Published in final edited form as:

Clin Trials. 2011 October ; 8(5): 581–590. doi:10.1177/1740774511414741.

Using Cure Models and Multiple Imputation to Utilize Recurrence as an Auxiliary Variable for Overall Survival

Anna S. C. Conlon, MS^a, Jeremy M. G. Taylor, PhD^a, Daniel J. Sargent, PhD^b, and Greg Yothers, PhD^{c,d}

^aDepartment of Biostatistics, School of Public Health, University of Michigan, 1420 Washington Heights, Ann Arbor, MI, USA

^bMayo Clinic, 200 First St. SW, Rochester, MN, USA

^cDept. of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA

^dNational Surgical Adjuvant Breast and Bowel Project (NSABP) Biostatistical Center, Pittsburgh, PA, USA

Abstract

Background—Intermediate outcome variables can often be used as auxiliary variables for the true outcome of interest in randomized clinical trials. For many cancers, time to recurrence is an informative marker in predicting a patient's overall survival outcome, and could provide auxiliary information for the analysis of survival times.

Purpose—To investigate whether models linking recurrence and death combined with a multiple imputation procedure for censored observations can result in efficiency gains in the estimation of treatment effects, and be used to shorten trial lengths.

Methods—Recurrence and death times are modeled using data from 12 trials in colorectal cancer. Multiple imputation is used as a strategy for handling missing values arising from censoring. The imputation procedure uses a cure model for time to recurrence and a time-dependent Weibull proportional hazards model for time to death. Recurrence times are imputed, and then death times are imputed conditionally on recurrence times. To illustrate these methods, trials are artificially censored 2-years after the last accrual, the imputation procedure is implemented, and a log-rank test and Cox model are used to analyze and compare these new data with the original data.

Results—The results show modest, but consistent gains in efficiency in the analysis by using the auxiliary information in recurrence times. Comparison of analyses show the treatment effect estimates and log rank test results from the 2-year censored imputed data to be in between the estimates from the original data and the artificially censored data, indicating that the procedure was able to recover some of the lost information due to censoring.

Limitations—The models used are all fully parametric, requiring distributional assumptions of the data.

Conclusions—The proposed models may be useful to improve the efficiency in estimation of treatment effects in cancer trials and shortening trial length.

Keywords

Auxiliary Variables; Colon Cancer; Cure Models; Multiple Imputation; Surrogate Endpoints

1. Introduction

There is an extensive literature on the use of intermediate outcome variables as either surrogate endpoints [1, 5, 22] or as auxiliary variables [8, 11] for the true outcome of interest in randomized clinical trials. A surrogate endpoint is one that is intended to replace the true endpoint in evaluating the therapy. An auxiliary variable is one that is intended to be used to improve the efficiency of the analysis of the true endpoint.

For randomized clinical trials in colorectal cancer overall survival (OS) has traditionally been considered the definitive endpoint. An alternative endpoint is disease free survival (DFS), defined as the time to the first event of either recurrence of the cancer or death, and is considered by some to be preferable [4]. It has been determined that DFS is a good surrogate endpoint for OS and is therefore considered an appropriate endpoint to use in clinical trials of colorectal cancer [20]. An alternative approach, which will be the focus of this paper, is to consider recurrence as an auxiliary variable, which can be used to enhance the efficiency of the analysis of the primary OS endpoint.

The data we consider are from 12 randomized trials of colorectal cancer. In such trials, patients are followed with cancer recurrences and deaths recorded as they occur. Recurrence is due to re-growth of the cancer, with the time of recurrence being when the cancer has grown to such a size or extent that it is detectable or causes symptoms. Deaths can occur either without prior recurrence or after a recurrence. The deaths that occur without prior recurrence are known not to be due to re-growth of the cancer, whereas deaths that occur after recurrence could be due to the cancer or to other causes. The exact cause of death for each person is either not known or known with certainty or known but not considered reliable. We do not consider cause of death. In this paper we will be developing models for time-to-recurrence and time-to-death. These models may be useful for understanding how these endpoints depend on covariates, including treatment, and for understanding the association between recurrence and death. They may also be useful for designing future trials, as they represent a way to simulate data and impose plausible magnitudes of treatment effects. However, we will primarily be utilizing these models to improve the efficiency of the analysis of the treatment effect on time to death.

In these data there are two censored event times of interest. The first is time to recurrence, and the second is time to death. An interesting feature of the recurrence outcome is that a proportion of the censored subjects may be cured and thus would have never experienced the event of interest even if they had been followed for longer. For other censored subjects, their event time would occur after their censoring time with longer follow-up and is therefore unobserved. For these types of data a cure model [6, 15, 21] would be applicable. For the death outcome, all subjects will eventually experience the event, and a cure model is therefore not likely to be appropriate.

Patients who are alive are right censored for death, which we will consider as a form of missing data. Relative to knowing the time of death, there is loss of information in a right censored observation. A patient's recurrence status prior to their censoring time is generally known, and those who experienced a recurrence are likely to die sooner than those who are recurrence free. Thus a patient's recurrence status at the censoring time may be useful in recovering some of the lost information due to censoring. The strategy that we will use to

recover some of this lost information is multiple imputation of death times based on regression models that include recurrence status as a time-dependent covariate. Multiple imputation is a popular general strategy for many missing data problems [17], and has been used to impute event times for censored observations in survival analysis [7, 10]. This strategy fills in missing values by drawing from the posterior predictive distribution of the missing data given the observed data. The procedure is then independently repeated M times, to produce separate datasets. These completed datasets can then be analyzed separately to give M estimates and standard errors for the quantity of interest. The results from each of these analyses are then combined following established rules [18] to give final conclusions.

We describe parametric models for recurrence and death that are used in the imputation strategy for the censored observations. For recurrence, a cure model is used and for death, a proportional hazards model with recurrence as a time-dependent covariate is employed. Each imputation is done in two steps. First, the cure model is used to model the time to cancer recurrence. This model is then used to construct the appropriate predictive distribution from which the missing values for recurrence times are drawn. Cure models assume that a proportion of subjects will never experience the event, thus some censored subjects are imputed as cured for recurrence, while others are given a recurrence time after their observed censoring time. Using these imputed recurrence times, a proportional hazard model is employed to model survival times, with recurrence as a time-dependent covariate. Death times for censored subjects are then imputed from the appropriate predictive distribution derived from this model. Once these death times are imputed the data is ready for subsequent analysis, specifically the analyses of interest in this paper are ones that assess the treatment effect on death, without adjusting for recurrence.

In the following sections of this paper we describe the data, and give details of the models and imputation strategy. We then present the results of fitting these models to the data. Finally, we illustrate the potential for this modeling approach to shorten the length of a clinical trial. Specifically, we artificially censor each trial two years after its final accrual, and compare the results from the modeling and imputation strategy applied to these data with simple analysis of the artificially censored data and the original data.

2. Data Description

The data consist of a total of 14,034 subjects from 12 randomized phase III adjuvant colorectal cancer clinical trials. Ten of the trials are included in the Sargent *et al.* (2005) publication, with two additional new trials. For each trial, one arm was defined as the control arm, and the other as the experimental arm. Each trial compared a different pair of treatments. Five of the trials (1, 2, 3, 6, and 7) compared surgery alone to surgery plus some form of chemotherapy. In the other seven trials, both arms contained surgery plus some form of chemotherapy and have been conducted in more recent years. Trial enrollment spanned from 1977 to 1999. One trial (7) included 210 patients with stage 1 cancer; these subjects were excluded from analysis. Due to differences in the long term follow-up practices between trials, subjects in all trials except trial 1 were censored at 8 years following the time of the last subject accrual. Subjects in trial 1 were censored 4.3 years from the last subject accrual due to a large number of patients administratively censored at this time. These data are described in Table 1. The median follow-up time for patients alive at their last follow up was 8.2 years. Subjects were followed with information on recurrence status, time to recurrence, survival time, age at treatment, cancer stage, and treatment group recorded. The censoring time for recurrence and death were not necessarily the same. For patients who were alive and censored for recurrence, the average proportion of patients with different censoring times for these two events was 11.6% across all trials, with a maximum of 38.2%

in trial 1 and a minimum of 0% in trial 3. Table 1 provides a summary of stage, age and treatment distributions in each study, as well as the number of recorded recurrences and deaths and longest follow-up time for each study. Recurrence rates for years 1 through 3 following randomization were 9.3%, 10.1%, and 5.2%, respectively. Kaplan-Meier plots of time to recurrence showed a clear leveling off with nearly all the recurrences happening before 5 years, which is characteristic of a cured group, and for which a cure model is appropriate.

3. Model Descriptions and Parameter Estimates

3.1. Model for Recurrence

Cancer recurrence is modeled using a lognormal cure model fit separately to each trial. This type of mixture model, containing two parts which can be separately interpreted, was first introduced by Berkson and Gage (1952). Cure models assume that a proportion of subjects will never experience the event, and are used to estimate and model the cured fraction in the population. For the uncured population, cure models provide information on the estimated time to event from the survival distribution. Let $R_1, \dots, R_n$ denote the times to recurrence for n subjects in a given study and $C_1, \dots, C_n$ the corresponding potential censoring times. Then $X_i = \min(R_i, C_i)$ and the event indicator for recurrence $\delta_i = I(R_i \leq C_i)$ are observed. For subjects who are censored for recurrence, some are administratively censored at the end of follow-up, while others are censored due to death from a non-cancer cause. We assume the censoring is non-informative. The cure model specifies the marginal distribution as $S(r) = 1 - p + pS_o(r)$, where $0 < p < 1$, and $S_o(r) = 1 - F_o(r)$, where $1 - p$ corresponds to the fraction of cured patients, and $F_o(r)$ represents the distribution of recurrence times amongst those who are not cured. Both p and $F_o(r)$ can depend on covariates and the likelihood contribution for subject i is:

$$\begin{aligned} & p_i f_o(x_i) && \text{for subjects with recurrence} \\ 1 - p_i + p_i S_o(x_i) && \text{for subjects censored for recurrence} \end{aligned}$$

Assuming logistic and lognormal models:

$$p_i = \frac{\exp(Z_i' \alpha)}{1 + \exp(Z_i' \alpha)} \quad (1)$$

$$f_o(x_i; W_i, \sigma) = \frac{1}{x_i \sigma \sqrt{2\pi}} \exp\left(-\frac{(\ln(x_i) - W_i' \beta)^2}{2\sigma^2}\right) \quad (2)$$

$$S_o(x_i; W_i, \sigma) = 1 - F_o(x_i) = 1 - \Phi\left(\frac{\ln(x_i) - W_i' \beta}{\sigma}\right)$$

Where Φ is the standard normal CDF, Z are the covariates which include treatment group and cancer stage, and W are the covariates including treatment group, cancer stage and age. Age was excluded from the logistic component of the model due to its lack of association with the probability of cure in most trials.

Table 2 provides cure model estimates from the 12 trials. These results were obtained using the R program, version 2.9.1, and the gfcure package from Peng, Dear, and Denham (1998).

The results show a consistent and strong effect of stage on the probability of recurrence. The final column provides the estimated probability of cure for stage 3 patients in the control group. The 5 trials where the control arm was surgery alone (trials 1, 2, 3, 6 and 7) tend to have lower cure rates on the control arm than the trials where the control arm included chemotherapy, illustrating the benefit of adding chemotherapy. There is also some consistency of the effect of stage and age on the time to recurrence, with higher stage and younger people tending to recur earlier. The effect of age on recurrence is interesting, in that it is not associated with whether someone is cured, but is associated with when they recur, with the direction of association being the opposite of what is seen when considering time to death, for which older people die sooner. The effect of treatment varies from one trial to the next. This is to be expected as the treatments in each trial differ. The median time to recurrence for the recurred group, shown in the last column, is between 12 and 22 months, and tends to be longer for later trials.

This model was used to impute recurrence times for subjects alive and censored for recurrence. The imputation procedure is described in section 4. These imputed values were then used as covariates in the models for the survival times. Twenty-five imputed data sets were created, with each data set only differing in imputed recurrence times.

The appropriateness of using the log-normal cure model instead of other parametric cure models was assessed using log-likelihood and AIC values. The log-normal was preferred to Weibull and gamma cure models for 10 out of 12 and 8 out of 12 datasets respectively.

3.2. Model for Death

Survival time was modeled using a proportional hazards model with a Weibull baseline hazard function. Let $T_1, \dots, T_n$ denote the survival times for n subjects in a given study, and $V_1, \dots, V_n$ the corresponding potential censoring times. Then $Y_i = \min(T_i, V_i)$ and the event indicator $\Delta_i = \mathbb{I}(T_i \leq V_i)$ are observed. The hazard model for subject i is given by:

$$\lambda(t; X_i, R_i) = \left(\frac{\rho}{\lambda}\right) \left(\frac{t}{\lambda}\right)^{\rho-1} \exp(X_i' \gamma + \theta I(t \geq R_i)) \quad (3)$$

where X are the covariates that include treatment group, cancer stage, and age and $\mathbb{I}(t \geq R_i)$ is a time dependent binary indicator for recurrence. The Weibull baseline hazard was chosen because it was considered flexible enough to describe these data. This model was fit to the survival data to obtain parameter estimates. Twenty-five different parameter values were obtained (using the procedure described in section 4), one from each data set with imputed recurrence values, and then combined using the standard combining rules for multiple imputation, to give the final parameter estimates shown in Table 3. The results show a very strong effect of recurrence on the hazard of death in all trials. Most trials also show a strong association of both stage and age on the time to death, with higher stage and older people dying earlier. The treatment coefficients vary from one trial to the next, and are mostly quite small in magnitude, likely because these estimates are adjusted for recurrence.

Diagnostic checks were made to assess whether the proportional hazards assumption was appropriate for age, stage and recurrence covariates. There was no evidence of substantial or consistent departures from the proportional hazards assumption across the 12 datasets for age and recurrence. For stage there was evidence that the hazard ratio between stage 3 and stage 2 increased with longer follow-up time for some of the datasets. A test of non-proportional hazards was significant in 6 of the 12 trials. We investigated using a hazard model stratified by cancer stage and performing the imputation procedure based on this model, but this did not lead to any conclusions that were substantially different, thus we present results for the simpler, unstratified model.

We note that the imputation of recurrence times was necessary to obtain the results in table 3 only because a small fraction of subjects were censored for recurrence earlier than they were censored for death.

4. Multiple Imputation of Recurrence and Death Times

Here we describe the two step strategy of multiple imputation using the models described above. The first step in the procedure of each imputation involves drawing new recurrence times using the cure models for subjects who were censored for recurrence. Given this new dataset the second step, consists of drawing new death times from the Weibull proportional hazards model for those censored for death. This procedure is independently repeated M times to obtain multiple imputed data sets for subsequent analysis.

4.1 Imputation of Recurrence Times

Imputed recurrence times were drawn from residual time distributions $P(R > c_i + a_i | R > c_i, \theta)$, where c_i is the censoring time for recurrence and θ are the parameters in Table 2. Specifically, imputation was performed as follows: First, a random number, $U_i \sim U(0, 1)$ was generated for each subject without recurrence. From the cure model, the conditional distribution of the time to recurrence for censored subject i is given by:

$$P(R_i > c_i + a_i | R_i > c_i) = \frac{1 - p_i + p_i S_o(c_i + a_i)}{1 - p_i + p_i S_o(c_i)}$$

where a_i is the imputed time to recurrence after c_i . Setting this residual time distribution equal to the U_i and solving for a_i , we obtain:

$$a_i = F_o^{-1} \left[1 - \frac{U_i [1 - p_i + p_i S_o(c_i)] - (1 - p_i)}{p_i} \right] - c_i$$

Where F_o is the lognormal CDF.

Subjects with $U_i \leq \frac{1 - p_i}{1 - p_i + p_i S_o(c_i)}$ were considered cured ($\delta_i^* = 0$). For these subjects, the quantity $1 - \frac{U_i [1 - p_i + p_i S_o(c_i)] - (1 - p_i)}{p_i}$ was taken to be 1. The end result was a new pair of variables, (x_i^*, δ_i^*) for each subject who was originally censored for recurrence. For subjects who were not cured, $\delta_i^* = 1$, and x_i^* was equal to the imputed recurrence time ($c_i + a_i$). Imputed subjects who were considered cured were given $\delta_i^* = 0$, and x_i^* was set to 50,000 years, an arbitrary number larger than any human life span. In order to avoid extrapolation of recurrence times outside of the range of the observed data, imputed recurrence times that were greater than the longest follow-up time for a given trial were relabeled as censored ($\delta_i^* = 0$) at the largest follow-up time. To check the validity of the imputed recurrence times, Kaplan-Meier plots using the new imputed recurrence times were compared to Kaplan-Meier plots from the original data. The two plots were very similar, suggesting that no distortion was being introduced by the imputation. These new imputed recurrence times together with the observed recurrence times, were then used in the procedure for the imputation of death times for those who were alive at their last follow-up.

4.2 Imputation of Death Times

We impute death times from the residual survival distribution $P(T < v_j + u | T > v_j, r_j, \phi)$, where v_j is the recurrence time and ϕ are the parameters in Table 3. To achieve this imputation of death times we first generated $U_i \sim U(0, 1)$, then this value was then set equal to the residual survival function for each censored subject. Specifically,

$$U_i = \exp \left(- \int_{v_i}^{v_i + b_i} \lambda(u) du \right)$$

was solved for b_i , where $\lambda(u)$ is given by equation 3. The value of b_i is the imputed time to death after v_i .

The specific details of solving for b_i are as follows. For subjects who had a recurrence at time r_j before their time of censoring for death, v_j , the following equation was solved:

$$b_i = F^{-1} \left[- \frac{\ln(U_i)}{\exp(X_i' \gamma + \theta)} + F(v_i) \right] - v_i$$

where $F(v) = \left(\frac{v}{\lambda} \right)^p$. For subjects who were cured, or had imputed recurrence times after their original censoring time for death, the following equation was first solved.

$$b_i = F^{-1} \left[- \frac{\ln(U_i)}{\exp(X_i' \gamma)} + F(v_i) \right] - v_i \quad (4)$$

If the imputed death time was still less than the recurrence time (r_j), it was kept as the new death time. If the imputed death time was greater than the recurrence time, then b_i from equation 4 was replaced by the solution to:

$$b_i = F^{-1} \left[\exp(-\theta) \left(- \frac{\ln(U_i)}{\exp(X_i' \gamma)} + F(v_i) - F(r_i) \right) + F(r_i) \right] - v_i$$

To avoid extrapolation beyond the range of observable data imputed death times were then censored at the minimum of the longest follow-up time for a trial and 10 years.

5. Application of the Model for Efficiency Gains

One potential use of this modeling approach is to shorten the length of a clinical trial while still keeping the primary endpoint of overall survival. To illustrate this, we artificially censored each of the 12 trials two years after the last patient accrual, and performed the multiple imputation procedure described above to obtain new recurrence and death times for censored subjects. Censoring at 2 years resulted in an overall 20.7% reduction in the number of recurrences across all trials compared to the original data with a maximum of 26.9% in trial 11 and a minimum of 4.8% in trial 3. The overall reduction in the number of deaths after 2-year censoring was 48.8% with a maximum of 58.0% in trial 11 and a minimum of 30.2% in trial 1. With death as the endpoint of interest, these new data were then analyzed and compared to analyses of the original data to assess efficiency gains. The questions of interest here are (i) if the trial had stopped earlier, would the modeling and imputation strategy have lead to conclusions that were more accurate than simple analysis of survival

from the available data, and (ii) would the conclusions have been compatible with the final results obtained from the full data.

The imputation strategy was performed 25 times on the artificially 2-year censored data. Each of these 25 imputed data sets were then analyzed using a log-rank test for treatment effect and also using a Cox proportional hazards model, $\lambda_0(t)\exp(X_i'\omega)$, where X includes treatment, stage and age as covariates but excludes recurrence as a covariate. The results from these analyses were then combined using the established rules for multiple imputation, and compared to analyses performed on the original data. The rules established by Rubin (1987) were used in combining parameter values and standard errors. Log-Rank test Chi-Square statistics were combined using the methods of Li *et al.* (1991).

Table 4 provides treatment estimates, and standard errors from the Cox proportional hazards models for the original data, the 2-year censored data without imputation, and the 2-year censored data with imputation. The point estimates of the treatment effect from the imputed data tend to be in between those of the original data and the 2-year censored data. The standard errors from the imputed data are smaller than from the 2-year censored data and in some cases are quite similar to those obtained from the original data. Table 4 also provides the results of the log-rank tests for treatment effect from the original, 2-year censored, and imputed 2-year censored data. The results obtained from the imputed 2-year censored data are closer to the original results than the 2-year censored data without imputation, indicating that some information that was lost through censoring the trials early was correctly recovered through imputation. Assessing the crude agreement of the results for the endpoints (based on p-values < 0.05) [2], the conclusions for the imputed data sets agreed with those of the original data for 10 of the 12 trials. In the censored datasets with no imputation, 9 of the 12 trials reached the same conclusion as the original data. This provides good support for the potential of this approach to allow shorter length clinical trials while preserving overall survival as the primary endpoint.

6. Simulation Study

One natural concern with the modeling and imputation approach is whether it preserves type I error under a null hypothesis of no treatment effect. Multiple imputation has a Bayesian justification and thus is not primarily concerned with frequentist properties such as the size of tests and coverage rates of confidence intervals. However, in numerous simulation studies frequentist properties following multiple imputation have been excellent when appropriate models have been used. To address the statistical properties of the methodology described in the current paper, we undertook a small simulation study. Recurrence times and death times were simulated from the models described in sections 4.1 and 4.2 to give us 'original data'. These times were then censored at two years after the last accrual date to give us '2 year censored data'. We then used the modeling and imputation strategy to impute recurrence and death times for the '2 year censored data' to give us the 'imputed data'. For each of the three datasets we assessed the treatment effect on time to death. We considered the hypothesis test of no treatment effect using the log-rank test, and estimates and standard errors of the relative hazard from a Cox model, $\lambda_0(t)\exp(X_i'\omega)$.

We designed the trials to have 150 subjects in each arm accrued over a 900 day period with follow-up extending to five years after the end of the accrual period. One trial was designed with 300 subjects per arm for comparison. We created 1000 such trials. Recurrence times were generated from equations 1 and 2, with the only covariate being treatment group. Death times were generated from equation 3 with the only covariates being treatment group and time-dependent recurrence. Under the null hypothesis ($\alpha_1 = \beta_1 = \gamma_1 = 0$), we varied the other parameters ($\alpha_0, \beta_0, \sigma, \rho, \lambda$, and θ) in four different scenarios. The results in Table 5

provide the size of the log-rank test, the average of the estimated treatment effect, $\hat{\omega}_1$, from the Cox model, the empirical SD of $\hat{\omega}_1$, and the average of $SE(\hat{\omega}_1)$. The results demonstrate for the situations considered, that the imputation strategy does preserve type 1 error, does not introduce bias into the estimate of ω_1 , that the SE's of $\hat{\omega}_1$ appear valid and there is some gain in efficiency in analyzing the 'imputed data' compared to the '2 year censored data'.

7. Discussion

The modeling described in this paper is used to assess whether recurrence is an auxiliary variable that can be used to improve the efficiency of the analysis while keeping overall survival as the endpoint. The results show modest, but consistent and clear gains in efficiency, as measured by smaller standard errors, due to using the recurrence information. We believe this could be useful, either for shortening the planned length of the trial or aiding data safety and monitoring boards, at the time of interim analyses, in deciding whether to end a trial by predicting the efficiency gain likely to be achieved by extending trial follow-up.

The multiple imputation approaches we use serve two purposes. The imputation of recurrence times is necessary only so that recurrence can be easily included as a time-dependent variable in the hazard model for death, for those whose censoring times for recurrence and death differ. The imputation of death times is used explicitly to gain efficiency in the marginal analysis of time to death. We note that the quantity of interest in this marginal analysis does not correspond to any of the parameters in the cure and Weibull models that were used for imputation. Thus this is an example where the desired analysis model differs from the imputation model.

Adaptations of the modeling and estimation approach we have taken are possible. For example, we have modeled data from 12 trials separately, without any attempt to use models to link the results. A natural way to extend the models we have used would be to make them hierarchical, with a set of random effects for each trial. Recent papers provide methodology for doing this [12, 16]. We have used fully parametric models in this paper. While there are certainly semi-parametric alternatives for the cure and Weibull models we have used, we thought it preferable to be imputing recurrence and death times from continuous rather than from discrete distributions. We performed the parameter estimation separately for the cure and Weibull models, using readily available software. We note that it would be possible to simultaneously estimate all the parameters using multistate models with a latent variable to represent the cured group. However, as far as we are aware, software to do this is not available.

We do make the potentially strong assumption of non-informative censoring of the recurrence times, specifically, that those who are censored due to death from another cause would have been no more likely to recur if they had not died than those who were administratively censored given their follow-up and covariate values. We cannot check this assumption, but in the context of colorectal cancers, we have no reason to believe it is not approximately valid.

We have used a number of different modeling and imputation strategies to aid in the analysis of data from randomized clinical trials. In the modeling, we consider recurrence and death separately because they are distinct events which may be impacted by covariates in different ways. The parameter estimates we provide for the models could also be useful for those designing future clinical trials, as they provide a way to simulate realistic data while hypothesizing different magnitudes of a treatment effect. The cure model for recurrence has two components, one for determining if a subject was cured and one for determining when

the recurrence happens given a subject is not cured. Consideration of these distinct components may also be useful for designing studies depending on whether the treatment under consideration is thought to be curative or thought to delay recurrence.

We believe that separate models for recurrence and death, with a cured component in the model for recurrence are natural in this setting and could be useful for other cancers, such as head and neck and breast cancer, for which a cured fraction is known to exist. In addition to having the potential for more informative interpretations of the data, such models can also be useful both at the design stage and at the analysis stage for improving efficiency.

References

1. Alonso A, Molenberghs G. Evaluating Time to Event Cancer Recurrence as a Surrogate Marker for Survival from an Information Theory Perspective. *Statistical Methods in Medical Research*. 2008; 17:497–504. [PubMed: 18285443]
2. Begg CB, Leung DHY. On the Use of Surrogate End Points in Randomized Trials. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*. 2000; 163(1):15–28.
3. Berkson J, Gage RP. Survival Curve for Cancer Patients Following Treatment. *Journal of the American Statistical Association*. 1952; 47:501–515.
4. Broglio KR, Berry DA. Detecting an Overall Survival Benefit that is Derived From Progression-Free Survival. *Journal of the National Cancer Institute*. 2009; 101:1642–1649. [PubMed: 19903805]
5. Buyse M, Molenberghs G. Criteria for the Validation of Surrogate Endpoints in Randomized Experiments. *Biometrics*. 1998; 54:1014–1029. [PubMed: 9840970]
6. Farewell VT. The Use of Mixture Models for the Analysis of Survival Data with Long-Term Survivors. *Biometrics*. 1982; 38:1041–1046. [PubMed: 7168793]
7. Faucett CL, Schenker N, Taylor JMG. Survival Analysis Using Auxiliary Variables Via Multiple Imputation, with Application to AIDS Clinical Trial Data. *Biometrics*. 2002; 58:37–47. [PubMed: 11892691]
8. Fleming TR, Prentice RL, Pepe MS, Glidden D. Surrogate and Auxiliary Endpoints in Clinical Trials with Potential Applications in Cancer and AIDS Research. *Statistics in Medicine*. 1994; 13:955–968. [PubMed: 8047747]
9. Freedman LS, Graubard BI, Schatzkin A. Statistical Validation of Intermediate Endpoints for Chronic Disease. *Statistics in Medicine*. 1992; 11:167–178. [PubMed: 1579756]
10. Hsu C-H, Taylor JMG, Murray S, Commenges D. Survival Analysis Using Auxiliary Variables Via Non-Parametric Multiple Imputation. *Statistics in Medicine*. 2006; 25:3503–3517. [PubMed: 16345047]
11. Kosorok M, Fleming T. Using Surrogate Failure Time Data to Increase Cost Effectiveness in Clinical Trials. *Biometrika*. 1993; 80:823–833.
12. Lai X, Yau KK. Long-term Survivor Model with Bivariate Random Effects: Applications to Bone Marrow Transplant and Carcinoma Study Data. *Statistics in Medicine*. 2009; 27:5692–5708. [PubMed: 18712761]
13. Li K-H, Meng X-L, Raghunathan TE, Rubin DB. Significance Levels from Repeated p-values with Multiply-Imputed Data. *Statistica Sinica*. 1991; 1(1):65–92.
14. Peng Y, Dear KB, Denham JW. A generalized F mixture model for cure rate estimation. *Statistics in Medicine*. 1998; 17:813–830. [PubMed: 9595613]
15. Peng Y, Dear KBG. A Non-Parametric Mixture Model for Cure Rate Estimation. *Biometrics*. 2000; 56:237–243. [PubMed: 10783801]
16. Peng Y, Taylor JMG. Mixture Cure Model with Random Effects for the Analysis of a Multi-Center Tonsil Cancer Study. *Statistics in Medicine*. 2010 In Press.
17. Rubin DB. Multiple Imputation in Sample Surveys—a Phenomenological Bayesian Approach to Nonresponse. *Journal of the American Statistical Association*. 1978:20–34.
18. Rubin, DB. Multiple Imputation for Nonresponse in Surveys. New York: Wiley; 1987.

19. Sargent DJ, Wieand HS, Haller DG, et al. Disease-Free Versus Overall Survival As a Primary End Point for Adjuvant Colon Cancer Studies: Individual Patient Data From 20,898 Patients on 18 Randomized Trials. *Journal of Clinical Oncology*. 2005; 23(34):8664–8670. [PubMed: 16260700]
20. Sargent DJ, Patiyil S, Yothers G, et al. End Points for Colon Cancer Adjuvant Trials: Observations and Recommendations Based on Individual Patient Data from 20,898 Patients Enrolled onto 18 Randomized Trials from the ACCENT Group. *Journal of Clinical Oncology*. 2007; 25:4569–4574. [PubMed: 17876008]
21. Sy JP, Taylor JMG. Estimation in a Cox Proportional Hazards Cure Model. *Biometrics*. 2000; 56:227–236. [PubMed: 10783800]
22. Wang Y, Taylor JMG. A Measure of the Proportion of Treatment Effect Explained by a Surrogate Marker. *Biometrics*. 2003; 58:803–812. [PubMed: 12495134]

Table 1

Data Summary

Study	N	Recurrences	Deaths Without Recur	Total Deaths	Longest Follow-Up (months)	Stage 2	Stage 3	Control	Treat	age (mean)
1-NC-784852	247	120	13	115	226	85	162	126	121	60.3
2-NC-874651	408	140	44	172	136	75	333	153	255	61.1
3-INT-0035	926	377	76	422	158	314	612	469	457	60.2
4-NC-894651	914	382	106	450	164	160	754	227	687	62.7
5-NC-914653	878	297	74	338	154	230	648	441	437	61.2
6-NSA-C01	724	278	132	397	242	313	411	375	349	59.8
7-NSA-C02	686	209	129	303	195	389	297	344	342	63.3
8-NSA-C03	1042	367	67	387	193	291	751	523	519	56.1
9-NSA-C04	2083	615	176	724	168	860	1223	693	1390	57.0
10-NSA-C05	2136	577	192	700	139	945	1191	1070	1066	58.0
11-NSA-C06	1556	395	115	438	96	725	831	776	780	60.5
12-NSA-C07	2434	627	106	544	72	701	1733	1222	1212	57.9
Total	14034	4384	1230	4990	242	5088	8946	6420	7614	59.1

Abbreviations: NSA, National Surgical Adjuvant Breast and Bowel Project; NC, North Central Cancer Treatment Group; INT, Intergroup.

Parameter Estimates from Lognormal Cure Models for Recurrence

Clin Trials. Author manuscript; available in PMC 2012 October 01.

Study	Logistic Model				Cured Fraction *		
	Treatment	Stage	Intercept				
11	-0.02	0.09	-0.28	0.11	0.04	6.52	0.26
12	0.07	0.09	-0.22	0.15	-0.01	6.70	0.27
					0.04		-0.04
							0.05
							0.05

* cured fraction and median days to recurrence for control group, stage 3 cancer, age 61

** estimates given per 10 year increase in age

Table 3

Parameter Estimates from Weibull Proportional Hazards Models for Survival

Study	Treatment		Stage		Log(age)		Recur		Log(Scale)		Log(Shape)	
	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
1	0.17	0.20	0.83	0.25	0.67	0.47	3.88	0.30	13.72	2.17	-0.01	0.13
2	-0.04	0.16	0.11	0.24	1.03	0.49	3.40	0.18	14.05	2.04	0.02	0.09
3	0.24	0.10	0.35	0.12	0.79	0.23	3.49	0.13	13.88	1.01	-0.04	0.04
4	0.08	0.11	0.36	0.15	1.27	0.27	3.33	0.11	17.11	1.53	-0.22	0.06
5	0.18	0.11	0.41	0.15	1.42	0.32	3.68	0.14	20.44	2.24	-0.39	0.08
6	0.11	0.11	0.15	0.11	1.15	0.32	2.95	0.11	14.92	1.55	-0.12	0.06
7	-0.12	0.12	0.30	0.12	2.71	0.42	2.64	0.13	23.39	2.27	-0.18	0.07
8	-0.11	0.11	0.31	0.14	0.32	0.29	3.81	0.14	12.90	1.37	-0.25	0.08
9	0.004	0.08	0.36	0.09	0.99	0.21	3.61	0.09	15.82	1.14	-0.20	0.05
10	0.0006	0.08	0.32	0.09	1.33	0.21	3.50	0.09	17.27	1.18	-0.18	0.05
11	-0.10	0.10	0.42	0.11	1.45	0.28	3.49	0.12	18.84	1.68	-0.25	0.07
12	0.10	0.09	0.12	0.12	1.43	0.24	3.77	0.12	20.04	1.63	-0.34	0.07

Table 4

Analysis of the effect of treatment on survival, from original data, censored data and censored data with imputation

Study	Data	Cox model		Log-Rank
		Log Hazard Ratio	SE	P-Value
1	Original	−0.28	0.19	0.22
	2-Year Censored	−0.45	0.21	0.06
	Imputed 2-Yr Censored	−0.29	0.19	0.22
2	Original	−0.25	0.16	0.09
	2-Year Censored	−0.24	0.25	0.34
	Imputed 2-Yr Censored	−0.17	0.21	0.42
3	Original	−0.31	0.10	0.004
	2-Year Censored	−0.27	0.13	0.05
	Imputed 2-Yr Censored	−0.26	0.11	0.04
4	Original	0.06	0.11	0.64
	2-Year Censored	0.02	0.16	0.94
	Imputed 2-Yr Censored	−0.02	0.14	0.71
5	Original	0.09	0.11	0.35
	2-Year Censored	0.10	0.13	0.44
	Imputed 2-Yr Censored	0.09	0.12	0.43
6	Original	−0.04	0.10	0.80
	2-Year Censored	−0.08	0.13	0.62
	Imputed 2-Yr Censored	−0.05	0.12	0.72
7	Original	−0.19	0.12	0.07
	2-Year Censored	−0.33	0.16	0.04
	Imputed 2-Yr Censored	−0.30	0.14	0.04
8	Original	−0.37	0.10	0.0003
	2-Year Censored	−0.41	0.16	0.007
	Imputed 2-Yr Censored	−0.31	0.12	0.01
9	Original	−0.16	0.08	0.04
	2-Year Censored	−0.16	0.12	0.17
	Imputed 2-Yr Censored	−0.14	0.10	0.15
10	Original	−0.02	0.08	0.86
	2-Year Censored	−0.15	0.11	0.19
	Imputed 2-Yr Censored	−0.07	0.09	0.43
11	Original	0.008	0.10	0.84
	2-Year Censored	−0.12	0.15	0.34

Study	Data	Cox model		Log-Rank
		Log Hazard Ratio	SE	P-Value
	Imputed 2-Yr Censored	−0.07	0.12	0.49
12	Original	−0.14	0.09	0.09
	2-Year Censored	−0.05	0.11	0.57
	Imputed 2-Yr Censored	−0.08	0.09	0.32

Table 5

Simulation Study Results

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Parameter values				
β_0	5.5	4	4	4
$\log(\rho)$	-0.2	-0.1	-0.2	-0.1
$\log(\lambda)$	9	8	7	8
n	300	300	300	600
Size of log-rank test				
Original	0.043	0.046	0.043	0.047
2-Year Censored	0.050	0.036	0.039	0.037
Imputed	0.039	0.035	0.036	0.033
Average $\hat{\omega}_1$				
Original	-0.009	-0.002	-0.002	-0.001
2-Year Censored	-0.009	-0.001	-0.0003	-0.002
Imputed	-0.009	-0.005	-0.0009	-0.008
Empirical SD $\hat{\omega}_1$				
Original	0.161	0.138	0.120	0.096
2-Year Censored	0.186	0.148	0.130	0.104
Imputed	0.173	0.140	0.126	0.098
Average SE $\hat{\omega}_1$				
Original	0.164	0.137	0.123	0.097
2-Year Censored	0.188	0.152	0.134	0.107
Imputed	0.183	0.148	0.133	0.104

Other parameter values used for generating data from equations 1, 2 and 3: $\alpha_0 = -0.4$, $\alpha_1 = 0$, $\beta_1 = 0$, $\sigma = 1$, $\theta = 3$, $\gamma_1 = 0$