating the mean. The principle demonstrated by (6), however, does apply in most cases.

One particular factor that can affect greatly the variability of the imputed values is the typical width of the intervals in which the missing values lie. If an interval is narrow, then the imputed value within the interval cannot have much variability even with no other information available; conversely if the interval is wide, then the imputed variable may have a high variance. This is an important factor in the radical-prostatectomy example, in which the ratio $(1 + M^{-1})B/\bar{U}$ reached 22 per cent. The PSA assay first became available in 1985 whereas the surgeries in the study date as far back as 1977; thus in some cases the first PSA reading occurred as late as eight years after surgery and led to some very wide intervals.

A possible reason for the major difference seen in Section 3 between the Cox regression analysis and the Kaplan–Meier survival analysis with respect to the fractional increase in variance due to imputation is as follows. The partial likelihood used in estimating the Cox model receives a contribution from each member of the risk set for every failure. The particular value imputed for an interval-censored case can only change the contribution to the partial likelihood for those failure times that fall inside the interval. For the hip-replacement data, let a_i denote the number of intervals containing the i th revision-surgery time and let b_i denote the size of the risk set for that time. The ratio $\Sigma a_i/\Sigma b_i$ is 11 per cent, indicating that only 11 per cent of the contributions to the partial likelihood had a chance of being affected by the imputation procedure, compared with the 27 per cent interval-censoring rate. Depending on where in the interval each failure time falls and the variability of the imputation method within the interval, the probabilities of affecting the partial likelihood are even smaller.

Further study of the fraction of missing information in interval-censoring problems is of theoretical interest. A useful feature of multiple imputation, however, is that the imputation variance is automatically included in the analysis without the need for special additional analyses. Once we had constructed reasonable imputation models for the two examples in this paper, the task of carrying out the multiple-imputation analysis was straightforward, particularly compared with the effort needed to develop and program full maximum-likelihood methods.

ACKNOWLEDGEMENTS

The authors' names appear in alphabetical order. This research was supported in part by National Institute of Mental Health Grant MH-37188. The authors thank Theodore Colton and two referees for helpful comments.

REFERENCES

1. Peto, R. 'Experimental survival curves for interval-censored data', *Applied Statistics*, **22**, 86–91 (1973).
2. Turnbull, B. W. 'The empirical distribution function with arbitrarily grouped, censored and truncated data', *Journal of the Royal Statistical Society, Series B*, **38**, 290–295 (1976).
3. Finkelstein, D. M. and Wolfe, R. A. 'A semiparametric model for regression analysis of interval-censored failure time data', *Biometrics*, **41**, 933–945 (1985).

4. Finkelstein, D. M. 'A proportional hazards model for interval-censored failure time data', *Biometrics*, **42**, 845–854 (1986).
5. Bacchetti, P. and Jewell, N. P. 'Nonparametric estimation of the incubation period of AIDS based on a prevalent cohort with unknown infection times', *Biometrics*, **47**, 947–960 (1991).
6. Little, R. J. A. 'Incomplete data in event history analysis', in Trussell, J., Harkinson, R. and Tilton, J. (eds.) *Demographic Applications of Event History Analysis*, Clarendon Press, Oxford, 1992.
7. Muñoz, A., Wang, M. C., Bass, S., Taylor, J. M. G., Kingsley, L. A., Chmiel, J. S. and Plk, F. 'Acquired immunodeficiency syndrome (AIDS) — free time after human immunodeficiency virus type 1 (HIV-1) seroconversion in homosexual men', *American Journal of Epidemiology*, **130**, 530–539 (1989).
8. Taylor, J. M. G., Muñoz, A., Bass, S., Saah, A., Chmiel, J. S., Kingsley, L. A. and the Multicenter AIDS Cohort Study. 'Estimating the distribution of times from HIV seroconversion to AIDS using multiple imputation', *Statistics in Medicine*, **9**, 505–514 (1990).
9. Heitjan, D. F. and Rubin, D. B. 'Ignorability and coarse data', *Annals of Statistics*, **19**, 2244–2253 (1991).
10. Rubin, D. B. and Schenker, N. 'Multiple imputation for interval estimation from simple random samples with ignorable nonresponse', *Journal of the American Statistical Association*, **81**, 366–374 (1986).
11. Rubin, D. B. and Schenker, N. 'Multiple imputation in health-care data bases: an overview and some applications', *Statistics in Medicine*, **10**, 585–598 (1991).
12. Rubin, D. B. 'Multiple imputations in sample surveys – a phenomenological Bayesian approach to nonresponse', *American Statistical Association 1978 Proceedings of the Section on Survey Research Methods*, 1978, pp. 20–34.
13. Rubin, D. B. *Multiple Imputation for Nonresponse in Surveys*, Wiley, New York, 1987.
14. Rubin, D. B. and Schenker, N. 'Interval estimation from multiply-imputed data: a case study using census agriculture industry codes', *Journal of Official Statistics*, **3**, 375–387 (1987).
15. Rubin, D. B. 'An overview of multiple imputation', *American Statistical Association 1988 Proceedings of the Section on Survey Research Methods*, 1988, pp. 79–84.
16. Little, R. J. A. and Rubin, D. B. 'The analysis of social science data with missing values', *Sociological Methods and Research*, **18**, 292–325 (1990).
17. Heitjan, D. F. and Rubin, D. B. 'Inference from coarse data via multiple imputation with application to age heaping', *Journal of the American Statistical Association*, **85**, 304–314 (1990).
18. Efron, B. 'Bootstrap methods: another look at the jackknife', *Annals of Statistics*, **7**, 1–26 (1979).
19. Heitjan, D. F. and Little, R. J. A. 'Multiple imputation for the fatal accident reporting system', *Applied Statistics*, **40**, 13–29 (1991).
20. Snedecor, G. W. and Cochran, W. G. *Statistical Methods*, 7th Edn., Iowa State University Press, Ames, 1980.
21. Draper, N. R. and Smith, H. *Applied Regression Analysis*, 2nd Edn., Wiley, New York, 1981.
22. Box, G. E. P. and Tiao, G. C. *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading, 1973.