

Multi-state models for colon cancer recurrence and death with a cured fraction

A. S. C. Conlon,^{a*†} J. M. G. Taylor^a and D. J. Sargent^b

In cancer clinical trials, patients often experience a recurrence of disease prior to the outcome of interest, overall survival. Additionally, for many cancers, there is a cured fraction of the population who will never experience a recurrence. There is often interest in how different covariates affect the probability of being cured of disease and the time to recurrence, time to death, and time to death after recurrence. We propose a multi-state Markov model with an incorporated cured fraction to jointly model recurrence and death in colon cancer. A Bayesian estimation strategy is used to obtain parameter estimates. The model can be used to assess how individual covariates affect the probability of being cured and each of the transition rates. Checks for the adequacy of the model fit and for the functional forms of covariates are explored. The methods are applied to data from 12 randomized trials in colon cancer, where we show common effects of specific covariates across the trials. Copyright © 2013 John Wiley & Sons, Ltd.

Keywords: colon cancer; cure model; Cox–Snell residuals; deviance residuals; multi-state model

1. Introduction

In longitudinal medical studies with a time-to-event final outcome, patients may experience multiple disease progression events prior to the event of interest. Examples include CD4 lymphocyte counts in the progression to HIV infection [1] and cancer progression prior to death in survival studies [2]. The data we examine come from 12 phase III randomized trials in colon cancer where there is interest in building a joint model for the two event times of interest, time to recurrence, and time to death and investigating how treatment and other covariates are associated with the event times. A common way to jointly model these disease progression events is through the use of multi-state models [3–5], which describe the progressions and transitions over time to the various disease states. Common forms of multi-state models include the ‘progressive three-state model’, in which subjects enter some intermediate disease state prior to entering the absorbing state, and the ‘illness-death model’, where healthy subjects may enter a diseased state prior to death, or die without disease. In these models, transition intensities, which can include covariates, provide the hazards for moving between states. Each disease state that a subject occupies is either a transient state that can be left, such as cancer recurrence, or an absorbing state, such as death, that can never be left once it is entered.

Different model assumptions can be made about the dependence of the transition intensities and time. One approach is to take $t = 0$ as the start of the study and then all subsequent times t refer to the time since the beginning of the study. Klein *et al.* [6] made this assumption in their analysis of relapse and death in bone marrow transplant patients. A second approach is to set the clock back to 0 upon entry into a new state. This approach assumes that the hazard for entry into each state depends on the entry time into that state. This type of model, termed a semi-Markov model, has been explored by Dabrowska *et al.* [7] and Lagakos *et al.* [8]. Additionally, in the semi-Markov model, the hazard for entry into a new state could depend on the time at which the current state was entered [9]. In our data analysis, we

^aDepartment of Biostatistics, University of Michigan, Ann Arbor, MI, U.S.A.

^bDivision of Biostatistics, Mayo Clinic, University of Minnesota School of Public Health, Rochester, MN, U.S.A.

*Correspondence to: Anna S. C. Conlon, Department of Biostatistics, University of Michigan, Ann Arbor, MI, U.S.A.

†E-mail: achern@umich.edu

use a semi-Markov model with recurrence time as a covariate in the hazard model for the transition from recurrence to death. The hazard for moving between states can be modeled either parametrically or semi-parametrically. Putter *et al.* [2] explored the use of the semi-parametric Cox model in their analysis of recurrence and survival in breast cancer. Foucher *et al.* [1] used a generalized Weibull model for the hazard of transitioning between states. Here, we use a proportional hazards model with a parametric Weibull baseline hazard for each of the transition rates. There is interest in using these semi-Markov multi-state models to jointly model disease progression events as they can be used to assess how individual covariates affect each of the progression rates, and to estimate overall survival (OS), given the disease history.

We propose a semi-Markov model with an incorporated latent cured state to model colon cancer recurrence and survival. This model structure is motivated by the disease process of colon cancer. Cure models have been used to model many different types of cancer where there is known to be a significant proportion of patients whose tumors are completely eliminated by the treatment and so will never experience a clinical recurrence. These patients are considered to be cured of the disease. We use the mixture model formulation of the cure model, introduced by Berkson and Gage [10]. This model assumes that a proportion of the population, p , will never experience the event of interest and are therefore cured. The mixture cure model has been widely discussed in the literature. Yamaguchi [11] explored the use of a cure model with a logistic mixture probability model and an accelerated failure time model with a generalized gamma distribution. Taylor [12] used a logistic model for the cure probability and a completely unspecified failure time process. Sy and Taylor [13] and Peng and Dear [14] explored the estimation for a semi-parametric Cox proportional hazards model for the failure time process.

One issue that arises with the use of the cure model is identifiability due to censoring before the end of the follow-up period [15]. Therefore, it can be difficult to distinguish models with a large population of uncured individuals and long tails of the failure time process from those with small populations of uncured individuals and short tails of the failure time process. In general, in order to justify the use of the cure model, there must be sufficient follow-up and a large number of censored observations after the last event. Problems with identifiability are likely to arise if the Kaplan–Meier survival plot of all data does not show a clear level plateau. In the models we propose, the joint modeling of survival time and recurrence time may aid in the identifiability as subjects with survival times greater than the last observed recurrence time are likely to be cured of disease. Additionally, the appropriateness of the cure fraction in the multi-state model can be assessed through a goodness of fit comparison with a model that does not model the cured fraction.

The multi-state model and cure model have each been considered with both nonparametric and parametric assumptions. Here, our proposed model combines aspects of both of these models providing insight into the role of covariates on both the curing of the disease and the disease process, as well as the association of the two endpoints of interest, recurrence and death. We apply our model separately to each of the 12 colon cancer trials in order to explore which covariates have similar effects on certain disease aspects across trials performed in different settings. A Bayesian estimation strategy is used to estimate the parameters of the Weibull model, the covariate effects for each of the transition times, and the covariate effects for the logistic model for each trial. As the cured state is only partially observed, we place informative priors on some model parameters to aid in estimation where there is a scientific rationale for these parameters to be close to zero. The adequacy of the model fit is assessed through the use of Cox–Snell residual plots, and deviance residual plots are used to determine the proper functional form of covariates. Although these would be natural to consider for multi-state models [16], we are not aware of any literature on using them in cure models. The remainder of the paper is organized as follows: Section 2 describes the data, and Section 3 describes the proposed model. In Section 4, the estimation procedure is outlined and results of the application of the model to the data from 12 colon cancer clinical trials are provided. Model checking procedures are described in Section 5. Section 6 provides simulation results, and Section 7 concludes with a discussion.

2. Data description

The data we consider consist of a total of 13,983 subjects from 12 randomized phase III adjuvant trials of locally advanced colon cancer. Ten of the trials are included in the Sargent *et al.* [17] publication, with two additional new trials. These 12 trials were previously analyzed by Conlon *et al.* [18], where a separate cure model and Weibull model were used to model time to recurrence and death, respectively. Trial enrollment spanned from 1977 to 1999. One trial (7) included 210 patients with stage 1 cancer; these

Table I. Data summary.

Study	N	Recurrences	Recurrence without death	Death without recur	Total deaths	Longest follow-up (years)	% with stage 3 cancer	% in treatment group	Age (mean), years
1	247	116	14	13	115	9.9	65.6%	49.0%	60.3
2	408	139	11	44	172	9.1	81.6%	62.5%	61.1
3	926	377	31	76	422	11.4	66.1%	49.4%	60.2
4	914	380	36	106	450	9.9	82.5%	75.2%	62.7
5	878	297	33	74	338	12.6	73.8%	49.8%	61.2
6	724	275	10	132	397	13.2	56.8%	48.2%	59.8
7	683	206	32	129	303	12.9	43.4%	50.1%	63.3
8	1040	356	36	67	387	9.7	72.1%	49.8%	56.1
9	2077	605	57	176	724	9.4	58.7%	66.7%	57.0
10	2128	574	66	192	700	10.3	55.9%	49.8%	58.0
11	1549	394	71	115	438	8	53.5%	50.3%	60.5
12	2409	627	189	106	544	6	71.1%	49.8%	57.9
Total	13983	4346	586	1230	4990	13.2	63.8%	54.3%	59.1

subjects were excluded from this analysis. Because of 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 because of a large number of patients administratively censored at this time. The median follow-up time for subjects alive at their last follow-up was 8.2 years. In each trial, subjects were followed, with a specific protocol, with cancer recurrences and deaths recorded as they occurred. We therefore have two censored event times of interest, recurrence and death. The censoring times for these two events are not necessarily the same, as ascertaining a recurrence time requires active follow-up usually involving a scan, whereas obtaining a dead or alive status could be obtained at a later time by other means, such as through the National Death Index, with no further information on recurrence status obtained. The average proportion of subjects censored for recurrence prior to their last follow-up was 9.5% across all trials, with a maximum of 16.9% in trial 2 and a minimum of 0% in trial 3. Of the 4346 observed recurrences, 3448 (79.3%) occurred within 3 years, 4075 (93.8%) occurred within 5 years, and 4281 (98.5%) occurred within 7 years. This type of event time data where very few events happen after a fixed window of time is a characteristic of a cured group and for which a cure model is appropriate. Kaplan–Meier plots of time to recurrence for the 12 trials, provided in Appendix A in Supporting information, show a clear leveling off providing a strong empirical rationale for the use of a cure model. Baseline covariates of interest include age, cancer stage, and treatment arm. Each trial compared a different pair of treatments, with one defined as the control arm and the other as the experimental arm. 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. The primary goal of all 12 trials was to compare OS between the pairs of treatments. On the basis of a simple log-rank test, three of the trials (3, 8, and 9) showed a statistically significant treatment effect. Table I provides a summary of stage, age, and treatment distributions in each trial, as well as the number of recorded recurrences and deaths and longest follow-up time for each trial. As each of these trials compared different pairs of treatments, we fit the model to each trial separately and then assess which of the covariate effects on the various diseases processes are similar across the 12 trials.

3. Multi-state model

We model time to recurrence and time to death using a multi-state model for the semi-Markov process. We also incorporate a latent cured fraction into the model for subjects who will never experience a recurrence. The cure model is applied to the recurrence event and assumes that there is a zero probability that some subjects will recur. In this setting, the treatment may eliminate the cancer, resulting in a cured group of patients. Curing of the cancer happens at the time of treatment but is not immediately observable. If the cancer is not eliminated, the patient is not cured of disease and will experience a recurrence when the tumor has regrown to a detectable size. The observed data provide information about whether the

patient was cured by the treatment. Patients with observed recurrences are known to be in the uncured group. Patients who do not have an observed recurrence may or may not be cured. For patients censored for recurrence, the model assumes that a proportion of these subjects would have never experienced a recurrence even if they had been followed longer and are therefore cured. Additionally, some subjects who were censored for recurrence could have experienced a recurrence after their censoring time had there been longer follow-up and were therefore in the uncured group with an unobserved recurrence time [19]. Deaths can occur either without a prior recurrence or following a recurrence. The deaths that occur without a prior recurrence are known not to be directly due to the regrowth of the cancer, whereas deaths following a recurrence may be due to the cancer or other causes. We do not consider cause of death in our models. We use the multi-state model to model four transition intensities, which include the transition from the uncured group to death, the transition from the cured group to death, the transition from the uncured group to recurrence, and the transition from recurrence to death.

3.1. Notation and model specifications

Let C_{ir} and T_{ir} be the censoring and event times for recurrence and let C_{id} and T_{id} be the censoring and event times for death for the i th subject, $i = 1, \dots, n$. Then $Y_{ir} = \min(C_{ir}, T_{ir})$ and the event indicator for recurrence, $\delta_{ir} = I(T_{ir} \leq C_{ir})$, and $Y_{id} = \min(C_{id}, T_{id})$ and the event indicator for death, $\delta_{id} = I(T_{id} \leq C_{id})$, are observed. Let Z_i , S_i , and A_i represent the baseline values of treatment group, cancer stage, and age for each subject, respectively. Each of these covariates is approximately centered prior to estimation so that $Z_i = -0.5$ or 0.5 for control or treatment, respectively, $S_i = -0.5$ or 0.5 for stage 2 or stage 3 cancer, respectively, and A_i is age, centered at the mean age for a given study in units of 10 years. Let p_i be the probability that the i th subject is cured of disease. We define state 1 to be alive and cured of disease, state 2 to be alive and not cured, state 3 to be alive with recurrence, and state 4 to be dead, as illustrated in Figure 1. States 1, 2, and 3 are transient states, whereas state 4 is an absorbing state. We model four transition times, $1 \rightarrow 4$, $2 \rightarrow 3$, $2 \rightarrow 4$, and $3 \rightarrow 4$. The true state progressions for many subjects are not fully observed. Specifically, those who were censored for recurrence and alive were either in state 1 or state 2 at the end of their follow-up, and subjects censored for recurrence and dead either made a $1 \rightarrow 4$ or a $2 \rightarrow 4$ transition at their time of death. Subjects with an observed recurrence transitioned from $2 \rightarrow 3$ at their recurrence time and then remained in state 3 if they were censored for death or made a $3 \rightarrow 4$ transition at their death time.

In the standard setting for cure models, there is one event time of interest, and the mixture model formulation assumes that a proportion of the population, p , will never experience the event of interest, in this case recurrence, and is therefore cured. For the uncured population, the cure model provides information on the estimated time to event from the survival distribution. The marginal survival function for recurrence, $S(t)$, for the entire population is given by $S(t) = p + (1 - p)S_0(t)$, where $S_0(t)$ is the conditional survival function for recurrence for the uncured group. It is common to use a logistic model, which includes time independent covariates for the incidence model. Common choices for $S_0(t)$ are the exponential and Weibull distributions. Nonparametric choices for $S_0(t)$ have also been explored. In our more complicated situation with two event times, the structure of the model is more involved, but we retain similar elements, that of a cured fraction described by a logistic model, and distributions of event times given cured status described by proportional hazards models with Weibull baseline hazard functions.

The proposed model can be used to assess how different covariates affect both the probability of being cured and the hazard of transitioning to recurrence or death. Other quantities of interest can also

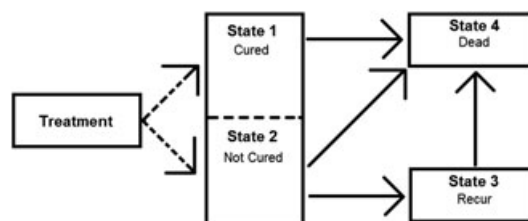


Figure 1. Multi-state cure model structure: dashed lines represent effect of treatment and solid lines represent transitions between states.

be derived from the model, such as 5-year survival within each of the treatment arms. Both the models for the time of entry into each state and p can depend on covariates. The probability of being cured, p_i , is modeled using a logistic link function given by

$$p_i = \frac{\exp(\gamma X_i)}{1 + \exp(\gamma X_i)}.$$

The multi-state process is characterized through transition intensities defined as

$$\lambda_{kj}(t) = \lim_{\Delta t \rightarrow 0} p_{kj}(t, t + \Delta t) / \Delta t$$

where $p_{kj}(s, t) = P(M(t) = j | M(s) = k, H_{s-})$ is the probability of being in state j at time t , given that the process was in state k at time s and the history of the process H_{s-} , for states $M(t)$ and $M(s)$ occupied at times t and s , respectively, and $s \leq t$. $\lambda_{kj}(t)$ is then the instantaneous hazard of entering state j , given that the previous state occupied was state k . From this hazard, we can define the survival distributions for each transient state. The survival distributions for remaining in states 1, 2, and 3 are as follows:

$$\begin{aligned} S_1(t) &= \exp\left(-\int_0^t \lambda_{14}(u) du\right) \\ S_2(t) &= \exp\left(-\int_0^t \lambda_{23}(u) du - \int_0^t \lambda_{24}(u) du\right) \\ S_3(t|t_r) &= \exp\left(-\int_0^{t-t_r} \lambda_{34}(u) du\right) \end{aligned}$$

where t_r is the entry time into state 3.

We use a proportional hazards model with a Weibull baseline hazard function to model the distribution of waiting times. Specifically, the hazard for transition kj for subject i is given by

$$\lambda_{kj}(t_i; X_i) = \left(\frac{\rho_{kj}}{\lambda_{kj}}\right) \left(\frac{t_i}{\lambda_{kj}}\right)^{\rho_{kj}-1} \exp(\beta_{kj} X_i)$$

For transitions $1 \rightarrow 4$ and $2 \rightarrow 4$, t_i is a death time. For transition $2 \rightarrow 3$, t_i is a recurrence time and for transition $3 \rightarrow 4$, t_i is the gap time between entry into the recurred state and death. X_i represents a vector of subject specific covariates. For transitions $1 \rightarrow 4$, $2 \rightarrow 3$, and $2 \rightarrow 4$, and for the probability of cure, we include the centered covariates age, treatment group, and stage. For transition $3 \rightarrow 4$, we include these variables as covariates plus time to recurrence, centered at the mean time to recurrence for those who recur in a given study in units of years. These covariates can be expected to be associated with each of the model components in differing ways, as each part of the model describes a different aspect of the disease process. The probability of being cured provides information about the tumor and the cell killing effect of the treatment. Transitions $1 \rightarrow 4$ and $2 \rightarrow 4$ give information about the person as opposed to the cancer, and transition $2 \rightarrow 3$ provides information about the tumor regrowth. Finally, transition $3 \rightarrow 4$ provides information about both the person and the regrowth of the tumor. Six distinct types of people contribute to the likelihood: those who recur and are either dead or alive at a later time, those censored for recurrence at Y_{id} and either dead or alive at Y_{id} , and those censored for recurrence at Y_{ir} prior to death or censoring at a later time Y_{id} . These likelihood contributions can be described by the following three equations:

Recur, dead, or alive:

$$(1 - p_i) \lambda_{23}(Y_{ir}) S_2(Y_{ir}) \lambda_{34}(Y_{id} - Y_{ir})^{\delta_{id}} S_3(Y_{id} | Y_{ir})$$

Censored for recurrence at \mathbf{Y}_{id} , dead or alive:

$$(1 - p_i) \lambda_{24}(Y_{id})^{\delta_{id}} S_2(Y_{id}) + p_i \lambda_{14}(Y_{id})^{\delta_{id}} S_1(Y_{id})$$

Censored for recurrence at \mathbf{Y}_{ir} prior to \mathbf{Y}_{id} , dead or alive:

$$\begin{aligned} &(1 - p_i) \lambda_{24}(Y_{id})^{\delta_{id}} S_2(Y_{id}) + \\ &(1 - p_i) \int_{Y_{ir}}^{Y_{id}} \lambda_{23}(u) S_2(u) \lambda_{34}(Y_{id} - u)^{\delta_{id}} S_3(Y_{id} | u) du + \\ &p_i \lambda_{14}(Y_{id})^{\delta_{id}} S_1(Y_{id}) \end{aligned}$$

In the 12 trials we examine, a small proportion of subjects (0.7% across all trials) had a recurrence at the same date as their time of death. As these subjects likely truly recurred prior to this date but for administrative or other reasons, these recurrences were not detected prior to death, their recurrence times were treated as interval censored and their contribution to the likelihood was

$$\begin{aligned} & \text{Recur at } Y_{id}, \text{ dead :} \\ & (1 - p_i) \int_0^{Y_{id}} \lambda_{23}(u) S_2(u) \lambda_{34}(Y_{id} - u)^{\delta_{id}} S_3(Y_{id} | u) du \end{aligned}$$

4. Estimation

We use a Bayesian MCMC technique to estimate the parameters of the multi-state cure model. The parameters for each of the 12 trials are estimated separately, with no mixing of patients across the different trials. There are a total of 25 parameters to estimate for each of the trials, which include a shape (ρ) and scale (λ) parameter from the Weibull model for each of the hazard rates, covariate effects for each of the hazard models, and covariate effects in the logistic model for the probability of cure. We place informative Normal(0, 0.25²) priors on the treatment and stage coefficients in transition 1 \rightarrow 4 as whereas treatment group and cancer stage may affect the likelihood of cure, they are unlikely to affect the survival of patients conditional on being cured of disease. We place Normal(0, 2²) priors on the log(λ)'s and gamma priors with mean 1 and standard deviation 0.6 on the ρ 's. Normal(0, 1) priors are placed on all of the remaining covariate coefficients in the hazard models and in the logistic model. To aid in estimation, at each iteration of the chain, the latent variable representing cured status is simulated, specifically subjects without recurrence are placed in either the cured or uncured group by drawing a Bernoulli random variable with probability of cure

$$c_i = \frac{p_i \lambda_{14}(Y_{id})^{\delta_{id}} S_1(Y_{id})}{p_i \lambda_{14}(Y_{id})^{\delta_{id}} S_1(Y_{id}) + (1 - p_i) \lambda_{24}(Y_{id})^{\delta_{id}} S_2(Y_{id})}$$

for those with censored for recurrence at Y_{id} and

$$c_i = \frac{p_i \lambda_{14}(Y_{id})^{\delta_{id}} S_1(Y_{id})}{p_i \lambda_{14}(Y_{id})^{\delta_{id}} S_1(Y_{id}) + (1 - p_i) \lambda_{24}(Y_{id})^{\delta_{id}} S_2(Y_{id}) + (1 - p_i) \int_{Y_{ir}}^{Y_{id}} \lambda_{23}(u) S_2(u) \lambda_{34}(Y_{id} - u)^{\delta_{id}} S_3(Y_{id} | u) du}$$

or those censored for recurrence at Y_{ir} prior to Y_{id} . As some integrals in the likelihood do not have closed form solutions, numeric integration was used. Specifically, the following integrals were calculated by adaptive quadrature using the 'integrate' function in R:

$$\begin{aligned} & \int_0^{Y_{ir}} \lambda_{23}(u) S_2(u) \lambda_{34}(Y_{id} - u)^{\delta_{id}} S_3(Y_{id} | u) du \\ & \int_{Y_{ir}}^{Y_{id}} \lambda_{23}(u) S_2(u) \lambda_{34}(Y_{id} - u)^{\delta_{id}} S_3(Y_{id} | u) du \end{aligned}$$

All covariates are centered prior to estimation, as described in Section 3. The posterior distributions for all of the parameters can be obtained from the product of the observed data likelihood and prior distributions. The Metropolis Hastings algorithm is used for parameter estimation. Appendix B in the Supporting information provides the full data likelihood and details of the algorithm.

For each parameter, we obtain 5000 draws from its posterior distribution. Table II provides the posterior mean and standard deviation estimates for all model parameters for each of the 12 trials. Covariate effect estimates that are greater than two times the posterior standard deviation are shown in bold. We estimate the parameters of each of the 12 trials separately and compare the estimates of each parameter across trials in order to examine which aspects of the disease process have common covariate effects across all of the trials and which have varying effects. The results show very consistent effects of most covariates on the probability of cure and on each of the transition times across all 12 studies. Stage is seen to have a consistent and strong effect on the probability of being cured and on the time to recurrence and a modest effect on the time to death after recurrence, with higher stage people more likely to recur, recurring earlier and dying sooner after recurrence. The effects of stage on time to death for those who do not recur are smaller in the cured group, with mixed effects across trials. Age has a strong effect on

Table II. Model parameter estimates and posterior standard deviations for all trials.

	Trial											
	1	2	3	4	5	6	7	8	9	10	11	12
Cure												
Logistic model												
Intercept	-0.07 (0.23)	0.79 (0.18)	0.48 (0.09)	0.40 (0.11)	0.73 (0.09)	0.38 (0.09)	0.56 (0.09)	0.83 (0.09)	0.85 (0.06)	0.99 (0.05)	1.01 (0.07)	1.17 (0.08)
$\gamma_{Treatment}$	0.27 (0.32)	0.16 (0.24)	0.55 (0.14)	-0.05 (0.17)	-0.07 (0.15)	0.04 (0.16)	0.14 (0.17)	0.38 (0.14)	0.17 (0.11)	0.11 (0.10)	0.02 (0.12)	0.20 (0.10)
γ_{Stage}	-0.56 (0.44)	-0.96 (0.31)	-1.04 (0.16)	-0.78 (0.20)	-0.86 (0.19)	-1.03 (0.16)	-1.13 (0.17)	-1.11 (0.17)	-0.98 (0.11)	-1.25 (0.11)	-0.99 (0.12)	-1.15 (0.15)
γ_{Age} (10 years)	0.11 (0.15)	0.14 (0.13)	0.006 (0.06)	0.005 (0.07)	0.02 (0.07)	-0.19 (0.08)	-0.09 (0.09)	-0.01 (0.07)	0.007 (0.05)	0.03 (0.05)	-0.07 (0.06)	-0.04 (0.04)
Time to recur												
2 \rightarrow 3 transition												
$\log(\lambda_{23})$	1.11 (0.19)	0.82 (0.22)	0.79 (0.07)	1.09 (0.10)	0.99 (0.08)	0.91 (0.06)	0.97 (0.07)	1.04 (0.07)	0.83 (0.04)	0.93 (0.04)	1.02 (0.05)	0.94 (0.08)
ρ_{23}	1.15 (0.10)	1.13 (0.10)	1.12 (0.05)	1.22 (0.05)	1.39 (0.07)	1.25 (0.06)	1.30 (0.07)	1.50 (0.06)	1.50 (0.05)	1.49 (0.05)	1.49 (0.06)	1.46 (0.05)
$\beta_{Treatment_{23}}$	-0.68 (0.26)	-0.28 (0.26)	-0.11 (0.12)	0.08 (0.15)	0.03 (0.13)	-0.18 (0.13)	-0.21 (0.16)	-0.31 (0.11)	-0.11 (0.09)	0.08 (0.09)	0.11 (0.11)	0.02 (0.10)
$\beta_{Stage_{23}}$	0.79 (0.38)	-0.12 (0.39)	0.39 (0.16)	0.44 (0.24)	0.37 (0.23)	0.29 (0.16)	0.53 (0.18)	0.72 (0.19)	0.24 (0.10)	0.13 (0.11)	0.47 (0.14)	0.50 (0.22)
$\beta_{Age_{23}}$ (10 years)	-0.07 (0.12)	0.07 (0.15)	-0.10 (0.05)	-0.008 (0.06)	-0.13 (0.07)	-0.17 (0.07)	-0.01 (0.07)	-0.05 (0.05)	-0.13 (0.04)	-0.18 (0.04)	-0.04 (0.06)	-0.05 (0.04)
Cure to death												
1 \rightarrow 4 transition												
$\log(\lambda_{14})$	4.26 (0.77)	3.54 (0.34)	3.28 (0.21)	3.00 (0.16)	3.47 (0.21)	3.08 (0.13)	3.16 (0.12)	4.18 (0.52)	3.56 (0.22)	3.42 (0.13)	3.85 (0.37)	3.43 (0.33)
ρ_{14}	1.59 (0.63)	1.58 (0.41)	2.36 (0.47)	2.72 (0.48)	2.46 (0.38)	1.95 (0.33)	1.86 (0.23)	1.45 (0.43)	1.89 (0.29)	2.25 (0.22)	1.80 (0.37)	2.35 (0.52)
$\beta_{Treatment_{14}}$	0.05 (0.24)	-0.18 (0.22)	-0.05 (0.20)	0.05 (0.19)	-0.01 (0.19)	0.01 (0.17)	-0.22 (0.17)	-0.20 (0.20)	-0.21 (0.16)	0.10 (0.14)	0.08 (0.18)	-0.19 (0.20)
$\beta_{Stage_{14}}$	0.04 (0.24)	0.05 (0.22)	0.08 (0.19)	0.04 (0.19)	0.06 (0.19)	-0.02 (0.16)	0.16 (0.16)	-0.01 (0.21)	0.20 (0.16)	-0.13 (0.15)	0.29 (0.18)	-0.10 (0.21)
$\beta_{Age_{14}}$ (10 years)	1.23 (0.47)	0.53 (0.31)	0.75 (0.16)	1.05 (0.19)	1.45 (0.23)	0.65 (0.12)	0.79 (0.14)	0.49 (0.20)	0.75 (0.14)	0.90 (0.13)	1.00 (0.17)	0.78 (0.17)
Noncure to death												
2 \rightarrow 4 transition												
$\log(\lambda_{24})$	4.05 (0.90)	3.58 (0.87)	3.96 (0.71)	4.62 (0.78)	5.61 (0.96)	3.70 (0.66)	5.64 (0.91)	4.42 (0.90)	4.42 (0.77)	4.49 (0.59)	4.70 (0.80)	5.16 (0.80)
ρ_{24}	1.73 (0.64)	1.43 (0.54)	1.06 (0.20)	0.74 (0.12)	0.72 (0.13)	1.07 (0.20)	0.52 (0.08)	1.45 (0.64)	0.90 (0.14)	0.88 (0.12)	0.87 (0.27)	0.67 (0.10)
$\beta_{Treatment_{24}}$	0.10 (0.88)	-0.69 (0.83)	0.71 (0.54)	-0.26 (0.44)	0.04 (0.56)	-0.57 (0.51)	0.10 (0.40)	0.22 (0.82)	0.31 (0.48)	0.05 (0.35)	-0.43 (0.57)	-0.02 (0.34)
$\beta_{Stage_{24}}$	0.11 (0.90)	-0.19 (0.85)	-0.33 (0.53)	-0.48 (0.47)	-0.98 (0.56)	0.33 (0.53)	-0.56 (0.38)	0.07 (0.86)	-0.34 (0.44)	-0.48 (0.38)	-0.65 (0.55)	-0.68 (0.40)
$\beta_{Age_{24}}$ (10 years)	0.64 (0.78)	-0.13 (0.61)	0.65 (0.30)	0.82 (0.24)	0.92 (0.38)	0.66 (0.31)	0.87 (0.30)	0.50 (0.61)	0.62 (0.36)	0.99 (0.24)	0.64 (0.37)	0.74 (0.18)
Recur to death												
3 \rightarrow 4 transition												
$\log(\lambda_{34})$	0.85 (0.17)	0.40 (0.27)	1.19 (0.19)	1.45 (0.16)	1.51 (0.23)	0.72 (0.19)	1.26 (0.26)	1.13 (0.20)	1.17 (0.18)	1.26 (0.16)	1.64 (0.21)	1.60 (0.18)
ρ_{34}	1.15 (0.09)	0.99 (0.07)	0.92 (0.04)	1.01 (0.04)	0.81 (0.04)	0.89 (0.04)	0.84 (0.05)	0.91 (0.04)	0.90 (0.03)	0.88 (0.03)	0.87 (0.04)	0.96 (0.04)
$\beta_{Treatment_{34}}$	0.16 (0.21)	0.14 (0.18)	0.27 (0.11)	0.16 (0.13)	0.18 (0.13)	0.25 (0.13)	-0.005 (0.15)	-0.06 (0.11)	0.03 (0.09)	-0.07 (0.09)	-0.15 (0.11)	0.18 (0.09)
$\beta_{Stage_{34}}$	1.05 (0.26)	-0.01 (0.29)	0.39 (0.14)	0.59 (0.18)	0.50 (0.18)	0.13 (0.14)	0.34 (0.16)	0.32 (0.16)	0.33 (0.10)	0.41 (0.11)	0.38 (0.13)	0.16 (0.15)
$\beta_{Age_{34}}$ (10 years)	0.06 (0.09)	0.16 (0.10)	0.07 (0.04)	0.12 (0.05)	0.12 (0.06)	0.03 (0.07)	0.30 (0.08)	0.01 (0.05)	0.12 (0.04)	0.14 (0.04)	0.13 (0.05)	0.20 (0.04)
β_{T_r} (Years)	-0.07 (0.07)	0.04 (0.06)	-0.11 (0.04)	-0.23 (0.04)	-0.17 (0.05)	-0.10 (0.04)	-0.07 (0.05)	-0.11 (0.05)	-0.12 (0.04)	-0.10 (0.03)	-0.17 (0.04)	-0.22 (0.05)

Regression parameter estimates greater than two times the posterior standard deviation are shown in bold.

time to death for both those who are cured and those who are not cured but die before recurrence, with older people dying sooner. There is a mild effect of age on time to recurrence and on time to death after recurrence, with younger people recurring earlier and older people dying sooner after recurrence. There is also a consistent effect of time to recurrence on time to death, with those recurring quickly dying sooner. The shape parameter, ρ , and scale parameter, λ , of the Weibull model are also consistent across trials within each transition. The shape parameter of the Weibull distribution determines the shape of the density curve. When $\rho < 1$, the failure rate decreases with time. Weibull distributions with ρ close to or equal to 1 have a fairly constant failure rate and distributions with $\rho > 1$ have a failure rate that increases with time. For all of the trials, ρ is greater than 1 in the transition from no cure to recur, indicating a short tail in the distributions of recurrence times. This is a characteristic of an event with a cured group, as events become unlikely after a certain amount of time. The intercept in the logistic model for the probability of cure indicates that the trials where the control arm was surgery alone tend to have lower cure rates on the control arm than the trials where the control arm included chemotherapy.

As the 12 trials compared different combinations of treatments, the treatment effects vary for one trial to the next. Based on log-rank p -values for OS, trials 3, 8, and 9 had a significant effect of treatment on OS, with log-rank p -values equal to 0.004, 0.0003, and 0.04, respectively. Trials 2, 7, and 12 had near significant treatment effects on OS with log-rank p -values of 0.09, 0.07, and 0.09, respectively. Additionally, trials 1, 3, 8, 9, and 12 had a significant treatment effect on time to recurrence, with log-rank p -values of 0.01, < 0.0001 , 0.001, 0.04, and 0.03, respectively. The treatment effect estimates from the Markov model show some consistencies with these results, with effects of treatment primarily seen on the probability of cure and time to recurrence. There is a significant effect of treatment on the probability of cure for trials 3 and 8. Trials 8 and 1 both have a significant treatment effect on time to recurrence. All trials, except trial 3, show no effects of treatment on time to death after a recurrence or on time to death without recurrence for either the cured or uncured group. There is a small adverse effect of treatment on the time to death after a recurrence in trial 3. Appendix C in the Supporting information provides plots of each of the treatment effect estimates across the trials.

5. Model checking and model extensions

5.1. Checking goodness of fit of the model

Once parameter estimates for the model have been obtained, the model can be used to estimate 5-year OS, traditionally the final endpoint of interest in trials of locally advanced colon cancer. Similarly, 3-year disease free survival (DFS), which is the minimum of recurrence and death times and has been shown to be an alternative endpoint to 5-year OS in these types of trials [20], can also be estimated from the model. The point estimates for 5-year OS and 3-year DFS for each treatment arm can be compared with the Kaplan–Meier estimate to check the model fit and the standard error estimate can be compared with that of the Kaplan–Meier estimate to assess gains in efficiency through use of the multi-state model. For each subject, we can calculate their 5-year OS probability as

$$P(T_i > 5 \mid X_i, \theta) = p_i S_1(5) + (1 - p_i) S_2(5) + (1 - p_i) \int_0^5 S_2(u) \lambda_{23}(u) S_3(5 \mid u) du$$

and three year DFS probability as

$$P(DFS_i > 3 \mid X_i, \theta) = p_i S_1(3) + (1 - p_i) S_2(3)$$

where θ is the vector of parameter values, and the probability is averaged across the age and stage covariate values for each subject and across all of the parameter draws. Table III provides the Kaplan–Meier and multi-state model 5-year OS and 3-year DFS estimates and standard errors. For the 12 trials considered, both the 5-year OS and 3-year DFS estimates from the multi-state model are similar to the Kaplan–Meier estimates, indicating that the model is an appropriate fit to the data as measured by predictions of OS at 5 years and DFS at 3 years. There is also some efficiency gained in estimating these quantities, as seen by the smaller standard errors in the multi-state model estimates as compared with the Kaplan–Meier estimates. Also noticeable from the estimates are the general improvement in 5-year OS from the chronologically early trials (trials 1, 2, and 3) to the more recent trials (10, 11, and 12).

Table III. Five-year overall survival and 3-year disease free survival estimates: Kaplan–Meier and multi-state model.

	$P(T > 5 Z_i = 1, X_i, \theta)$		$P(DFS > 3 Z_i = 1, X_i, \theta)$		$P(T > 5 Z_i = 0, X_i, \theta)$		$P(DFS > 3 Z_i = 0, X_i, \theta)$	
	Kaplan–Meier estimates (SE)	Multi-state model estimates (SE)	Kaplan–Meier estimates (SE)	Multi-state model estimates (SE)	Kaplan–Meier estimates (SE)	Multi-state model estimates (SE)	Kaplan–Meier estimates (SE)	Multi-state model estimates (SE)
Trial 1	0.651 (0.044)	0.657 (0.037)	0.669 (0.043)	0.690 (0.036)	0.561 (0.044)	0.552 (0.036)	0.548 (0.044)	0.540 (0.038)
Trial 2	0.695 (0.029)	0.705 (0.025)	0.721 (0.028)	0.727 (0.025)	0.638 (0.039)	0.653 (0.034)	0.665 (0.038)	0.659 (0.035)
Trial 3	0.693 (0.022)	0.694 (0.019)	0.700 (0.021)	0.702 (0.020)	0.620 (0.022)	0.622 (0.020)	0.590 (0.023)	0.589 (0.021)
Trial 4	0.635 (0.018)	0.632 (0.017)	0.641 (0.018)	0.650 (0.017)	0.658 (0.032)	0.656 (0.029)	0.637 (0.032)	0.665 (0.031)
Trial 5	0.693 (0.022)	0.698 (0.019)	0.697 (0.022)	0.705 (0.020)	0.716 (0.022)	0.730 (0.019)	0.712 (0.022)	0.722 (0.020)
Trial 6	0.639 (0.026)	0.626 (0.022)	0.682 (0.025)	0.675 (0.022)	0.601 (0.025)	0.618 (0.021)	0.634 (0.025)	0.642 (0.021)
Trial 7	0.748 (0.024)	0.735 (0.020)	0.731 (0.024)	0.731 (0.021)	0.668 (0.026)	0.697 (0.021)	0.693 (0.025)	0.691 (0.023)
Trial 8	0.762 (0.019)	0.756 (0.016)	0.749 (0.019)	0.759 (0.017)	0.656 (0.021)	0.661 (0.019)	0.665 (0.021)	0.657 (0.019)
Trial 9	0.738 (0.012)	0.746 (0.011)	0.743 (0.012)	0.747 (0.012)	0.704 (0.017)	0.712 (0.016)	0.711 (0.017)	0.707 (0.016)
Trial 10	0.765 (0.013)	0.775 (0.011)	0.767 (0.013)	0.770 (0.012)	0.761 (0.013)	0.762 (0.012)	0.760 (0.013)	0.760 (0.012)
Trial 11	0.788 (0.015)	0.794 (0.013)	0.778 (0.015)	0.779 (0.013)	0.788 (0.015)	0.783 (0.013)	0.783 (0.015)	0.783 (0.014)
Trial 12	0.802 (0.012)	0.797 (0.011)	0.780 (0.012)	0.775 (0.011)	0.783 (0.012)	0.781 (0.011)	0.748 (0.013)	0.743 (0.012)

SE, standard error.

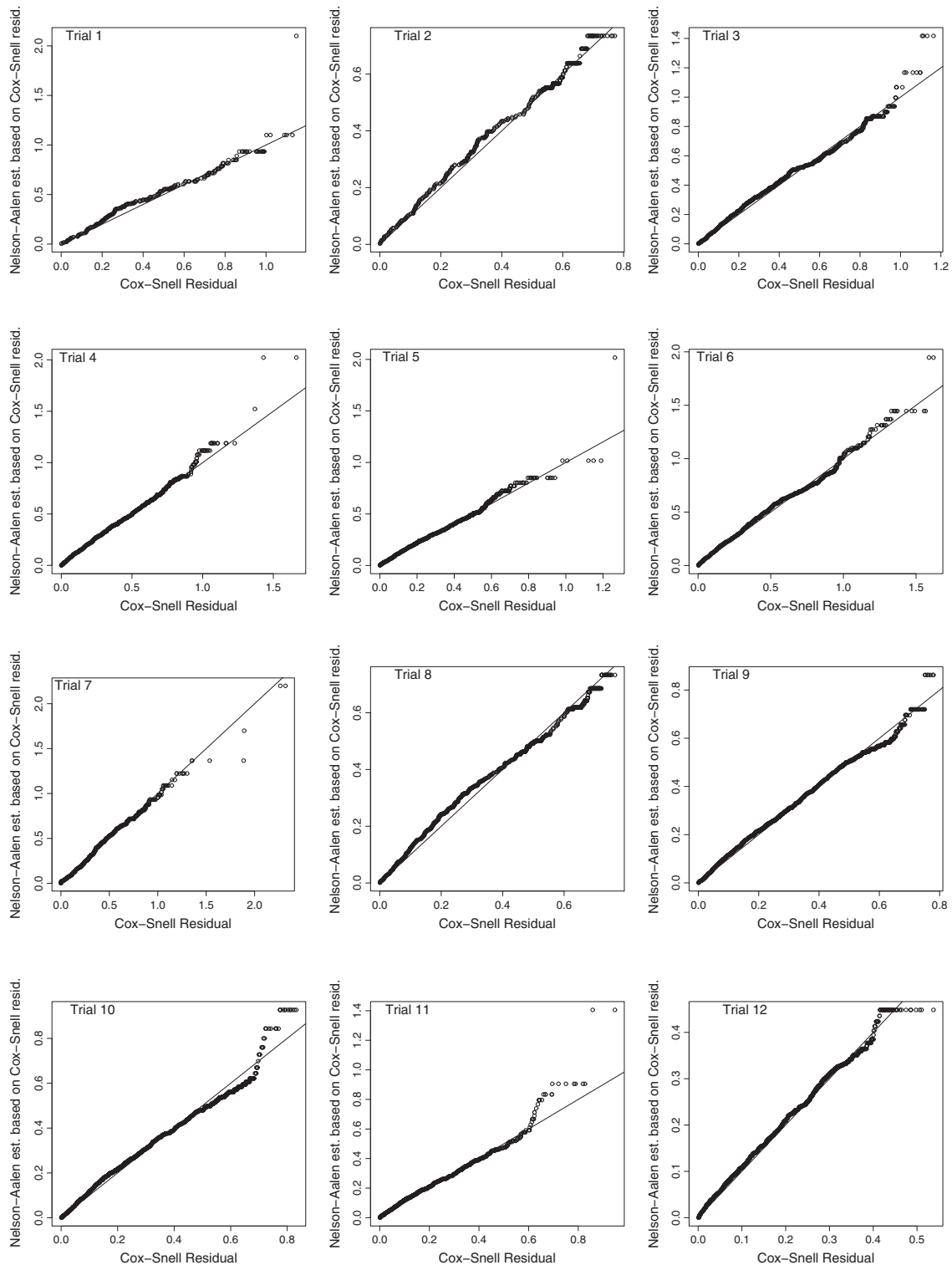


Figure 2. Cox–Snell residual plots for time to death. Results from 12 trials.

Additional model fit assessments can be made by examining Cox–Snell residual plots. For each subject, we calculate the Cox–Snell residual for time to death. Let $w_i = -\log \hat{S}(Y_{id})$, where

$$\hat{S}(Y_{id}) = p_i S_1(Y_{id}) + (1 - p_i) S_2(Y_{id}) + (1 - p_i) \int_0^{Y_{id}} S_2(u) \lambda_{23}(u) S_3(Y_{id} | u) du.$$

If the model is correct, the pairs (w_i, δ_{id}) should be like a sample from a censored $Exp(1)$ distribution. Therefore, a plot of w_i versus the Nelson–Aalen estimator for the pairs of (w_i, δ_{id}) should yield a straight line through the origin with slope 1. Figure 2 provides the Cox–Snell residual plots for OS for the 12 trials. For most of the trials, the proposed model appears to provide an adequate fit to the data. The aforementioned Cox–Snell residuals involve all five model components. Cox–Snell residuals can also be used to check the model fit for some selected aspects of the overall model, specifically for the time to recurrence and for the transition from recur to death. For the time to recurrence, the residual is $w_i = -\log \hat{S}(Y_{ir})$, with the pair (w_i, δ_{ir}) compared with the $Exp(1)$ distribution, where

$$\hat{S}(Y_{ir}) = p_i + (1 - p_i) \exp \left(- \left(\frac{Y_{ir}}{\lambda_{23}} \right)^{\rho_{23}} \exp(\beta_{trt_{23}} Z_i + \beta_{st_{23}} S_i + \beta_{age_{23}} A_i) \right)$$

and for the recur to death transition the residual is $w_i = -\log \hat{S}(Y_{id} - Y_{ir})$, with the pair (w_i, δ_{id}) compared with the $Exp(1)$ distribution, where

$$\hat{S}(Y_{id} - Y_{ir}) = \exp \left(- \left(\frac{Y_{id} - Y_{ir}}{\lambda_{34}} \right)^{\rho_{34}} \exp(\beta_{trt_{34}} Z_i + \beta_{st_{34}} S_i + \beta_{age_{34}} A_i + \beta_{Tr_{34}} Y_{ir}) \right)$$

where w_i is calculated only for those who recur. Appendix D in the supporting information provides the Cox–Snell residual plots for the time to recurrence transition and the recurrence to death transition. The residual plots for the individual transitions show that the model fits for the time to recurrence fairly well, but there may be some lack of fit in the transition from recurrence to death in some trials.

Deviance residuals are a standard way of assessing the functional form of covariates in survival analysis models. Here, we adapt them to the multi-state cure model. For each person, the Martingale residual for OS is defined as $r_i = \delta_{id} - \hat{\Lambda}(Y_{id})$, where

$$\begin{aligned} \hat{\Lambda}(Y_{id}) = & -\log \left(p_i S_1(Y_{id}) + (1 - p_i) S_2(Y_{id}) \right) \\ & + (1 - p_i) \int_0^{Y_{id}} S_2(u) \lambda_{23}(u) \exp \left(- \int_0^{Y_{id}-u} \lambda_{34}(v) dv \right) du \end{aligned}$$

The Martingale residuals can be viewed as the difference between the observed number of deaths for subject i between time 0 and Y_{id} and the expected number based on the model. As the Martingale residuals have an asymmetric range, the deviance residuals are often preferred. The deviance residuals are defined as $D_i = \text{sign}(r_i) \sqrt{-2[r_i + \delta_{id} \log(\delta_{id} - r_i)]}$. Plots of the deviance residuals against covariates should be symmetric about a horizontal line at 0.

In addition to checking the deviance residuals for OS, we can also use them to assess the functional forms of covariates for time to recurrence and for time to death after recurrence. In the first case, we have $r_i = \delta_{ir} - \hat{\Lambda}(Y_{ir})$, where

$$\begin{aligned} \hat{\Lambda}(Y_{ir}) = & -\log \left(p_i + (1 - p_i) \exp \left(- \left(\frac{Y_{ir}}{\lambda_{23}} \right)^{\rho_{23}} \exp(\beta_{trt_{23}} Z_i + \beta_{st_{23}} S_i + \beta_{age_{23}} A_i) \right) \right) \\ D_i = & \text{sign}(r_i) \sqrt{-2[r_i + \delta_{ir} \log(\delta_{ir} - r_i)]} \end{aligned}$$

and in the second case we have $r_i = \delta_{id} - \hat{\Lambda}(Y_{id})$, where

$$\begin{aligned} \hat{\Lambda}(Y_{id}) = & \left(\frac{Y_{id} - Y_{ir}}{\lambda_{34}} \right)^{\rho_{34}} \exp(\beta_{trt_{34}} Z_i + \beta_{st_{34}} S_i + \beta_{age_{34}} A_i + \beta_{Tr_{34}} Y_{ir}) \\ D_i = & \text{sign}(r_i) \sqrt{-2[r_i + \delta_{id} \log(\delta_{id} - r_i)]} \end{aligned}$$

where r_i is only calculated for those who recur. Results of these deviance residuals are shown in Appendix E in the Supporting information. Deviance residuals plots are in general hard to interpret. The plots suggest that a linear function of age is not inadequate for both the recurrence transition and the death after recurrence transition and for the overall model fit. The covariate for recurrence time in the death after recurrence transition also appears to be adequately modeled by a linear function. The graphs suggest that there is no obvious, consistent departure across the 12 trials.

Table IV. Model comparison by DIC values.

	Complete model	No cured fraction	Parameters in 1 → 4, 2 → 4 same
# of parameters	25	16	20
	DIC		
Trial 1	1093.7	1130.8	1099.4
Trial 2	1789.4	2106.3	1793.5
Trial 3	4206.5	4375.7	4212.0
Trial 4	4392.7	4507.4	4419.6
Trial 5	3572.5	3694.5	3591.7
Trial 6	3607.8	3740.2	3615.0
Trial 7	3176.4	3301.0	3209.4
Trial 8	4084.6	4303.0	4096.4
Trial 9	7677.2	8083.2	7684.4
Trial 10	7706.4	8061.3	7721.2
Trial 11	5225.7	5407.6	5238.6
Trial 12	7172.9	7349.0	7178.5

Results for multi-state cure model, model without a cured fraction, and model with parameters in transitions 1 → 4 and 2 → 4 constrained to be the same. The DIC value of the best fitting model for each trial is shown in bold.

5.2. Model adaptations

Although we have shown a good fit of the multi-state cure model with meaningful interpretation of the parameters, a natural question is could we have obtained an adequate fit with a simpler model. To compare the full multi-state model to simpler models, we use the deviance information criterion (DIC) statistic [21]. The *DIC* value for each study is calculated as $DIC = 2[D(\hat{\theta})] - D(\bar{\theta})$, where $D(\hat{\theta})$ is -2 times the log-likelihood calculated at the mean value of each parameter, and $E[D(\hat{\theta})]$ is the mean of -2 times the log-likelihood across all parameter draws.

A multi-state model is fit to these data without including a cured group. This results in a simpler model with 16 as opposed to 25 parameters, and may be a better choice if there is uncertainty about the existence of the cured population and may give an adequate fit. In this case, the multi-state model would be fit without modeling the cured fraction, and there would be only one path from the alive state to the death state for subjects who died without recurrence. Therefore, we only consider transitions 2 → 3, 2 → 4, and 3 → 4. Another simpler model that may provide an adequate fit to the data is one in which all of the parameters in the 1 → 4 transition and in the 2 → 4 transition are forced to be the same. Subjects making a 1 → 4 transition are those who die after being cured of disease and those who make a 2 → 4 transition are those who are not cured of disease but die prior to recurrence. In both of these cases, the subjects are dying from causes other than the cancer, so it is plausible that the parameters of these two hazard rates may be similar. This simpler model has five fewer parameters to estimate than the full model. The fit of the models without a cured fraction and with the parameters in transitions 1 → 4 and 2 → 4 constrained to be the same can be compared with the full model through a comparison of *DIC* values. Table IV provides the *DIC* values for each of these models for the 12 trials, with the *DIC* of the best fitting model in bold. In all 12 trials, the model with a cured group and different parameter values for the 1 → 4 and 2 → 4 transitions is preferred over the simpler models, with the model having no cured fraction providing the worst fit. The Cox–Snell residual plots for time to death can also be compared between the two models to visually assess the adequacy of the model fits. Appendix F in the supporting information provides the Cox–Snell residual plots for the model with no cured fraction. These plots show a larger deviation from the line through the origin than the plots made using the complete model, indicating a poorer fit to the data.

5.3. Recurrence only model

The multi-state Markov model provides a convenient way to deal with the competing risk nature of the recurrence and death events. In colon cancer, recurrence substantially changes the risk of death, and some noncured patients die from other causes prior to experiencing a recurrence. Therefore, for these patients,

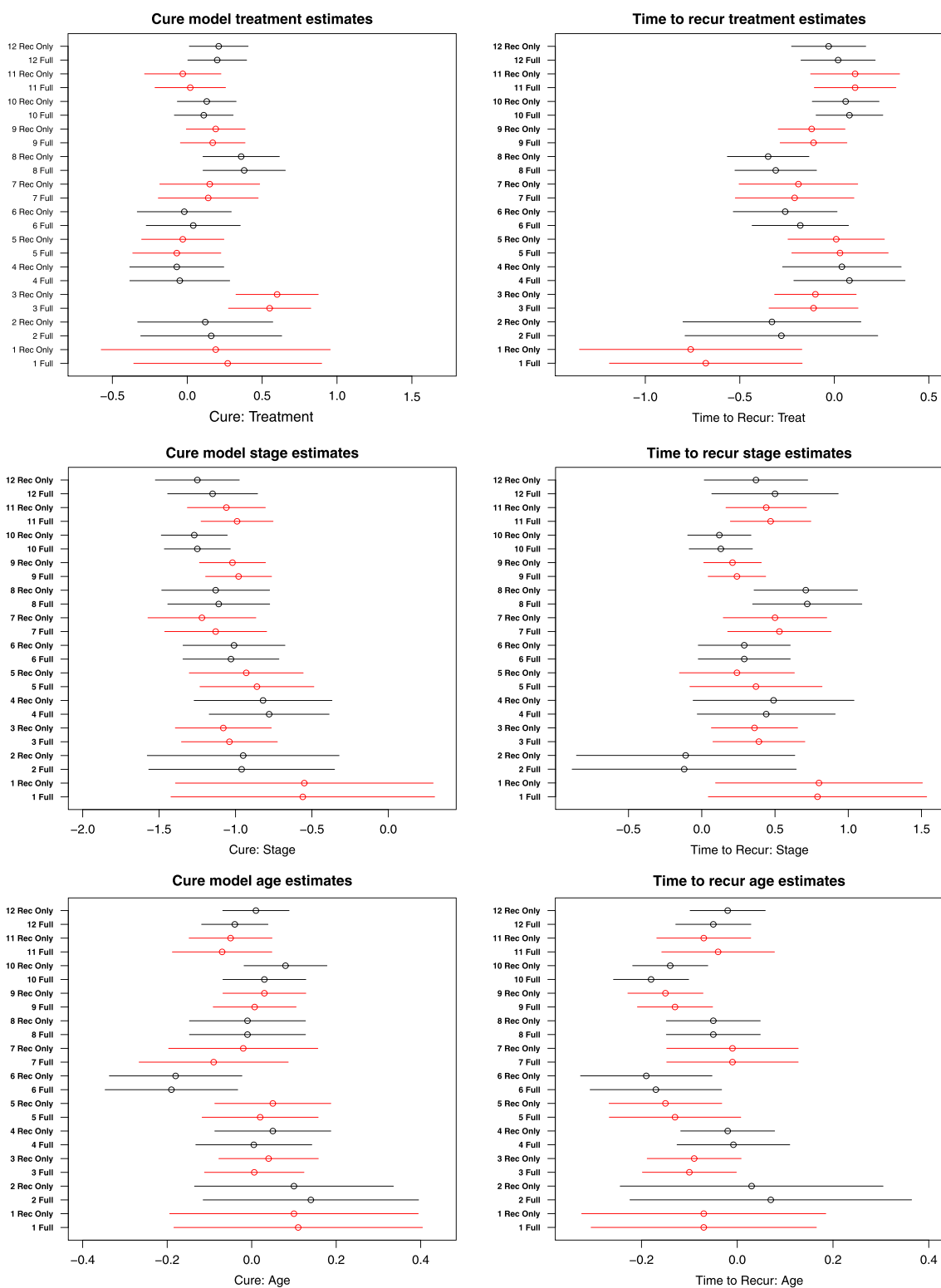


Figure 3. Comparison of estimates to recurrence only model. Results from 12 trials. The lower line for each trial is the posterior mean and 95% credible interval for the coefficient from the full multi-state cure model. The upper line is the posterior mean and 95% credible interval for the coefficient from the cure model that does not incorporate death times.

their recurrence time is unobservable and some caution must be taken in interpreting the probability of recurrence in the presence of death as a competing risk. If we are only interested in the recurrence event, we can empirically examine the effect of ignoring death by comparing the estimates obtained from the proposed multi-state model to those from a standard cure model for recurrence with a marginal proportional hazard model with Weibull baseline function for the hazard of recurrence. In this model, subjects who die without recurrence are censored for recurrence at their death time. A comparison of the estimates from the multi-state cure model and the simpler cure model provides insight into whether or not it is necessary to build the entire joint multi-state model for recurrence and death if we are only interested in the recurrence event. Figure 3 provides plots of the parameter estimates for treatment, stage, and age in the logistic model and hazard model for recurrence from both the multi-state model and from a standard cure model for recurrence. The plots show similar estimates of covariate effects with similar standard errors of effect estimates for both parts of the model, indicating that if we are only interested in the recurrence end point, a simpler model that ignores the time to death endpoint provides adequate estimation of parameters for these data sets. It can be shown algebraically that if we assume that $\lambda_{14}(t) = \lambda_{24}(t)$ then the maximum likelihood estimates from the full multi-state cure model are equivalent to the maximum likelihood estimates from the simple cure model, if the censoring times for recurrence and death do not differ. The deaths in the $1 \rightarrow 4$ and $2 \rightarrow 4$ transitions are both primarily due to causes other than cancer, so it might be reasonable to expect $\lambda_{14}(t)$ and $\lambda_{24}(t)$ to be similar, although in the previous section, we showed better *DIC* values for the full model compared with the model with $\lambda_{14}(t) = \lambda_{24}(t)$. Although the points estimates from the data for transitions $1 \rightarrow 4$ and $2 \rightarrow 4$ are dissimilar, there is wide uncertainty in the estimation of these quantities. Thus, it is not too surprising that there is considerable similarity in the estimates for the logistic model and for the $\lambda_{23}(t)$ parameters from the full multi-state cure model and simple cure model.

6. Simulations

A small simulation study was conducted to assess the performance of the estimation procedure and the impact of the prior distributions and sample size. The first simulation uses the same prior distributions as were used in the estimation of the colon cancer data, with a sample size of 1000. In the second simulation, the variance of all prior distributions is increased. The third simulation has the same prior distributions as the first simulation with the sample size decreased to 500 to assess how the prior distributions used in our data analysis performed among the varying trial sizes. In each of these three cases, half of the subjects were assigned to be in the treatment arm and half were assigned to be in the control arm. Two-thirds of the subjects were assigned to have stage three disease, and the remaining one-third were assigned to have stage two disease. Subjects were accrued over a 4-year period with 6 years of additional follow-up.

The probability of being cured of disease is first generated using $p_i = \frac{\exp(\gamma_0 + \gamma_{trt} Z_i + \gamma_{stage} S_i)}{1 + \exp(\gamma_0 + \gamma_{trt} Z_i + \gamma_{stage} S_i)}$, where Z_i denotes treatment group and S_i denotes stage. Each of these covariates is centered at 0 so that Z_i is equal to -0.5 for the control group and 0.5 for the treatment group and S_i is equal to -0.75 for stage two disease and 0.25 for stage three diseases. For those who are cured of disease, we then generate a death time using the hazard model for transition $1 \rightarrow 4$ with treatment as the only covariate. For those who are not cured of disease, we generate a recurrence time using the hazard model for transition $2 \rightarrow 3$ and a death time using the hazard model for transition $2 \rightarrow 4$ with treatment as the only covariate in each of these hazard rates. If the death time for uncured subjects is less than the recurrence time, then a $2 \rightarrow 4$ transition is made at the death time. If the recurrence time is less than the death time, then a $2 \rightarrow 3$ transition is made at that time. For those who recur, the time between their recurrence and death is generated using the hazard model for transition $3 \rightarrow 4$ with treatment and recurrence time as covariates. Subjects are censored 6 years after the last accrual.

Table V provides posterior means, standard deviations, average posterior standard deviations ($P\bar{S}D$), and coverage rates from 200 simulations. As the parameters in the hazards for transition for subjects with a recurrence are fully identified, the parameters in transitions $2 \rightarrow 3$ and $3 \rightarrow 4$ are estimated with little bias in all three simulations. Parameters in the logistic model for the probability of being cured are also consistently estimated. When the sample size remains the same, but the variance on the priors is increased, the parameters in transition $1 \rightarrow 4$ have slightly lower coverage rates and slightly higher standard deviations and $P\bar{S}D$'s. The parameters in transition $2 \rightarrow 4$ also have slightly lower coverage rates and noticeably higher standard deviations and $P\bar{S}D$'s. When the priors are kept the same but the

Table V. Simulation results: simulations 1 and 3 have informative prior distributions and simulation 2 uses weakly informative priors.

Parameter	Simulation 1 ($n = 1000$)					Simulation 2 ($n = 1000$)					Simulation 3 ($n = 500$)				
	Prior	True value	Estimate (SD)	PSD	Coverage	Prior	True value	Estimate (SD)	PSD	Coverage	Prior	True value	Estimate (SD)	PSD	Coverage
$\log(\lambda_{14})$	N(0,4)	1.8	1.79 (0.04)	0.04	0.94	N(0,25)	1.8	1.79 (0.04)	0.04	0.92	N(0,4)	1.8	1.79 (0.05)	0.05	0.98
ρ_{14}	G(1,0.4)	2	1.94 (0.16)	0.16	0.94	G(1,1.6)	2	1.94 (0.18)	0.17	0.93	G(1,0.4)	2	1.94 (0.19)	0.20	0.95
β_{trr14}	N(0,0.06)	0	0.004 (0.10)	0.11	0.97	N(0,4)	0	0.006 (0.12)	0.12	0.94	N(0,0.06)	0	-0.0001 (0.10)	0.12	0.995
$\log(\lambda_{23})$	N(0,4)	0.8	0.80 (0.07)	0.07	0.96	N(0,25)	0.8	0.79 (0.07)	0.07	0.95	N(0,4)	0.8	0.80 (0.09)	0.10	0.94
ρ_{23}	G(1,0.4)	1.5	1.49 (0.08)	0.08	0.93	G(1,1.6)	1.5	1.49 (0.08)	0.08	0.93	G(1,0.4)	1.5	1.47 (0.10)	0.10	0.94
β_{trr23}	N(0,1)	-0.5	-0.49 (0.16)	0.17	0.95	N(0,4)	-0.5	-0.49 (0.17)	0.17	0.93	N(0,1)	-0.5	-0.50 (0.19)	0.21	0.97
$\log(\lambda_{24})$	N(0,4)	1.8	2.06 (0.35)	0.41	0.95	N(0,25)	1.8	2.25 (0.60)	0.61	0.91	N(0,4)	1.8	2.13 (0.37)	0.55	0.97
ρ_{24}	G(1,0.4)	1	0.98 (0.13)	0.13	0.95	G(1,1.6)	1	0.96 (0.16)	0.16	0.92	G(1,0.4)	1	1.01 (0.15)	0.19	0.98
β_{trr24}	N(0,1)	0	-0.02 (0.25)	0.30	0.97	N(0,4)	0	-0.08 (0.36)	0.40	0.96	N(0,1)	0	-0.07 (0.34)	0.43	0.98
$\log(\lambda_{34})$	N(0,4)	0.9	0.89 (0.08)	0.08	0.95	N(0,25)	0.9	0.89 (0.08)	0.08	0.95	N(0,4)	0.9	0.92 (0.09)	0.10	0.95
ρ_{34}	G(1,0.4)	0.9	0.91 (0.05)	0.05	0.95	G(1,1.6)	0.9	0.91 (0.05)	0.05	0.94	G(1,0.4)	0.9	0.90 (0.06)	0.06	0.95
β_{trr34}	N(0,1)	0	0.007 (0.13)	0.14	0.98	N(0,4)	0	0.005 (0.13)	0.14	0.97	N(0,1)	0	0.002 (0.18)	0.18	0.95
β_{TT34}	N(0,1)	-0.1	-0.10 (0.06)	0.07	0.95	N(0,4)	-0.1	-0.10 (0.06)	0.07	0.96	N(0,1)	-0.1	-0.11 (0.08)	0.08	0.95
γ_0	N(0,1)	0.5	0.52 (0.13)	0.14	0.94	N(0,4)	0.5	0.54 (0.15)	0.14	0.92	N(0,1)	0.5	0.53 (0.15)	0.18	0.97
γ_{trt}	N(0,1)	0.6	0.60 (0.16)	0.18	0.98	N(0,4)	0.6	0.61 (0.17)	0.19	0.97	N(0,1)	0.6	0.58 (0.22)	0.24	0.97
γ_{stage}	N(0,1)	-0.9	-0.90 (0.16)	0.17	0.97	N(0,4)	-0.9	-0.92 (0.16)	0.17	0.96	N(0,1)	-0.9	-0.87 (0.20)	0.23	0.98

N(x,y), Normal(mean, variance); G(x,y), Gamma(mean, variance); SD denotes standard deviation of posterior means; PSD denotes average of the posterior standard deviations.

sample size is decreased, the parameter estimates remain similar, but the standard deviations and $P\bar{S}D$'s are increased across all parameters. Simulation 1, which uses the priors that were used in obtaining estimates for our 12 trials and a large sample size, appears to provide the best estimation across all parameters. We note, however, that even in this simulation, there is some bias and large uncertainty in estimating the parameters for transition $2 \rightarrow 4$ because of the small number of subjects who are uncured of disease but die prior to a recurrence.

7. Discussion

In this paper, we have used a multi-state Markov model to formulate a joint model for recurrence and death in colon cancer with an incorporated cured fraction. The proposed model is complex with a large number of parameters to estimate; however, it is well motivated by the context of the disease process of colon cancer, where it is likely that a proportion of the population will be cured of disease, and recurrence is known to influence survival time. The parameter estimates obtained from the model provide meaningful interpretations as to how different covariates affect the various disease elements. We presented methods for assessing the adequacy of the model fit and the functional form of covariates, both for the overall model and individual model components, which can aid in choosing an appropriate model. A Bayesian estimation strategy was used to estimate parameters, with informative priors placed on some parameters to aid in estimation. As our simulation results show, some parts of the model are more sensitive to the choice of the prior due to a lack of identifiability.

Adaptations to this model are possible. For example, we have used fully parametric models in this article. Semi-parametric alternatives for the Weibull model could be explored [22]; however, the Weibull model appeared to provide an adequate fit to the data in this setting, except, possibly for the transition from recurrence to death. To explore this, we fit the proposed model with a generalized Weibull baseline hazard, as described by Foucher *et al.* [1] for the $3 \rightarrow 4$ transition. Based on a comparison of DIC values, this model provided a slightly better overall fit to the data for all 12 trials, and the Cox–Snell residual plots for the $3 \rightarrow 4$ transition appeared to provide a better fit. However, the covariate coefficient estimates and PSD for this model were nearly identical to those obtained by the model presented in this paper, indicating that the proposed model is somewhat robust to slight model misspecification.

Maximum likelihood estimation could be used to obtain parameter estimates as opposed to our Bayesian approach; however, this is computationally more difficult, and the Bayesian estimation approach facilitates placing informative priors on selected parameters. In our model formulation, recurrence times are treated as known, when more realistically they occur sometime within an interval of scheduled clinic visits, but the left hand end of this interval is unknown to us. However, if this information was available, the models could be formulated to reflect this by treating recurrence time as interval censored. An additional useful extension to the proposed model for current clinical trial practice would be the ability to return to the disease-free state after recurrence. This now appears to occur for about 10% of patients who recur but then live for a long time after the recurrence, presumably due to subsequent therapy.

As the models demonstrated, there were common effects of age, stage, and the Weibull shape parameter across studies within the logistic model and within the transition rates. Therefore, information from these covariates could be borrowed across trials with estimates shrunk towards common values by fitting a Hierarchical model to the 12 trials examined. Estimates from the hierarchical model could then be applied to data from a new trial during follow-up to aid in the estimation and analysis of treatment effects on OS. This will be described in future research.

The proposed model also has the potential to use recurrence as an auxiliary variable for OS. As recurrence time is often an informative marker in predicting a patient's OS time, recurrence information along with the parameter estimates from the joint model could be used to impute death times for censored subjects and potentially improve the efficiency in the analysis of OS. This strategy could also result in the shortening of the length of the trial, if the information lost due to early censoring could be correctly recovered by the model. The effectiveness of this strategy will be examined in future work.

Acknowledgements

This research was partially supported by National Institutes of Health grants CA083654 and CA129102.

References

1. Foucher Y, Mathieu E, Saint-Pierre P, Durand JF, Daurés JP. A semi-Markov model based on generalized Weibull distribution with an illustration for HIV disease. *Biometrical Journal* 2005; **47**(6):1–9.
2. Putter H, van der Hage J, de Bock GH, Elgalta R, van de Velde CJH. Estimation and prediction in a multi-state model for breast cancer. *Biometrical Journal* 2006; **48**:366–380.
3. Andersen PK, Keiding N. Multi-state models for event history analysis. *Statistical Methods in Medical Research* 2002; **11**:91–115.
4. Meira-Machado L, de Uña-Álvarez J, Cadarso-Suárez C, Andersen PK. Multi-state models for the analysis of time-to-event data. *Statistical Methods in Medical Research* 2009; **18**:195–222.
5. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine* 2007; **26**:2389–2430.
6. Klein JP, Keiding N, Copelan EA. Plotting summary predictions in multi-state survival models-probabilities of relapse and death in remission for bone-marrow transplantation patients. *Statistics in Medicine* 1993; **12**:2315–2332.
7. Dabrowska DM, Sun GW, Horowitz MM. Cox regression in a Markov renewal model: an application to the analysis of bone-marrow transplant data. *Journal of the American Statistical Association* 1994; **89**:867–877.
8. Lagakos SW, Sommer CJ, Zelen M. Semi-Markov models for partially censored data. *Biometrics* 1978; **65**:311–317.
9. Andersen PK, Esbjerg S, Sørensen TIA. Multi-state models for bleeding episodes and mortality in liver cirrhosis. *Statistics in Medicine* 2000; **19**:587–599.
10. Berkson J, Gage RP. Survival curve for cancer patients following treatment. *Journal of the American Statistical Association* 1952; **47**:501–515.
11. Yamaguchi K. Accelerated failure-time regression models with a regression model of surviving fraction: an application to the analysis of “permanent employment” in Japan. *Journal of the American Statistical Association* 1992; **87**:284–292.
12. Taylor JMG. Semi-parametric estimation in failure time mixture models. *Biometrics* 1995; **51**:899–907.
13. Sy JP, Taylor JMG. Estimation in a Cox proportional hazards cure model. *Biometrics* 2000; **56**:227–236.
14. Peng Y, Dear KBG. A nonparametric mixture model for cure rate estimation. *Biometrics* 2000; **56**:237–243.
15. Farewell VT. Mixture models in survival analysis: are they worth the risk?. *Canadian Journal of Statistics* 1986; **14**:257–262.
16. Kneib T, Hennerfeind A. Bayesian semi parametric multi-state models. *Statistical Modelling* 2008; **8**:169–198.
17. Sargent DJ, Wieand HS, Haller DG, Gray R, Benedetti JK, Buyse M, Labianca R, Seitz JF, O’Callaghan CJ, Francini G, Grothey A, O’Connell M, Catalano PJ, Blanke CD, Kerr D, Green E, Wolmark N, Andre T, Goldberg RM, De Gramont A. 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.
18. Conlon ASC, Taylor JMG, Sargent DJ, Yothers G. Using cure models and multiple imputation to utilize recurrence as an auxiliary variable for overall survival. *Clinical Trials* 2011; **8**:581–590.
19. Farewell VT. The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* 1982; **38**:1041–1046.
20. Sargent DJ, Patiyil S, Yothers G, Haller DG, Gray R, Benedetti J, Buyse M, Labianca R, Seitz JF, O’Callaghan CJ, Francini G, Grothey A, O’Connell M, Catalano PJ, Kerr D, Green E, Wieand HS, Goldberg RM, de Gramont A, ACCENT Group. 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.
21. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B* 2002; **64**(4):583–639.
22. Satten G, Sternberg M. Fitting semi-markov models to interval-censored data with unknown initiation times. *Biometrics* 1999; **55**:507–513.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher’s web site.