1　_Article_

2　# Using a Counting Process Method to Impute

3　# Censored Follow-up Time Data

4　**Jimmy T. Efird[1,*], Charulata Jindal[1]**

5　[1]　Centre for Clinical Epidemiology and Biostatistics (CCEB), School of Medicine and Public Health, The
6　　　University of Newcastle (UoN), Australia; charulata.jindal@newcastle.edu.au (C. J.)
7　*　Correspondence: jimmy.efird@stanfordalumni.org; Tel.: +1.650.248.8282

8

9　**Abstract:** Censoring occurs when complete follow-up time information is unavailable for patients
10　enrolled in a clinical study. The process is considered to be informative (nonignorable) if the
11　likelihood function for the censoring model cannot be partitioned into a set of response parameters
12　that are independent of the censoring parameters. In such cases, estimated survival time
13　probabilities may be biased, prompting the need for special statistical methods to remedy the
14　situation. The problem is especially salient when censoring is skewed toward the early phase of a
15　study. In this paper, we describe a method to impute censored follow-up times using a counting
16　process method.
17
18　**Keywords:** counting process; censoring; Cox proportional-hazard regression; Kaplan-Meier;
19　imputation; survival analysis

20

21　**1. Introduction**

22　　　Ideally, censoring in a survival analysis should be non-informative and not related to any aspect
23　of the study that could bias results [1-7]. For example, toxic side effects of an investigational drug
24　may prompt the most ill patients to withdraw early from the study. Other patients may opt to leave
25　before the intended end of a trial if the treatment is effective and they feel well. Even when censoring
26　is non-informative (e.g., relocation to another city because of plant closure), by chance alone, it may
27　still have a serious effect on estimated survival probabilities, especially if the dropouts occur early in
28　the study.

29　　　In this paper, we present an example of early censoring to illustrate how the resulting survival
30　probabilities may be biased. We then describe a method to impute censored follow-up times by
31　rearranging the data as a counting process and generating jump-point plots.

32　**2. Materials and Methods**

33　_2.1 Imputing Censored Follow-up Times_

34　　　Let $\xi_{(j)}$ denote the follow-up time for the jth ordered observation, given a total of
35　(n) observations and (k) integer valued time (t) points

$$(t_2 = t_1 + 1, t_3 = t_2 + 1, \dots, t_k = t_{k-1} + 1; t_k \geq n). \tag{1}$$

36　Accordingly,

$$\xi_{(1)} \leq \xi_{(2)} \dots \leq \xi_{(n)} \tag{2}$$

37　and

$$t_1 \leq \xi_{(1)} \leq \xi_{(2)} \dots \leq \xi_{(n)} \leq t_k. \tag{3}$$

38 Then,

$$\forall \xi_{(j)} \wedge t_i \ni: j = 1 \; to \; n, i = 1 \; to \; k, \tag{4}$$

39 the counting process "indicator" variable (ζ) assumes the value 0 if $t_i \leq \xi_{(j)}$ and $\xi_{(j)}$ corresponds

40 to a censored follow-up time (i.e., the outcome event, such as death, has yet to occurred by time $\xi_{(j)}$).

41 Otherwise, when the observation denotes an event, (ζ) assumes the value 0 if $t_i < \xi_{(j)}$ and the

42 value 1 if $t_i \geq \xi_{(j)}$ [8].

43   Next, we fit a multiple logistic regression model with $(p+1)$ covariates to the above counting

44 process data, i.e.,

$$P(\zeta = 1|\beta_0, \beta_1, \beta_2, \dots \beta_{p+1}) = \frac{1}{1 + e^{-\{\beta_0 + \beta_1(t_i) + \beta_2(x_1) + \beta_3(x_2) + \dots + \beta_{p+1}(x_p)\}}} \tag{5}$$

45 where the first covariate denotes time (t_i) and the remaining (p) covariates $(x_1, x_2, \dots, x_p)$

46 represent outcome related risk factors. Note that the selection of this model is arbitrary and any other

47 binary predictor equation could be substituted at this step. For example, the addition of a higher

48 order term to the model or taking the logarithm of a covariate might be desirable. Other possible

49 options include using a log-binomial or negative-binomial equation to model the data [9-10]. Model

50 fit may be assessed by examining diagnostic leverage and residual plots, in lieu of standard goodness

51 of fit tests, which would assume that the underlying counting process data are independently

52 distributed [11-13].

53   A jump-point plot for a particular covariate pattern corresponding to a censored follow-up time

54 $(\xi_{(j)}^c)$ may be obtained by plotting the model predicted values $(\hat{p})$ against the time variable(t_i). Let

55 $\tilde{\xi}_{(j)}^c$ denote the value of (t_i) that maps to $\hat{p} = .50$ (i.e., maximum likelihood estimate of the mean

56 jump-point observation). The imputed censored follow-up time is given as

$$\widetilde{\text{imp}}\xi_{(j)}^c = inf\{sup(\tilde{\xi}_{(j)}^c, \xi_{(j)}^c), \xi_{max}\}, \tag{6}$$

57 where (ξ_{max}) is the natural upper limit for a patient's follow-up time. For example, in the case of

58 cancer therapy, the maximum life expectancy of a patient rarely exceeds 101 years of age. For a patient

59 aged 89 at diagnosis, we see that ξ_{max} is computed as $(101 - 89)$. The original censored follow-up

60 times are then replaced with $\widetilde{\text{imp}}\xi_{(j)}^c$. However, it is important to note that these follow-up times are

61 still treated as censored values rather than events when computing survival probability estimates.

62   Assuming Martingale independence and considering the event times to be binomially

63 distributed [i.e., the probability (π_v) of an event (ς_v) occurring in a particular risk set (v) is equal

64 to its expectation $E(\varsigma_v)$ divided by the risk set size n_v (accounting for censored and event times),

65 with corresponding variance equal to $n_v \pi_v (1 - \pi_v)$], it follows that the resulting data be may

66 analyzed using standard methods for handling censored time-to-event observations (e.g., Kaplan-

67 Meier and Cox proportional-hazards models) [14-15]. Here, the sampling variability for estimating

68 the imputed censored follow-up times is absorbed into the sampling variability for each binomially

69 distributed event time, given that the estimates are asymptotically consistent and adhere to certain

70 regularity conditions.

71 **3. Results**

72 *3.1 Kaplan-Meier (Product-Limit) Example*

73   Consider the data in Appendix A, which provides the event and censored follow-up times for

74   250 cancer patients undergoing treatment and their simulated complete values (for illustrative

75   purposes). Approximately 10% of the follow-up times were censored, with the majority of these

76   values occurring early in the study. Figure 1 shows a Kaplan-Meir (KM) plot comparing the original

77   censored data with the dataset of complete follow-up times. The probability of surviving 5 years (60

78   months) was ~26% for the original censored data compared with ~18% for the complete dataset. In

79   Figure 2, we compare the complete dataset with the imputed censored dataset. At 5 years, rounding

80   to the nearest whole number, we see that the survival times are identical (i.e., 26%). Indeed, the

81   survival curves are similar until ~90 months, at which point the survival times for the imputed

82   censored time curve are divergently lower than those for the complete dataset. This lack of fit at the

83   extreme end of the KM curves is evident when examining the panel of diagnostic plots in Figure 3.
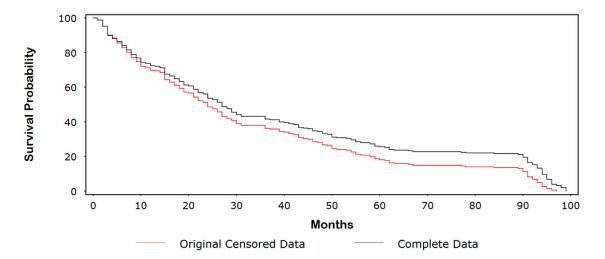

84   Figure 1. Kaplan-Meir (KM) plot comparing the original censored data with the dataset of complete follow-up times.
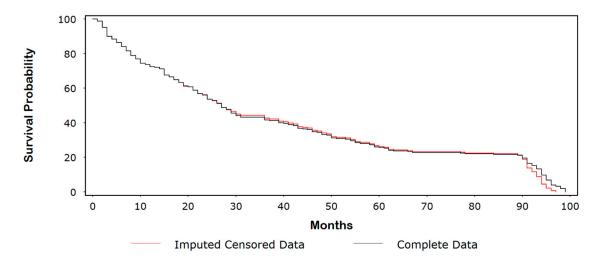

85   Figure 2. Kaplan-Meir (KM) plot comparing the imputed censored data with the dataset

86   of complete follow-up times.

87

○ Raw + Projected

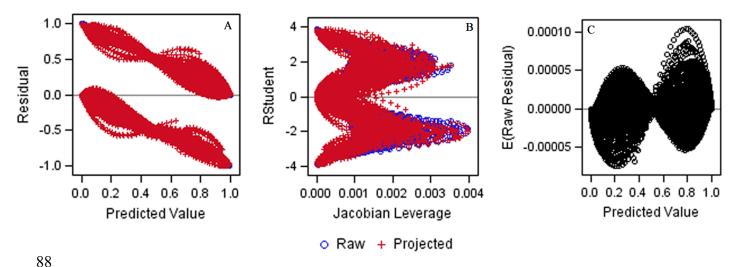88   Figure 3. Diagnostic plots. A: Residual by predicted values; B: RStudent by Jacobian
89   leverage; C: Expected raw residuals by predicted values.

90   *3.2 Generating The Jump-Point Plot*

91   The jump-point plot corresponding to covariate pattern

$$(x_1 = 76, x_2 = 3, \ x_3 = 0, x_4 = 0, x_5 = 1, x_6 = 0) \tag{7}$$

92   is shown in Figure 4. Here, the variables $(x_1, x_2, \dots, x_6)$ correspond to age (years), grade (I–IV), lymph
93   node invasion (1=yes, 0=no), positive margins (1=yes, 0=no), no hormone therapy (1=yes, 0=no), and
94   no radiation therapy (1=yes, 0=no), respectively. We see that the imputed censored follow-up time of
95   29.33 months closely matches the actual event time of 29 months. This observation was originally
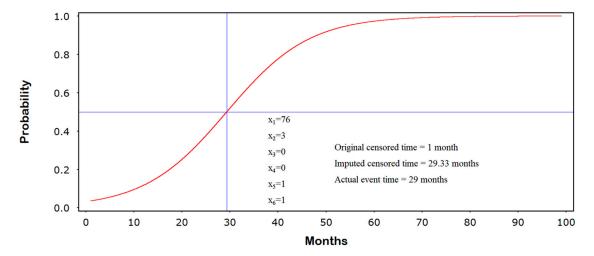96   censored at 1 month.

97   Figure 4. Jump-point plot.

98   *3.3 SAS Code*

99   The SAS code use to generate the jump-point plot is shown in Figure 5. In this code, it is assumed
100   that the data contained in Appendix A has been previously read into the dataset "a". The counting
101   process variables are created in dataset "b" and then modeled using the "PROC LOGISTIC"

102   procedure. The predicted probabilities generated by this procedure are plotted against time (ranging
103   from 1 to 99 months) to obtain the jump-point plot. Analyses were performed using SAS Version 9.4.
104   (Cary, NC).

```
data b;
   set a;
      do j=1 to 99;
         if os_censor=1 then do;
                              if j<os_tm        then dead=0;
                                 else if j>=os_tm then dead=1;
                              tm=j;
                              output;
                           end;
         else if os_censor=0 then do;
                                    if j<=os_tm then do;
                                                        tm=j;
                                                        dead=0;
                                                        output;
                                                     end;
                                 end;
      end;
ods output "Parameter Estimates"=parest;
proc logistic data=b descending;
   model dead=tm agex gradex nstage posmarg hormno rtno/risklimits clodds=pl;
run;
quit;
ods output close;
proc transpose data=parest out=out1;
   var estimate;
data flipparest (keep=col1-col8 rename=(col1=beta0 col2=beta1 col3=beta2 col4=beta3
                  col5=beta4 col6=beta5 col7=beta6 col8=beta7));
   set out1;
data c;
   set flipparest;
      agex=76;
      gradex=3;
      nstage=0;
      posmarg=0;
      hormno=1;
      rtno=1;
      do tm=1 to 99 by 0.0001;
         pdead=1/(1+exp(-(beta0+beta1*tm+beta2*agex+beta3*gradex+beta4*nstage+beta5*posmarg
               +beta6*hormno+beta7*rtno)));
      output;
      end;
proc plot data=c;
   plot pdead*tm;
run;
```

105   Figure 5. SAS code used to generate jump-point plot.

106   **4. Discussion**

107   In this manuscript we have introduced a simple method to partially correct for nonignorable
108   early censoring. By rearranging the data as a counting process we are able to account for the follow-
109   up times of all patients, regardless of their censoring status. For example, if a patient is censored at
110   50 months, the counting process creates 50 observations corresponding to each month and
111   accordingly assigns the value of 0 to an indicator variable. On the other hand, if the patient had an
112   event at 50 months, the counting process would create 49 observations with the indicator variable set

113   to 0 and similarly create new observations for each month thereafter until the last month of the study,
114   but instead assigning the value of 1 to the indicator variable.

115   These counting responses are then analyzed with logistic regression, or another appropriate
116   model for fitting binary data. In addition to including outcome related covariates, a variable denoting
117   the time (e.g., month) associated with the indicator variable is added to the model. The predicted
118   value generated by the model for a particular covariate pattern (associated with a censored
119   observation) is then plotted against the time variable (spanning each month of the study) to obtain a
120   jump-point plot. Dropping a line from the midpoint of this plot to the x-axis gives an imputed
121   censored follow-up time. We then replace the original censored time with this value if it larger of the
122   two values. Also, the imputed value is constrained by a natural upper bound to prevent impossible
123   censored follow-up times (e.g., the value must be consistent with a patient's maximum biologic
124   lifespan).

125   An important aspect of this technique is identifying a well-fitting model for the counting process
126   data so that it is able to accurately predict if the outcome is 0 or 1. Understanding the dynamics of the
127   disease or process under study will aid in the selection of appropriate outcome-related covariates.
128   However, formally testing for model goodness of fit is not practical given the highly correlated nature
129   of the counting process data. In theory, while it may be possible to assess model goodness of fit for
130   dependent data using a robust "Huber-White" approach, the regularity conditions for such estimates
131   are quite stringent [16-18]. Instead, we recommend using leverage and residual diagnostic plots to
132   rule out ill-fitting models [11-13]. In some cases, including power and trigonometric terms into the
133   model may potentially improve the efficiency of the fitting algorithm.

134   The advantage of using a counting process approach is that the imputed censored follow-up
135   times, when appropriately constructed, will better reflect the survival prospects of those who
136   continued in the study. However, because the method is modeled based, it may not be suitable for
137   small datasets or those lacking a set of reasonably predictive covariates. Additionally, it may not
138   always be possible to identify a well-fitting model if there are abrupt changes in the hazard function
139   of the underlying data. The method at hand should not be used if patients who enroll later in study
140   survive longer (e.g., treatment improves over time) or if enrollment criteria change over the course
141   of the study (e.g., worst patients are excluded midway through recruitment) [5].

142   **5. Conclusions**

143   Overall, the best means for handling informative censoring is to avoid the problem in the first
144   place. Careful planning at the study design stage, routine patient monitoring, and implementing
145   proactive strategies to minimize patient dropout are some important steps to ensure the fidelity of a
146   survival time study.

147   While it was beyond the scope of the current manuscript, it will be informative in future analyses
148   to compare our method with other approaches for dealing with censored values, especially
149   highlighting best and worst-case scenarios [3, 19-24].

152 **Author Contributions:** Jimmy T. Efird and Charulata Jindal conceive and design the project, analyze the data,
153 and wrote the manuscript. Both authors approved the final version of the manuscript.

154 **Conflicts of Interest:** The authors declared no conflict of interest.

155

156

152 **Author Contributions:** Jimmy T. Efird and Charulata Jindal conceive and design the project, analyze the data,
153 and wrote the manuscript. Both authors approved the final version of the manuscript.

157   **Appendix A. Example cancer dataset (N=250)**

A	B	C	D	E	F	G	H	I	J	K
1	93	4	1	1	1	1	1	1	--	1
2	90	3	1	1	1	1	1	1	--	1
3	89	4	1	1	1	1	1	1	--	1
4	89	4	1	0	1	1	2	2	--	1
5	86	3	1	1	1	0	2	2	--	1
6	86	3	1	0	1	1	2	2	--	1
7	95	4	1	1	1	1	2	2	--	1
8	90	3	1	1	1	1	2	2	--	1
9	86	3	0	1	1	1	2	2	--	1
10	84	4	1	0	0	1	2	2	--	1
11	82	4	1	0	1	1	2	2	--	1
12	91	2	1	1	1	1	2	2	--	1
13	96	4	1	0	0	1	3	3	--	1
14	95	3	1	0	1	1	3	3	--	1
15	89	2	1	1	1	1	3	3	--	1
16	88	4	1	0	1	0	3	3	--	1
17	91	3	1	1	0	1	3	3	--	1
18	91	3	1	1	1	1	3	3	--	1
19	90	4	1	1	1	1	3	3	--	1
20	98	4	1	0	1	1	3	3	--	1
21	94	3	1	1	1	1	3	3	--	1
22	96	3	1	0	1	1	3	3	--	1
23	88	3	1	0	1	1	3	3	--	1
24	95	3	0	1	1	1	3	3	--	1
25	83	3	1	0	1	1	3	3	--	1
26	88	3	1	0	1	1	4	4	--	1
27	95	4	1	1	1	1	4	4	--	1
28	95	2	1	1	0	1	4	4	--	1
29	92	3	1	1	0	0	4	4	--	1
30	96	4	1	0	1	1	5	5	--	1
31	96	3	1	0	1	0	5	5	--	1
32	90	3	1	1	0	1	5	5	--	1
33	89	4	1	1	1	1	5	5	--	1
34	87	3	1	0	1	1	5	5	--	1
35	97	2	1	1	1	1	6	6	--	1
36	96	4	1	0	1	1	6	6	--	1
37	93	3	1	0	1	1	6	6	--	1
38	94	3	1	0	0	1	6	6	--	1
39	93	3	1	1	1	1	6	6	--	1
40	97	3	1	1	1	1	6	6	--	1
41	97	3	1	0	1	1	7	7	--	1
42	97	2	1	0	1	1	7	7	--	1
43	86	3	1	1	1	1	7	7	--	1
44	85	3	1	0	1	1	7	7	--	1
45	91	3	1	0	1	1	7	7	--	1
46	94	3	1	1	1	1	7	7	--	1
47	86	3	1	0	1	1	8	8	--	1
48	86	4	1	0	1	1	8	8	--	1
49	83	1	1	1	0	1	8	8	--	1
50	96	3	1	1	1	1	8	8	--	1
51	94	4	1	0	0	1	8	8	--	1
52	91	3	1	1	1	1	8	8	--	1
53	99	3	1	1	1	1	8	8	--	1
54	93	4	1	1	0	0	9	9	--	1
55	92	2	1	1	0	1	9	9	--	1
56	98	2	1	0	1	1	9	9	--	1
57	83	4	1	1	0	1	9	9	--	1
58	48	3	1	0	1	1	9	9	--	1
59	87	2	1	1	1	1	10	10	--	1
60	83	4	1	1	0	1	10	10	--	1
61	83	3	1	0	0	0	10	10	--	1
62	88	4	0	1	1	1	10	10	--	1
63	97	2	1	0	1	1	10	10	--	1
64	87	3	1	1	1	1	10	10	--	1
65	89	4	1	1	1	1	11	11	--	1
66	89	2	1	1	1	1	11	11	--	1
67	84	2	0	0	0	1	12	12	--	1
68	94	4	1	0	0	1	12	12	--	1
69	97	3	1	1	0	1	12	12	--	1
70	96	2	1	1	0	1	13	13	--	1
71	99	3	0	0	1	1	14	14	--	1
72	83	2	1	1	1	1	14	14	--	1
73	85	4	1	1	0	1	15	15	--	1
74	83	3	1	1	0	1	15	15	--	1
75	81	3	1	0	1	1	15	15	--	1
76	98	1	1	0	0	1	15	15	--	1
77	94	3	1	0	0	1	15	15	--	1
78	93	4	1	0	0	0	15	15	--	1
79	90	2	1	0	1	1	15	15	--	1
80	88	3	1	0	1	1	15	15	--	1
81	82	3	1	0	1	1	15	15	--	1
82	85	4	1	1	0	0	16	16	--	1
83	88	2	1	0	0	0	16	16	--	1
84	82	3	1	0	1	0	16	16	--	1
85	84	4	1	1	0	0	17	17	--	1
86	90	2	1	1	0	1	17	17	--	1
87	93	3	1	0	1	0	17	17	--	1
88	90	4	1	1	1	1	17	17	--	1
89	88	3	0	0	1	1	18	18	--	1
90	97	3	1	1	1	1	18	18	--	1
91	94	4	1	1	0	0	18	18	--	1
92	85	2	1	0	1	1	18	18	--	1
93	84	4	1	0	1	1	19	19	--	1
94	83	2	1	1	1	1	19	19	--	1
95	99	4	1	0	0	1	19	19	--	1
96	93	3	1	1	0	1	19	19	--	1
97	83	2	1	0	0	0	19	19	--	1
98	81	3	0	0	1	1	20	20	--	1
99	87	3	1	1	0	1	21	21	--	1
100	85	4	1	1	1	1	21	21	--	1
101	85	3	1	1	0	1	21	21	--	1
102	82	3	1	0	0	1	21	21	--	1
103	80	3	1	1	1	1	21	21	--	1
104	81	2	1	1	1	1	22	22	--	1
105	80	4	1	0	0	1	22	22	--	1
106	91	3	0	1	0	1	22	22	--	1
107	75	3	1	0	1	0	22	22	--	1
108	72	4	0	0	1	1	22	22	--	1
109	83	3	1	1	0	0	23	23	--	1
110	79	3	1	1	1	1	23	23	--	1
111	79	4	1	0	0	1	24	24	--	1
112	78	3	1	1	0	1	24	24	--	1
113	75	4	1	0	0	1	24	24	--	1
114	84	3	1	0	0	1	24	24	--	1
115	78	2	1	1	1	1	24	24	--	1
116	78	2	1	0	1	1	24	24	--	1
117	79	3	1	0	0	1	25	25	--	1
118	73	3	1	1	1	1	25	25	--	1
119	73	4	1	0	0	1	26	26	--	1
120	72	2	1	0	0	0	26	26	--	1
121	75	2	1	0	0	0	26	26	--	1
122	69	3	1	1	1	0	26	26	--	1
123	89	2	1	0	0	1	27	27	--	1
124	78	3	1	0	0	0	27	27	--	1
125	78	4	1	1	1	1	27	27	--	1
126	72	3	1	1	0	1	27	27	--	1
127	63	2	1	0	1	1	27	27	--	1
128	62	4	1	1	0	1	27	27	--	1
129	79	3	1	0	0	1	28	28	--	1
130	73	3	1	0	1	0	28	28	--	1
131	73	2	1	0	0	0	28	28	--	1
132	76	3	0	0	1	1	29	1	29.33	0
133	79	4	1	0	0	1	29	4	18.88	0
134	84	4	1	0	0	1	29	29	--	1
135	82	2	0	0	0	1	29	29	--	1
136	78	2	1	0	1	1	29	3	26.73	0
137	78	3	0	0	0	1	30	30	--	1
138	73	4	0	0	0	1	30	30	--	1
139	60	3	1	0	0	0	30	30	--	1
140	71	2	1	1	0	0	30	30	--	1
141	78	2	0	0	0	0	31	31	--	1
142	64	4	1	0	0	0	31	31	--	1
143	69	3	1	0	1	1	36	36	--	1
144	68	2	1	0	0	0	36	36	--	1
145	62	2	1	0	0	0	36	36	--	1
146	58	3	1	0	1	1	36	36	--	1
147	69	4	1	0	0	1	37	37	--	1
148	68	3	0	0	0	1	39	39	--	1
149	65	3	1	0	0	1	39	39	--	1
150	65	3	1	1	0	0	39	39	--	1
151	68	3	1	1	1	0	40	40	--	1
152	68	4	1	0	1	1	41	41	--	1
153	63	3	1	1	0	1	41	41	--	1
154	62	3	0	0	1	0	42	42	--	1
155	68	1	1	0	1	1	43	43	--	1
156	68	2	0	0	0	0	43	43	--	1
157	69	2	0	0	1	1	43	43	--	1
158	62	3	1	0	1	1	43	43	--	1
159	67	3	0	1	1	1	44	44	--	1
160	68	2	1	0	1	1	45	45	--	1
161	64	3	1	0	0	1	46	46	--	1
162	62	2	1	1	0	0	46	46	--	1
163	64	3	0	0	0	0	46	46	--	1
164	61	3	0	0	1	1	47	47	--	1
165	58	4	1	1	0	0	48	48	--	1
166	52	3	0	0	0	1	48	48	--	1
167	51	2	1	0	0	1	48	48	--	1
168	51	2	1	0	1	1	49	49	--	1
169	53	3	1	0	0	1	50	50	--	1
170	81	3	0	0	0	1	50	50	--	1
171	77	3	1	1	0	0	50	50	--	1
172	44	3	1	0	1	1	50	50	--	1
173	68	3	1	0	0	1	51	51	--	1
174	72	3	0	1	0	0	53	53	--	1
175	87	3	0	1	1	1	54	54	--	1
176	74	4	1	0	0	1	54	54	--	1
177	52	1	0	0	0	1	55	55	--	1
178	69	3	1	0	0	0	55	55	--	1
179	41	3	0	0	1	1	55	55	--	1
180	64	3	1	0	0	1	56	56	--	1
181	64	2	1	0	1	1	58	58	--	1
182	58	3	1	0	1	1	58	58	--	1
183	53	2	0	1	1	1	59	59	--	1
184	53	3	1	0	0	1	59	59	--	1
185	58	1	0	0	0	0	59	59	--	1
186	54	3	1	0	0	1	60	60	--	1
187	54	2	0	0	1	1	61	61	--	1
188	67	3	1	1	0	0	62	62	--	1
189	66	2	1	0	0	1	62	62	--	1
190	53	2	0	0	0	0	62	62	--	1
191	56	3	1	1	1	1	63	63	--	1
192	53	2	1	0	0	1	66	66	--	1
193	49	3	1	1	1	1	67	67	--	1
194	69	2	1	0	1	1	77	77	--	1
195	69	2	0	1	0	0	78	78	--	1
196	54	2	1	0	1	1	84	84	--	1
197	49	2	1	0	0	1	89	89	--	1
198	40	3	0	1	1	1	90	90	--	1
199	40	3	1	1	0	0	90	90	--	1
200	40	2	1	0	1	1	90	90	--	1
201	45	2	1	1	1	1	90	90	--	1
202	46	3	1	1	1	1	91	2	73.22	0
203	47	2	0	0	0	1	91	91	--	1
204	46	2	1	0	0	1	91	91	--	1
205	49	1	1	0	1	1	91	91	--	1
206	50	1	0	1	0	1	91	91	--	1
207	47	2	1	1	1	1	91	91	--	1
208	41	2	1	0	1	0	91	91	--	1
209	46	2	0	1	0	0	91	91	--	1
210	46	2	0	0	1	1	92	92	--	1

158   **Appendix A (Continued)**

159

160

A	B	C	D	E	F	G	H	I	J	K	A	B	C	D	E	F	G	H	I	J	K
211	49	3	1	0	1	1	92	92	--	1	246	47	1	0	0	1	1	99	3	85.42	0
212	52	2	1	0	1	1	92	92	--	1	247	43	2	0	0	1	1	99	10	86.12	0
213	47	2	0	0	1	1	93	4	79.9	0	248	42	1	1	0	1	1	99	5	88.18	0
214	48	2	1	0	0	0	93	93	--	1	249	42	1	1	0	0	0	99	2	90.24	0
215	44	2	1	1	0	1	93	93	--	1	250	46	1	0	0	0	1	99	4	91.71	0
216	50	3	1	0	0	1	93	93	--	1											
217	40	1	0	0	0	1	93	93	--	1											
218	40	3	1	0	1	1	94	2	80.3	0											
219	56	2	1	1	0	1	94	94	--	1											
220	51	2	1	0	1	1	94	94	--	1											
221	48	2	0	0	0	1	94	94	--	1											
222	51	1	0	0	1	1	94	94	--	1											
223	51	2	1	1	1	1	94	94	--	1											
224	52	1	1	1	1	1	94	94	--	1											
225	48	2	0	1	0	1	94	6	85.4	0											
226	42	2	1	0	0	1	94	1	87.4	0											
227	49	1	0	0	1	1	95	95	--	1											
228	49	2	1	1	1	1	95	95	--	1											
229	51	2	1	0	1	1	95	2	68.7	0											
230	51	3	0	0	1	1	95	95	--	1											
231	45	1	1	0	0	0	95	8	85.6	0											
232	45	1	1	1	0	1	95	4	90.5	0											
233	43	2	0	0	0	0	95	2	88.2	0											
234	47	1	1	1	0	1	96	3	87.4	0											
235	42	1	1	0	1	1	96	96	--	1											
236	46	3	0	0	1	1	96	1	76	0											
237	47	2	1	1	1	1	96	96	--	1											
238	45	2	0	0	1	1	96	4	83	0											
239	46	1	1	1	1	1	96	7	84.2	0											
240	46	2	0	1	0	1	96	1	88.5	0											
241	46	1	1	0	0	1	97	7	86.7	0											
242	46	2	1	1	1	1	97	97	0	1											
243	46	3	0	1	1	1	98	4	78.2	0											
244	46	2	0	0	1	1	98	5	81.5	0											
245	46	3	1	0	0	1	98	2	75.7	0											

161

162

163

164

165

166

167

168

169

170

171

172

173   A=Observation number, B=Age, C=Grade, D= Lymph node invasion, E=Positive margin, F=No

174   hormonal therapy, G=No radiation therapy, H=Actual follow-up time, I=Original censored time,

175   J=Imputed censored time, K=Censoring variable.

176

177   **References**

178   1. Ranganathan, P.; Pramesh, C. S., Censoring in survival analysis: Potential for bias. Perspect Clin Res 2012, 3

179   (1), 40.

180   2. Lagakos, S. W., General right censoring and its impact on the analysis of survival data. Biometrics 1979, 35

181   (1), 139-56.

182   3. Wu, M. C.; Carroll, R. J., Estimation and Comparison of Changes in the Presence of Informative Right

183   Censoring byModeling the Censoring Process. Biometrics 1988, 44 (1), 175-188.

184   4. Leung, K. M.; Elashoff, R. M.; Afifi, A. A., Censoring issues in survival analysis. Annu Rev Public Health

185   1997, 18, 83-104.

186   5. Zhang, J.; Heitjan, D. F., Nonignorable censoring in randomized clinical trials. Clin Trials 2005, 2 (6), 488-96.

187   6. Shih, W., Problems in dealing with missing data and informative censoring in clinical trials. Curr Control

188   Trials Cardiovasc Med 2002, 3 (1), 4.

189   7. Miettinen, O. S., Survival analysis: up from Kaplan-Meier-Greenwood. Eur J Epidemiol 2008, 23 (9), 585-92.

190   8. Collett, D., Modelling Survival Data in Medical Research. Third ed.; Taylor & Francis group: Florida, USA,

191   2015.

192   9.   Blizzard, L.; Hosmer, D. W., Parameter estimation and goodness-of-fit in log binomial regression. Biom J
193   2006, 48 (1), 5-22.

194   10. Hilbe, J. M., Negative Binomial Regression. Second ed.; Cambridge University Press: 2011.

195   11. Laurent, R. T. S.; Cook, R. D., Leverages and Superleverages in Nonlinear Regression. J Am Stat Assoc 1992,
196   87 (40), 985-990.

197   12. Laurent, R. T. S.; Cook, R. D., Leverages, Local Influence, and Curvature in Nonlinear Regression. Biometrika
198   1993, 80 (1), 99-106.

199   13. Pregibon, D., logistic regression diagnostic Ann. Stat. 1981, 9 (4), 705-724.

200   14. Williams, D., Probability with Martingales. Cambridge University Press: New York, 1991.

201   15. Crowder, M. J.; Kimber, A. C.; Smith, R. L.; Sweeting, T. J., Statistical Analysis of Reliability Data. CHAPMAN
202   & HALL: London, 1991.

203   16. Huber, P. J. (1967). "The Behavior of Maximum Likelihood Estimates under Nonstandard Conditions,"
204   Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. I, pp. 221–33,
205   1967.

206   17. White, H., Maximum likelihood estimation of misspecified models. Econometrica 1982, 50 (1), 1-25.

207   18. Freedman, D. A., On The So-Called "Huber Sandwich Estimator" and "Robust Standard Errors".
208   https://www.stat.berkeley.edu/~census/mlesan.pdf. 2006.

209   19. Diggle, P.; Kenward, M. G., Informative drop-out in longitudinal data analysis. J R Stat Soc Ser C Appl Stat
210   1994, 43 (1), 49-93.

211   20. Plante, J. F., About an adaptively weighted Kaplan-Meier estimate. Lifetime Data Anal 2009, 15 (3), 295-315.

212   21. Hogan, J. W.; Laird, N. M., Model-based approaches to analysing incomplete longitudinal and failure time
213   data. Stat Med 1997, 16 (1-3), 259-72.

214   22. Wu, M. C.; Albert, P. S.; Wu, B. U., Adjusting for drop-out in clinical trials with repeated measures: design
215   and analysis issues. Stat Med 2001, 20 (1), 93-108.

216   23. Wei, L.; Shih, W. J., Partial imputation approach to analysis of repeated measurements with dependent drop-
217   outs. Stat Med 2001, 20 (8), 1197-214.

218   24. Little, R. J. A., Modeling the Drop-Out Mechanism in Repeated-Measures Studies. J Am Stat Assoc. 1995, 90
219   (431), 1112-1121.

220