Estimating Adverse Outcome Incidence Rate Imputing Partial Status with Application to a Phase IV Cancer Trial

Shesh N Rai[1], Jianmin Pan[1], Xiaoyong Wu[1], Pradeep Singh[2],
Melissa M Hudson[3] and Deo K Srivastava[3]
[1] University of Louisville, Louisville, KY, USA,
[2]Southeast Missouri State University, Cape Girardeau, MO, USA
[3] St. Jude Children's Research Hospital, Memphis, TN, USA

Abstract
Phase IV clinical trials are designed to monitor the long-term toxic effects of drugs in cancer survivors. Evaluations to study the long-term effects of the cancer treatment are often made in cross-sectional surveys. In addition to finding prognostic factors for log-term survival outcome, estimating the cumulative incidence rates of adverse outcomes is also desired. Such data pose many issues: incomplete data, competing risks and selection bias. For example, one such study was designed to study the effect of anthracyclines exposure, received as part of treatment for childhood cancer, to cardiotoxicity.[1] In this paper, we resort to imputing the missing current status using regression method, under some parametric assumptions, and then combining with methods previously described in Rai et al.[2] to estimate the cumulative incidence rates in an illness-death/failure model. A comprehensive simulation study suggests that the results obtained using the imputation approach is significantly more efficient than those obtained without imputation. We further apply the proposed approach to the data reported in Rai et al.[2] and compare the results reported there to our approach that utilizes imputation.

Keywords
Phase IV clinical trial, Imputation, Cross-section survey data, Interval censored data, K-M Method, Missing Value

1. Introduction

Recent advance for treatment of childhood cancer through clinical studies has benefited patients and has significantly improved their long term survival. However, in addition to monitoring improved survival the focus of phase IV clinical trials is also to monitor for long-term side effects of primary treatment of cancer. For example, anthracyclines remain a critical component of the treatment for many pediatric malignancies because of their favorable therapeutic benefits[1,3-4]. However, it is well-known that one of the long-term complications of the treatment with anthracyclines is the increased risk of cardiotoxicity. There are different types of drug toxicity, such as cytotoxicity, carcinogenicity, mutagenicity, teratogenicity and cardiotoxicity. Just as drugs come in many varieties and with varying beneficial effects, there are various ways in which they can cause potentially life threatening toxicities. It is not just the drug itself that causes the toxicity, but a breakdown product which remains active and has unwanted effects. There are even more complex situations, such as in the case of thalidomide, where the drug taken is broken down by the body into two components, one of which acts to reduce morning sickness for pregnant women, while the other causes malformation of the fetus.

Cardiotoxicity is the occurrence of heart electrophysiology dysfunction or/and muscle damage. The heart becomes weaker and is not as efficient in pumping and therefore circulating blood. There are many measures of electrophysiology dysfunction

or/and muscle damage, including shortening fraction, afterload, QTc interval, ejection fraction. It is not economic and feasible to evaluate patients very frequently to estimate the onset time of cardiotoxicity, and hence estimate the incidence rates. Usually patients are followed longitudinally in the clinics, but not all follow a routine pattern. Therefore, it is convenient to design cross-sectional surveys for estimating the effect of long-term side effect of treatments and its predictors. We only know the current status of the patient with onset prior to current status but not the actual onset times of these events. These types of incomplete data are referred to as interval censored data since the actual onset time of the events are unknown[2-3] and our interest is in estimating the onset rate or the cumulative incidence rate.

Nonparametric procedures for analyzing interval censored failure time data have been extensively studied and discussed in the literature.[5-9] Another issue in the cross-section survey study is that results need to be generalized to the specific population. There can be competing toxic effects from the same drug. Sun[9] provides an extensive survey of non-parametric methods of estimation using EM algorithm in studies involving interval censored data. In this paper, we have the same interest, as Rai, et al.[2], in estimating the cumulative incidence rates in a parametric setting but focus on improving the accuracy by imputing the missing observations using multivariable regression method.

In application, most investigators exclude observations with missing values, incomplete cases. While using only complete cases has its simplicity, one may lose the important information in the incomplete cases and ignore the possible systematic difference between the complete and incomplete cases. Hence, the resulting inference may not be applicable to the population of all cases, especially with a smaller number of complete cases. It is well known that imputation is a widely used method for handling missing data. Little and Rubin[10] provide an excellent overview of the methods for conducting analyses with missing data. For further information on multiple imputations in particular, see Rubin.[11-12] Rubin, Stern, and Vehovar[13] give a simple example about discrete data. King et al.[14] review many of the practical costs and benefits of multiple imputations. For routine imputation of missing data, Schafer[15] presents a method for the multivariate normal distribution. Liu[16] uses the *t* distribution, and Van Buuren, Boshuizen, and Knook[17] use interlocking regressions. Abayomi, Gelman, and Levy[18] discuss problems for fitting imputation models. Furthermore, Troxel, Ma, and Heitjan[19] present a method to study the sensitivity of inferences to missing-data assumptions.

This paper is organized as follows. In Section 2, we give a motivational example to introduce the problem. In Section 3, we give a brief description of the procedure introduced in Rai, et al.[2] and construct corresponding likelihood function. The data from the motivation example is analyzed by imputing the missing values and compared with the results obtained without imputation in Section 4. The results of an extensive simulation experiment to study the performance of the imputation approach are summarized in Section 5. Section 6 is devoted to conclusions and miscellaneous remarks. Appendix can be found in Section 7.

2. Motivation Example

Study participants were recruited from the population of a long-term follow-up clinic that monitors childhood cancer survivors on an annual basis. Enrollment targeted five diagnostic groups treated with cardiotoxic antineoplastic agents: acute leukemia, a group receiving low cumulative dosages of anthracycline (100 mg/m^2); mediastinal lymphoma, a group with potential cardiovascular injury from anthracyclines and/or thoracic irradiation; sarcoma, a group receiving relatively high cumulative dosages of anthracyclines (300 to 500 mg/m^2); and neuroblastoma, a group with cardiac immaturity

at anthracycline administration. A control survivor group was also recruited (acute lymphoblastic leukemia, Wilms' tumor, and germ cell tumors) that was not exposed to anthracyclines or thoracic irradiation. Survivors with a history of congenital heart disease, chronic systemic illness requiring ongoing medical treatment, trisomy 21, and anemia (hematocrit < 28%) were excluded from participation. Consecutive patients meeting the eligibility criteria were identified and invited to participate. Enrollment was closed for the specific diagnostic groups after reaching the protocol-specified targeted accrual yielding a convenience cohort of study participants. Survivors were designated At Risk (AR) or No Risk (NR) based on treatment with potentially cardiotoxic agents.

The study methods involved a one-time clinical cardiac assessment by the primary oncologist to evaluate for signs of heart failure and assignment to the New York Heart Association class. Noninvasive cardiac testing was performed within 24 hours of the clinical assessment and included 12-lead ECG and echocardiography.

Cardiotoxicity is the occurrence of heart electrophysiology dysfunction or/and muscle damage. The heart becomes weaker and is not as efficient in pumping and therefore circulating blood. There are many measures of electrophysiology dysfunction or/and muscle damage, including shortening fraction, afterload, QTc interval, ejection fraction.[1, 4, 20] When these cardiac measures are out of the normal range, patients are declared to have cardiotoxicity. Following Hudson et al.[1], we consider two outcome measures, fractional shortening (*FS*) and afterload (*AF*). The main measure is defined as $FS = (LVEdD - LVEsD)/LVEdD$, where *LVEdD* is the left ventricular end-diastolic diameter, *LVEsD* is the left ventricular end-systolic diameter. The other measure *AF* can be described as the pressure that the chamber of the heart has to generate in order to eject blood out of the chamber. The study was planned to enroll almost equal number of patients from each group to detect a medium effect size[21] increase in mean *AF* (or decrease in mean *FS*) at $\alpha = 0.05$ and $\beta = 0.20$, without adjusting for multiple outcomes or multiple comparisons.

A short summary about the cohort is summarized in Table 1. Further description is available in Hudson et al.[1]. The scatter plot between AF and FS displayed in Figure 1 does not show any clear missing pattern; the missing values of AF are in the entire range of values of FS. Also note that the missing values of AF are almost same proportions in the NR and AR groups.

Using actual measures of these dependent variables *FS* and *AF*, the threshold values were used to classify patients as abnormal if ($FS < 0.28$) or ($AF > 74$ g/cm^2). Threshold values for FS and AF were determined based on published normative data[25-26]; these are well accepted norms. Let *AFS* and *AAF* denote the indicators of these abnormalities. The 278 patients participated in the study, representing 22% of the clinic population of 1,268 patients; 223 were designated AR and 55 were designated NR based on treatment. Data on each individual also included demographics, date of cancer diagnosis, time since treatment completion, disease related variables (such as type, histology, and stage of cancer), treatment related variables (such as chemotherapy drugs, doses and irradiation). In the AR group, noninvasive assessment identified subclinical dysfunction with *FS* in 37 (13.6%) of 272 and *AF* in 33 (13.9%) of 238; prolonged QTc interval in 11 (4.0%) of 273. These are the estimates of prevalence of cardiac abnormalities. Among others, one main objective of the study is to estimate cumulative incidence rates of *AFS* and *AAF*.

In this study echocardiography was performed as a research measure and not in response to clinical symptoms. Individuals with previously established cardiac disease were excluded from participation. As formally assessed by New York Heart Association class, none of the study participants reported clinical symptoms of cardiac dysfunction at enrollment. The imaging quality in echocardiography is dependent on obtaining a clear

acoustic window from which ventricular volumes are estimated based on geometric assumptions. Operator experience and variations in thoracic structures can contribute to difficulties in obtaining technically satisfactory data in a given study. These factors randomly contributed to missing data among study participants. AF is not a standardly used assessment in clinical practice, thus, despite training of ultrasonographers for this study, this factor may have contributed to a higher prevalence of missing AF measurements compared to FS. Thus, we feel that there is no selection bias related to those with and without abnormal FS and AF identified as part of the study. The missing values of AF are displayed in Figure 1, which does not show any pattern.

Table 1. Characteristics of 278 Patients Enrolled Onto the Noninvasive Cardiac Study

	At-Risk Group (n=223)		No-Risk Group (n=55)		Total (n=278)	
Demographics	**N**	**%**	**N**	**%**	**N**	**%**
Sex						
Male	108	51.6	31	56.4	139	50
Female	115	48.4	24	43.6	139	50
Treatment group						
Anthracycline	157	70.4	0	0	157	56.5
Anthracycline + Radiation	60	26.9	0	0	60	21.6
Radiation	6	2.7	0	0	6	2.1
None	0	0	55	100	55	19.8
Race/ethnicity						
White	183	82.1	44	8.0	227	81.7
Black	30	13.5	11	2.0	41	14.7
Other	10	4.4	0	0	10	3.6
Diagnosis						
Leukemia	67	30.0	10	18.2	77	27.7
Sarcomas	60	26.9	14	25.4	74	26.6
Lymphoma	54	34.2	2	3.6	56	20.1
Embryonal tumors	42	18.8	29	52.7	71	25.6
Age at Cancer Diagnosis, Yrs						
N	223		55		278	
Mean	7.37		5.77		7.05	
Median	5.46		3.11		4.68	
Range	0.01-23.56		0.29-20.06		0.01-23.56	
Afterload						
N	191		47		238	
Mean	57.50		45.73		55.18	
Median	55.43		42.18		51.88	
Range	15.38-147.32		25.66-95.02		15.38-147.32	
Fractional Shortening						
N	218		54		272	
Mean	0.33		0.36		0.34	
Median	0.33		0.36		0.34	
Range	0.20-0.57		0.24-0.49		0.20-0.57	

A crude approach to estimating the incidence rates and obtaining confidence intervals is to apply the Kaplan-Meier estimator with the assumption of the follow-up as the onset time. Then, incidence rate of each type of toxicity is estimated. Some of these toxicity measures are missing. In this paper, we impute the missing measurements using regression method first and then use a parametric approach to estimate incidence rates of specific toxicity. For the cardiotoxicity data in the motivation example, there are 40 missing *AF* values out of 278 (14%), and 6 (2%) of which missing both *AF* and *FS*. The 34 missing *AF* values are imputed using a multivariable regression about *FS* and other covariates, such as age and race. The estimates of cumulative incidence rates were derived based on the data after imputation and then compared with those derived in Rai, et al.[2] without imputation.

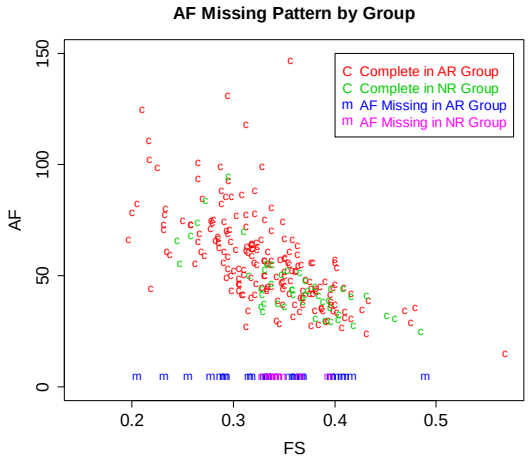

Figure 1: Scatter plot between AF and FS within AR and NR group of patients. Complete data (leveled c) displays strong correlation between AF and FS. Missing values of AF are in almost entire range of FS values (leveled m).

3. Cardio-Measures Abnormality and Imputation Model

In Hudson et al.,[1] the subjects who died or had cardiac failure during the treatment or during the follow-up after completion of therapy were excluded; however, this information is available from the medical record abstraction. Hence, we present the general theory here for a cross-sectional data with indicators of cardiac abnormality and death/cardiac failure, and time since the treatment to the survey or the death/cardiac failure, as depicted in Figure 2. We also assume cardiac abnormality is the precursor for cardiac failure.

3.1 Abnormality Model

Let stochastic process $\{X(t)\}$ identify the state occupied by a patient at time t. For simplicity, we suppose that n patients in state 1 at time $t = 0$ are those who are identified with different disease groups and are planned for treatment, where we have assumed that no patient has cardiac abnormality at time $t = 0$. Let the random variable T denote the observation time (survey, death/cardiac failure) from the study evaluation and U the time of *AFS* or *AAF* from the study evaluation. Thus, at any time t, $X(t) = 1, X(t) = 2$ and $X(t) = 3$ indicate the patient alive with normal cardiac measure, alive with abnormal cardiac measure and died or had cardiac failure with or without cardiac abnormality, respectively. We also assume that the development of abnormality is an irreversible event without the treatment for cardiotoxicity, and therefore transitions from state 2 to state 1 do not occur, as illustrated in Figure 2. According to

practice in this study, the patients are chosen for survey independent of their health status, which ensures that the survey results can be regarded as independent of the times of the events of interest. The intensities $\lambda_1(u)$, $\lambda_2(t)$ and $\lambda_3(t \mid u)$, shown in Figure 2, are corresponding transitions rates, where t is the observation time and u is the time of *AAF* or *AFS*.

The survival function and the cardiac abnormality prevalence function are derived in Rai et al.[2] as follows

$$S(t) = Q(t) + \int_0^t \lambda_1(u)Q(u)Q_3(t \mid u)du \quad \text{and} \quad \pi(t) = \frac{\int_0^t \lambda_1(u)Q(u)Q_3(t \mid u)du}{S(t)},$$

where

$$Q_i(t) = \exp\{-\int_0^t \lambda_i(v)dv\}$$

for $i = 1$ and 2, and

$$Q_3(t \mid u) = \exp\{-\int_u^t \lambda_3(v \mid u)dv\}.$$

are pseudo-survival functions corresponding to the intensities $\lambda_1(u)$, $\lambda_2(t)$ and $\lambda_3(t \mid u)$ and $Q(t) = Q_1(t)Q_2(t)$ is the probability that the time to the first event— alive with abnormal value or death with normal value — exceeds t.

We are interested in estimating $\Lambda_1(t) = \int_0^t \lambda_1(u)du$, the cumulative incidence function (CIF), but focus on the comparison between CIFs based on the original data and imputation data.

Table 2 identifies the various types of observations which occur in this illness-death/failure model and the corresponding contribution to the likelihood, denoted as $L_1(t)$ to $L_4(t)$, which are functionals of intensities and pseudo-survival functions. Rai et al.[2] derive the explicit form of $L_1(t)$ to $L_4(t)$ for both constant and piecewise Exponential model and the likelihood functions, which are summarized in appendix.

3.2 Imputation Model

Let the cardiac measure, such as *FS* or *AF*, be denoted by *Y*. Assume that $Y = (Y_1, Y_2)^T$ be a $n \times 1$ response vector with Y_1 $(n_1 \times 1)$ observed and Y_2 $(n_2 \times 1)$ missed, and $X = (X_1, X_2)^T$ be corresponding $n \times p$ matrix comprised of covariates including other cardiac measures (other response variables). A multivariable regression model $y = x\beta + \varepsilon$ is fitted based on the complete data Y_1 and X_1, and the least squared estimator $\hat{\beta}$ of β $(p \times 1)$ is derived; then the missing information in Y_2 is imputed as $\hat{Y}_2 = X_2\hat{\beta}$, assuming no covariate information missing. It is natural to use the imputation data $(Y_1, \hat{Y}_2)^T$ instead of only Y_1 and is anticipated that the imputed information in $\hat{Y}_2$ will improve the related results in statistical analysis.

4. Application: Cancer Survivor Study

In this section we obtain the imputed data for the cardiotoxicity example and then apply the theory for the Exponential model described in appendix to evaluate the effect of anthracyclines on cardiotoxicity. Furthermore, the results obtained using the imputation approach are then compared with those obtained without imputation and reported in Rai

et al.[2], under the assumption of no deaths/cardiac failures. The Simplest model is the Parametric-1, which is one parameter Exponential model. Since there are very few events before 5 years and after 10 years, we also fit two piecewise Exponential models based on two incidence rates one up-to year five and the second for year 5 and above, and three incidence rates one up-to year 5, second between years 5 and 10 and the last one for year 10 and above, denoted as Parametric-2 and Parametric-3 for models with two and three incidence rates (see Figure 3).

Table 2: Likelihood Contributions for Anthracycline Cardiac Toxicity Study

Observation Type	Outcome	Likelihood Contribution
Death with No Cardiac Abnormality	$T = t, X(t^-) = 1$	$L_1(t) = \lambda_2(t)Q(t)$
Alive with No Cardiac Abnormality	$T > t, X(t) = 1$	$L_2(t) = Q(t)$
Death/Cardiac Failure with Cardiac Abnormality	$T = t, X(t^-) = 2$	$L_3(t) = \int_0^t \lambda_1(u)Q(u)\lambda_3(t \mid u)Q_3(t \mid u)du$
Alive with Cardiac Abnormality	$T > t, X(t) = 2$	$L_4(t) = \int_0^t \lambda_1(u)Q(u)Q_3(t \mid u)du$

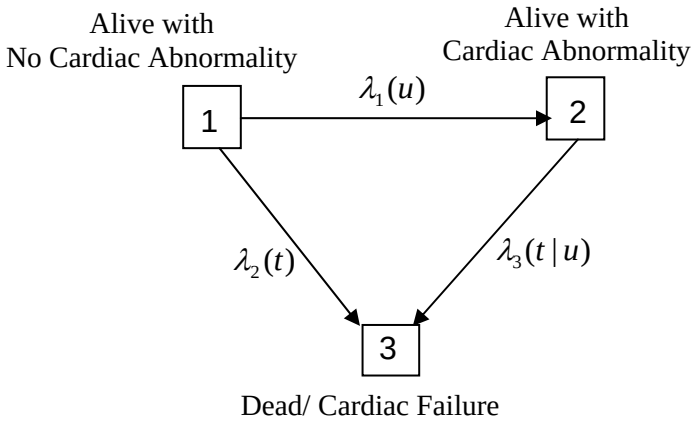

Figure 2: An abnormal cardiac measure-death/cardiac Failure model involving three states. State 1 corresponds to patients who are alive with normal value. Patients who are alive with abnormal value are in state 2. State 3 is an absorbing state and corresponds to death or Cardiac Failure.

In the cardiotoxicity example, there are 34 measurements missing for *AF* and 6 were missing in both measurements, but no covariate information missing. Hence, we employ the multivariable regression method to estimate the 34 missing measurements in *AF* based on the values of *FS* and corresponding covariates, like age at diagnosis, race, gender, BMI, QTC, diagnosis group and risk group (AR/NR)[22-23] . Based on 4 diagnosis groups, Leukemia (n=77), Sarcoma (n=74), Lymphoma (n=56) and Embryonal (n=71), we define three dummy variables as follows:

$$\text{Diag1} = \begin{cases} 1 & \text{Leukemia} \\ 0 & \text{Otherwise,} \end{cases} \quad \text{Diag2} = \begin{cases} 1 & \text{Sarcoma} \\ 0 & \text{Otherwise,} \end{cases} \quad \text{and} \quad \text{Diag3} = \begin{cases} 1 & \text{Lymphoma} \\ 0 & \text{Otherwise.} \end{cases}$$

Before conducting the regression analysis, the Shapiro-Wilk test of normality was applied to original *AF* and *FS* measurements and a few commonly used transformations for making the underlying distributions of *AF* and *FS* more normal. Log(AF) was better behaved normally distributed (p=0.701) than original AF (p<0.001). On the other hand, Log(FS), Logit(FS) and FS were not normally distributed with p

values <0.001, 0.007 and 0.026, respectively. This suggests fitting the regression model using logarithm transformation of *AF* and original *FS*. The significant predictors with coefficients, their p-values and R^2 in the model are presented in Table 3a. That is,

$$\log(AF) = 5.437 - 3.891FS + 0.008\,\text{Age} - 0.081\,\text{Risk} + 0.098\,\text{Diag2} - 0.010\,\text{BMI} + \varepsilon,$$

where Risk=1 for patient in AR group and 0 in NR group. Based on the regression model, the values of *FS* and the covariates, the 34 missing *AF* values are imputed. Thus, after imputing the missing values the total sample size is 172.

Then, the methods described in Appendix were applied to the imputed data sets for each group, AR and NR, and both groups combined. Then, based on the likelihood ratio test, the corresponding p-value for group effect for the variable *AF* is 0.015. On the other hand, the p-value without imputation is 0.020^2. It is obvious that the results with imputation are a little more sensitive for the group effect in *AF* compared to that without imputation. Note that without using the threshold model and no imputation the group effects was only marginally significant (p=0.065)[1-2], which led to better understanding of this data and developing these methods. A summary of the results is given in Table 3b.

The cumulative incidence function (CIF) was derived for exponential and piecewise exponential models for the imputed data and were compared to those based on the original data (without imputation). The standard error (SE) of the estimates of CIF were computed using bootstrap approach (bootstrap size B=5000) for each case. The cumulative incidences and standard errors are presented in Table 4 for *AF*. The relative gain in efficiency due to imputation was obtained as $RE = \frac{\hat{\sigma}_N^2 - \hat{\sigma}_Y^2}{\hat{\sigma}_N^2} \times 100$, where $\hat{\sigma}_Y$ and $\hat{\sigma}_N$ are the standard errors of cumulative incidences corresponding to imputed and non-imputed data. We also present the cumulative incidences in Figure 4 (left panel for Parametric-1, the middle one for Parametric-2 and right panel for Parametric-3) for *AF*. From Table 4, it is clear that the RE is greater than 20%, which implies that the imputation approach is consistently more efficient. It is also seen that the standard error with imputation approach is consistently smaller than those without imputation, which indicates that the confidence intervals of CIF after imputing the missing values are narrower than those without imputation, that is, the estimate based on the imputation data is more accurate compared to that based on the original data.

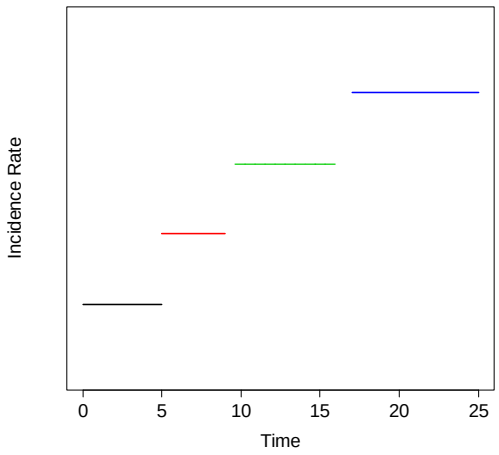

Figure 3: Piecewise Incidence Rate

5. Simulation Study

A limited simulation experiment was conducted to study the performance of using imputation for *AF* in Anthracycline Cardiac Toxicity data here. Firstly, we compute the CIFs and their standard errors for *AF* based on the original data, including complete observations only (n=238). Then we randomly select to delete 10%, 15% or 20% from the 238 observations and calculate the CIFs and their standard errors based on the data after deletion. Finally, we impute these deleted observations as if we do not know these values, using the multivariable regression method, and compute the CIFs and standard errors based on imputation data. All estimates are derived using bootstrap method for B=5000. The relative efficiencies are also calculated as follows: $RE_1 = \frac{\hat{\sigma}_2^2 - \hat{\sigma}_1^2}{\hat{\sigma}_2^2} \times 100$ and $RE_2 = \frac{\hat{\sigma}_3^2 - \hat{\sigma}_2^2}{\hat{\sigma}_3^2} \times 100$, $\hat{\sigma}_1$, $\hat{\sigma}_2$ and $\hat{\sigma}_3$ are the standard errors of cumulative incidences corresponding to the method based on the complete data, with imputing and without imputation after deletion. All results are provided in Table 5. It is seen that all $RE_1 < 2\%$ and all RE_2 are between 10% and 20%. The small values of RE_1 imply that we can almost make-up all missing information using imputation method compared to the complete data set, and the relative larger values of RE_2 indicate that there will be a big loss if we do not impute the missing values.

6. Discussion

In this paper, we employed a well-established methodology in illness-death/Failure model[2] and imputed the missing observations for a phase IV clinical trial study as an example. Although, we assumed a very simple parametric model, it is straightforward to expand to other parametric or semi-parametric models. From a clinician's point of view the simple approaches such as log rank test and KM survival curves[24] are most commonly used and understood. Using the similar logic, our approach is simple and still robust as displayed by the data analyses and simulation for estimating fixed-term cumulative incidence function without imputation[2] or with imputation proposed here.

When studying the long-term effects of treatment, there can be multiple unwanted events identified at the time of observation. Some of these events can be competing and others are not correlated. This leads to multivariate time-to-event data. One simple approach is to study the incidence of first event and then incidence of specific event. In our example, cardiotoxicity measures included abnormal *AF* and *FS* but there are some other measures to evaluate cardiotoxicity. For some reason, not all patients had both measures and the models based on bi-variate time-to-event outcomes would include only those patients who have data on both outcomes and this would reduce the sample size and potentially ignore important information. Based on this consideration, the multivariable regression method was used to impute the missing observations and to apply the parametric method to the imputed data and compare the results with those obtained without imputation. We further studied the properties of our method in a limited simulation study involving the same data based on the procedure advocated by Efron and Tibshirani (1986)[30].

As stated before, the problem of evaluating possible toxic effects of cancer therapies in a Phase IV trial setting is an important problem. Among the many issues in such studies, missing data is a key aspect that can influence the inferences. It is also important to understand the nature of impact of missing data on the analysis and the interpretation of the study data. Also note that imputation increases the sample size, and thus increases statistical power to detect the same effect size. But if the model

assumptions are not correct, the inference may not be valid. Thus, it is recommended to report the p-values with and without imputation. However, with higher absolute correlations between two outcome measures (the primary outcome measure, AF, with higher missing and the secondary outcome measure, FS, with little or no missing), produced efficient results, a rigorous simulation study with different amount of correlations between two outcome measures, amount of missing and model uncertainty is underway to consider this aspect and will be reported elsewhere; this is along the lines our work for a randomized clinical study[28].

Table 3a: Coefficients and P-values in Regression Model			
Variable	**Estimator**	**SE**	**p-value**
Intercept	5.437	0.110	<0.001
FS	-3.891	0.274	<0.001
Age at Diagnosis	0.008	0.003	0.005
Risk Group	-0.081	0.038	0.033
Diag2	0.098	0.035	0.006
BMI	-0.010	0.003	<0.001
R^2	0.586		

Table 3b. AF p-values for Group Effect using Likelihood Ratio Test		
Procedure	Without Imputation	With Imputation
Logistic	0.065	0.046
Interval Censored-1	0.055	0.034
Interval Censored-2	0.041	0.023
Parameter-1	0.020	0.017
Parameter-2	0.012	0.010
Parameter-3	0.078	0.081

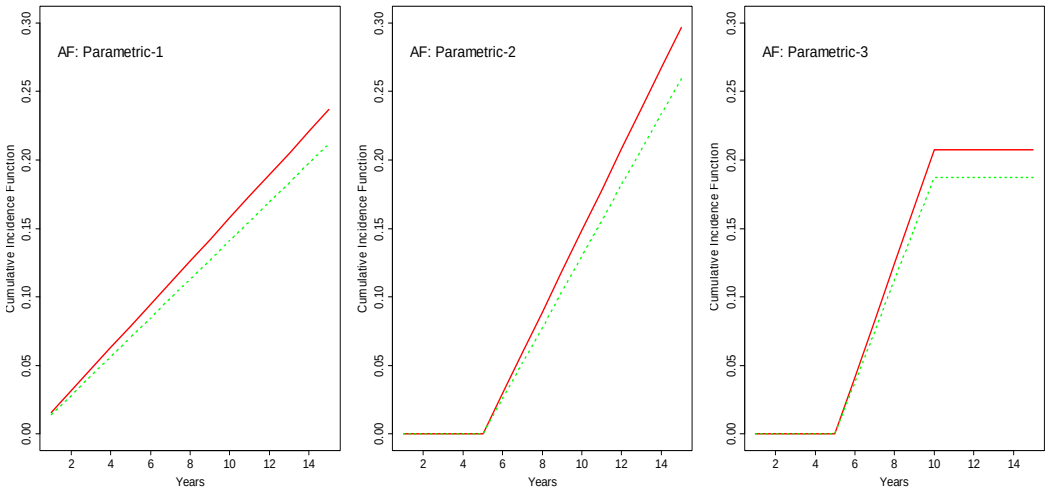

Figure 4: Cumulative Incidence Comparison for AF. (Solid line: CIFs Based on Original Data, Dotted Line: CIFs Based on Imputation Data; Parametric-1: single constant incidence rate; Parametric-2: mixture of two constant incidence rates; Parameter-3: mixture of three constant incidence rates).

The study involved all the patients visiting the clinic in a pre-specified time frame (such as 1 year of accrual) and represents a somewhat unbiased survey of patients. Since the outcome measure may depend on disease type, an almost equal allocation was used to enroll patients. It has been reported in Hudson et al. [1] that the prevalence depends on disease type; hence it will be another research direction to adjust the sampling allocation and variability due to sampling and modeling for generalizing the results for the entire patient population.[25]

Another limitation of this study is that this is a cross-sectional survey to estimate the long-term effect of cardiotoxicity of the primary treatment of cancer. Dodge (2003)[29] defined cross-sectional survey "A method of data collection whereby a battery of questions is asked of participation at one single point or in a relatively small interval of time. Inferences about a population must be anchored to the time period in which the sample was taken. Data from cross-sectional surveys are typically unable to be used to prove the existence of cause-and-effect relationships." Even though this is based on enrolling consecutive eligible patients in a very homogeneous environment (St. Jude Children's Hospital treats patients without charge to patients), effect of this limitation is minimized but cannot be reduced to zero. Generalization of the results to a general population should be done with caution.

7. Appendix

Due to the actual time of onset of abnormality, U, is not known, the observed quantity for each patient includes the observation time, T, and two indicators of status, δ and γ, at the time of survey, where δ is an indicator of patient alive with no cardiac failure or dead/cardiac failure, and γ is an indicator of patient with a normal or abnormal value. Let t_i be the observation time (death, cardiac failure or survey) for the i^{th} subject. That is,

$$\delta_i = \begin{cases} 1, & \text{if unit } i \text{ dead or cardiac failure at time } t_i \\ 0, & \text{if unit } i \text{ alive and no cardiac failure at time } t_i, \end{cases}$$

and

$$\gamma_i = \begin{cases} 1, & \text{if unit } i \text{ with abnormal value at time } t_i \\ 0, & \text{if unit } i \text{ with normal value at time } t_i. \end{cases}$$

The simplified forms of intensities $\lambda_i(t) = \lambda_i$ for $i = 1 \text{ or } 2$ and $\lambda_3(t \mid u) = \lambda_3$ lead to $Q_i(t) = e^{-\lambda_i t}$ for $i = 1 \text{ or } 2$, $Q_3(t \mid u) = e^{-\lambda_3 (t-u)}$ and $Q(t) = e^{-(\lambda_1 + \lambda_2)t}$. We derive the corresponding likelihood contributions from $L_1(t)$ to $L_4(t)$ for the four observation types in Table 2 and then the log-likelihood function as follows

$$\begin{aligned} l(\lambda_1, \lambda_2, \lambda_3) = & \sum_{i=1}^{n} a_i [\log \lambda_2 - (\lambda_1 + \lambda_2) t_i] - \sum_{i=1}^{n} b_i (\lambda_1 + \lambda_2) t_i] \\ & + \sum_{i=1}^{n} c_i [\log \lambda_1 + \log \lambda_3 - \log(\lambda_1 + \lambda_2 - \lambda_3) + \log(e^{-\lambda_3 t_i} - e^{-(\lambda_1 + \lambda_2) t_i}] \\ & + \sum_{i=1}^{n} d_i [\log \lambda_1 - \log(\lambda_1 + \lambda_2 - \lambda_3) + \log(e^{-\lambda_3 t_i} - e^{-(\lambda_1 + \lambda_2) t_i}], \end{aligned}$$

where $a_i = \delta_i (1 - \gamma_i), \mathrm{b}_i = (1 - \delta_i)(1 - \gamma_i), c_i = \delta_i \gamma_i$ and $\mathrm{d}_i = (1 - \delta_i)\gamma_i$ are the indicators corresponding to observation type 1 to type 4 in Table 2. Then the maximum

likelihood estimators $\hat{\lambda}_1, \hat{\lambda}_2$ and $\hat{\lambda}_3$ of λ_1, λ_2 and λ_3 are derived from the following equations[2].

$$\frac{D_3 + D_4}{\lambda_1} - \frac{D_3 + D_4}{\lambda_1 + \lambda_2 - \lambda_3} - T_2 + \sum_{i=1}^{n} \frac{(c_i + d_i)t_i}{e^{(\lambda_1+\lambda_2-\lambda_3)t_i} - 1} = 0$$

$$\lambda_2 = \frac{D_1}{D_3 + D_4}\lambda_1$$

$$\lambda_3 = \frac{D_3}{T_1\lambda_1 - D_3 - D_4}\lambda_1$$

where

$$D_1 = \sum_{i=1}^{n} a_i, D_2 = \sum_{i=1}^{n} b_i, D_3 = \sum_{i=1}^{n} c_i, D_4 = \sum_{i=1}^{n} d_i$$

$$T_1 = \sum_{i=1}^{n} (a_i + b_i + c_i + d_i)t_i \text{ and } T_2 = \sum_{i=1}^{n} (a_i + b_i)t_i.$$

It is further extended to the model to allow the intensity λ_1 with piecewise constant[2]. Assume two intervals: less than t_c years and above t_c (including t_c) years (say, $t_c = 5$) and let these two rates be λ_{11} and λ_{12}. Then the log-likelihood function is derived as

$$l(\lambda_{11}, \lambda_{12}) = \sum_{i=1}^{n} [b_i \log L_2(t_i) + d_i \log L_4(t_i)]$$

$$= \sum_{i \in S_1} [-b_i t_i \lambda_{11} + d_i \log(1 - e^{-t_i\lambda_{11}})] + \sum_{i \in S_2} \{-b_i [t_c\lambda_{11} + (t_i - t_c)\lambda_{12}] + d_i \log(1 - e^{-t_c\lambda_{11} - (t_i - t_c)\lambda_{12}})\},$$

for a special case with no deaths/cardiac failures, that is, $a_i = c_i = 0$, $\lambda_2 = \lambda_3 = 0$, where $S_1 = \{i : \mathrm{t_i} < t_c\}$ and $S_2 = \{i : \mathrm{t_i} \geq t_c\}$. Hence, the estimates of λ_{11} and λ_{12} can be derived easily from following score equations

$$\sum_{i \in S_1} b_i t_i - \sum_{i \in S_1} \frac{d_i t_i}{e^{\lambda_{11} t_i} - 1} + \sum_{i \in S_2} b_i t_c - \sum_{i \in S_1} \frac{d_i t_c}{e^{\lambda_{11} t_c + (t_i - t_c)\lambda_{12}} - 1} = 0$$

$$\sum_{i \in S_2} b_i (t_i - t_c) - \sum_{i \in S_2} \frac{d_i (t_i - t_c)}{e^{\lambda_{11} t_c + (t_i - t_c)\lambda_{12}} - 1} = 0.$$

For a general model with piecewise constants in parameter λ_1, the log-likelihood function is

$$l(\lambda_{11}, \lambda_{12}, \lambda_2, \lambda_3) = \sum_{i=1}^{n} [a_i \log L_1(t_i) + b_i \log L_2(t_i) + c_i \log L_3(t_i) + d_i \log L_4(t_i)],$$

where $L_1(t) = \lambda_2 Q(t)$, $L_2(t) = Q(t)$, $L_4(t) = L_3(t) / \lambda_3$ and if $t < t_c$,

$$L_3(t) = \frac{\lambda_{11}\lambda_3}{\lambda_{11} + \lambda_2 - \lambda_3} e^{-\lambda_3 t} (1 - e^{-(\lambda_{11} + \lambda_2 - \lambda_3)t})$$

if $t \geq t_c$,

$$L_3(t) = \frac{\lambda_{11}\lambda_3}{\lambda_{11}+\lambda_2-\lambda_3} e^{-\lambda_3 t_c}(1-e^{-(\lambda_{11}+\lambda_2-\lambda_3)t_c}) + \frac{\lambda_{12}\lambda_3}{\lambda_{12}+\lambda_2-\lambda_3} e^{-(\lambda_{11}-\lambda_{12})t_c-\lambda_3 t}(e^{-(\lambda_{12}+\lambda_2-\lambda_3)t_c} - e^{-(\lambda_{12}+\lambda_2-\lambda_3)t}),$$

in which $Q_1(t) = e^{-\lambda_{11}t}$ if $t < t_c$, and $Q_1(t) = e^{-(\lambda_{11}-\lambda_{12})t_c-\lambda_{12}t}$ if $t \geq t_c$; $Q_2(t) = e^{-\lambda_2 t}$, $Q_3(t \mid u) = e^{-\lambda_3(t-u)}$ and $Q(t) = Q_1(t)Q_2(t)$. Based on the log-likelihood equation, the maximum likelihood estimates of $\lambda_{11}, \lambda_{12}, \lambda_2$ and λ_3 can be computed from the score equations using numerical method. It is similar to derive the likelihood function for exponential model if λ_1 has three or more pieces.

Acknowledgements

We thank Dr. Donald Miller for his support for this research work. Partial funding for Dr. Rai's work was provided by Wendell Cherry Chair in Clinical Trial Research.

References

1. Hudson MM, Rai SN, Nunez C, Merchant TE, Marina NM, Zalamea N, Cox C, Phipps S, Pompeu R and Rosenthal D. Noninvasive evaluation of late anthracycline cardiac toxicity in childhood cancer survivors. *Journal of Clinical Oncology* 2007; **25**: 3635-3643.

2. Rai SN, Pan J, Sun J, Hudson MM and Srivastava Dk. Estimating incidence rate on current status data with application to a Phase IV cancer trial. *Communications in Statistics: Theory and Methods* 2013; **42**:2417-2433.

3. Krischer JP, Epstein S, Cuthbertson DD, et al. Clinical cardiotoxicity following anthracycline treatment for childhood cancer: The Pediatric Oncology Group experience. *Journal of Clinical Oncology* 1997; **15**: 1544-1552.

4. Sorensen K, Levitt GA, Bull C, et al. Late anthracycline cardiotoxicity after childhood cancer: A perspective longitudinal study. *Cancer* 2003; **97**: 1991-1998.

5. Ayer M, Brunk HD, Ewing, GM, Reid WT, and Silverman E. An empirical distribution function for sampling with incomplete information. *Ann. Math. Statist*.1955; **26**: 641-647.

6. Van Eeden C. Maximum likelihood estimation of ordered probabilities. *Indagationes Mathematicae* 1956; **18**: 444-455.

7. Peto R. Experimental survival curves for interval-censored data. *Appl. Statist*. 1973; **22**: 86-91.

8. Turnbull BW. The empirical distribution function with arbitrarily grouped, censored and truncated data. *J. R. Statist*. Soc. B 1976; **38**: 290-295.

9. Sun J. *The Statistical Analysis of Interval-Censored Failure Time Data*. Springer: New York, 2006.

10. Little RJ, Rubin DB. *Statistical Analysis with Missing Data*, 2nd edition, New York: John Wiley, 2002.

11. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. J. Wiley & Sons, New York, 1987.

12. Rubin DB. Multiple imputation after 18+ years (with discussion). *Journal of the American Statistical Association,* **91**: 473-489.
13. Rubin DB, Stern H, and Vehovar V. (1995), Handing "don't know" survey responses: The case of the Slovenian plebiscite. *Journal of the American Statistical Association* 1995; **90**: 822-828.
14. King G, Honaker J, Joseph A, and Scheve K. Analyzing incomplete political science data: an alternative algorithm for multiple imputation. *The American Political Science Review* 2001; **95**: 49–69.
15. Schafer JL. *Analysis of Incomplete Multivariate Data.* Chapman & Hall, London, 1997.
16. Liu C. Missing data imputation using the multivariate t distribution. *Journal of Multivariate Analysis* 1995; **53**: 139-158.
17. Abayomi K, Gelman A, Levy M. Diagnostics for Multivariate Imputations. *Journal of the Royal Statistical Society*: Series C (Applied Statistics) 2008; **57**: 273-291.
18. van Buuren S, Boshuizen HC and Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine* 1999; **18**: 681-694.
19. Troxel A, Guoguang M and Heitjan DF. An index of local sensitivity to nonignorability. *Statistica Sinica* 2004; **14**: 1221-1237.
20. Pein F, Sakiroglu O, Dahan M, et al. Cardiac abnormalities 15 years and more after adriamycin therapy in 229 childhood survivors of a solid tumor at the Institut Gustave Roussy. British *Journal of Cancer* 2004; **91**: 37-44.
21. Cohen, J. *Statistical power analysis for the behavioral sciences,* 2nd edition. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.
22. Raghunathan TE, Lepkowski JM, Van Hoewyk J, and Solenberger P. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology* 2001; **27**: 85–95.
23. Schenker N, Raghunathan TE, Chiu PL, Makuc DM, Zhang G, and Cohen AJ. Multiple imputation of missing income data in the National Health Interview Survey. *Journal of the American Statistical Association* 2006; **101**: 924-933.
24. Kaplan EL and Meier P. Nonparametric estimation from incomplete observation. *Journal of the American Statistical Association* 1958; **53**: 457-481.
25. Kovacevic MS and Rai SN. Log-Linear Modeling of Change Using Longitudinal Survey Data. *Communications in Statistics—Theory and Methods* 2002; **31**: 1815-1835.
26. Henry WL, Gardin JM and Ware JH. Echocardiographic measurements in normal subjects from infancy to old age. *Circulation* 1980; 62:1054-1061.
27. Colan SD, Parness IA, Spevak PJ, et al. Developmental modulation of myocardial mechanics: Age- and growth-related alterations in afterload and contractility. *Journal of the American College of Cardiology*, 1992; 19:619-629.
28. Fong DY, Rai SN and Lam KS. Estimating the effect of multiple imputation on incomplete longitudinal data with application a randomized clinical study. *Journal of Biopharmaceutical Statistics* 2013; 23:1004-1022.
29. Dodge Y. *The Oxford Dictionary of Statistical Terms.* The International Statistical Institute, Oxford Press, 2003.
30. Efron B and Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1:54-75, 1986.

Table 4: Cumulative Incidence Function and Standard Error for *AF*												
Method	**Parametric-1**				**Parametric-2**				**Parametric-3**			
Sample Size	**Imputation**				**Imputation**				**Imputation**			
	No		**Yes**		**No**		**Yes**		**No**		**Yes**	
Year	**CIF**	**SE**	**CIF**	**SE**	**CIF**	**SE**	**CIF**	**SE**	**CIF**	**SE**	**CIF**	**SE**
1	0.016	0.003	0.014	0.002								
2	0.032	0.005	0.028	0.005								
3	0.047	0.008	0.042	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.063	0.010	0.057	0.009								
5	0.079	0.013	0.071	0.011								
6	0.095	0.016	0.085	0.014	0.030	0.005	0.026	0.004	0.041	0.007	0.038	0.006
7	0.110	0.018	0.099	0.016	0.059	0.010	0.052	0.009	0.083	0.014	0.075	0.012
8	0.126	0.021	0.113	0.018	0.089	0.015	0.078	0.013	0.124	0.020	0.113	0.018
9	0.142	0.023	0.127	0.020	0.119	0.020	0.104	0.017	0.166	0.027	0.150	0.024
10	0.158	0.026	0.141	0.023	0.149	0.025	0.130	0.021	0.207	0.034	0.187	0.030
11	0.174	0.029	0.155	0.025	0.178	0.030	0.156	0.026	0.207	0.034	0.187	0.030
12	0.189	0.031	0.170	0.027	0.208	0.035	0.182	0.030	0.207	0.034	0.187	0.030
13	0.205	0.034	0.184	0.029	0.238	0.040	0.208	0.034	0.207	0.034	0.187	0.030
14	0.221	0.037	0.198	0.032	0.267	0.045	0.234	0.038	0.207	0.034	0.187	0.030
15	0.237	0.039	0.212	0.034	0.297	0.050	0.260	0.043	0.207	0.034	0.187	0.030
20	0.316	0.052	0.283	0.045	0.446	0.075	0.389	0.064	0.207	0.034	0.187	0.030
Relative Efficiency	RE = 25.28				RE = 27.59				RE = 21.64			

Table 5: Simulation Results for Cumulative Incidence Function and Standard Error for *AF* Using Exponential Model														
Missing %	**0%**		**10%**				**15%**				**20%**			
Imputation	**NA**		**Yes**		**No**		**Yes**		**No**		**Yes**		**No**	
Year	**CIF**	**SE**	**CIF**	**SE**	**CIF**	**SE**	**CIF**	**SE**	**CIF**	**SE**	**CIF**	**SE**	**CIF**	**SE**
1	0.016	0.003	0.015	0.003	0.016	0.003	0.015	0.003	0.016	0.003	0.015	0.003	0.016	0.003
2	0.032	0.005	0.030	0.005	0.032	0.006	0.030	0.005	0.032	0.006	0.029	0.005	0.032	0.006
3	0.047	0.008	0.046	0.008	0.047	0.008	0.045	0.008	0.048	0.009	0.044	0.008	0.048	0.009
4	0.063	0.011	0.061	0.010	0.063	0.011	0.060	0.011	0.064	0.012	0.059	0.011	0.064	0.012
5	0.079	0.013	0.076	0.013	0.079	0.014	0.075	0.013	0.079	0.014	0.074	0.013	0.080	0.015
6	0.095	0.016	0.091	0.016	0.095	0.017	0.090	0.016	0.095	0.017	0.088	0.016	0.096	0.018
7	0.110	0.018	0.106	0.018	0.111	0.019	0.105	0.019	0.111	0.020	0.103	0.019	0.112	0.021
8	0.126	0.021	0.122	0.021	0.126	0.022	0.120	0.021	0.127	0.023	0.118	0.021	0.128	0.024
9	0.142	0.024	0.137	0.024	0.142	0.025	0.135	0.024	0.143	0.026	0.133	0.024	0.144	0.026
10	0.158	0.026	0.152	0.026	0.158	0.028	0.150	0.027	0.159	0.029	0.147	0.026	0.159	0.029
11	0.173	0.029	0.167	0.029	0.174	0.030	0.165	0.030	0.175	0.032	0.162	0.029	0.175	0.032
12	0.189	0.032	0.182	0.031	0.189	0.033	0.180	0.032	0.191	0.035	0.177	0.032	0.191	0.035
13	0.205	0.034	0.198	0.034	0.205	0.036	0.195	0.035	0.206	0.037	0.191	0.034	0.207	0.038
14	0.221	0.037	0.213	0.037	0.221	0.039	0.210	0.038	0.222	0.040	0.206	0.037	0.223	0.041
15	0.236	0.039	0.228	0.039	0.237	0.041	0.225	0.040	0.238	0.043	0.221	0.040	0.239	0.044
20	0.315	0.053	0.304	0.052	0.316	0.055	0.300	0.054	0.317	0.058	0.295	0.053	0.319	0.059
Relative Efficiency	NA		RE_1=1.57		RE_2=9.77		RE_1=1.84		RE_2=13.63		RE_1=1.65		RE_2=13.73	