

Prediction Model for Active Surveillance of Prostate Cancer

R Yates Coley, Scott Zeger, Aaron Fisher, Ken Pienta, H Ballentine Carter

(*can discuss inclusion, order*)

July 6, 2015

1 Introduction

Prostate cancer is the most commonly diagnosed non-skin cancer in men in the United States [19]. Upon diagnosis, early curative treatment with surgery, radiation, or androgen deprivation therapy is common [8, 38]. Prostate cancer treatments can be physically, emotionally, and financially taxing for patients. In particular, one-month mortality after surgery is as high as 0.5%, and at least 20-30% of men experience urinary incontinence and erectile dysfunction after surgery or radiotherapy [6, 7].

Active surveillance with curative intent offers an alternative to early treatment for individuals with lower risk disease detected [5, 21, 22, 33, 34, 36, 37]. Though active surveillance regimes vary, the approach generally entails regular biopsies (e.g., annually) with curative intervention recommended upon detection of higher risk histological features, as determined by the Gleason grading system. Biopsies with a Gleason score of 6 indicate low risk disease while a Gleason score of 7 or above is considered *reclassification* to a higher risk. Prostate-specific antigen (PSA) is also routinely measured and may be used to recommend biopsies.

The success of active surveillance programs depend on clinicians' ability to identify tumors with metastatic potential with sufficient time for curative intervention to be effective. Biopsies used to characterize tumors, however, are only informative about the sampled tissue and, moreover, have imperfect sensitivity and specificity[12]. As a result, biopsy results do not indicate the true state of an individual's cancer, but rather are measurements made with error. Existing clinical support tools that predict biopsy outcomes for active surveillance patients, including, most recently, Ankerst et al. 2015, contribute valuable information to guide decisions about biopsy timing and frequency but are insufficient to address patients' primary concerns. Instead, patients and clinicians would prefer prediction of the pathological make-up of the entire prostate, including any presence of higher risk features, in order to guide decision-making.

With this motivation in mind, we have developed a Bayesian hierarchical model that enables prediction of an individual's underlying disease status via joint modeling of repeated PSA measurements and biopsies. Specifically, we predict a binary cancer state- *indolent* or *aggressive*-

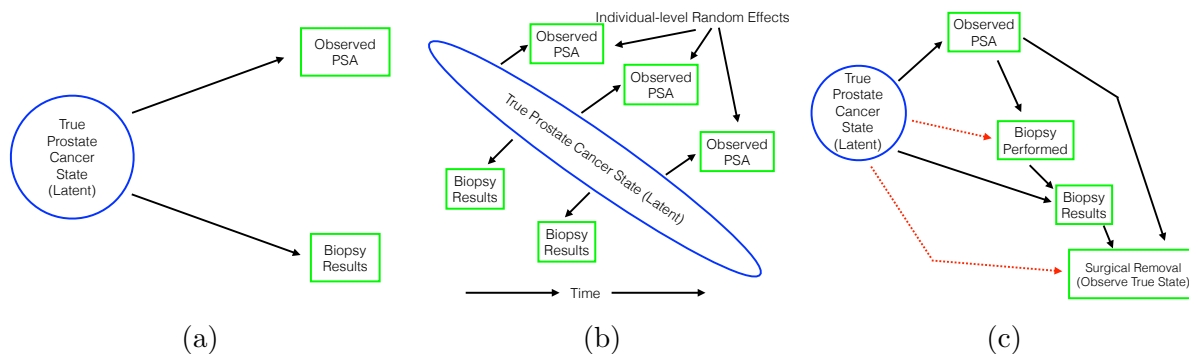


Figure 1: DAGs describing the relationships between latent class (circled) and observed measurements (squared) (a) at a single point in time, (b) over time, and (c) in the presence of an informative observation process.

with the latter defined as a true (i.e., measured without error) Gleason score of 7 or higher. Predictions are informed by a subset of patients for whom the true state is observed, that is, patients who underwent elective prostatectomy (either before or after reclassification) and for whom results of the pathological analysis of the removed tissue is available. In this sense, cancer state operates as a partially-latent class in the proposed model [40].

As reflected through a joint modeling framework, an individual’s cancer state influences both the level and trajectory of PSA measurements as well as the outcome of repeated biopsies. These relationships are illustrated by the directed acyclic graph (DAG) in Figure 1(a) for a single point in time and in Figure 1(b) for multiple time points. In the model we are proposing, PSA measurements follow a mixed effects model with mean effects varying across latent classes [23]. Then, biopsy outcomes, a binary indicator of reclassification at each year of follow-up, are modeled with a pooled logistic regression model under the assumption that biopsy results are independent conditional on cancer state and covariates (age, time since diagnosis, etc.) [9, 10]. As codified in the Figure 1(b) DAG, PSA and biopsy results are also assumed to be conditionally independent given latent class.

Our proposed model is similar to Lin et al. in 2002, which first proposed a latent class approach to joint analysis of longitudinal PSA and time-to-diagnosis of prostate cancer, extending developments in joint models by Schluchter (1992), DeGruttola and Tu (1994), and Henderson, Diggle, and Dobson (2000). This joint latent class model (JLCM) has since been applied in many settings, including an extension of the method by Proust-Lima and Taylor (2009) to develop a dynamic prognostic tool for prostate cancer recurrence after radiation therapy. Use of JLCM is motivated by interest in modeling differential risk and disease progression across a population and classifying individuals with similar outcomes, but, unlike our approach, typically does not

involve a priori specification of classes of interest.

In contrast to the model we are proposing, the JLCM model of Lin et al. assumes that prostate cancer diagnoses are both correct – ignoring measurement error in diagnostic biopsies – and a reflection of time of onset. On the contrary, diagnosis of low risk disease typically follows a physician’s recommendation to perform a biopsy rather than presentation of clinical symptoms. To this point, the conditional independence assumption is suspect as observation of high PSA typically triggers biopsy recommendations. The same concern is present for predicting cancer state in an active surveillance population; biopsies are not always performed annually and, as a patient continues in the program, PSA kinetics are increasingly relied on for biopsy recommendations. The proposed model addresses this limitation by modeling reclassification with a pooled logistic regression model [9, 10] and modifying the conditional independence assumption to include conditioning on the presence of a biopsy. Moreover, under the assumption that the choice to get a biopsy depends on observed factors (PSA, patient characteristics and past biopsy results), i.e., that biopsy results are missing at random (MAR), unbiased identification of latent classes in a likelihood-based Bayesian framework does not require specification of a probability model for biopsy occurrence [26]. A similar line of reasoning can be followed for the decision to undergo surgical removal of the prostate, resulting in observation of the true cancer state. Under the MAR assumption, it is unnecessary to specify a model for the probability of electing surgery.

We also offer a novel extension to the proposed model that allows for observation of biopsy results and the true cancer state to be missing not at random (MNAR). Consider the dotted lines in the DAG in Figure 1(c)– if true cancer state is associated with the choice to perform a biopsy or undergo surgery after conditioning on observed PSA and biopsy results, then informative missingness is present and predictions will be biased [39]. In response, we propose including cancer state as a predictor in regression models for the probability having a biopsy and having surgery at each annual interval. In this way, the dependence between observing the outcome and its value is accommodated. This approach is similar to the latent class dropout model, proposed by Roy (2003, 2007), a type of shared parameter model [14] with discrete random effect. Unlike Roy’s model for intermittent missingness which models latent class conditional on the observation process, we follow the model formulation outlined in Albert and Follmann (2009), specifying distributions for the outcome and the observation process conditional on cancer state.

Finally, in order to support decision-making in a clinical setting, we have developed an importance sampling algorithm to enable real-time updating of model predictions[16].

This paper is organized as follows. In Section 2, we describe data from an active surveillance cohort at Johns Hopkins. A joint hierarchical model for latent class prediction is outlined in

Section 3, including a description of Bayesian estimation procedures. In Section 4, we propose and implement additions to the model in order to account for likely informativeness of biopsy and cancer state observations. Section 5 presents an application of the model to the active surveillance cohort. Then, importance sampling methods to obtain quick predictions update are detailed and demonstrated in Section 6. We close with a discussion.

2 Active Surveillance Program for Prostate Cancer at Johns Hopkins

As of Month Year, the Johns Hopkins Active Surveillance cohort has enrolled XX prostate cancer patients since 1995 (CITE). *Still waiting for Mufaddal to send me this information so that we are consistent across papers.* In this study, patients with very-low-risk or low-risk prostate cancer diagnoses (according to criteria outlined in Epstein et al. (1994), including biopsy Gleason score of 6) who elect to delay curative intervention in favor of active surveillance are followed prospectively. Characteristics of diagnostic biopsy and results of all prior PSA tests are collected at enrollment. As part of the surveillance regimen, PSA tests are typically performed every six months, and biopsies are performed annually, though these intervals vary; the results of all tests are recorded. Curative intervention is recommended upon biopsy reclassification, that is, when the Gleason score assigned to a biopsy exceeds 6. Some patients also choose to undergo curative intervention prior to reclassification. For patients who elect surgical removal of the prostate (whether prior to or after reclassification), Gleason score determination based on full pathologic analysis of the prostate specimen is also recorded.

3 Joint Hierarchical Latent Class Model

We have developed a Bayesian, joint hierarchical model to predict the underlying cancer state of patients enrolled in active surveillance. Predictions are made by incorporating information from repeated PSA and biopsy measurements for all individuals, as well as true cancer state information observed in a subset of the cohort. In this section, we first present notation before specifying models for each component of the likelihood (partially-observed latent class, PSA, and biopsy data) and defining the full likelihood. Next, we complete Bayesian specification of the model by defining the joint posterior distribution and discussing prior distributions for model parameters.

3.1 Notation

We first define each individual's underlying true cancer state, η_i , as either indolent, $\eta_i = 0$, or aggressive, $\eta_i = 1$, for all cohort members $i = 1, \dots, n$. An individual's cancer state influences his observed PSA values, which we denote with M_i -length vector \mathbf{Y}_i , as well as his biopsy results. Biopsy data is categorized into discrete, annual time intervals with (B_{ij}, R_{ij}) denoting binary variables that indicate whether a biopsy was performed and reclassification observed, respectively, for individual i in year j . Finally, S_{ij} is an indicator of surgery for individual i in year j .

3.2 True Cancer State

We first define each individual's true cancer state, η_i , using a dichotomized summary of the Gleason score that would be assigned if his entire prostate were to be removed and pathologic analysis performed:

$$\eta_i = \begin{cases} 0 & \text{indolent; Gleason score for individual } i = 6 \\ 1 & \text{aggressive; Gleason score for individual } i \geq 7 \end{cases}$$

Since we observe this true cancer state on a subset of patients in active surveillance who choose surgical removal of the prostate, η is best described, in the context of this model, as a partially observed latent variable. Cancer state is then modeled as a Bernoulli random variable, $\eta_i \sim \text{Bern}(\rho)$, where we assume a shared underlying probability of aggressive cancer, ρ , as our data do not include any baseline predictors of cancer state (other than PSA and biopsy measurements), such as genetic markers or family history.

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 cancer categorization does not change over the time period under surveillance, i.e., that upgrading is due to sampling error at diagnosis rather than tumor dedifferentiation since diagnosis. 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 inception and it is assumed that, while tumor volume would likely differ, the same histologic category (i.e. Gleason = 6 vs. Gleason ≥ 7) would be observed if a full pathologic analysis was performed earlier.

3.3 Longitudinal PSA Measurements

Next, we specify a regression model for PSA, a time-varying biomarker dependent on an individual's true cancer state. PSA is also expected to increase with age and 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.

To reflect these characteristics, we use a mixed effects model to model the anticipated trajectory of an individual's PSA as they age [23]. In this model, mean effects for predictors are allowed to vary across groups defined by partially 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, and \mathbf{b}_i is the patient-specific vector of random effects. Following the specification of a Bayesian mixed effects model presented by Gelman and Hill (2006), unscaled random effects are centered at the mean effects for each latent class k , $\boldsymbol{\mu}_k$:

$$[\check{\mathbf{b}}_i | \eta_i = k] \sim MVN(\boldsymbol{\mu}_k, \Sigma_k), \quad k = 0, 1$$

where Σ_k is a covariance matrix that allows for correlation between random effects. Random effects are then scaled with parameter vector $\boldsymbol{\xi}$: $\mathbf{b}_i = \text{diag}(\check{\mathbf{b}}_i \boldsymbol{\xi}^T)$.

Lastly, 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.

3.4 Repeated Biopsy Outcomes

Finally, we consider the likelihood contribution for biopsy results. An individual's true prostate cancer state also influences, but does not determine, the findings of biopsies performed as a part

of the active surveillance regimen. While biopsies target areas of the prostate where tumors are expected to be located, as identified by pre-biopsy ultrasound or MRI, biopsied tissue constitutes only a sample of the prostate. (For example, the volume of tissue collected in a typical 12-core biopsy only includes, at most, 0.5% of the total prostate volume,) Moreover, sectioning of biopsied tissues for evaluation can introduce further uncertainty into histologic assessment, and overgrading is possible. As a result, biopsy-determined Gleason scores have imperfect sensitivity and specificity for predicting true cancer state.

Since the frequency of biopsies are typically dictated by AS protocol, we categorize biopsy times and results into discrete time intervals. If individual i has a biopsy during interval j , it is indicated by $B_{ij} = 1$. Then, we define R_{ij} as a binary outcome denoting reclassification during this interval. If a biopsy is performed, we use a logistic regression model to predict the probability of grade reclassification within that interval conditional on true cancer state and time-varying patient characteristics:

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

where $\mathbf{V}_{ij}(k)$, is a matrix of time-varying predictors including cancer state $\eta_i = k$ and $\boldsymbol{\gamma}$ is a parameter vector. If a biopsy is not performed, $B_{ij} = 0$, then reclassification cannot be observed and R_{ij} is undefined.

Using this logistic regression model within each interval where a biopsy is performed, we model time-until-reclassification using a pooled logistic regression model, that is, we take the product of the probability of the observed result over all biopsies performed [10]. Intervals without biopsies do not contribute to this product, and B_{ij} is only defined for intervals up to the time of reclassification or censoring, $j = 1, \dots, J_i$.

3.5 Likelihood

The likelihood for the proposed model can be written as:

$$\begin{aligned} L & \left(\rho, \boldsymbol{\beta}, (\boldsymbol{\mu}_k, \Sigma_k), k = 0, 1, \boldsymbol{\gamma}, (\eta_i, \mathbf{b}_i) \mid (\mathbf{Y}_i, \underline{\mathbf{X}}_i, \underline{\mathbf{Z}}_i, (B_{ij}, R_{ij}, \mathbf{V}_{ij}), j = 1, \dots, J_i), i = 1, \dots, n \right) \\ &= \prod_{i=1}^n \rho^{\eta_i} (1 - \rho)^{1-\eta_i} f(\mathbf{Y}_i | \eta_i, \underline{\mathbf{X}}_i, \underline{\mathbf{Z}}_i, \mathbf{b}_i, \boldsymbol{\beta}, \sigma^2) g(\mathbf{b}_i | \boldsymbol{\mu}_{\eta_i}, \Sigma_{\eta_i}) \\ & \quad \prod_{j=1}^{J_i} (P(R_{ij} = 1 | \eta_i, \mathbf{V}_{ij}, \boldsymbol{\gamma})^{R_{ij}} P(R_{ij} = 0 | \eta_i, \mathbf{V}_{ij}, \boldsymbol{\gamma})^{1-R_{ij}})^{B_{ij}} \end{aligned} \quad (2)$$

where 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 above, given covariance matrices $\underline{\mathbf{X}}_i = [X_{i1}, \dots, X_{iM_i}]$ and $\underline{\mathbf{Z}}_i = [Z_{i1}, \dots, Z_{iM_i}]$.

We note that dependence structures codified in the Figure 1 (a) and (b) DAGs are reflected in this likelihood. First, PSA measurements, \mathbf{Y}_i , are independent of biopsy results conditional on true cancer state. Second, biopsy results are also conditionally independent given true cancer state.

3.6 Bayesian Estimation

We use a Bayesian approach to estimate the proposed joint latent class model. Standard prior distributions are available for model parameters including, for example a beta prior on the probability of having aggressive cancer ($\eta = 1$) and normal priors on logistic regression model coefficients. Prior specification for the mixed effects model can follow the Bayesian estimation procedure outlined in Gelman and Hill (2006) for a linear mixed effects model with correlated random effects using a scaled inverse Wishart prior.

After specifying priors for all model parameters, the joint posterior distribution of the parameters and partially-latent classes is given by:

$$\begin{aligned} & p\left(\rho, \boldsymbol{\beta}, (\boldsymbol{\mu}_k, \Sigma_k), \boldsymbol{\gamma}, (\eta_i, \mathbf{b}_i) \mid (\mathbf{Y}_i, \underline{\mathbf{X}}_i, \underline{\mathbf{Z}}_i, (B_{ij}, R_{ij}, \mathbf{V}_{ij})); \boldsymbol{\Theta}\right) \\ & \propto L\left(\rho, \boldsymbol{\beta}, (\boldsymbol{\mu}_k, \Sigma_k), \boldsymbol{\gamma}, (\eta_i, \mathbf{b}_i) \mid (\mathbf{Y}_i, \underline{\mathbf{X}}_i, \underline{\mathbf{Z}}_i, (B_{ij}, R_{ij}, \mathbf{V}_{ij}))\right) \times \boldsymbol{\pi}(\rho, \boldsymbol{\beta}, (\boldsymbol{\mu}_k, \Sigma_k), \boldsymbol{\gamma} \mid \boldsymbol{\Theta}) \end{aligned}$$

where $\boldsymbol{\pi}(\cdot \mid \boldsymbol{\Theta})$ denotes the joint prior density for model parameters with hyper priors $\boldsymbol{\Theta}$ and indexing $k = 0, 1$, $j = 1, \dots, J_i$, and $i = 1, \dots, n$.

4 Informative Observation Process

In this section, we propose an addition to the joint latent class model presented above in order to account for informative missingness.

The model outlined in Section 3 assumes that unobserved biopsy results, i.e. R_{ij} for intervals where $B_{ij} = 0$, and true cancer state are either missing completely at random (MCAR) or missing at random (MAR) [26]. The likelihood given in (2) is consistent with both scenarios: if observations are MCAR, no modeling of the missingness mechanism is necessary and, if observations

are MAR, the missingness mechanism is known and can be explicitly modeled. In the latter case, terms modeling the probability of observing true cancer state or biopsy results are not needed in the likelihood for our Bayesian approach because they do not contain parameters needed for prediction and thus fall out when parsing the joint posterior into full conditional posteriors for parameters of interest.

The likelihood in (2) is not consistent, however, with the possibility that the true cancer state influences whether an individual has a biopsy or surgery performed (the latter resulting in observation of the cancer state) or, said another way, that the observation process itself is informative of true cancer state. This scenario is an example of observations missing not at random (MNAR) and is depicted in Figure 1(c). If missingness is indeed informative, the arrows pointing from cancer state to biopsy performance and surgical removal of the prostate and present and the model given in Section 3 will result in biased prediction of the true cancer state.

To allow for the possibility of an informative observation process (IOP), we propose explicitly modeling the presence of biopsies and surgery as dependent on the true cancer state. This approach could be described as a shared parameter model with a discrete random effect where the prediction and interpretation of the random effect is of interest [14].

First, we specify a model for the probability that patient i receives a biopsy in each time interval j conditional on latent class, η_i , and time-varying covariates, $\mathbf{U}_{ij}(k)$:

$$P(B_{ij} = 1 | \eta_i = k, \mathbf{U}_{ij}, \boldsymbol{\nu}) = \text{logit}^{-1}(\mathbf{U}_{ij}(k)\boldsymbol{\nu}) \quad (3)$$

where $\boldsymbol{\nu}$ is a parameter vector. Individual i contributes $P(B_{ij} = 1 | \eta_i = k, \mathbf{U}_{ij}, \boldsymbol{\nu})^{B_{ij}} P(B_{ij} = 0 | \eta_i = k, \mathbf{U}_{ij}, \boldsymbol{\nu})^{1-B_{ij}}$ to the likelihood for every time interval up until reclassification or censoring.

Next, we define S_{ij} , an indicator of surgical removal of the prostate for individual i during time j and model the time-to-surgery using a pooled logistic regression model with the probability of surgery in each time interval defined as:

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

where \mathbf{W}_{ij} is a matrix of time-varying predictors including cancer state $\eta_i = k$ and $\boldsymbol{\omega}$ is a parameter vector. The possibility of surgery extends beyond the time of reclassification as the decision to undergo prostatectomy is often informed by biopsy results. Accordingly, S_{ij} is defined for all time intervals for which a patient is under observation, which may continue past reclassification. We denote the final interval under observation as J_{S_i} for individual i .

For both IOP models, time-varying covariates (U_{ij} and W_{ij}) may include information from past intervals, including the number and results of past biopsies, which would likely influence future biopsy and surgery decisions.

Finally, the likelihood and joint posterior for the proposed model can be updated to include the IOP models:

$$\begin{aligned}
& L\left(\rho, \beta, (\boldsymbol{\mu}_k, \Sigma_k), \boldsymbol{\nu}, \gamma, \boldsymbol{\omega}, (\eta_i, \mathbf{b}_i) \mid (\mathbf{Y}_i, \underline{\mathbf{X}}_i, \underline{\mathbf{Z}}_i, (B_{ij}, \mathbf{U}_{ij}), (R_{ij}, \mathbf{V}_{ij}), (S_{ij}, \mathbf{W}_{ij})), i = 1, \dots, n\right) \\
&= \prod_{i=1}^n \rho^{\eta_i} (1 - \rho)^{1-\eta_i} f(\mathbf{Y}_i \mid \eta_i, \underline{\mathbf{X}}_i, \underline{\mathbf{Z}}_i, \mathbf{b}_i, \beta, \sigma^2) g(\mathbf{b}_i \mid \boldsymbol{\mu}_{\eta_i}, \Sigma_{\eta_i}) \\
&\quad \prod_{j=1}^{J_i} P(B_{ij} = 1 \mid \eta_i, \mathbf{U}_{ij}, \boldsymbol{\nu})^{B_{ij}} P(B_{ij} = 0 \mid \eta_i, \mathbf{U}_{ij}, \boldsymbol{\nu})^{1-B_{ij}} \\
&\quad \quad (P(R_{ij} = 1 \mid \eta_i, \mathbf{V}_{ij}, \gamma)^{R_{ij}} P(R_{ij} = 0 \mid \eta_i, \mathbf{V}_{ij}, \gamma)^{1-R_{ij}})^{B_{ij}} \\
&\quad \prod_{j=1}^{J_{S_i}} P(S_{ij} = 1 \mid \eta_i, \mathbf{W}_{ij}, \boldsymbol{\omega})^{S_{ij}} P(S_{ij} = 0 \mid \eta_i, \mathbf{W}_{ij}, \boldsymbol{\omega})^{1-S_{ij}}. \tag{4}
\end{aligned}$$

and

$$\begin{aligned}
& p\left(\rho, \beta, (\boldsymbol{\mu}_k, \Sigma_k), \boldsymbol{\nu}, \gamma, \boldsymbol{\omega}, (\eta_i, \mathbf{b}_i) \mid (\mathbf{Y}_i, \underline{\mathbf{X}}_i, \underline{\mathbf{Z}}_i, (B_{ij}, \mathbf{U}_{ij}), (R_{ij}, \mathbf{V}_{ij}), (S_{ij}, \mathbf{W}_{ij})); \boldsymbol{\Theta}\right) \\
&\propto L\left(\rho, \beta, (\boldsymbol{\mu}_k, \Sigma_k), \boldsymbol{\nu}, \gamma, \boldsymbol{\omega}, (\eta_i, \mathbf{b}_i) \mid (\mathbf{Y}_i, \underline{\mathbf{X}}_i, \underline{\mathbf{Z}}_i, (B_{ij}, \mathbf{U}_{ij}), (R_{ij}, \mathbf{V}_{ij}), (S_{ij}, \mathbf{W}_{ij}))\right) \\
&\quad \times \pi(\rho, \beta, (\boldsymbol{\mu}_k, \Sigma_k), \boldsymbol{\nu}, \gamma, \boldsymbol{\omega} \mid \boldsymbol{\Theta}). \tag{5}
\end{aligned}$$

5 Application

5.1 Methods

We applied both the unadjusted and the IOP model to data from the Johns Hopkins Active Surveillance Cohort. Patients who met Epstein et al. (1994) criteria for “very low risk”, had at least two PSA measurements, and at least one post-diagnosis biopsy as of October 2014 were included in the analysis. Patients still active in the program were censored at this date. Otherwise, observations on a patient were censored when he received curative intervention (including radical prostatectomy, radiation therapy, hormone therapy, or cryotherapy), died, or was lost to follow-up. Loss to follow-up was defined as two years without a PSA or biopsy after the last observation.

PSA trajectory was modeled with a linear mixed effects model, as described in Section 3.3. Random effects for intercept and age (\mathbf{Z}), centered at the mean intercept and slope for each partially-latent class, were estimated for each patient in addition to fixed effects for prostate volume (\mathbf{X}). A shared covariance matrix was assumed for the random effects, i.e. $\Sigma_0 = \Sigma_1$, in order to aid identifiability. The plausibility of this assumption was confirmed by fitting the mixed effects model in the subset of patients with cancer state observed.

Biopsy, reclassification, and surgery observations were categorized into annual intervals, since Johns Hopkins’ active surveillance protocol is to perform a biopsy every year. A small number of intervals contained two biopsies (approximately 1%). To accommodate this, we redefine the logistic regression model in Equation (3) as the probability of any biopsies during the year. (We also considered modeling the number of biopsies with a Poisson regression model, but this approach did not improve model fit or predictive accuracy.) Also, for intervals with two biopsies, both contributed separately to the pooled logistic regression model for time-to-reclassification.

In addition to cancer state, age, time since diagnosis, and calendar time were included as predictors in all logistic regression models. The number of previous biopsies was also included as a predictor in the biopsy and surgery submodels. The surgery submodel also included reclassification as well as other biopsy results (number of positive cores, maximum percent involvement of any core). Natural spline representations of predictors were used in place of linear representations when doing so improved model fit.

Model parameters and non-informative priors for the analysis are included in the model summary given in Figure 2. Posterior sampling was performed in **R2JAGS** with code available at <http://github.com/rycoley/XXX>. Simulated data (based on model estimates) is also provided.

Predictive accuracy was assessed in the following ways. First, the out-of-sample area-under-the-curve (AUC) and receiver operating characteristics (ROC) of posterior cancer state predictions were calculated among those participants for whom true state had been observed [18]. AUC and ROC were found for both the unadjusted model (as described in Section 3) and the model adjusted for a possibly informative observation process (IOP). Second, calibration plots were drawn and examined in order to compare the posterior probability of biopsies, reclassification, and surgery in each person-year to the observed rates of these outcomes via natural spline regression.

	Cancer State	PSA	Biopsy	Surgery
<u>Data</u>	Outcome Covariates	η_i \mathbf{Y}_i $\mathbf{X}_i, \mathbf{Z}_i$	$(B_{ij}, R_{ij}), j = 1, \dots, J_i;$ $(\mathbf{U}_{ij}, \mathbf{V}_{ij}), j = 1, \dots, J_i$	$S_{ij}, j = 1, \dots, J_{S_i}$ $\mathbf{W}_{ij}, j = 1, \dots, J_{S_i}$
<u>Model</u>	$\eta_i \sim \text{Bern}(\rho)$	$\mathbf{b}_i \eta_i = k \sim \text{MVN}(\boldsymbol{\mu}_k, \Sigma)$ $\mathbf{b}_i = \text{diag}(\mathbf{b}_i \boldsymbol{\xi}^T)$ $\mathbf{Y}_i \sim \text{MVN}(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i, \sigma^2 \mathbf{I}_{M_i})$	$B_{ij} \eta_i = k \sim \text{Bern}(P(B_{ij} = 1 \mathbf{U}_{ij}(k), \boldsymbol{\nu}))$ $R_{ij} \eta_i = k \sim \text{Bern}(P(R_{ij} = 1 \mathbf{V}_{ij}(k), \boldsymbol{\gamma}))$	$S_{ij} \eta_i = k \sim \text{Bern}(P(S_{ij} = 1 \mathbf{W}_{ij}(k), \boldsymbol{\omega}))$
<u>Priors</u>	$\rho \sim \text{Beta}(1, 1)$	$\boldsymbol{\mu}_k \sim \text{MVN}(0, 10^2 \times \mathbf{I}_{D_Z}), k = 0, 1$ $\Sigma \sim \text{InvWish}(\mathbf{I}_{D_Z}, D_Z + 1)$ $\xi_d \sim U(0, 10), d = 1, \dots, D_Z$ $\boldsymbol{\beta} \sim N(0, 10^2 \times \mathbf{I}_{D_X})$ $\sigma^2 \sim U(0, 10)$	$\boldsymbol{\nu} \sim \text{MVN}(0, 10^2 \times \mathbf{I}_{D_U})$ $\boldsymbol{\gamma} \sim \text{MVN}(0, 10^2 \times \mathbf{I}_{D_V})$	$\boldsymbol{\omega} \sim \text{MVN}(0, 10^2 \times \mathbf{I}_{D_W})$

Figure 2: Model Summary. D_X is the length of vector \mathbf{X} and \mathbf{I}_{D_X} is the identity matrix with dimension D_X .

	Total # observations	Median # per patient (IQR)
PSA	10,425	10 (6,16)
Biopsy	2,741	3 (1,4)
Years of follow-up (prior to reclassification)	4,980	5 (3,8)

Table 1: Summary of observations and follow-up time for $n=874$ patients included in analysis.

5.2 Results

874 patients from the Johns Hopkins Active Surveillance cohort met the criteria for inclusion in this analysis. The number of observations and years of follow-up available for analysis are summarized in Table 1. Among the 874 patients included in the analysis, reclassification was observed in 160 patients (18%). 167 patients (19%) elected for surgical removal of the prostate (78 after experiencing reclassification), of which 161 had post-surgery full prostate Gleason score determination available. Results of post-surgery pathologic analysis in comparison to pre-surgery reclassification are given in Table 2. A total of 318 patients (36%) were censored due to receiving some curative intervention, 130 (15%) were lost to follow-up, and nineteen (2.2%) were censored due to death. (No patients died of prostate cancer.) 407 patients (47%) remained active in the program at the time of data collection for this analysis.

Figure 3 shows the PSA and biopsy data available on a dozen patients in the cohort. Plotting circles represent observed PSA values, with the scale given on the lefthand y-axis. Triangles represent biopsies, with open triangles indicating no biopsy in an annual interval (and, thus, no

	Reclassification		Total
	$R=0$	$R=1$	
Indolent, $\eta = 0$	66 (69%)	17 (26%)	83
Aggressive, $\eta = 1$	30 (31%)	48 (74%)	78
Total	96	65	161

Table 2: Summary of post-surgery cancer state determination (η) compared to reclassification (R) on final biopsy with column percentages in subset of patients with true state observed ($n = 161$)

reclassification observed) and filled triangles indicating biopsy results; triangles at the bottom of the plot represent no grade reclassification while those at the top represent a Gleason score of 7 or higher on biopsy. Alongside the observed data, Figure 3 also gives posterior predictions from the proposed model about each patient’s true cancer state (above each plot), likely PSA trajectory, and risk of reclassification on a future biopsy (with scale given on righthand y-axis). Shaded credible intervals indicate the posterior distribution for these predictions with the darkest shading occurring at the posterior median. Note that no reclassification projections are given for those patients with grade reclassification observed, as they will not continue with biopsies.

The ROC curves and calibration plot in Figure 4 illustrate the proposed informative observation process (IOP) model’s ability to accurately predict an individual’s cancer state. The out-of-sample AUC among patients with true cancer state known is 0.75. We also see that the IOP model outperforms the unadjusted model, which has an AUC of 0.72. In the right-hand panel, the calibration plot shows that posterior predictions of underlying aggressive cancer generally follow the observed rates (i.e., track the $x=y$ axis), with the exception of very low risk predictions, where less data is available. In this plot, hashmarks at $y=0$ and $y=1$ show the observed cancer state for individuals against their posterior prediction on the x-axis.

Posterior estimates from the mixed effects model for PSA are displayed in Figure 5. In this plot, each plotting circle represents the scaled random intercept (x-axis) and slope (y-axis) estimates for a single patient. Filled circles represent patients for whom the true cancer state was observed, with red indicating potentially lethal cancer found after surgery and green indicating a determination of indolent cancer. The color of open circles reflects the posterior probability of aggressive cancer, ranging from low (green) to high (red), among patients for whom true state was not observed. Finally, credible ellipses show the posterior mean and covariance of random effects for individuals in each partially latent class. We see that there is a fair amount of overlap in these intervals, indicating that PSA level and trajectory does not provide information about the true state for most patients. PSA is more informative, however, for patients with particularly

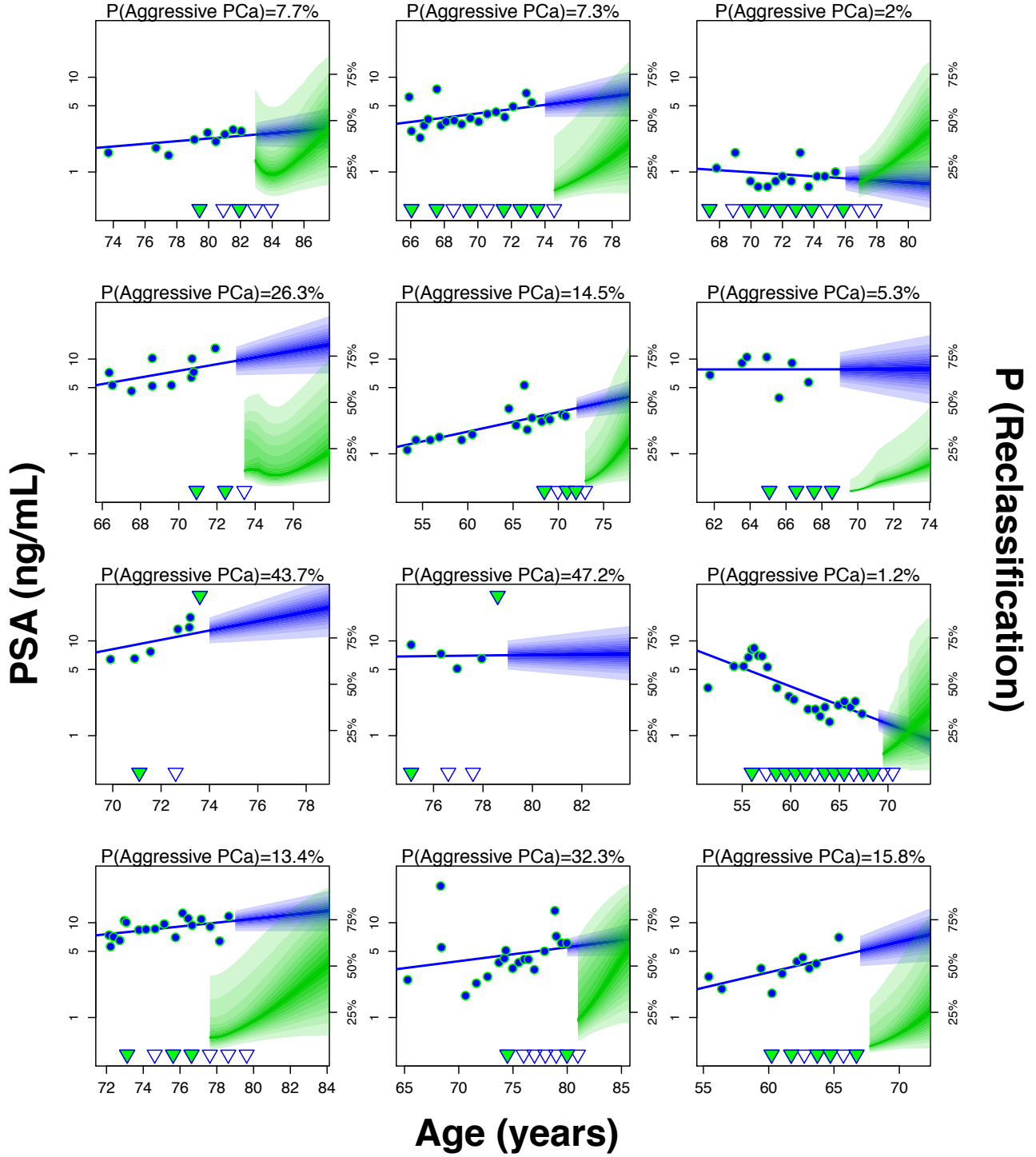


Figure 3: PSA (circles) and reclassification (triangle) data for a dozen patients. Posterior probabilities of having aggressive prostate cancer are above each patient's data. Shaded intervals show posterior credible intervals around projected PSA (blue) and reclassification (green) trajectories.

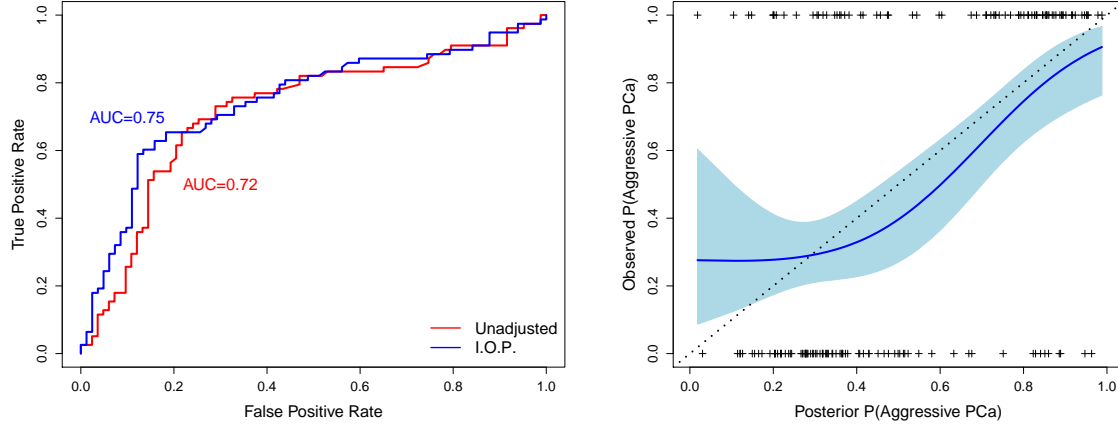


Figure 4: ROC curve (left panel) and calibration plot (right panel) for predictions of underlying cancer state, η , among those with true state observed

high or low levels and trajectories.

Calibration plots for posterior predictions of biopsy, reclassification, and surgery are given in Figure 6. Overall, posterior probabilities reflect observed rates of each outcome, especially in ranges with more data. Posterior predictions appear less accurate for reclassification but we note that the vast majority of person-years (87%) have a predicted risk of reclassification below 10%, where the model fits well. We suspect that the lack of strong predictors of reclassification limited our ability to accurately predict risk at higher ranges. (Only age, secular time, and time since diagnosis were available for inclusion in the reclassification model).

The model was also fit with informative priors on the effect of partially latent cancer state on the risk of surgery in order to assess robustness of posterior predictions to prior specification and under the concern that the relationship between an outcome's missingness and its value may not be identifiable from the likelihood alone. A range of informative priors (e.g., aggressive cancer increases, does not affect, or decreases risk of surgery) were found to have no impact on the model's predictive accuracy. (For this reason, results are not discussed here in more detail.) Informative priors were not explored for other model parameters as non-identifiability was not suspected.

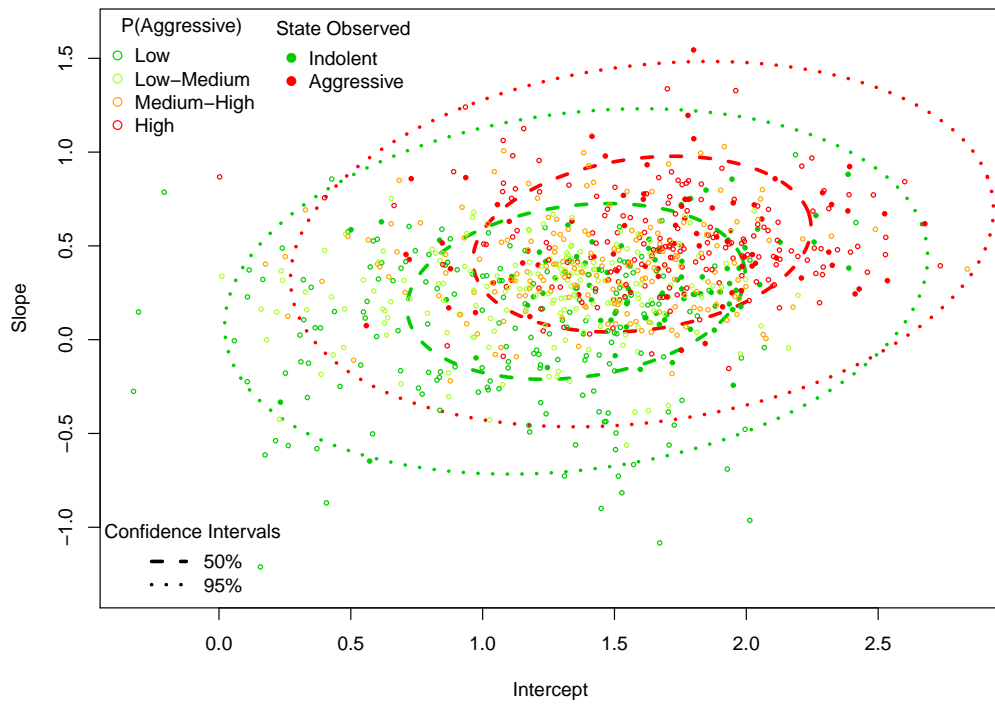


Figure 5: Random effects estimates from stratified mixed effects model for PSA.

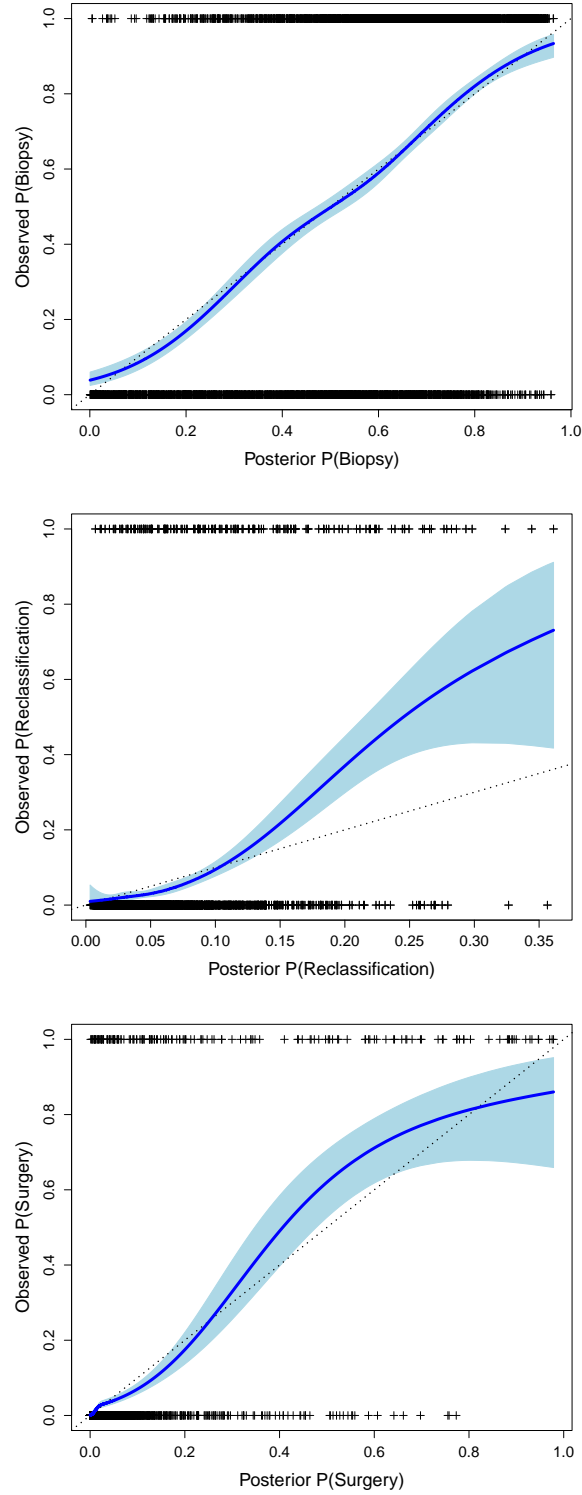


Figure 6: Calibration plots for predicting biopsies (top), reclassification (middle), and surgery (bottom) in annual intervals for all patients

6 Fast, individualized posterior estimates for new data

In order for the proposed prediction model to be useful in a clinical setting, it is necessary to be able to update posterior estimates quickly to incorporate new biopsy or PSA results during a patient’s visit. While a standard implementation of the proposed model would entail re-running MCMC to obtain updated posteriors of patient-specific variables, this approach could take hours to complete. Instead, we use importance sampling [4] to get fast prediction updates. After obtaining posterior estimates using current data, importance sampling to update these estimates for either a new patient or new data on an existing patient requires three steps: generating proposal values for the latent variables to be updated, calculating weights for proposed values, and weighting proposed values to estimate an updated posterior.

In order to generate proposal values for importance sampling for new patients without previous predictions, we start with samples of population-level parameters from the full joint posterior (Equation 5) based on previously observed data. Then, for each draw, we generate proposed values for patient-specific latent variables (true state, η , and random effects, \mathbf{b}) from their conditional posteriors. For patients with existing predictions that only need to be updated, posterior samples of latent variables based on earlier observations can be used as proposal values. Next, the importance weights for proposed values are proportional to the conditional likelihood of all patient data given population-level parameters and proposed latent variables. The weighted proposal distribution then constitutes the updated posterior for patient-specific latent variables. This particular importance sampling procedure can be more specifically described as a one-step version of a sequential importance resampler, also known as a particle filter [4]. We give further details of the proposed particle filter in the Appendix. Accompanying code is available at <http://github.com/rycoley/XXX>.

For implementation in clinical practice, proposals for new patients can be re-generated prior to actually observing new data, so that only weight calculating and re-weighting of the proposal distribution needs to be done in real-time. Then, predictions for each patient can be obtained in a matter of seconds. By random chance, some patients will have data such that very few of the pre-generated proposed latent values will receive high weights; this can cause instability in posterior means. However, such scenarios can be detected by monitoring the effective size of the posterior sample, also known as the effective number of particles. When this number drops below a pre-specified threshold (e.g., 500), the procedure can be repeated with a larger set of pre-generated proposals.

The proposed importance sampling approach still requires the periodic use of MCMC to

perform full model updates, similar to the procedure described in Lee and Chia (2002). While fully online updating of all posteriors (that is, updating not dependent on periodic MCMC) via sequential importance resampling would be ideal, such online methods are known to suffer from the problem of degeneracy when the model includes static parameters (such as the population-level parameters in our model), as explained in Andrieu et al. (2005, section II).

6.1 Application to clinical data

We assessed performance of the proposed importance sampling approach in the Johns Hopkins Active Surveillance cohort data. We compared the posterior probability of aggressive cancer, $P(\eta_i)$, obtained from MCMC performed in JAGS under two scenarios for each patient in the cohort without true cancer state observed. First, we examine agreement between the two models when each patient is completely new, that is, MCMC is performed in the absence of any of their data and, then, the particle filter is used to predict their true cancer using all of their available data. Second, we consider the scenario where additional data is observed for a patient who has already been included in the prediction model. Specifically, MCMC is performed including all data from a patient up until their penultimate biopsy (and all data from all other patients). Then, the particle filter is used to predict their true cancer state using the remaining data for that patient.

Results of this comparison are shown in Figure 7. On the left, we see agreement between methods when a patient is newly introduced. Here, correlation between posterior probabilities from JAGS and the particle filter is 0.9950. The righthand panel compares posterior predictions between the JAGS fit and the particle filter for each patient when the final biopsy is included. In this scenario, the correlation is 0.XX. From these results, we conclude that the particle filter is an appropriate substitute for full MCMC runs in order to provide real-time updates in a clinical setting.

7 Discussion

In this paper, we have presented a hierarchical Bayesian joint model for predicting latent cancer state among low risk prostate cancer patients. While many models exist for providing individualized predictions of biopsy results for this population, ours is the first model to propose prediction of the outcome of chief concern- the true underlying state of an individual’s prostate cancer. This model demonstrates the potential for new precision medicine approaches that provide individu-

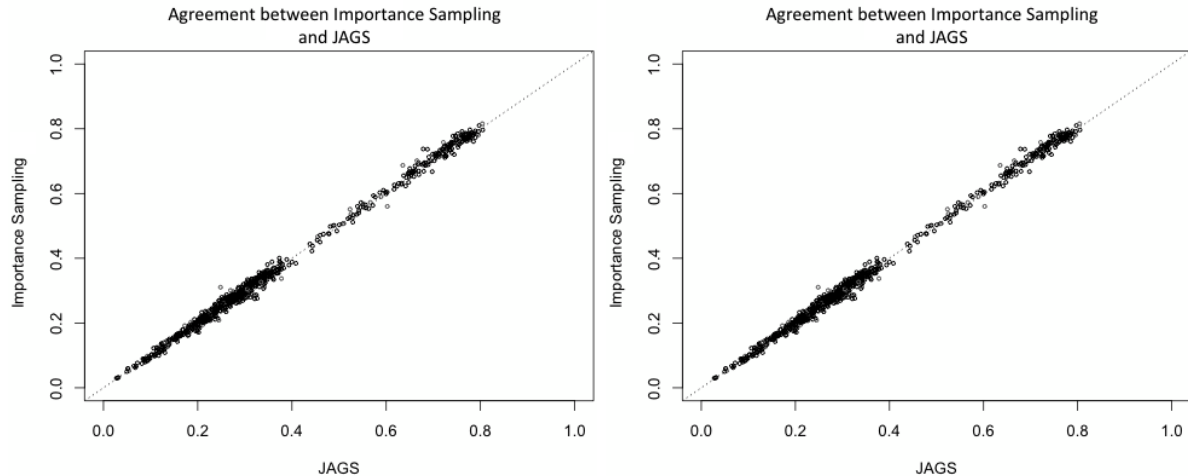


Figure 7: Agreement between importance sampling and JAGS posterior predictions of prostate cancer state for a new patient (left panel) and for a patient with updated data (right panel). (Placeholder figures)

alized predictions of latent health states and trajectories, instead of focusing only on error-prone measurements of the health state.

The proposed prediction model also exemplifies the statistical underpinnings of a learning health care system [32]. As more patients enroll in the Johns Hopkins Active Surveillance cohort, and as more information is collected on existing patients, our ability to predict underlying health states and the likely trajectory of clinical outcomes improves. Through a real-time importance sampling algorithm, updated predictions based on the most current information can be communicated to patients in a clinical setting to support decision making. At this time, real-time updating is only available for individual-level parameters; standard Bayesian estimation procedures can be routinely performed on refreshed data to update all model hyperparameters. Further research on cutting edge alternative methods (see Kantas et al. (2014)) will enable us to support fully online updating in the future.

As is typical in applied work, this model is limited by the assumptions it involves. In particular, we assume that an individual’s latent cancer state is constant during surveillance. The bias introduced by this assumption is likely small, as Porten et al. (2011) show that upgrading, particularly early on in active surveillance, is likely due to misclassification at diagnosis rather than tumor dedifferentiation. Moreover, the implications of previous research on the rate of progression among men with low risk diagnoses are limited because a distinction between progression of the underlying cancer state and missed diagnoses due to sampling error was notably absent [35]. More recently, however, Inoue et al. (2014), using a natural history approach to modeling

the rate of disease progression that also accounted for the error of biopsy grading, estimated a 12-24% rate of disease progression rate within ten years of enrollment among the Johns Hopkins AS cohort. In light of this research, we intend to extend this model to allow for time-varying health state.

Other assumptions in this model can be easily modified in response to the availability of additional data or an evolution in scientific understanding including, for example, the assumption of a shared covariance matrix for random effects across latent states. Moreover, this model can easily accommodate new predictors or outcomes. If, for example, a genetic marker for prostate cancer risk were identified, the probability distribution for latent state could be stratified by subgroups defined by this marker. Or, if additional biopsy results, such as the number of positive cores or maximum percent cancer involvement, were thought to be predictive of true cancer state, regressions of these repeated outcome measures on latent class could also be included in the joint model. (These outcomes were not associated with cancer state in our cohort and, as a result, were not included in the proposed model.) This ability to update the prediction model based on data availability and clinical expertise further reflects the characteristics of a learning health care system.

References

- [1] Paul S Albert and Dean A Follmann. Shared-parameter models. *Longitudinal data analysis*, pages 433–452, 2009.
- [2] Christophe Andrieu, Arnaud Doucet, and Vladislav B Tadic. On-line parameter estimation in general state-space models. In *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC’05. 44th IEEE Conference on*, pages 332–337. IEEE, 2005.
- [3] Donna P Ankerst, Jing Xia, Ian M Thompson, Josef Hoefler, Lisa F Newcomb, James D Brooks, Peter R Carroll, William J Ellis, Martin E Gleave, Raymond S Lance, et al. Precision medicine in active surveillance for prostate cancer: Development of the canary–early detection research network active surveillance biopsy risk calculator. *European urology*, 2015.
- [4] Christopher M Bishop et al. *Pattern recognition and machine learning*, volume 4. springer New York, 2006.
- [5] H.Ballentine Carter, Anna Kettermann, Christopher Warlick, E. Jeffrey Metter, Patricia Landis, Patrick C. Walsh, and Jonathan I. Epstein. Expectant management of prostate

cancer with curative intent: an update on the Johns Hopkins experience. *Journal of Urology*, 178:2359–2364, 2007.

- [6] Roger Chou, Jennifer M. Croswell, Dana Tracy, Christina Bougatsos, Ian Blazina, Rongwei Fu, Ken Gleitsmann, Helen C. Koenig, Clarence Lam, Ashley Maltz, J. Bruin Rugge, and Kenneth Lin. Screening for prostate cancer: a review of the evidence for the U.S. Preventive Services Task Force. *Annals of Internal Medicine*, 155:762–771, 2011.
- [7] Roger Chou, Tracy Dana, Christina Bougatsos, Rongwei Fu, Ian Blazina, Ken Gleitsmann, and J. Bruin Rugge. Treatments for Localized Prostate Cancer: Systematic Review to Update the 2002 U.S. Preventive Services Task Force. Evidence Synthesis No. 91. ARHQ Publication No. 12-0516-EF-2. Rockville, MD: Agency for Healthcare Research and Quality, 2011.
- [8] Matthew R. Cooperberg, Jeanette M. Broering, and Peter R. Carroll. Time Trends and Local Variation in Primary Treatment of Localized Prostate Cancer. *Journal of Clinical Oncology*, 28:1117–1123, 2010.
- [9] L Adrienne Cupples, Ralph B D’Agostino, Keaven Anderson, and William B Kannel. Comparison of baseline and repeated measure covariate techniques in the framingham heart study. *Statistics in medicine*, 7(1-2):205–218, 1988.
- [10] Ralph B D’Agostino, Mei-Ling Lee, Albert J Belanger, L Adrienne Cupples, Keaven Anderson, and William B Kannel. Relation of pooled logistic regression to time dependent cox regression analysis: the framingham heart study. *Statistics in medicine*, 9(12):1501–1515, 1990.
- [11] Victor DeGruttola and Xin Ming Tu. Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics*, 50:1003–1014, 1994.
- [12] Jonathan I Epstein, Zhaoyong Feng, Bruce J Trock, and Phillip M Pierorazio. Upgrading and downgrading of prostate cancer from biopsy to radical prostatectomy: incidence and predictive factors using the modified gleason grading system and factoring in tertiary grades. *European urology*, 61(5):1019–1024, 2012.
- [13] Jonathan I. Epstein, Patrick C. Walsh, Marné Carmichael, and Charles B. Brendler. Pathologic and clinical findings to predict tumor extent of non palpable (stage T1c) prostate cancer. *Journal of the American Medical Association*, 271:368–374, 1994.

- [14] Dean Follmann and Margaret Wu. An approximate generalized linear model with random effects for informative missing data. *Biometrics*, pages 151–168, 1995.
- [15] Andrew Gelman and Jennifer Hill. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press, 2006.
- [16] John Geweke. Bayesian inference in econometric models using monte carlo integration. *Econometrica: Journal of the Econometric Society*, pages 1317–1339, 1989.
- [17] Robin Henderson, Peter Diggle, and Angela Dobson. Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465–480, 2000.
- [18] David W Hosmer Jr, Stanley Lemeshow, and Rodney X Sturdivant. *Applied logistic regression*, volume 398. John Wiley & Sons, 2013.
- [19] N. Howlander, A.M. Noone, M. Krapcho, J. Garshell, D. Miller, S.F. Altekruse, C.L. Kosary, M. Yu, J. Ruhl, Z. Tatalovich, A. Mariotta, D.R. Lewis, H.S. Chen, E.J. Feuer, and K.A. (eds.) Cronin. SEER Cancer Statistics Review, 1975–2011. National Cancer Institute, Bethesda, MD, http://seer.cancer.gov/csr/1975_2011/, 2014. based on November 2013 SEER data submission, posted to the SEER web site, April 2014.
- [20] Nikolas Kantas, Arnaud Doucet, Sumeetpal S Singh, Jan M Maciejowski, and Nicolas Chopin. On particle methods for parameter estimation in state-space models. *arXiv preprint arXiv:1412.8695*, 2014.
- [21] Ali Khatami, Damber Jan-Erik, Lilja Hans, Lodding Pär, and Hugosson Jonas. Psa doubling time predicts the outcome after active surveillance in screening-detected prostate cancer: Results from the european randomized study of screening for prostate cancer, sweden section. *International journal of cancer*, 120(1):170–174, 2007.
- [22] L Klotz, L. Zhang, A. Lam, and et al. Clinical results of long-term follow-up of a large, active surveillance cohort with localized prostate cancer. *Journal of Clinical Oncology*, 2010:126–131, 2010.
- [23] Nan M Laird and James H Ware. Random-effects models for longitudinal data. *Biometrics*, pages 963–974, 1982.
- [24] Dominic S Lee and Nicholas KK Chia. A particle algorithm for sequential bayesian parameter estimation and model selection. *Signal Processing, IEEE Transactions on*, 50(2):326–336, 2002.

- [25] Haiqun Lin, Bruce W Turnbull, Charles E McCulloch, and Elizabeth H Slate. Latent class models for joint analysis of longitudinal biomarker and event process data: application to longitudinal prostate-specific antigen readings and prostate cancer. *Journal of the American Statistical Association*, 97(457):53–65, 2002.
- [26] Roderick JA Little and Donald B Rubin. *Statistical analysis with missing data*. John Wiley & Sons, 2014.
- [27] Sima P Porten, Jared M Whitson, Janet E Cowan, Matthew R Cooperberg, Katsuto Shinohara, Nannette Perez, Kirsten L Greene, Maxwell V Meng, and Peter R Carroll. Changes in prostate cancer grade on serial biopsy in men undergoing active surveillance. *Journal of Clinical Oncology*, 29(20):2795–2800, 2011.
- [28] Cécile Proust-Lima and Jeremy MG Taylor. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment psa: a joint modeling approach. *Biostatistics*, 10(3):535–549, 2009.
- [29] Jason Roy. Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. *Biometrics*, 59(4):829–836, 2003.
- [30] Jason Roy. Latent class models and their application to missing-data patterns in longitudinal studies. *Statistical methods in medical research*, 2007.
- [31] Mark D Schluchter. Methods for the analysis of informatively censored longitudinal data. *Statistics in medicine*, 11(14-15):1861–1870, 1992.
- [32] Mark Smith, Robert Saunders, Leigh Stuckhardt, J Michael McGinnis, et al. *Best care at lower cost: the path to continuously learning health care in America*. National Academies Press, 2013.
- [33] Mark S. Soloway, Cynthia T. Soloway, Steve Williams, Ranjinkanth Ayyathurai, Bruce Kava, and Murugesan Manoharan. Active surveillance; A reasonable management alternative for patients with prostate cancer: The Miami experience. *BJU International*, 101:165–169, 2008.
- [34] Jeffrey J. Tosoian, Bruce J. Trock, Patricia Landis, Zhaoyong Feng, Jonathan I. Epstein, and Alan W. Partin. Active Surveillance Program for Prostate Cancer: An Update of the Johns Hopkins Experience. *Journal of Clinical Oncology*, 29:2185–2190, 2011.

- [35] Kenneth S Tseng, Patricia Landis, Jonathan I Epstein, Bruce J Trock, and H Ballentine Carter. Risk stratification of men choosing surveillance for low risk prostate cancer. *The Journal of urology*, 183(5):1779–1785, 2010.
- [36] Nicholas J. van As, Andrew R. Norman, Karen Thomas, Vincent S. Khoo, Alan Thompson, Robert A. Huddart, Alan Horwich, David P. Dearnaley, and Christopher C. Parker. Predicting the probability of deferred radical treatment for localised prostate cancer managed by active surveillance. *European Urology*, 54:1297–1305, 2008.
- [37] Roderick C. N. van den Bergh, Stijn Roemeling, Monique J. Roobol, Gunner Aus, Jonas Hugosson, Antti S. Rannikko, Teuvo L. Tammela, Chris H. Bangma, and Schröder. Outcomes of men with screen-detected prostate cancer eligible for active surveillance who were managed expectantly. *European Urology*, 55:1–8, 2009.
- [38] H. Gilbert Welch and Peter C. Albertsen. Prostate cancer diagnosis and treatment after the introduction of prostate-specific antigen screening: 1986-2005. *Journal of the National Cancer Institute*, 101:1325–1329, 2009.
- [39] Margaret C Wu and Raymond J Carroll. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, pages 175–188, 1988.
- [40] Zhenke Wu, Maria Deloria-Knoll, Laura L Hammitt, and Scott L Zeger. Partially latent class models for case–control studies of childhood pneumonia aetiology. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 2015.

8 Appendix

8.1 Importance sampling procedure

For the purposes of this appendix, we introduce the following abbreviated form of the joint posterior given in 5:

$$p(\theta, b_{1:n} | y_{1:n}) \propto \prod_{i=1}^n [f(y_i | b_i, \theta) g(b_i | \theta)] \pi(\theta) \quad (6)$$

Where y_i is the vector of measurements for patient i , $y_{1:n}$ is the list of measurements for the first n patients, b_i is a vector of latent variables for patient i , $b_{1:n}$ is a list of latent variables for the first n patients, θ contains the population-level parameters, π is the prior for θ , and f and

g are multivariate distributions coming from the model likelihood in 4. We first illustrate how importance weighting can be used to quickly estimate latent variables for a new patient before showing how similar calculations can be done to incorporate newly measured data on existing patients in real-time.

8.1.1 Fast predictions for a new patient

Here, we focus on obtaining posteriors of latent variables for a new patient (indexed by $n + 1$). Our ultimate goal is to calculate expectations with respect to the posterior distribution based on all $n + 1$ patients (i.e. $p(\theta, b_{1:(n+1)}|y_{1:(n+1)})$). While we cannot immediately draw from this distribution, we can evaluate a function that is proportional to its density (Equation 6). To carry out importance sampling, we need to choose a proposal distribution q from which to generate candidate values of $(\theta, b_{1:(n+1)})$. We use the posterior distribution based on the first n patients as our proposal distribution. This approach is analogous to a one-step particle filter [4].

$$q(\theta, b_{1:(n+1)}) := g(b_{n+1}|\theta)p(\theta, b_{1:n}|y_{1:n})$$

Practically, this consists of taking J draws of θ and $b_{1:n}$ from the previously fitted posterior in Eq 6. Then, conditional on θ , we draw b_{n+1} from the distribution g . We index each of the resulting draws as $(\theta^{(j)}, b_{1:(n+1)}^{(j)})$, with $j = 1, \dots, J$. The importance weights w_j are then proportional to

$$\begin{aligned} w^{(j)} &\propto \frac{p(\theta^{(j)}, b_{1:(n+1)}^{(j)}|y_{1:(n+1)})}{q(\theta^{(j)}, b_{1:(n+1)}^{(j)})} \\ &\propto \frac{\prod_{i=1}^{n+1} [f(y_i|b_i^{(j)}, \theta^{(j)})g(b_i^{(j)}|\theta^{(j)})]\pi(\theta^{(j)})}{g(b_{n+1}^{(j)}|\theta^{(j)}) \prod_{i=1}^n [f(y_i|b_i^{(j)}, \theta^{(j)})g(b_i^{(j)}|\theta^{(j)})]\pi(\theta^{(j)})} \\ &= f(y_i|b_i^{(j)}, \theta^{(j)}) \end{aligned} \tag{7}$$

The final weights $w^{(j)}$ are standardized to sum to 1. The new posterior for $(\theta, b_{1:(n+1)})$ can then be represented as the mixture distribution satisfying $P(\theta = \theta^{(j)}, b_{1:(n+1)} = b_{1:(n+1)}^{(j)}) = w^{(j)}$. A posterior mean for $b_{(n+1)}$ can be calculated as $\sum_{j=1}^J w^{(j)} b_{(n+1)}^{(j)}$. The unstandardized weights can also be used in a rejection sampling procedure, although we found this approach to be less computationally efficient than importance sampling for our scenario.

8.1.2 Fast predictions for existing patients with new data

For a patient k with existing data, where we already have a posterior sample for their latent variable values, we instead use this posterior as our proposal distribution $q(\theta^{(j)}, b_{1:n}^{(j)})$, with $i \leq n$. Let $y_{1:n}^{k+}$ refer to the data set after incorporating new data on patient k , where $y_i^+ = y_i$ if $k \neq i$. The importance weights in Equation 7 then simplify to

$$\begin{aligned} w^{(j)} &\propto \frac{p(\theta^{(j)}, b_{1:n}^{(j)} | y_{1:n}^+)}{q(\theta^{(j)}, b_{1:n}^{(j)})} \\ &\propto \frac{\prod_{i=1}^n [f(y_i^+ | b_i^{(j)}, \theta^{(j)}) g(b_i^{(j)} | \theta^{(j)})] \pi(\theta^{(j)})}{\prod_{i=1}^n [f(y_i | b_i^{(j)}, \theta^{(j)}) g(b_i^{(j)} | \theta^{(j)})] \pi(\theta^{(j)})} \\ &= \frac{f(y_k^+ | b_k^{(j)}, \theta^{(j)})}{f(y_k | b_k^{(j)}, \theta^{(j)})} \end{aligned}$$

Let L_k denote that number of measurements for which we've previously fit a posterior for b_k , and let N_k denote the number of new measurements we wish to incorporate into this posterior. Then, y_k^+ can be expressed as the vector $y_k^+ = (y_{k[1]}, y_{k[2]}, \dots, y_{k[L_k]}, y_{k[L_k+1]}^+, \dots, y_{k[L_k+N_k]}^+)$, where $y_{k[l]}^+$ is the l^{th} measurement from patient k . If the repeated measures for each patient are independent conditional on b_i , as is the case in the proposed model, then the above ratio reduces to

$$\begin{aligned} w^{(j)} &\propto \frac{\prod_{l=1}^{L_k+N_k} f(y_{k[l]}^+ | b_k^{(j)}, \theta^{(j)})}{\prod_{l=1}^{L_k} f(y_{k[l]} | b_k^{(j)}, \theta^{(j)})} \\ &= \prod_{l=L_k+1}^{L_k+N_k} f(y_{k[l]}^+ | b_k^{(j)}, \theta^{(j)}) \end{aligned}$$

We then proceed as above to get a re-weighted posterior for the latent variables of patient k .