

Prediction Model for Active Surveillance of Prostate Cancer

Rebecca Yates Coley

March 27, 2015

1 Introduction

We have developed a Bayesian, joint hierarchical model that predicts the underlying cancer state, PSA trajectory, and future biopsy results for individuals in the Johns Hopkins Active Surveillance Cohort.

2 Model

The DAG in Figure 1 illustrates the assumed relationships between variables. Blue circles are used to denote the perfect measurement of these variables, if they could be obtained, and green circles indicate observations of the variables of interest that are made with error. We describe latent and observed variables below, as well as the statistical models we use to investigate their relationships.

2.1 Latent Class Model

We first define each man’s underlying true cancer state as the prognostic grade score that would be assigned if the entire prostate were to be removed and a pathological analysis performed. Since we only observe this true cancer state on the subset of patients in active surveillance who choose surgical removal of the prostate (approximately 15%), we model true cancer state as a partially observed latent variable, which we denote η_i . This definition assumes that there is no variability in grading for full prostate specimens and that all observed grade determinations are correct. It also assumes that an individual’s prostate cancer state does not change over the time period under consideration. Under this conceptual framework, a prostate cancer is assumed to manifest its true state over time. More lethal cancers are considered fundamentally different from indolent ones from their inception and it is assumed the same histology would be observed if a full pathological analysis was performed earlier.

For our initial model, we assume a binary latent cancer state: indolent (Gleason score 6 or

prognostic grade group 1) or potentially lethal (Gleason score 7 or above or prognostic grade group 2-5), and denote these classes with $\eta = 0$ and $\eta = 1$, respectively. We do not use covariates to model latent cancer state; instead we assume a shared, underlying probability distribution for latent class membership, ρ , i.e. $P(\eta_i = 1) = \rho$ for all individuals $i = 1, \dots, n$.

2.1.1 Regressions Models for Latent Cancer Status

In the current model, we aren't using models described in this subsection.

The choice of model for latent cancer state depends on how many levels the variable takes and the assumed relationship between latent cancer state and baseline covariates. For a binary latent variable (i.e., insolvent vs. lethal), a logistic regression model is natural:

$$P(\eta_i = k | \mathbf{V}_i) = \text{logit}^{-1}(\mathbf{V}_i \boldsymbol{\nu})$$

Here, \mathbf{V}_i is a vector of baseline covariates for individual i and $\boldsymbol{\nu}$ is a parameter vector modeling the association between \mathbf{V}_i and η_i . (Since latent class is assumed to be constant over time, only baseline covariates are used.)

In the future, we may model latent cancer state as an ordinal variable corresponding to each prognostic grade group, i.e. $\eta_i \in (1, 2, 3, 4, 5)$. Modeling options are more numerous for an ordinal outcome variable, but we will likely use a cumulative link model:

$$G^{-1}(P(\eta_i \leq k | \mathbf{V}_i)) = \alpha_k - \mathbf{V}_i \boldsymbol{\zeta}_k$$

where α_k is an intercept parameter satisfying $-\infty = \alpha_0 < \alpha_1 < \dots < \alpha_{K-1} < \alpha_K = \infty$ and $\boldsymbol{\zeta}_k$ is a parameter vector. It follows that

$$P(\eta_i = k | \mathbf{V}_i) = G^{-1}(P(\eta_i \leq k | \mathbf{V}_i)) - G^{-1}(P(\eta_i \leq k-1 | \mathbf{V}_i))$$

and parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\zeta}$ must be constrained such that $G(\alpha_k - \mathbf{V}_i \boldsymbol{\zeta}_k) > G(\alpha_{k-1} - \mathbf{V}_i \boldsymbol{\zeta}_{k-1})$. Note that the probabilities of classification in the lowest and highest prognostic grade groups are defined as:

$$\begin{aligned} P(\eta_i = 1 | \mathbf{V}_i) &= G(\alpha_1 - \mathbf{V}_i \boldsymbol{\zeta}_1) \\ P(\eta_i = K | \mathbf{V}_i) &= 1 - G(\alpha_{K-1} - \mathbf{V}_i \boldsymbol{\zeta}_{K-1}) \end{aligned}$$

and that ζ_k is only defined for $k = 1, \dots, K - 1$ since $\alpha_0 = -\infty$ and $\alpha_K = \infty$.

We may choose to simplify this more general model and use the proportional odds model, which is a cumulative link model with logit link and ζ_k equal for all $k = 1, \dots, K - 1$.

2.2 Longitudinal PSA Model

An individual's true prostate cancer state gives rise to his PSA levels (as indicated by the arrow from True Prostate Cancer Status to True PSA in Figure 1). PSA is independently influenced by additional factors including prostate volume and other sources of inflammation. We cannot observe an individual's true PSA; we only observe imprecise measurements of PSA at some points in time.

We use a mixed effects model to model the anticipated trajectory of an individual's PSA as they age. In this model, mean effects for predictors are allowed to vary across groups defined by latent class. Specifically, we expect the PSA of those with more aggressive cancer to have a higher level and steeper slope than those with indolent cancer (although we do not impose any restrictions on the model to enforce this expectation). Random effects allow for individuals within each group to have intercepts and slopes that differ from the latent class means. We specify the following mixed effects model:

$$[Y_{im} | \eta_i = k, \mathbf{X}_{im}, \mathbf{Z}_{im}] = \mathbf{X}_{im}\boldsymbol{\beta} + \mathbf{Z}_{im}\mathbf{b}_i + \epsilon_{im}$$

where Y_{im} is the observed (log-transformed) PSA and \mathbf{X}_{im} and \mathbf{Z}_{im} are vectors of covariates for individual i 's m th PSA measurement, $\boldsymbol{\beta}$ is a parameter vector for fixed effects, \mathbf{b}_i is the subject-specific vector of random effects, and residuals ϵ_{im} are assumed to follow a normal distribution with mean 0 and variance σ^2 . As a result, the vector of observed PSAs for individual i , $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iM_i})$, follows a multivariate normal distribution. (Note that the residual variance is assumed to be the same across latent classes.)

We specify a mixed effects model that allows for correlation between random effects, similar to the model proposed by Gelman and Hill (2007). In this specification, covariates do not appear in both \mathbf{X}_{im} and \mathbf{Z}_{im} and, consequently, random effects do not have a mean of zero but are instead centered at the mean effects for each latent class k , denoted here as $\boldsymbol{\mu}_k$:

$$\mathbf{b}_i | \eta_i = k \sim MVN(\boldsymbol{\mu}_k, \Sigma)$$

Note that, while the mean differs, the covariance matrix for the random effects, Σ , is assumed

to be the same across latent classes. This assumption aids model identifiability and convergence by reducing the number of parameters to be estimated and was made only after examining the covariance estimates obtained by fitting the mixed effects models in the subset of patients with latent class observed and confirming its plausibility.

We use age as a predictor, rather than time since diagnosis, since we recognize that diagnostic time is somewhat arbitrary for those with very low risk (non-symptomatic) prostate cancer and we expect PSA values to be more similar across individuals of the same age rather than those with the same time since diagnosis. We specify a natural cubic spline model for the relationship between age and log-PSA and allow random effects for each of the three basis functions. Following the alternative basis for the natural cubic spline given in Chapter 11 of Wakefield (2013), given below, these random effects correspond to a random intercept, random slope, and random curvature for age, z :

$$h_1(z) = 1, \quad h_2(z) = z, \quad h_3(z) = \frac{(z - \xi_1)_+^3 - (z - \xi_3)_+^3}{\xi_3 - \xi_1} - \frac{(z - \xi_2)_+^3 - (z - \xi_3)_+^3}{\xi_3 - \xi_2}$$

where (ξ_1, ξ_2, ξ_3) are knots set equal to the interior quartiles of the standardized age distribution. (Standardization is performed by subtracting the mean age of all individuals at all PSA observations and dividing by the standard deviation.) Thus, the row of design matrix \mathbf{Z} corresponding to the i th individual's m th PSA observation is $\mathbf{Z}_{im} = [1, age_{im}, h_3(age_{im})]$ where age_{im} is the standardized age of individual i at the time of his m th observation and $h_3(\cdot)$ is the natural cubic spline basis defined above.

We also include fixed effects for prostate volume and the interaction of prostate volume and latent class in the model, which allow the association between prostate volume and PSA to differ between classes. The row of design matrix \mathbf{X} corresponding to the i th individual's m th PSA observation is then $\mathbf{X}_{im} = [PV_i, PV_i \times \mathbf{1}_{[\eta_i=1]}]$, where PV is the prostate volume for individual i (averaged over all available measurements).

2.2.1 Model for PSA Outliers

One extension we may want to consider for the final model would be to assume that measurement errors follow a mixture of normal distributions (instead of a single normal distribution). This would model would better accommodate unexpectedly high PSA values due to prostatitis as well as unexpectedly small values due to PSA-lowering medications like finasteride. This could be

written as:

$$\epsilon_{ij} \sim \begin{cases} N(0, \sigma^2) & \text{with probability } p_{im} \\ N(0, c\sigma^2) & \text{with probability } 1 - p_{im} \end{cases}$$

where c is some large scalar (e.g., 10), and p_{im} has a beta prior with mean close to 1. Note that the mixing is specific to single observations within an individual, allowing for deviations from an individual’s expected trajectory consistent with episodic inflammation or medication. This model is not intended to account for artificially low PSA across all of an individual’s observations, as may happen if an individual was on PSA-lowering medication throughout the observation period. (Ideally, we will be able to get some data on medication history, but this data is not anticipated to be complete as it was not uniformly collected on active surveillance patients.)

2.3 Logistic Biopsy Model

An individual’s true prostate cancer state also influences, but does not determine, any biopsy results, as indicated in Figure 1. Biopsies are essentially samples of cells from the population of all prostate cancer cells and sampling targets areas of the prostate where tumors are expected to be located, as identified by pre-biopsy ultrasound or MRI. Moreover, grade determination from biopsy samples is somewhat subjective, which contributes additional error to biopsy results, and overgrading is possible.

While many tumor characteristics are quantified by biopsy results, including number of positive cores sampled and the maximum percent involvement of any positive cores, we focus on the outcome of grade reclassification (that is, biopsy determination of Gleason grade 7 or above) as that is the clinical endpoint of interest. Active surveillance with curative intent regimes are designed to monitor patients until reclassification, at which time curative intervention is recommended. Accordingly, patients in our cohort are censored upon reclassification.

While active surveillance protocol recommends repeating biopsies once per year, the frequency and timing of biopsies varies. We discretize biopsy times into six month-intervals where time is measured since diagnosis, e.g. 0-6 months, 6 months to 1 year, etc. If individual i has at least one biopsy during an interval j , it is indicated by $S_{ij} = 1$. Then, R_{ij} is a binary outcome denoting reclassification (at any biopsy) during this interval. S_{ij} is only defined for $j = 1, \dots, J_i$, where J_i is the last interval with a biopsy observed for individual i , and R_{ij} is only defined for j such that $S_{ij} = 1$. Only intervals with a biopsy contribute to the likelihood. (We discuss informative missingness and censoring below.)

We use a logistic regression model to predict the probability of grade reclassification conditional on true cancer state and time-varying patient characteristics. We write this model as:

$$P(R_{ij}|\eta_i = k, \mathbf{W}_{ij}) = \text{logit}^{-1}(\mathbf{W}_{ij}\boldsymbol{\gamma})$$

where \mathbf{W}_{ij} is a covariate vector at time j for individual i and $\boldsymbol{\gamma}$ is a parameter vector modeling the log-odds of reclassification conditional on covariates. In our model, the row vector of design matrix \mathbf{W} corresponding to the i th individual's j th biopsy after diagnosis is $\mathbf{W}_{ij} = [1, age_{ij}, time_{ij}, \mathbf{1}_{[\eta_i=1]}]$ where age_{ij} is the (standardized) age of individual i at the time of his j th biopsy and $time_{ij}$ is the calendar time since his initial diagnosis. Here, we assume an additive effect of underlying cancer state on the log-odds scale.

One feature of this model that is of note: biopsy results are independent of PSA measurements conditional on latent cancer status. All of an individual's biopsy results are also independent conditional on latent state.

2.3.1 Joint Model for Other Biopsy Results

Originally, we had intended to construct a joint model for the many determinations made from each biopsy, including the number of positive cores and the maximum percent cancer involvement in any positive cores. These biopsy results are associated with reclassification (citations) and presumed to be predictive of the true underlying cancer state. However, among the subset of our cohort for whom true underlying cancer state is observed, these particular biopsy results are actually negatively, though not significantly, associated with lethal prostate cancer. Thus, inclusion of models for the probability of these biopsy results conditional on latent cancer state in the broader model did not improve predictive performance and, instead, frequently led to poor mixing and identifiability concerns.

2.3.2 Missingness and Censoring in Biopsy Observations

As mentioned above, individuals only contribute biopsy information for the intervals at which they had a biopsy. While we do not expect biopsies to occur at random- on the contrary, we expect more frequent biopsies for those with higher PSAs, past biopsy results indicating higher risk, and those with greater "health-seeking behavior"- nor, do we assume that those who continue engagement with active surveillance are similar to those who drop-out of the cohort, we assume that intermittent missingness and censoring are independent of latent cancer state

conditional on observed variables. More to the point, we assume that the only factors influencing patients' (and clinicians') biopsy decisions are observed PSA and biopsy results. Furthermore, since this is a very low risk cohort and great effort is made to follow patients' medical records even after they stop regular follow-up, we assume that no individual is censored because of prostate cancer status, including death from prostate cancer. As a result, it is not necessary to include censoring and observation probabilities in the likelihood since they provide no information about an individual's latent state, future PSA, or biopsy results and we are not interested in understanding the association between clinical history and choosing to have a biopsy. We are also not interested in estimating causal effects of any variable on the probability of reclassification, which would require weighting observed biopsy results on the probability of observation.

In the future, we may decide to include a latent class for health-seeking behavior in the model. Presumably, individuals who are more inclined to get regular biopsies may differ from those that do not in a way that also influences their underlying cancer. For example, individuals who exercise regularly and eat well may be more likely to get biopsies and less likely to have more aggressive cancer. There are numerous potential complications with this approach. First, it would be hard to formulate a consistent definition of the association between health-seeking behavior and frequency of biopsies because men are encouraged to get annual biopsies during the first few years of active surveillance but many clinicians do not recommend annual biopsies after several years of follow-up. Thus, infrequent biopsies in those who have been in the cohort longer may actually be evidence in favor of health-seeking behavior, here, following the doctor's advice. Secondly, the causal diagram including health-seeking behavior would have to be carefully considered, as health-seeking behavior may both affect latent cancer state, but it is also possible that observed PSA and biopsy results trigger change in an individual's behavior. Likewise, health-seeking behavior may change throughout follow-up as new PSA and biopsy results are observed.

2.4 Likelihood

The likelihood for the initial model described above can be written as:

$$\begin{aligned}
L(\rho, \beta, \gamma, \underline{\mu}, \Sigma \mid (\eta_i, \mathbf{b}_i, \mathbf{Y}_i, \mathbf{S}_i, \mathbf{R}_i, \underline{\mathbf{X}}_i, \underline{\mathbf{Z}}_i, \underline{\mathbf{W}}_i), i = 1, \dots, n) \\
= \prod_{i=1}^n \rho^{\eta_i} (1 - \rho)^{1-\eta_i} f(\mathbf{Y}_i \mid \eta_i, \underline{\mathbf{X}}_i, \underline{\mathbf{Z}}_i, \mathbf{b}_i, \beta, \sigma^2) g(\mathbf{b}_i \mid \underline{\mu}_{\eta_i}, \Sigma) \\
\prod_{j=1}^{J_i} (P(R_{ij} = 1 \mid \eta_i, \mathbf{W}_{ij}, \gamma)^{R_{ij}} P(R_{ij} = 0 \mid \eta_i, \mathbf{W}_{ij}, \gamma)^{1-R_{ij}})^{S_{ij}}
\end{aligned}$$

where $\underline{\mathbf{X}}_i$, $\underline{\mathbf{Z}}_i$, and $\underline{\mathbf{W}}_i$ are covariate matrices with row entries \mathbf{X}_{im} , \mathbf{Z}_{im} , and \mathbf{W}_{ij} , respectively, for $m = 1, \dots, M_i$ and $j = 1, \dots, J_i$, $\underline{\boldsymbol{\mu}}$ is the parameter vector of mean random effects $[\mu_1, \mu_2]$, and f and g are multivariate normal densities for the vector of log-transformed PSAs, \mathbf{Y}_i , and random effects \mathbf{b}_i , respectively, for individual i , each with mean and covariance as defined in Section 2.2.

2.5 Priors

We specify the following prior distributions for model parameters.

First, we specify a flat beta prior on the probability of latent class membership $\eta = 1$, i.e., having lethal cancer:

- $\pi(\rho) = \text{Beta}(1, 1)$

Next, for the mixed effects model, we assume flat normal priors on the fixed effects and mean random effects for each latent class:

- $\pi(\boldsymbol{\beta}) = \text{MVN}(\mathbf{0}, 10^2 \times \mathbf{I})$
- $\pi(\boldsymbol{\mu}_k) = \text{MVN}(\mathbf{0}, 10^2 \times \mathbf{I}), k = 0, 1$

where \mathbf{I} is an identity matrix of dimension equal to the length of parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\mu}_k$, respectively. We then specify an inverse Wishart prior on the covariance matrix, Σ , for the random effects, which is a conjugate prior for and allows random effects to be correlated:

- $\pi(\Sigma) = \text{Wishart}^{-1}(\mathbf{I}_D, D + 1)$

where D is the length of parameter vector $\boldsymbol{\mu}$. (While a scaled inverse Wishart is preferable for the prior on the covariance matrix for random effects, we found that estimating a scale parameter could hinder model mixing and setting the scale equal to one did not change parameter estimates.) Lastly, we set a uniform prior on the residual standard deviation:

- $\pi(\sigma) = \text{Uniform}(0, 1)$.

Finally, for the logistic regression model of risk of reclassification we specify a diffuse normal prior on all components of parameter vector $\boldsymbol{\gamma}$:

- $\pi(\boldsymbol{\gamma}) = \text{MVN}(\mathbf{0}, 10^2 \mathbf{I})$

2.6 Posterior Estimation Procedure

After specifying priors for all model parameters, we define the joint posterior density of the parameters and latent variables as the product of each parameter's prior density and the model likelihood:

$$p(\rho, \beta, \gamma, \underline{\mu}, \Sigma, \underline{\mathbf{b}}, \boldsymbol{\eta} | (\mathbf{Y}_i, \mathbf{S}_i, \mathbf{R}_i, \underline{\mathbf{X}}_i, \underline{\mathbf{Z}}_i, \underline{\mathbf{W}}_i), i = 1, \dots, n; \boldsymbol{\Theta}) \\ \propto L(\rho, \beta, \gamma, \underline{\mu}, \Sigma, |(\eta_i, \mathbf{b}_i, \mathbf{Y}_i, \mathbf{S}_i, \mathbf{R}_i, \underline{\mathbf{X}}_i, \underline{\mathbf{Z}}_i, \underline{\mathbf{W}}_i), i = 1, \dots, n) \times \boldsymbol{\pi}(\rho, \beta, \gamma, \underline{\mathbf{b}} | \boldsymbol{\Theta})$$

where $\boldsymbol{\pi}(\cdot | \boldsymbol{\Theta})$ denotes the joint prior density for all model parameters with hyper priors $\boldsymbol{\Theta}$, $\underline{\mathbf{b}}$ is a matrix containing all subject-specific random effect vectors, and $\boldsymbol{\eta}$ is the vector $[\eta_1, \dots, \eta_n]$.

Posterior sampling is performed in RJAGS.

2.7 Cross-Validation

2.8 Sequential Bayes Updating

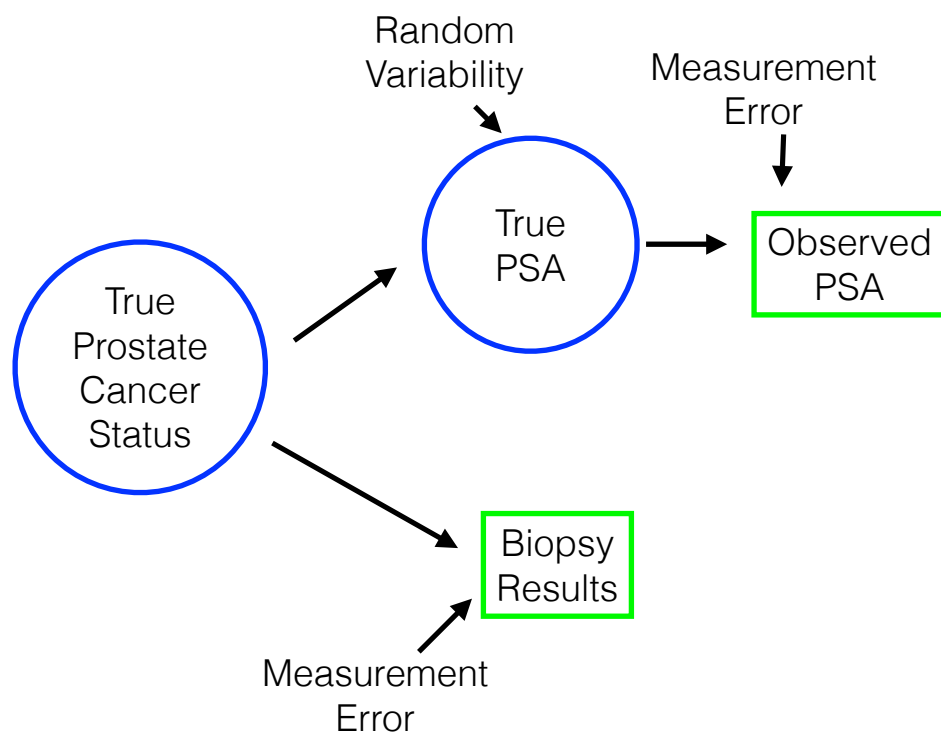


Figure 1: Directed acyclic graph