Research Article

A New Method for Imputing Censored Values in Crossover Designs with Time-to-Event Outcomes Using Median Residual Life

Maryam Jalali [ID],[1] **Zahra Bagheri,**[1] **Najaf Zare,**[2] **and Seyyed Mohammad Taghi Ayatollahi** [ID][1]

[1]*Department of Biostatistics, Medical Schools, Shiraz University of Medical Sciences, Shiraz, Iran*
[2]*Department of Biostatistics, Infertility Research Center, Shiraz University of Medical Sciences, Shiraz, Iran*

Correspondence should be addressed to Seyyed Mohammad Taghi Ayatollahi; ayatolahim@sums.ac.ir

Crossover designs are commonly applied in research due to efficiency and subject parsimony compared to parallel studies. Baseline measurements would improve the power of comparison. For time to event outcomes, the sample size is reduced due to censorship, if they are ignored; thus, applying traditional regression models will be limited. A logical solution is to impute the censored observation and apply common analytical models for analyzing the data. Nevertheless, techniques to impute censored data in time-to-event outcomes in crossover designs are not practiced as much. Accordingly, we propose a method to impute the censored observation using median residual life regression and then analyze the data using analyses of covariance (ANCOVA), considering the difference of period-specific baselines as covariate. We used simulation to show the favorable performance of our method relative to a recently proposed method, multiple imputation with model averaging and ANCOVA (MI$^{\mathrm{MI}}$). Specifically, the censored observations were multiply-imputed using prespecified parametric event time models, and then, the methods were applied to a real data example.

1. Introductions

In clinical trial designs, to compare the effect of treatment on the same subject over different treatment periods, crossover designs are appropriate due to higher accuracy and subject parsimony [1]. Two-treatment two-period crossover is the simplest design with two treatments, A and B, where subjects are randomly allocated to AB or BA sequences [2].

Based on previous studies, consideration of the baseline measurement in the analysis of crossover design, which is gathered before receiving treatment, can increase the power of comparisons [3–5]. The literature suggests that the analysis of covariance (ANCOVA), considering within subject differences in baseline responses as a covariate, has a good performance in crossover design [3, 4, 6–9].

Time-to-event endpoints in crossover clinical designs are also commonly investigated. For example, in a crossover trial of pregabalin for neurogenic claudication, the outcome variable was "the time to first moderate pain symptoms during a 15-minute treadmill test." Markman et al. studied the efficacy of an oral spray involving treadmill exercise testing [10, 11]. In both mentioned studies, baseline measurements were gathered, but not considered in the analysis. Xu et al. examined a crossover trial in a treadmill walking test. The recorded time to a specific cardiopulmonary event was the main outcome, and the participants who walked more than 10 min were considered right censored observation. Every participant had four responses: baseline and posttreatment responses for periods 1 and 2. They applied analysis of covariance (ANCOVA) to consider the baseline measurements in estimating the drug effect [1].

For ethical reasons, it is essential to keep the number of patients in a clinical trial as low as possible. As evidenced by extensive research publications, crossover design can be a useful and powerful tool to reduce the number of patients required for a parallel group design [12, 13]. An explicit statistical issue in regression modeling is that adequate statistical power cannot be reached in a small sample size

(which is a characteristic of crossover designs) [14]. Apparently, in time-to-event endpoints, censorship is inevitable; thus, ignoring censored responses can lead to fewer sample sizes and critical conditions. The alternative is to impute the censored values as real as possible before data analysis.

Xu et al. incorporated multiple imputation (MI) of censored values and ANCOVA to consider baseline measurements in analyzing 2×2 crossover studies with censored time-to-event outcomes. The superiority of their method over the hierarchical rank test was demonstrated through simulations [1, 15]. Their imputation method was based on fitting parametric models and using MI. Finally, they applied Rubin's rule to estimate the treatment effect. In spite of satisfactory results, the computation and imputation process was cumbersome and time-consuming [1].

Although matching is done in crossover design, there are some variables that can affect the duration of a time-to-event outcome before receiving any treatment. For example, regardless of the study type, in a treadmill test, age is an effective factor on the walking duration. Hence, it would be beneficial to apply a method that incorporates demographic variables in the imputation step. In addition, imputation would be more precise, if we consider the censored times and use them as a predictive input when imputing them. However, these were not included in the aforementioned method.

The use of advanced statistical indices would improve statistical analysis. The mean residual life function and the quantile (median) residual life function are functional indices that are usually applied to summarize the survival experience of patients [16]. The mean residual life function has some limitations, though. Firstly, it cannot be estimated reliably in the presence of censored observation and secondly; as it extremely depends on outliers, it is very unstable even in completed data [16].

In this paper, we develop a new approach that can be used directly to infer the effect of covariates on the median residual lifetime (MERL) at any specific time point, to predict actual lifetimes more accurately.

After imputing the predicted censored observations in crossover time-to-event data based on MERL regression models, then, we apply ANCOVA to consider the baseline measurements and treatment effect. Specifically, two models will be discussed: (i) the imputation model which is the main aim of this study which is applied for imputing censored values and (ii) the analytical model, ANCOVA, which is performed similar to the Xu et al. study [1].

To impute the censored observations in the first part, we applied the median residual life regression model, which has not been used for this purpose in crossover designs and is free from the mentioned limitations.

Section 2 presents details of the proposed method. In Section 3, we contrast the numerical performance of our proposed method with the multiple imputation method through simulation studies. Section 4 presents the results from applying the different methods to a real data example. Section 5 deals with discussion.

2. Methods

We assumed a 2×2 crossover trial with two treatments (A and B) and two sequences (AB or BA) with a wash-out period between periods 1 and 2. The baseline and posttreatment responses are indicated as b_{ijk} and Y_{ijk}, where $i = 1, 2$ represents periods; $j = 1, 2, \cdots, n$ denotes subjects; and $k = 1, 2$ shows sequences. It can be assumed that after logarithm transformation, $(b_{1j1}, Y_{1j1}, b_{2j1}, Y_{2j1})^T$ and $(b_{1j2}, Y_{1j2}, b_{2j2}, Y_{2j2})^T$ follow a multivariate distribution with different means and the same variance-covariance structure Σ. If a subject dose not experience the event during the follow-up time, he will be assumed as a censored observation at the end of period time which is shown by time τ. Another assumption is that there is no censoring observation at baseline.

In our method, first, the censored observations are imputed by applying the predicted median residual life regression added to the censored time, which is the main objective of our study. Then, similar to Xu et al.'s study, ANCOVA is done to estimate "the ratio of geometric means of the event times for treatment relative to B," which is represented by θ.

Next, the null hypothesis H0:$\theta = 1$ is tested. For symmetric distributions (on the log scale), the geometric mean is equivalent to the median. Thus, this parameter can be used for estimating the ratio of median survival of the two treatments, which is usually of interest in survival analysis.

2.1. Imputation Model

2.1.1. Imputing Censors by Median Residual Life Regression.
The median residual life function is as follows:

$$m(t_0) = \text{median}(\text{Ti} - t_0 \mid \text{Ti} > t_0), \tag{1}$$

This is interpreted as "the remaining lifetimes among survivors beyond time t_0".

To adjust for confounding variables and consider them as covariates, regression modeling is a proper choice.

Suppose that z_{1i} is a bivariate covariate that 0 and 1 are assigned to the control and treatment groups, respectively. Applying the invariance property of the median, the median residual life repression equation is written as follows [16]:

$$\text{Median}(\text{Ti} - t_0 \mid \text{Ti} > t_0, z_{1i}) = \exp\left(\beta_{t_0}^{(0)} + \beta_{t_0}^{(1)} z_{1i}\right). \tag{2}$$

Now, in our study, we impute the censored times using median residual life regression. First, this is done for the censored observations in period 1 and afterwards, censored times in period 2.

We fit the above model and use baseline information in period 1, age and treatment indicators as covariates, and then compute the median residual life regression for every censored observation:

$$\text{Median} \left(\text{Ti} - t_0 \mid \text{Ti} > t_0, z_{1jk}, \text{age}_{1jk}, b_{1jk} \right)$$
$$= \exp \left(\beta_{t_0}^{(0)} + \beta_{t_0}^{(1)} z_{1i}k + \beta_{t_0}^{(2)} A_{1ik} + \beta_{t_0}^{(3)} b_{1ik} \right), \quad j = 1, 2, \cdots, n, k = 1, 2 \tag{3}$$

where z_{1jk} is the treatment indicator, A_{1ik} represents the age of each subject in period 1, and b_{1ik} denotes the baseline measurement of each subject. The corresponding complete set of uncensored post treatment values in period 1 is denoted by Y_{1jk}.

After completing the censored time in period 1, the censored values in period 2 are imputed by the regression equation ahead:

$$\text{Median} \left(\text{Ti} - t_0 \mid \text{Ti} > t_0, z_{2jk}, \text{age}_{2jk}, b_{2jk}, Y_{1jk} \right)$$
$$= \exp \left(\beta_{t_0}^{(0)} + \beta_{t_0}^{(1)} z_{2ik} + \beta_{t_0}^{(2)} A_{2ik} + \beta_{t_0}^{(3)} b_{2ik} + \beta_{t_0}^{(3)} b_{1ik} + \beta_{t_0}^{(4)} Y_{1jk} \right). \tag{4}$$

In addition to the mentioned variables, Y_{1jk} also is added to the equation. The corresponding complete set of uncensored post treatment values in period 2 is represented by Y_{2jk}.

For each censored value, t_0 is set to the censorship time itself as the idea is to find the residual life time after t_0. After computing the median residual time for every censored time, the final imputed time is obtained through summing the median residual life time and the censored time for that person.

2.2. Analysis of Covariance. After imputing censored times, ANCOVA is applied to the completed data on the log scale. The difference between posttreatment event times $\Delta_{jk} = \log (Y_{1jk}) - \log (Y_{2jk})$ is regressed on the difference between baseline values $D_{jk} = \log (b_{1jk}) - \log (b_{2jk})$ and the sequence indicator Q_j:

$$\Delta_{jk} = \gamma_0 + \gamma_1 D_{jk} + \gamma_2 Q_j + \varepsilon_{jk}, \quad \varepsilon_{jk} \sim N \left(0, \eta^2 \right). \tag{5}$$

The point estimator for the geometric mean for treatment A relative to B is $\log \widehat{\theta} = \widehat{\gamma_2}/2$.

The P value for the hypothesis H0:$\theta = 1$ is obtained from the P value of the test H0:$\log \widehat{\gamma_2} = 0$ in the ANCOVA output.

3. Simulation

3.1. Simulation Set-Up. We carried out a simulation study to compare the performance of our proposed and multiple imputation method. Type 1 error and power are reported for both methods. The bias and 95% confidence interval coverage probability for θ are reported too.

The distributions considered for event time in the simulations were *Exponential*, *Weibull*, and *Gamma*, which are commonly applied for survival data.

Multivariate correlated event times from *Exponential* were generated with mean $(.3, .3\theta, .3, .3)^T$ for AB sequence and $(.3, .3, .3, .3\theta)^T$ for BA sequence with common variance-covariance structure and correlation coefficients.

Weibull distribution with shape parameter s and decay parameter d has the density of

$$f(x) = s \times d \times x^{(s-1)} \times \exp (-d \times x^s),$$
$$\text{For } x \geq 0 \,; s, d > 0. \tag{6}$$

For the Weibull distribution, the shape parameter was $(0.95, 0.95, 0.95, 0.95)^T$ for both sequences and the decay parameter was $(0.29, 0.29\theta, 0.29, 0.29)^T$ for AB sequence and $(0.29, 0.29, 0.29, 0.29\theta)^T$ for BA sequence [17].

The third distribution, *Gamma*, was generated for better presenting the performance of our method. The correlated event time for Gamma distribution was generated with the same shape $(3.2, .3.2, 3.2, 3.2)^T$ for both sequences and rate $(.7, .7\theta, .7, .7)^T$ for AB sequence and $(.7, .7, .7, .7\theta)^T$ for BA sequence with common variance-covariance structure and correlation coefficients [17].

The parameters for generating multivariate distributions were selected such that the generated data would be approximately similar to the motivating data example in Xu et al.'s study. They were selected such that the data would fall within the range of 0-10 with a mean of about 5 for each period [1].

We considered the first-order autoregressive (AR(1)) for the correlation structure as it is usually similar to the real data correlation structure. The AR(1) correlation structure is as follows:

$$\begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}. \tag{7}$$

The mean correlation coefficient $\bar{\rho}$ was considered 0.5 and 0.7: $\rho = 0.7$ for $\bar{\rho} = 0.5$ and $\rho = 0.83$ for $\bar{\rho} = 0.7$. These are the same as Xu et al.'s study. These means of correlation are used to keep a reasonable level of intersubject correlation in the generated data.

We assumed that censorship did not occur at baseline time and the case of censorship was right-censoring in the posttreatment time at τ. The parameter of study, θ, is the ratio of the geometric means of the event times for treatment A and treatment B.

The performance of different models was compared in distinct scenarios of sample size, correlation coefficients, and percentage of censoring.

The sample size per sequence was considered $N = 12, 24, 48$, and the percentage of censoring was 30% and 50% for the total sample. Each scenario was repeated 1000 times:

TABLE 1: True θ values used in simulation study under the alternative hypothesis for each combination of distribution, mean pairwise correlation coefficient, percentage of censoring, and sample size.

Mean pairwise correlation	$\bar{\rho} = 0.5$						$\bar{\rho} = 0.7$					
Percentage of censoring	30%			50%			30%			50%		
N	12	24	48	12	24	48	12	24	48	12	24	48
Dist												
Exp	2.33	1.59	1.6	2.3	1.8	1.5	2.1	1.7	1.5	2	1.52	1.35
Weibull	2.3	1.93	1.68	2.3	1.69	1.51	2	1.78	1.56	2	1.64	1.28
Gamma	2.1	1.4	1.3	1.95	1.65	1.2	2	1.4	1.24	2.1	1.3	1.14

$\theta = 1$, under the null hypothesis. Under the alternative hypothesis, the value of θ was chosen such that the power was about 80% for the multiple imputation method proposed by Xu et al. in different scenarios of simulations [1]. The true θ values used in the simulation are shown in Table 1. All analyses and data generation were done in R software. lcmix package was applied for generating correlated multivariate datasets. In this package, the construction of multivariate distributions from univariate marginal distributions using normal copulas is discussed.

3.2. Simulation Results. Type 1 error for the proposed and multiple imputation method is shown in Table 2. The errors in all scenarios for the two methods were less than 5% and well controlled. Obviously, the Weibull distribution controlled the error more than Exponential.

For the Gamma distribution, the type one errors had a noticeable reduction in comparison to the Exponential and Weibull distribution, especially in 30% censoring.

Powers of the mentioned methods are shown in Table 2. In all distribution, powers of the proposed method were more or equal to the multiple imputation method. On average, the power of our proposed method was 11% more than the multiple imputation method in each mean pairwise correlation, irrespective of sample size, percentage of censoring, and distribution. When censoring was 30%, the average power of our method was 12% more than multiple imputation, irrespective of sample size, distribution, and mean pairwise correlation. This figure was 14% and 12% for the Gamma and Weibull distribution, respectively, irrespective of other scenarios of simulation.

For our proposed method, in each percentage of censoring, the power increased with increase in the sample size. The power for the Gamma distribution was greater than that of the other two distributions. Further, the power was higher when the mean pairwise correlation was 0.7 in comparison to 0.5 and the percentage of censoring was 30% in comparison to 50%.

The percentage of bias under the null and alternative hypothesis and 95% confidence interval (C.I.) coverage probability for log (θ) under the alternative hypothesis are reported in Table 3. The bias was not larger than 10% in all scenarios and was ignorable in all scenarios.

Further, the 95% C.I. coverage probability was kept at or more than the nominal level in every scenario.

4. Data Application

The type of data in this study is rare; hence, finding a suitable data that fulfills the condition of crossover design (time to event outcome) is very limited. Thus, we applied the data set from Xu et al.'s study [1]. This dataset is a 2×2 crossover clinical trial, where the effect of a drug was investigated in a treadmill test as discussed in the introduction. The data and its details are presented in Table 4.

The sample size was 40 in total, of which 20 were randomly allocated to the placebo-drug sequence and the other 20 to drug-placebo sequence, which was a treadmill walking test. Also, the time until an event related to cardio was considered the outcome. The follow-up time was 10 min. The baseline measurements were performed for each participant before receiving any treatment in both periods. The raw data have been presented in Xu et al.'s article [1]. After imputing the censored data by our proposed method and applying ANCOVA, there was a significant difference between the drug and placebo group (P value = 0.004). Our results were in line with the multiple imputation method (P value = 0.005). In addition, the ratio of geometric mean of time to the mentioned event (θ) was 1.78 (SE = 0.16), with 95% C.I. of (1.47, 2.09). This ratio was 1.67 (SE = 0.25), with 95% C.I. of (1.18, 2.35) in Xu et al.'s study [1].

Table 5 outlines the estimated survival time and standard errors (SE) of the censored observations in the treadmill test for the multiple imputation method and the proposed method with and without adding age as a covariate. The estimates for our method are more accurate than the multiple imputation method. Also, adding age to the model led to enhanced accuracy.

5. Discussion

To our knowledge, methods for imputing censored observations in time-to-event outcomes in small sample size data have been seldom discussed in the literature. One discussed method for imputing censored observation in time-to-event outcomes is restricted mean survival time. For example, Liu et al. as well as Grover and Gupta proposed a method based on multiple imputations to impute censored outcomes, using restricted mean survival time in small sample size data. Their method was similar, but the superiority of the second one was that the upper limit for imputation of censored survival time was taken as the highest survival time within the study, where

TABLE 2: Type 1 error (target is 5%) and power (%) for the multiple imputation with model averaging and analysis of covariance (MI^{MA}) and our proposed method (1000 simulations).

| Mean pairwise correlation | | | $\bar{\rho} = 0.5$ | | | | | | $\bar{\rho} = 0.7$ | | | | | |
| Censoring percentage | | | 30% | | | 50% | | | 30% | | | 50% | | |
Distribution		N/seq Method	12	24	48	12	24	48	12	24	48	12	24	48
Exp	Error	MI^{MA}	4.8	4	4.9	5	4.5	5	4.5	4.6	4.9	4.8	5	5
		MERL	4.6	4	4.8	4.9	4.1	5	3.9	4	4.4	3.8	4.1	4.5
	Power	MI^{MA}	78	80	81	69	79	80	81	79	84	78	80	81
		MERL	81	87	91	78	80	83	82	94	85	84	86	87
Weibull	Error	MI^{MA}	3.8	4	4.3	4	5	4	4.1	3	4.9	4.2	3.4	4.8
		MERL	3.5	4.1	4	4.2	4	4.3	3.8	3.1	4	4	3.5	4.2
	Power	MI^{MA}	79	80	81	72	80	81	79	80	78	81	79	80
		MERL	85	90	91	75	81	90	89	95	96	89	90	91
Gamma	Error	MI^{MA}	1.3	1.5	2.7	4.5	3.1	4.6	2.2	2.9	3.4	1.8	2.4	4.1
		MERL	1	1.2	2.1	4.1	4.1	5	1.9	2	3.4	2	2.1	3.9
	Power	MI^{MA}	77	80	79	75	78	80	81	81.9	79	81	83	82
		MERL	88	95	93	80	90	95	90	95	96	88	92	94

TABLE 3: Percentage of bias (%) under the null and alternative hypothesis and 95% confidence interval coverage probability under the alternative hypothesis $H_1:\theta \neq 1$ for estimating $\log\theta$ for our proposed method when the covariance structure is AR (1) and the distributions are exponential (E), Weibull (W), and Gamma (G).

| Mean pairwise correlation | | | $\bar{\rho} = 0.5$ | | | | | | $\bar{\rho} = 0.7$ | | | | | |
| Percentage of censoring | | | 30% | | | 50% | | | 30% | | | 50% | | |
Dist	N /seq		12	24	48	12	24	48	12	24	48	12	24	48
E	Bias (H_0)		-3×10^{-5}	-4×10^{-5}	3×10^{-5}	-2×10^{-5}	-3×10^{-5}	-3×10^{-5}	45×10^{-5}	-5×10^{-5}	71×10^{-5}	2×10^{-5}	2×10^{-5}	-3×10^{-5}
	Bias (H_1)		0.80	1	0.6	3	3.1	2.9	1.6	1.7	1.2	3	5	4.9
	Coverage		94.5	94.9	95	97	95	94.4	95.3	95	96.5	96	95.5	97
W	Bias (H_0)		4×10^{-5}	3×10^{-5}	68×10^{-6}	-1×10^{-5}	-2×10^{-5}	3×10^{-5}	4×10^{-5}	-3×10^{-5}	5×10^{-5}	5×10^{-5}	2×10^{-5}	-2×10^{-5}
	Bias (H_1)		0.6	0.82	0.5	2.5	2.7	2.9	1.5	1.7	1.3	2.8	4.9	5
	Coverage		95	95.1	96	96	96.7	95	96	97	96.7	97	96	94.5
G	Bias (H_0)		1×10^{-5}	2×10^{-5}	-3×10^{-5}	-5×10^{-5}	1×10^{-5}	4×10^{-5}	2×10^{-5}	-2×10^{-5}	5×10^{-5}	0×10^{-5}	0×10^{-5}	3×10^{-5}
	Bias (H_1)		-6.1	-5.5	-4.9	-8.2	-6.5	-5.5	-5.4	-4.1	-3.5	-7.9	-8.5	-9
	Coverage		95.1	94	95	95.3	94.2	95.1	95.5	94.4	95.2	96	95.3	95.1

Liu et al. set the upper limit as some reasonable value [18, 19]. Based on Liu et al.'s simulations, their method was much better than its competitor in respect of bias and efficiency. Survival estimates of restricted lifetime model parameters and marginal survival estimates had better precision.

Based on Grover's method, the estimates were better and the standard errors were smaller in comparison to the analysis of the data having censored observations in outcome but ignoring cases with missing covariates and censored outcomes. In both mentioned studies, the final achievements were similar to ours in terms of reducing bias and increasing precision, but there are some weakness, although statistical inferences based on mean survival time are appealing, but as discussed before, it has some limitations. Another aspect that should be noted is that if the observed survival time is too long or the follow-up time is very short, the mean survival time is not a good choice [20]. Multiple imputation leads to more complicated computation and work. Obviously in regression modeling, using covariates that are correlated to the response variable leads to the improvement of predictions [18]. Although, in these studies, demographic information was imbedded in the model of imputation, the superiority of our new method has been use of median residual life time instead of mean residual life time

TABLE 4: Event times (min) for a 10 min treadmill test in a 2 × 2 crossover clinical trial.

| | Placebo-drug sequence | | | | Drug-placebo sequence | | | |
| | Period 1 (placebo) | | Period 2 (drug) | | | Period 1 (drug) | | Period 2 (placebo) | |
Subject (age)	X_1	Y_1	X_2	Y_2	Subject (age)	X_1	Y_1	X_2	Y_2
1 (52.91)	1.5	1	1	1.5	2 (65)	1	1	1	2.5
3 (45.79)	6	4	3.5	>10	4 (40)	6	>10	2.5	2.5
5 (64.37)	1	1	1.5	4.5	6 (59.08)	3	2	1	.5
7 (54.13)	3.5	1.5	.5	3	8 (63.46)	2.5	2.5	1.5	2
9 (61.14)	.5	1	3.5	8	10 (51.78)	2	2.5	2.5	3
11 (47.59)	6	10	6	>10	12 (58.59)	1.5	4.5	2.5	1
13 (70)	.5	.5	1	>10	14 (55.08)	3.5	5.5	4.5	9.5
15 (57.28)	1	1	1	2.5	16 (65.16)	1	2	2	>10
17 (59.75)	1.5	1	.5	.5	18 (41.34)	6	>10	5	3.5
19 (67.77)	1	1.5	2	4	20 (59.44)	2	3	1.5	1.5
21 (42.91)	5	5.5	3	1.5	22 (65.54)	1.5	2.5	1.5	.5
23 (50.72)	2.5	5	6	4.5	24 (70)	1.5	3.5	2.5	3
25 (47.01)	5	5.5	4.5	6	26 (55.63)	3.5	9	6	6
27 (62.26)	1	2	2.5	8.5	28 (62.93)	2	5.5	3.5	8
29 (40)	5	5.5	3.5	2	30 (56.67)	2.5	2.5	1	.5
31 (66.27)	.5	1	2	7.5	32 (63.13)	2.5	3.5	2.5	4
33 (48.16)	5	4	2	2	34 (41.62)	5.5	3	1	.5
35 (65.04)	.5	.5	1	1.5	36 (55.70)	3	5.5	5	.5
37 (66.49)	1.5	2	3	3	38 (66.62)	.5	1	1	5.4
39 (43.19)	6	4	1.5	.5	40 (53.10)	2.5	5	2.5	.5
Median	1.5	1.75	2	3.5	Median	2.5	3.25	2.5	2.5

X_1: baseline response in period 1; Y_1: posttreatment response in period 1; X_2: baseline response in period 2; Y_2: posttreatment response in period 2.

TABLE 5: Estimated survival time of censored observations in treadmill test for the multiple imputation method, the proposed method, and the proposed method by adding age to the model.

| | Placebo-drug sequence | | | Drug-placebo sequence | | |
Model	Subject (age)	Period 2 (drug) Y_2	Subject	Period 1 (drug) Y_1	Period 2 (placebo) Y_2
	3	14.14 ± 0.35	4	14.41 ± 0.39	_
Multiple imputation	11	21.46 ± 0.40	16	_	10.63 ± 0.37
	13	10.62 ± 0.41	18	23.34 ± 0.51	_
	3	15 ± 0.30	4	15.1 ± 0.33	_
MERL	11	23 ± 0.38	16	_	11.47 ± 0.32
	13	11.5 ± 0.40	18	24.58 ± 0.48	_
	3 (45.79)	16.1 ± 0.28	4 (40)	16.1 ± 0.31	_
MERL (adding age)	11 (47.58)	24 ± 0.31	16 (65.16)	_	12.5 ± 0.30
	13 (70)	12.2 ± 0.35	18 (41.34)	25.1 ± 0.43	_

Y_1: posttreatment response in period 1; Y_2: posttreatment response in period 2.

which is more robust. [16]. Additionally, the other advantage of our proposed method has been applying the information of censoring time, which was ignored in the mentioned studies.

We proposed a new method to impute censored observation in time-to-event outcomes in crossover designs. To our knowledge, this is the first study in which the median residual life regression has been applied to impute censored survival data based on correlated covariates. Our proposed method, median residual life regression, is free of the stated limitations. In addition to resolving the mentioned limitations, our method has two additional advantages. It considers the

effect of censored time when imputing the life times for censored observations, by adding the time of censoring to the predicted median residual life, as well as its simplicity.

We showed that our method generated more efficient results in comparison to the recently presented method "multiple imputation" by Xu et al. [1], through different combinations of sample size, percentage of censoring, distribution, and mean pairwise correlation coefficient.

Based on our simulations, we found that our method had equal or higher power than the multiple imputation method proposed by Xu et al. In addition, there were no inflated type 1 errors and the entire bias was negligible under the null and alternative hypothesis. This is a notable achievement of our study. The reason is that Xu et al.'s method is based on considerable computation and multiple imputations. Furthermore, they applied complicated rules to combine the datasets in imputation steps, while our proposed method is straightforward and not that time-consuming. In addition, we pointed out some practical issues, which were ignored in the mentioned study. First of all, consideration of the confounding and effective factors in computing the median residual life might lead to a more accurate prediction of the censored observations. As there are many effective factors, a penalized method [21] can be applied to select the most important variables. Another advantage of our method is that the imputed time is added to the censored time, but the role of censoring time is not considered in the multiple imputation method.

The comparison of our study results with the results of multiple imputation method [1] showed that the results of both methods were almost the same and the P value of both methods when testing the effect of treatment was significant. The estimation of the parameters for both methods was close to each other, but the estimates from our method had less standard error and were more accurate. The comparison of imputed times for censored observations revealed that the imputation method based on our proposed model which used the age variable in imputation was more accurate than the multiple imputation model. It was even more accurate than our proposed method regardless of age and led to more accurate predictions.

Crossover designs can be performed in small sample sizes, such as 4 or 6. One limitation in our method is that it cannot be converged in the mentioned sample sizes, similar to Xu et al.'s method. Additionally, we considered 30% and 50% of censoring, while in real data a higher censoring percentage can be encountered. Another issue that should be pointed out is that censoring was independent of covariates in our study. Further studies are warranted to evaluate the performance of the discussed methods at higher percentages of censoring, for censoring which is not independent of covariates and for evaluating median residual life regression in imputing censors in all survival studies besides crossover designs.

6. Conclusion

Our proposed method outperformed the multiple imputation method. All the powers were greater than 80%, and there was no bias in estimating the parameter of interest. In addition, all type 1 errors were less than 5%. Thus, it is suggested for imputing censored observations given its simplicity and not very heavy computational work as well as its efficiency.

Data Availability

The [table] data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] R. Xu, D. V. Mehrotra, and P. A. Shaw, "Incorporating baseline measurements into the analysis of crossover trials with time-to-event endpoints," *Statistics in Medicine*, vol. 37, no. 23, pp. 3280–3292, 2018.

[2] D. A. Berry, *Statistical methodology in the pharmaceutical sciences*, CRC Press, 2016.

[3] X. Chen, Z. Meng, and J. Zhang, "Handling of baseline measurements in the analysis of crossover trials," *Statistics in Medicine*, vol. 31, no. 17, pp. 1791–1803, 2012.

[4] M. G. Kenward and J. H. Roger, "The use of baseline covariates in crossover studies," *Biostatistics*, vol. 11, no. 1, pp. 1–17, 2010.

[5] S. S. Senn and S. Senn, *Cross-over trials in clinical research*, vol. 5, John Wiley & Sons, 2002.

[6] M. Hills and P. Armitage, "The two-period cross-over clinical trial," *British Journal of Clinical Pharmacology*, vol. 8, no. 1, pp. 7–20, 1979.

[7] C. Metcalfe, "The analysis of cross-over trials with baseline measurements," *Statistics in Medicine*, vol. 29, no. 30, pp. 3211–3218, 2010.

[8] Z. Yan, "The impact of baseline covariates on the efficiency of statistical analyses of crossover designs," *Statistics in Medicine*, vol. 32, no. 6, pp. 956–963, 2013.

[9] D. V. Mehrotra, "A recommended analysis for 2× 2 crossover trials with baseline measurements," *Pharmaceutical Statistics*, vol. 13, no. 6, pp. 376–387, 2014.

[10] A. Kimchi, G. Lee, E. Amsterdam, K. Fujii, P. Krieg, and D. T. Mason, "Increased exercise tolerance after nitroglycerin oral spray: a new and effective therapeutic modality in angina pectoris," *Circulation*, vol. 67, no. 1, pp. 124–127, 1983.

[11] J. D. Markman, M. E. Frazer, S. A. Rast et al., "Double-blind, randomized, controlled, crossover trial of pregabalin for neurogenic claudication," *Neurology*, vol. 84, no. 3, pp. 265–272, 2015.

[12] A. P. Grieve, "Crossover versus parallel designs," in *Statistical methodology in the pharmaceutical sciences*, pp. 253–284, CRC Press, 2016.

[13] J. A. Lewis, "Migraine trials: crossover or parallel group?," *Neuroepidemiology*, vol. 6, no. 4, pp. 198–208, 2004.

[14] G. Van Belle, *Statistical rules of thumb*, vol. 699, John Wiley & Sons, 2011.

[15] E. Brittain and D. Follmann, "A hierarchical rank test for crossover trials with censored data," *Statistics in Medicine*, vol. 30, no. 30, pp. 3507–3519, 2011.

[16] J.-H. Jeong, *Statistical inference on residual life*, Springer, 2014.

[17] P. Xue-Kun Song, "Multivariate dispersion models generated from Gaussian copula," *Scandinavian Journal of Statistics*, vol. 27, no. 2, pp. 305–320, 2000.

[18] G. Grover and V. K. Gupta, "Multiple imputation of censored survival data in the presence of missing covariates using restricted mean survival time," *Journal of Applied Statistics*, vol. 42, no. 4, pp. 817–827, 2014.

[19] L. X. Liu, S. Murray, and A. Tsodikov, "Multiple imputation based on restricted mean model for censored data," *Statistics in Medicine*, vol. 30, no. 12, pp. 1339–1350, 2011.

[20] L. Sun, X. Song, and Z. Zhang, "Mean residual life models with time-dependent coefficients under right censoring," *Biometrika*, vol. 99, no. 1, pp. 185–197, 2012.

[21] S. E. Ahmed, *Penalty, shrinkage and pretest strategies: variable selection and estimation*, Springer, 2014.