Supplementary Materials to "Bayesian variable selection and shrinkage strategies in a complicated modeling setting with missing data: a case study using multistate models"

Lauren J Beesley and Jeremy M G Taylor
University of Michigan, Department of Biostatistics

Contents

S1 Generalized multistate cure model structure

S1.1 Model formulation

As described in the main paper **Section 2**, we model the head and neck cancer data as follows:

$$
\begin{aligned}
&\text{P(Persistent)}: && \text{logit}(P(G_i = 2|X_i)) = \omega_0 + X_i\omega_1 \\
&\text{P(Not Cured | Not Persistent)}: && \text{logit}(P(G_i = 1|X_i, G_i \neq 2)) = \alpha_0 + X_i\alpha_1 \\
&\text{Treatment} \rightarrow \text{Death } (G = 2): && \lambda_{54}(t) = \lambda^0_{54}(t)\exp(X_i\beta_{54}), t > 0 \\
&\text{Treatment} \rightarrow \text{Death } (G = 1): && \lambda_{14}(t) = \lambda^0_{14}(t)\exp(X_i\beta_{14}), t > 0 \\
&\text{Treatment} \rightarrow \text{Death } (G = 0): && \lambda_{24}(t) = \lambda^0_{24}(t)\exp(X_i\beta_{24}), t > 0 \\
&\text{Treatment} \rightarrow \text{Recurrence}: && \lambda_{13}(t) = \lambda^0_{13}(t)\exp(X_i\beta_{13}), t > 0 \\
&\text{Recurrence} \rightarrow \text{Death}: && \lambda_{34}(t - T_{ir}) = \lambda^0_{34}(t - T_{ir})\exp(X_i\beta_{34}), t > T_{ir}
\end{aligned}
$$

We call these different Cox and logistic regression models the "submodels," which together form the proposed multistate cure model. This model is an extension of the standard multistate cure model in Beesley and Taylor (2018) and Conlon et al. (2013) to incorporate disease persistence. As shown in Conlon et al. (2013), the "standard" multistate cure model without the baseline persistence category is identifiable for data with sufficient follow-up for death and recurrence. Improved parameter identifiability results from assuming $\beta_{14} = \beta_{24}$ (Beesley and Taylor, 2018). Since our model is an extension of the multistate cure model to incorporate an additional baseline category and since it is known for all patients whether they are persistent or non-persistent, our proposed model is also identifiable given sufficient follow-up and sufficient numbers of persistent patients. In the model structure above, we leave the forms of the baseline hazards unspecified, but we will assume parametric Weibull baseline hazards in our data analysis.

Now, we will describe the forms of the likelihood function for this multistate cure model. First, we note that

$$
\begin{aligned}
P(G_i = 2|X_i) =& \frac{e^{\omega_0 + X_i\omega_1}}{1 + e^{\omega_0 + X_i\omega_1}} \\
P(G_i = 1|X_i) =& \frac{1}{1 + e^{\omega_0 + X_i\omega_1}} \times \frac{e^{\alpha_0 + X_i\alpha_1}}{1 + e^{\alpha_0 + X_i\alpha_1}} \\
P(G_i = 0|X_i) =& \frac{1}{1 + e^{\omega_0 + X_i\omega_1}} \times \frac{1}{1 + e^{\alpha_0 + X_i\alpha_1}}
\end{aligned}
$$

Let $S_j(t)$ to denote the probability of remaining in state j for time t. Let $\Lambda_{jk}(t)$ denote the cumulative hazard for the $j \rightarrow k$ transition. The multistate modeling assumptions result in the following complete data likelihood:

$$
\begin{aligned}
L^{(com)}(\Theta) \propto \prod_{i=1}^{n} &\left[\frac{e^{\omega_0 + \omega_1^T X_i}}{1 + e^{\omega_0 + \omega_1^T X_i}} S_5(Y_{id})\lambda_{54}(Y_{id})^{\delta_{id}}\right]^{\mathbb{I}(G_i=2)} \left[\frac{1}{1 + e^{\omega_0 + \omega_1^T X_i}}\right]^{\mathbb{I}(G_i \neq 2)} \\
&\times \left[\frac{1}{1 + e^{\alpha_0 + \alpha_1^T X_i}} S_2(Y_{id})\lambda_{24}(Y_{id})^{\delta_{id}}\right]^{\mathbb{I}(G_i=0)}
\end{aligned} \tag{S1.1}
$$

$$\times \left[\frac{e^{\alpha_0+\alpha_1^T X_i}}{1+e^{\alpha_0+\alpha_1^T X_i}} S_1(Y_{id}) \lambda_{14}(Y_{id})^{\delta_{id}} \right]^{\mathbb{I}(G_i=1,\delta_{ir}=0)}$$

$$\times \left[\frac{e^{\alpha_0+\alpha_1^T X_i}}{1+e^{\alpha_0+\alpha_1^T X_i}} S_1(Y_{ir}) \lambda_{13}(Y_{ir}) S_3(Y_{id}-Y_{ir}) \lambda_{34}(Y_{id}-Y_{ir})^{\delta_{id}} \right]^{\mathbb{I}(G_i=1,\delta_{ir}=1)}$$

In reality, baseline status G is not always known. In particular, we cannot always know whether patients are cured or non-cured. The observed data likelihood as a function of observed baseline state information is as follows:

$$L^{(obs)}(\Theta) \propto \prod_{i=1}^{n} \left[\frac{e^{\omega_0+\omega_1^T X_i}}{1+e^{\omega_0+\omega_1^T X_i}} S_5(Y_{id}) \lambda_{54}(Y_{id})^{\delta_{id}} \right]^{\mathbb{I}(G_i=2)} \left[\frac{1}{1+e^{\omega_0+\omega_1^T X_i}} \right]^{\mathbb{I}(G_i\neq 2)} \tag{S1.2}$$

$$\times \left[\frac{1}{1+e^{\alpha_0+\alpha_1^T X_i}} S_2(Y_{id}) \lambda_{24}(Y_{id})^{\delta_{id}} + \frac{e^{\alpha_0+\alpha_1^T X_i}}{1+e^{\alpha_0+\alpha_1^T X_i}} S_1(Y_{id}) \lambda_{14}(Y_{id})^{\delta_{id}} \right]^{\mathbb{I}(G_i\neq 2,\delta_{ir}=0)}$$

$$\times \left[\frac{e^{\alpha_0+\alpha_1^T X_i}}{1+e^{\alpha_0+\alpha_1^T X_i}} S_1(Y_{ir}) \lambda_{13}(Y_{ir}) S_3(Y_{id}-Y_{ir}) \lambda_{34}(Y_{id}-Y_{ir})^{\delta_{id}} \right]^{\mathbb{I}(G_i=1,\delta_{ir}=1)}$$

In a Bayesian formulation of the above model, we can estimate parameters using an iterative MCMC algorithm. Since the form of the posterior distributions will be messy, we will need to perform many Metropolis-Hastings steps within a Gibbs sampling algorithm. Rather than maximizing S1.2 directly, we propose an augmentation step in which we draw missing values of G_i at each step of the MCMC algorithm. This allows us to then draw parameters using the simpler likelihood form in S1.1.

Given observed data, we know whether or not $G_i = 2$ for all patients. However, for non-persistent patients censored for recurrence, we do not know whether $G_i = 0$ or $G_i = 1$. For these patients, we can impute $G_i = 1$ with probability

$$\begin{aligned} &P(G_i = 1 | X_i, Y_{id}, Y_{ir}, \delta_{id}, \delta_{ir} = 0, G_i \neq 2) \qquad \text{(S1.3)} \\ &= \frac{P(G_i = 1 | G_i \neq 2) \lambda_{14}(Y_{id})^{\delta_{id}} S_1(Y_{id})}{P(G_i = 1 | G_i \neq 2) \lambda_{14}(Y_{id})^{\delta_{id}} S_1(Y_{id}) + P(G_i = 0 | G_i \neq 2) \lambda_{24}(Y_{id})^{\delta_{id}} S_2(Y_{id})} \\ &= \frac{e^{\alpha_0 + X_i \alpha_1} \lambda_{14}(Y_{id})^{\delta_{id}} e^{-\Lambda_{13}(Y_{id}) - \Lambda_{14}(Y_{id})}}{e^{\alpha_0 + X_i \alpha_1} \lambda_{14}(Y_{id})^{\delta_{id}} e^{-\Lambda_{13}(Y_{id}) - \Lambda_{14}(Y_{id})} + \lambda_{24}(Y_{id})^{\delta_{id}} e^{-\Lambda_{24}(Y_{id})}} \end{aligned}$$

S1.2 Estimation with additional missing data

In practice, we may often have additional forms of missing data that must be handled within the estimation algorithm. In our head and neck cancer data analysis example, we have covariate missingness, missingness in latent cure status, and missing data in recurrence information resulting from longer follow-up for death. In this more complicated setting, we perform a three-step imputation procedure at each iteration of the MCMC algorithm in which we impute each type of missing data fixing the most recent draw of Θ.

First, we comment on underlying assumptions about the missingness. Our analysis does rely on the assumption that missingness is missing at random (MAR) (Little and Rubin, 2002). Although we have explored the possibility of missingness related to underlying cure status for these data in Beesley et al. (2019), this paper does assume MAR. The main types of missingness in this paper are as follows: (1) cure status, (2) unequal censoring of the outcomes, (3) missingness in covariates and particularly HPV status. Missingness in cure status can be entirely explained by the observed outcome information (cure status missing if and only if patient is non-persistent and censored for recurrence), so its missingness is MAR. Under an assumption of uninformative censoring of recurrence given covariates, we can also treat the unequal censoring problem as MAR. Therefore, our main concern regarding the MAR assumption is in missingness of HPV status.

It is possible that HPV status missingness may be related to calendar time, which is not included as an adjustment factor in the main model. In general, however, our goal is to use this head and neck cancer example as a case study for implementation and comparison of the various methods in a complicated modeling setting. Therefore, we do not expect minor violations of MAR for HPV status missingness to substantially impact the main takeaways of this data analysis.

(1) Imputation for unequal censoring of recurrence and death outcomes
In many multistate modeling settings, we may have different censoring times for different time-to-event outcomes. In our head and neck cancer example, our outcomes are time to death and time to cancer recurrence. In order to follow patients for death, researchers can contact patients (e.g. by phone) and also consult national death records. In contrast, follow-up for cancer recurrence requires regular doctor's appointments and evaluation of recurrence through some diagnostic test. It is very common, therefore, for us to have information about death status long after patients are no longer followed for recurrence. This creates a type of missing data where death status is known for a given patient up to time Y_{id}, but recurrence status is only known to be censored at some time prior to Y_{id}. Recurrence status is unknown in the intermediate time.

First, we build up some notation. Let C_r be the censoring time for recurrence and C_d be the censoring time for death. We assume that $C_r \leq C_d$ and for some patients, $C_{ir} < C_{id}$. Let T_{ir} and T_{id} represent the true recurrence and censoring times, where T_{ir} is defined as infinity for cured and persistent subjects. For all patients, we observe C_{ir}-censored recurrence information, $Y_{ir}^0 = \min(T_{ir}, C_{ir}, T_{id})$ and $\delta_{ir}^0 = \mathbb{I}(Y_{ir} = T_{ir})$. We also observed C_{id}-censored death information, $Y_{id} = \min(T_{id}, C_{id})$ and $\delta_{id} = \mathbb{I}(Y_{id} = T_{id})$. When $\delta_{ir}^0 = 0$ and $Y_{ir}^0 < Y_{id}$, we have a problem with missing data.

Define $\mathbb{Z} = (Y_{ir}^0, \delta_{ir}^0, Y_{id}, \delta_{id}, G_i, X_i)$. Our goal is to impute values of $Y_{ir} = \min(T_{ir}, C_{id}, T_{id})$ and $\delta_{ir} = \mathbb{I}(Y_{ir} = T_{ir})$ that would have been observed if we had followed patients for recur-

rence as long as we followed them for death. We use the method proposed in Beesley and Taylor (2018) to accomplish this task. Without going into too many details, this method involves first imputing whether a recurrence occurred in the time interval (Y_{ir}^0, Y_{id}) and, if it did, when it occurred. First, we impute missing δ_{ir} from a Bernoulli distribution with probability

$$P(\delta_{ir} = 1|\mathbb{Z}) = \frac{\int_{Y_{ir}^0}^{Y_{id}} \lambda_{13}(t)S_1(t)S_3(Y_{id} - t)\lambda_{34}(Y_{id} - t)^{\delta_{id}}dt}{\int_{Y_{ir}^0}^{Y_{id}} \lambda_{13}(t)S_1(t)S_3(Y_{id} - t)\lambda_{34}(Y_{id} - t)^{\delta_{id}}dt + \lambda_{14}^{\delta_{id}}(Y_{id})S_1(Y_{id})}$$

If imputed $\delta_{ir} = 0$, we set $Y_{ir} = Y_{id}$. Otherwise, we draw $Y_{ir} = T_{ir}$ from

$$f(T_{ir} = t|\delta_{ir} = 1, \mathbb{Z}) \propto \lambda_{13}(t)S_1(t)S_3(Y_{id} - t)\lambda_{34}(Y_{id} - t)^{\delta_{id}}\mathbb{I}(Y_{ir}^0 < t < Y_{id})$$

(2) Imputation of cure status

Fixing imputed Y_{ir} from step 1, we can then impute missing cure status G_i using S1.3. In Beesley and Taylor (2018), however, we found improved performance when we imputed missing $G_i = 1$ for patients missing G_i with probability

$$P(G_i = 1|X_i, Y_{id}, Y_{ir}^0, \delta_{id}, \delta_{ir}^0 = 0, G_i \neq 2) =$$
$$\frac{P(G_i = 1|G_i \neq 2)\left[\lambda_{14}(Y_{id})^{\delta_{id}}S_1(Y_{id}) + K_i\right]}{P(G_i = 1|G_i \neq 2)\left[\lambda_{14}(Y_{id})^{\delta_{id}}S_1(Y_{id}) + K_i\right] + P(G_i = 0|G_i \neq 2)\lambda_{24}(Y_{id})^{\delta_{id}}S_2(Y_{id})}$$
$$\text{where } K_i = \mathbb{I}(Y_{ir}^0 < Y_{id} \text{ and } \delta_{ir}^0 = 0)\int_{Y_{ir}^0}^{Y_{id}} \lambda_{13}(t)S_1(t)S_3(Y_{id} - t)\lambda_{34}(Y_{id} - t)^{\delta_{id}}dt$$

This expression is similar to S1.3 except for the inclusion of term K_i. If we don't have unequal censoring of the two event time outcomes, then $K_i = 0$ for all patients, and we recover S1.3.

(3) Imputation of missing covariate values

We can use any method we want to impute the missing covariate values at this point, conditional on the most recent imputations for the other variables and the most recent drawn Θ. For example, we might specify a chained equations imputation procedure. Covariate imputation in Beesley and Taylor (2018) and in our data analysis proceeds using the substantive model compatible chained equations imputation method proposed in Bartlett et al. (2014), which involves imputing missing covariates iteratively from a distribution that incorporates both the outcome model distribution (in this case, the generalized multistate cure model) and a model for the particular covariate given the other covariates.

In particular, let $X^{(p)}$ be a covariate with missingness and $X^{(-p)}$ be the other covariates in X. We impute each missing covariate $X^{(p)}$ one-by-one from distribution $f(X^{(p)}|O, X^{(-p)})$, where O represents outcome information. In usual chained equations, imputation proceeds assuming a regression model form for $f(X^{(p)}|O, X^{(-p)})$. For continuous $X^{(p)}$, for example, we might specify a linear regression with predictors $X^{(-p)}$ and outcome information. In substantive model compatible chained equations, however, we specify the structure of $f(X^{(p)}|O, X^{(-p)})$ so that it is compatible with the outcome model we wish to fit at the end of the day. We impute from distribution $f(X^{(p)}|O, X^{(-p)})$ proportional to $f(O|X)f(X^{(p)}|X^{(-p)})$. $f(O|X)$ is our multistate model of interest in S1.1,

and $f(X^{(p)}|X^{(-p)})$ is an assumed regression model for the given covariate given all other covariates. For categorical variables, the exact form of $f(X^{(p)}|O, X^{(-p)})$ can be easily worked out, and imputation from $f(X^{(p)}|O, X^{(-p)})$ is straightforward. In other settings, $f(X^{(p)}|O, X^{(-p)})$ is more complicated and we might impute using sampling methods such as Metropolis-Hastings.

In order to perform the covariate imputation step in the head and neck cancer data analysis, we need to specify regression model $f(X^{(p)}|X^{(-p)})$ for each covariate with missingness. We have missingness in gender, ACE27 comorbidities (none, mild, moderate, severe), smoking status (never, current, former), and HPV status. The models $f(X^{(p)}|X^{(-p)})$ are specified using a multinomial or logistic regression model form including all other variables in X as predictors. Additional information not included in the outcome model may be useful for imputation (for example, number of sexual partners for imputing HPV status), but we focus only on the variables included in our outcome model of interest for this analysis.

S2 Reversible jump proposal distributions

In the main paper, we describe a reversible jump algorithm for drawing parameters under the point mass at zero prior. Here, we provide some additional information about the proposal distributions used in those Metropolis-Hastings steps for drawing θ_s and γ_s, where θ_s and γ_s correspond to the parts of θ and γ from the s^{th} submodel in the multistate model as a whole.

We partition to joint proposal distribution as follows:

$$q(\gamma_s^*, \theta_s^* | \gamma_s, \theta_s) = q_1(\theta_s^* | \gamma_s^*, \gamma_s, \theta_s) q_2(\gamma_s^* | \gamma_s)$$

Here, we provide one strategy for specifying q_1 and q_2, although other strategies could also be used.

Defining q_2: Given the current value of γ_s, we first propose γ_s^* using the following rules (which imply the form of q_2). At each iteration of the reversible jump algorithm, we perform one of the following moves: (a) add a covariate into the submodel, (b) remove a covariate, (c) swap in a covariate for one already in the submodel, and (d) keep the covariate set the same (called a null move). Note that we can incorporate grouping by performing the following for *groups* of covariates rather than individual covariates. At each iteration of the algorithm, we choose our move type as in **Table S1**. In particular, we first draw the move type (a-d) with probabilities as in **Table S1**. For moves involving changes to the covariate set, we chose which covariates to include/exclude with equal probability among candidate covariates. For example, if γ_s has all but two covariates included in the model and we determine we want to add a covariate to the model on γ_s^* (drawn with probability 1/6), we would choose which covariate to include with probability 1/2. Moves in which we swap covariates require us to determine which included covariate will be dropped and which excluded covariate will be added by drawing with equal probability from candidate covariates in each step. The probabilities in **Table S1** for each of the moves follow Newcombe et al. (2014).

Table S1: Proposal distribution $q_2(\gamma_s^* | \gamma_s)$*

γ_s	Possible moves for γ_s^*	Probability
No covariates included	Add a covariate	1/6
	Null (keep the same)	5/6
All covariates included	Subtract a covariate	1/6
	Null (keep the same)	5/6
else	Add a covariate	1/6
	Subtract a covariate	1/6
	Swap covariates	1/6
	Null (keep the same)	1/2

** For non-null moves, randomly select which covariate/s to add/substract with equal probabilities among candidate covariates to obtain γ^*.*

Defining q_1: Now, we propose $\theta_s^* | \gamma_s^*, \gamma_s, \theta_s$. If we subtract a covariate in γ_s^* relative to γ_s, we then set the corresponding elements of θ_s^* to zero. If we add a covariate, we draw the corresponding element of θ_s^* from a normal distribution centered at the previous nonzero value. Swapping a covariate will involve both subtracting and adding a covariate

following the above rules. For null moves, we obtain θ_s^* by re-drawing all nonzero elements from normal distributions centered at their current values in θ_s. The variance of these normal proposal distributions is a tuning parameter.

S3 Additional information for head and neck cancer data analysis

In this section, we present some additional analyses conducted comparing the variable selection/shrinkage strategies using the head and neck cancer data. **Table S2** provides some descriptive information about the observed covariate information, missingness rates, and outcome information.

As described in **Section 2**, we break the patients at baseline into three states: cured of cancer, never appeared to clear their cancer (persistent), and non-cured/non-persistent. It is only subjects in this final category that are eligible to experience a cancer recurrence. We model the data using **Figure 1**.

Table S2: Characteristics of N = 1692 patients with head and neck cancer

Characteristic	N (%) or Mean (SD)	Missing N (%)	Characteristic	N (%) or Mean (SD)	Missing N (%)
Baseline variables					
Age at diagnosis	60.3 (11.6)		ACE27 comorbidities		6 (0.4)
Cancer stage		0 (0)	None	418 (24.7)	
I/Cis	249 (14.7)		Mild	718 (42.4)	
II	190 (11.2)		Moderate	374 (22.1)	
III	234 (13.8)		Severe	176 (10.4)	
IV	1019 (60.2)		Cancer site		0 (0)
Cigarette use		34 (2.0)	Larynx	373 (22.0)	
Never	377 (22.3)		Hypopharynx	79 (4.7)	
Current	759 (44.9)		Oral cavity	569 (33.6)	
Former	522 (30.9)		Oropharynx	671 (39.7)	
HPV status		880 (52.0)	Gender		1 (0.05)
Negative	444 (26.2)		Female	428 (25.3)	
Positive	368 (21.7)		Male	1263 (74.6)	
Additional information					
Baseline state		1005 (59.3)	Outcomes		-
Persistent	173 (10.2)		Death	723 (42.7)	
Non-cured/non-persistent	354 (20.9)		Recurrence	354 (20.9)	
Assumed cured	160 (9.4)		Unequal censoring	813 (48.0)	

S3.1 Model selection and posterior inclusion probabilities

In this section, we explore the posterior distribution of γ and corresponding posterior inclusion probabilities for the spike and slab priors. Recall, these probabilities are calculated as the proportion of MCMC iterations in which a given covariate or covariate group is assigned $\gamma_g = 1$ for a given submodel. Although the horseshoe prior does not perform explicit variable selection, we can also obtain a pseudo posterior inclusion probability for each covariate in each submodel. This prior did not incorporate grouping of predictors. We will compare the posterior inclusion probabilities across the various selection/shrinkage methods. We will also explore the posterior distribution of *vector* γ_s

corresponding to sets of covariates included in a given submodel.

Table S3 shows the exact posterior inclusion probabilities for groups of variables under the two spike and slab priors considered. These probabilities are calculated as the proportion of iterations of the MCMC algorithm in which the group of variables was assigned $\gamma_g = 1$. Groups included in at least 50% of the MCMC iterations are shaded in gray.

Table S1 provides similar measures for the horseshoe prior. For this prior, posterior inclusion probabilities are not readily available. However, a rough measure of the inclusion rate for each θ_k can be estimated as the posterior mean of $1/(1 + \lambda_k^2 \tau_s)$ where τ_s corresponds to the submodel containing θ_k. Results are generally similar to the spike and slab priors, but some notable difference can be seen in the model for the probability of being non-cured given non-persistent. The spike and slab priors indicate only HPV status is important, while the horseshoe prior also identifies stage and potentially comorbidities as being important predictors in this model. This result may be driven by the grouped nature of the spike and slab prior selection we applied, where dummy variables with small associations may draw the inclusion probability for the entire group down toward zero.

Another use of the draws of γ_g is to determine the proportion of iterations in which *combinations* of γ_g are draw for a particular submodel. This provides some evidence regarding which covariates tend to be included together in the model. The final column in **Table S3** presents the proportion of MCMC iterations in which the shaded *combination* of variables was included in the corresponding submodel. The two priors tend to give similar results, but this is not always the case. The shaded covariate combinations tend to be chosen for a large proportion of iterations of the MCMC algorithm, with a larger proportion of iterations indicating stronger evidence in the "best" combination of predictors for each submodel.

Figure S2 visualizes the (cumulative) proportion of MCMC iterations each combination of predictors was chosen, where the filled-in area corresponds to the chosen covariates and the height of the filled-in area corresponds to the proportion of iterations in which that combination of covariates was chosen. The various model formulations chosen by the MCMC algorithm for each submodel were sorted from most to least often chosen. We note that the proportion of colored area above a particular variable in these plots corresponds to the posterior inclusion probability for that variable in the corresponding submodel. This type of plot was used in Chipman (1996) as a way to visualize the highest posterior models. For example, consider the first plot under the transition from treatment to recurrence in **Figure S2** under a mixture of normals prior. The results suggest that the combination of HPV and cancer stage was the model chosen for roughly 60% of the MCMC iterations. The cancer subsite variables (hypopharynx, larynx, oropharynx) were included in only a few of the iterations. Cancer stage was included in all iterations. In contrast, the point mass at zero prior indicates that cancer subsite variables should be included in the model since that group is included in at least 60% of the MCMC iterations. Overall, there are many instances in which the posterior weight put on different combinations of variables differs between the two priors. For some transitions, however, there is strong agreement between the two priors (e.g. that HPV status should be included in the model for death for persistent patients).

Table S3: Posterior inclusion probabilities for spike and slab priors in head and neck cancer model*

Mixture of normals

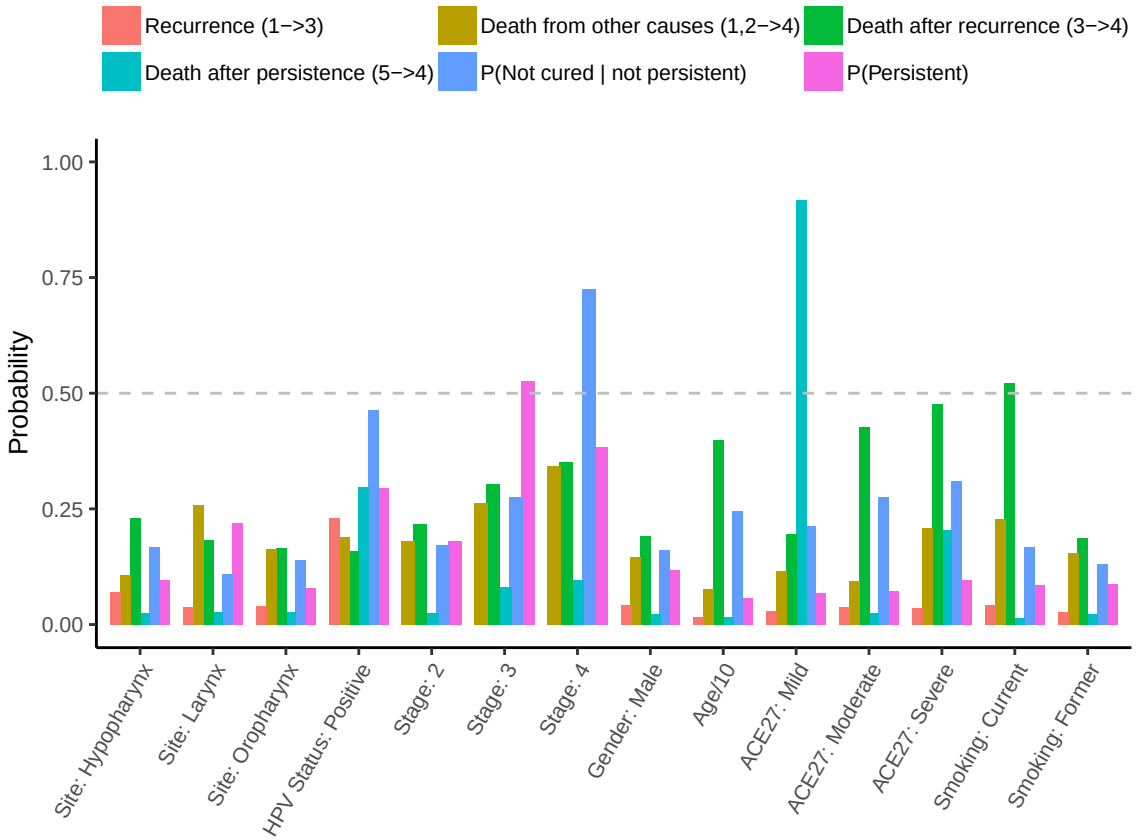
Transition	Site	HPV	Stage	Gender	Age	ACE27	Smoking	Prop.**
Recurrence	0.14	0.88	1.00	0.29	0.06	0.00	0.01	0.51
Death from other causes	0.02	0.14	0.34	0.24	1.00	0.93	1.00	0.41
Death after recurrence	0.50	0.47	0.96	0.29	0.09	0.01	0.37	0.12
Death after persistence	0.01	0.89	0.00	0.16	0.06	0.00	0.02	0.69
P(Not cured | not persistent)	0.01	0.99	0.12	0.10	0.06	0.00	0.01	0.71
P(Persistent)	0.01	1.00	1.00	0.14	0.28	0.94	0.03	0.57

Point mass at zero

Transition	Site	HPV	Stage	Gender	Age	ACE27	Smoking	Prop.**
Recurrence	0.67	0.93	1.00	0.19	0.04	0.00	0.01	0.49
Death from other causes	0.02	0.13	0.86	0.26	1.00	1.00	1.00	0.54
Death after recurrence	0.69	0.54	1.00	0.45	0.37	0.02	0.52	0.05
Death after persistence	0.01	0.92	0.00	0.15	0.04	0.00	0.01	0.75
P(Not cured | not persistent)	0.00	0.99	0.30	0.10	0.08	0.03	0.01	0.55
P(Persistent)	0.01	1.00	1.00	0.13	0.80	1.00	0.02	0.66

* Predictors with posterior inclusion probability > 0.5 are shaded for each submodel.
** Proportion of iterations in which that shaded *combination* of predictors was selected

Figure S1: Rough posterior inclusion probabilities for the horseshoe prior in head and neck cancer model*

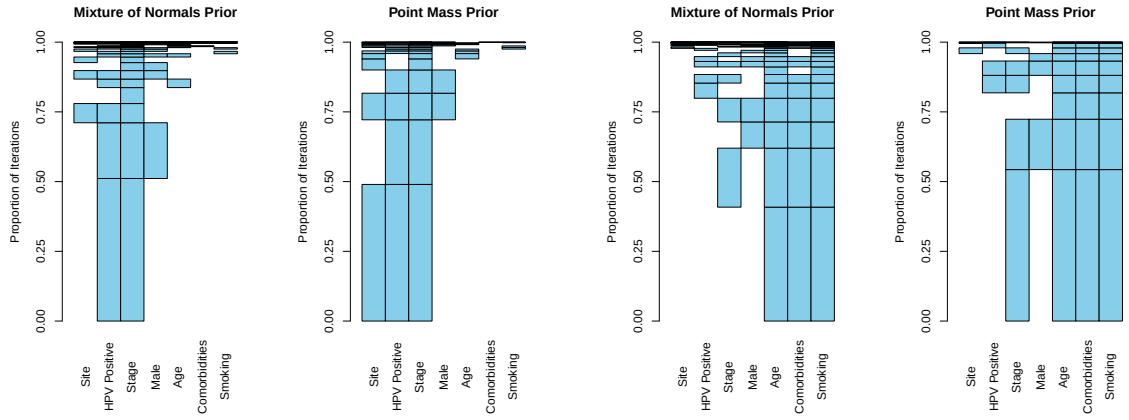


*The line height for a predictor indicates the posterior mean of $1/(1 + \lambda_k^2\tau^2)$ for that predictor, which is a measure of the covariate importance

Figure S2: Posterior probabilities for different covariate combinations in head and neck cancer model*

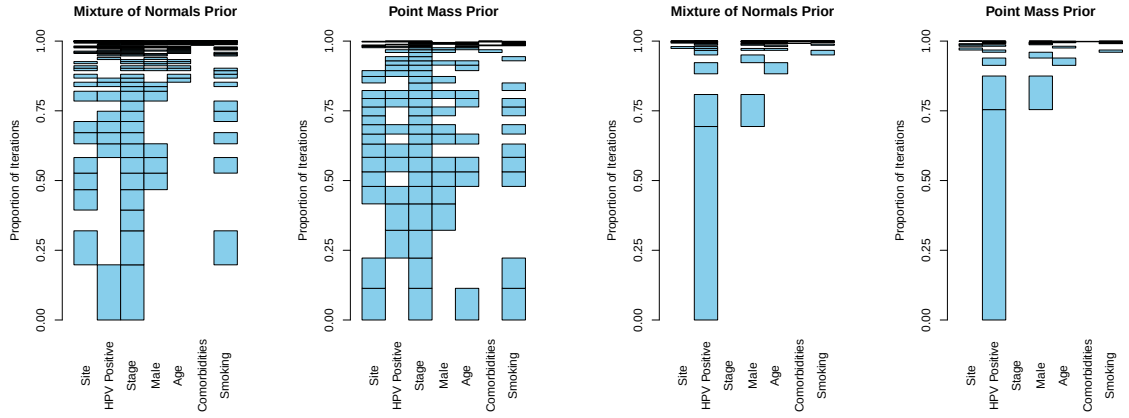
(a) Tx → Recur

(b) Tx → Death



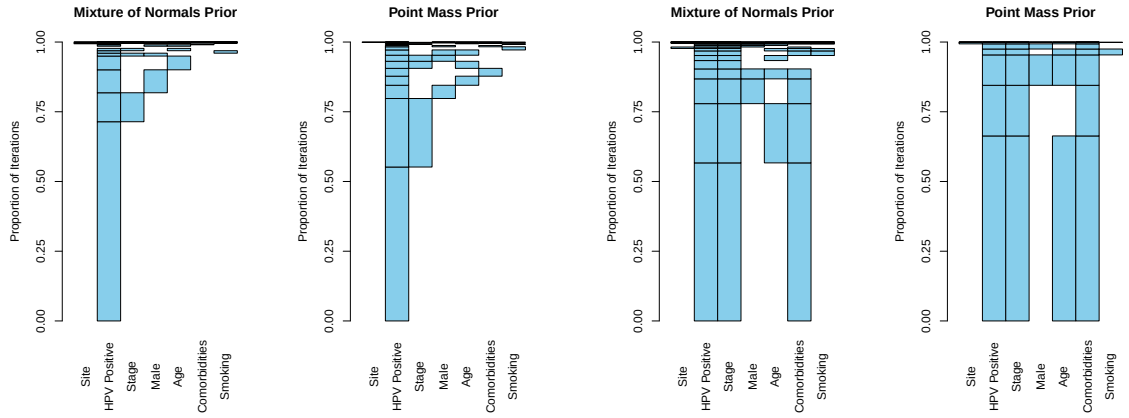
(c) Recur → Death

(d) Persist → Death



(e) P(Not cured | not persistent)

(f) P(Not persistent)



*The proportion of MCMC iterations in which each combination of covariates was included. A filled-in rectangle indicates that the covariate group was included in the model. The height of the rectangle indicates the proportion of iterations.

S3.2 Correlation of draws across transitions under multistate cure model

Figure S3 shows the correlations of the posterior draws of γ across iterations for both the mixture of normals and the point mass at zero priors. The first column of plots corresponds to submodels that are related to cure status and will be impacted by missing values in the cure status variable. We have shown elsewhere that parameter estimates across these transitions can be correlated even with fully observed covariate data. The second column of plots corresponds to the other submodels contained in the multistate cure model. We do not expect as many problems related to correlation *across* submodels for these transitions.

We generally see that the point mass at zero prior tends to have more strongly correlated draws of γ across iterations than seen for the mixture of normals prior. This may be a result of the reversible jump parameter sampling scheme used to fit the model with the point mass at zero prior. This algorithm draws the entire γ vector for a given submodel jointly, and the resulting algorithm may not as easily explore the posterior space of γ.

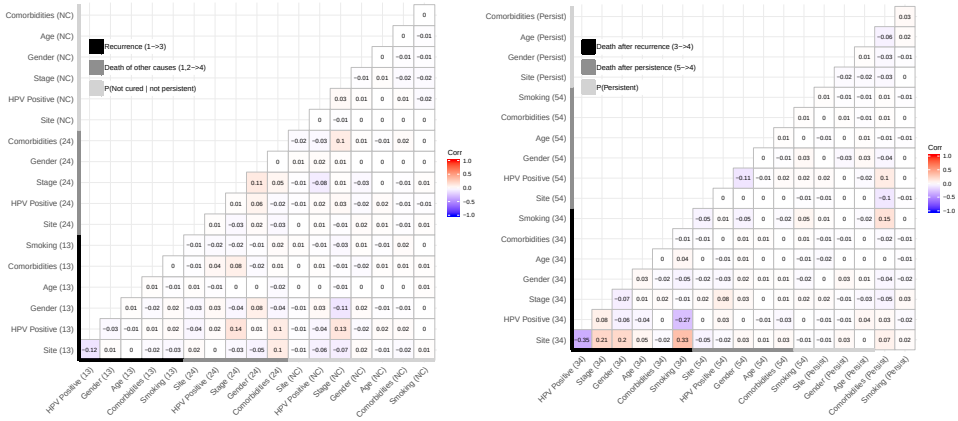
For both of these priors, inclusion and exclusion is drawn separately for parameters in different submodels. However, **Figure S3** provides evidence of correlation of the γ draws across submodels. For example, the parameter related to cancer site in the $1 \rightarrow 3$ transition has non-negligible correlation with the parameter for stage in the transition to death from other causes. In the main paper, we expressed a concern about correlation of the inclusion/exclusion indicators for a particular group across submodels. In this particular data analysis, we do not see strong cause for concern, but the potential for cross-model correlation may prove to be a greater problem in other applications.

In addition to correlation for γ directly, we are interested to see if we have evidence of strong correlation in draws of Θ across transitions. **Figure S4** shows the correlations of the posterior draws of Θ. Under both spike and slab priors, we see evidence of correlation in the draws of Θ related to stage across several submodels. Again, this may not be a huge problem in our current setting, but this highlights the potential for a problem related to MCMC parameter estimation in future data analyses.

Figure S3: Correlations of γ across iterations of the MCMC algorithm for head and neck cancer model*

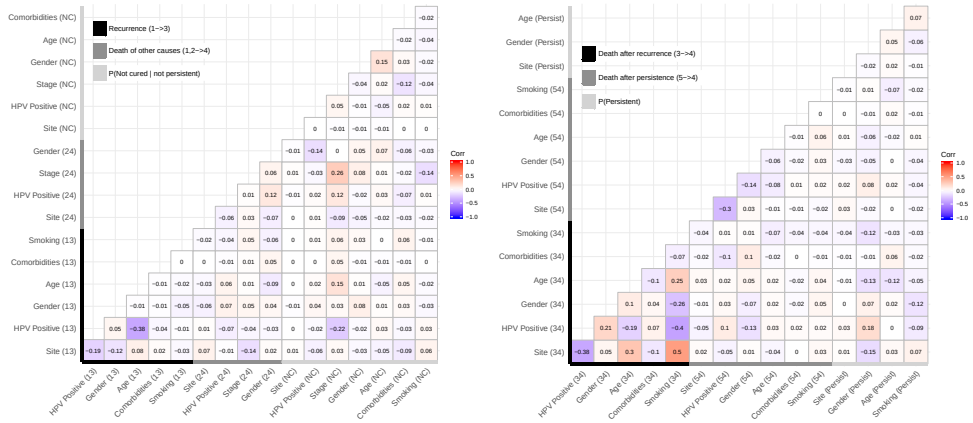
Mixture of normals:

(a) Cure-related submodels (b) Other submodels



Point mass at zero:

(c) Cure-related submodels (d) Other submodels

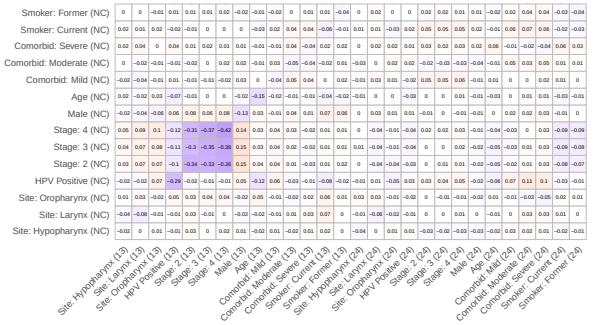


*This figure presents correlations for γ indicators across MCMC iterations. If γ is always 0 or 1, the variable is not listed.

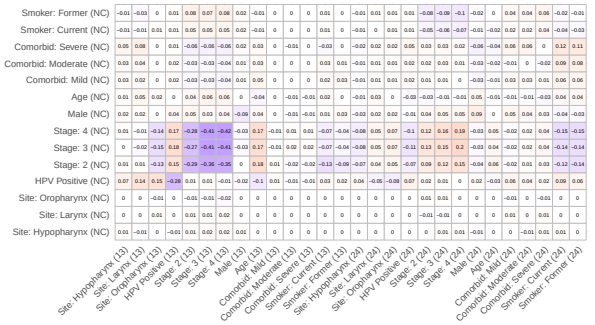
Label "NC" indicates parameters in the logistic regression for probability of being cured given non-persistent. Label "Persist" indicates parameters in the logistic regression for the probability of being persistent. Labels "13", "24", "34", and "54" indicate parameters in the $1 \to 3$, $1 \to 4/2 \to 4$, $3 \to 4$, and $5 \to 4$ transitions respectively

Figure S4: Correlations of θ across iterations of the MCMC algorithm for head and neck cancer model*

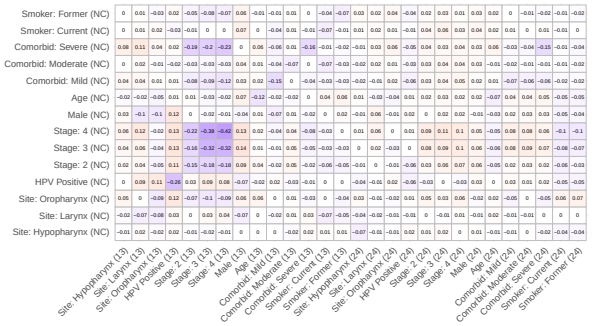
Mixture of normals:



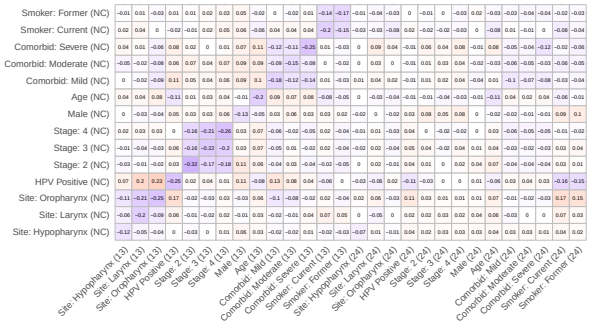
Point mass at zero:



Horseshoe:



No Shrinkage:



*This figure presents correlations for Θ across MCMC iterations. If Θ is always 0, the variable is not listed. Label "NC" indicates parameters in the logistic regression for probability of being cured given non-persistent. Label "Persist" indicates parameters in the logistic regression for the probability of being persistent. Labels "13", "24", "34", and "54" indicate parameters in the $1 \to 3$, $1 \to 4/2 \to 4$, $3 \to 4$, and $5 \to 4$ transitions respectively

S3.3 Comparison of multistate cure model and illness-death model

In this section, we compare the impact of fitting a simpler multistate model structure. In particular, we might wonder whether we need to separate out the cured and non-cured non-persistent patients. Instead, we might not believe patients can be cured and that seemingly-cured patients will eventually have a cancer recurrence or died from other causes. We compare our multistate cure model that treats cured and non-cured patients separately with an illness-death model structure, where the cured and non-cured, non-persistent patients are merged. Results are presented in **Figure S5**.

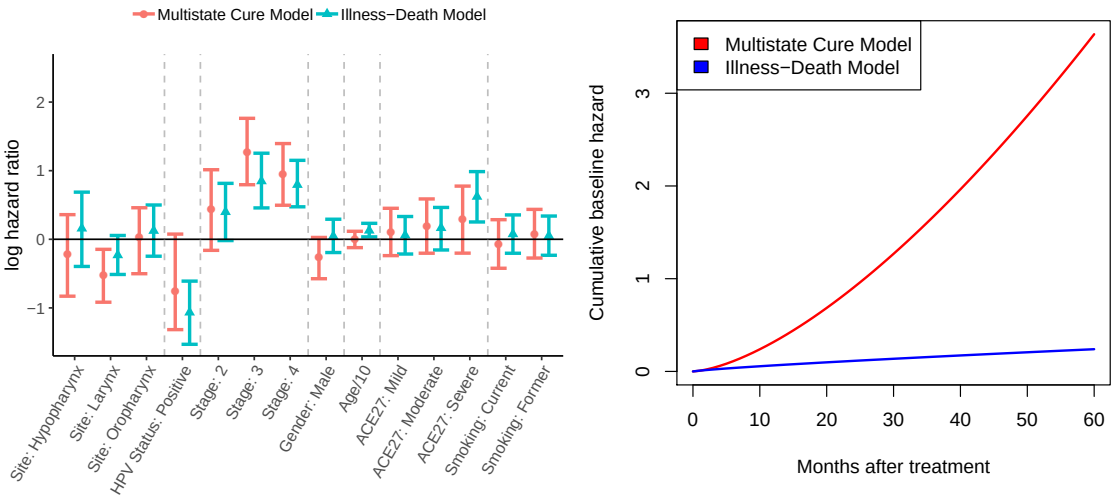
In the left panel, we provide credible intervals for the parameters in a model for the transition to recurrence in our multistate cure model and in an illness-death model (where $G = 1$ and $G = 2$ categories are merged). The resulting posterior means are not strikingly different, but there are some specific parameters for which the results do differ. For example, the multistate cure model suggests that there may be some association between gender and the rate of recurrence among non-cured patients, while the illness-death model does not provide evidence of this. In the right panel of the figure, we plot the posterior estimate for the cumulative baseline hazard function for recurrence estimated in each of the two modeling settings. The multistate cure model results in a markedly higher cumulative hazard estimate relative to the illness-death model. This is because presumably cured patients are treated the same as non-cured patients in the illness-death model, watering down the resulting estimated recurrence event rates.

Ultimately, we have strong prior beliefs that the separate cure category is important for accurately modeling head and neck cancer recurrence, and these results demonstrate that failing to separate out this category can negatively impact statistical inference.

Figure S5: Comparison of multistate cure model and illness-death model

(a) Posterior 95% credible intervals for recurrence model parameters

(b) Baseline hazards for recurrence



f

References

Bartlett, J. W., Seaman, S. R., White, I. R., and Carpenter, J. R. (2014). Multiple imputation of covariates by fully conditional specification: accomodating the substantive model. *Statistical Methods in Medical Research*, 24(4):462–487.

Beesley, L. J. and Taylor, J. M. G. (2018). EM Algorithms for Fitting Multistate Cure Models. *Biostatistics*, 20(3):416–432.

Beesley, L. J., Taylor, J. M. G., and Little, R. J. A. (2019). Sequential imputation for models with latent variables assuming latent ignorability. *Australian and New Zealand Journal of Statistics*, 61(2):213–233.

Chipman, H. (1996). Bayesian variable selection with related predictors. *The Canadian Journal of Statistics*, 24(1):17–36.

Conlon, A. S. C., Taylor, J. M. G., and Sargent, D. J. (2013). Multi-state Models for Colon Cancer Recurrence and Death with a Cured Fraction. *Statistics in Medicine*, 33(10):1750–1766.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical analysis with missing data*. John Wiley and Sons, Inc, Hoboken, NJ, 2nd edition.

Newcombe, P. J., Ali, H. R., Blows, F. M., Provenzano, E., Pharoah, P. D., Caldas, C., and Richardson, S. (2014). Weibull regression with Bayesian variable selection to identify prognostic tumour markers of breast cancer survival. *Statistical Methods in Medical Research*, 26(1):414–436.