Prediction Model for Active Surveillance of Prostate Cancer

Rebecca Yates Coley

June 24, 2015

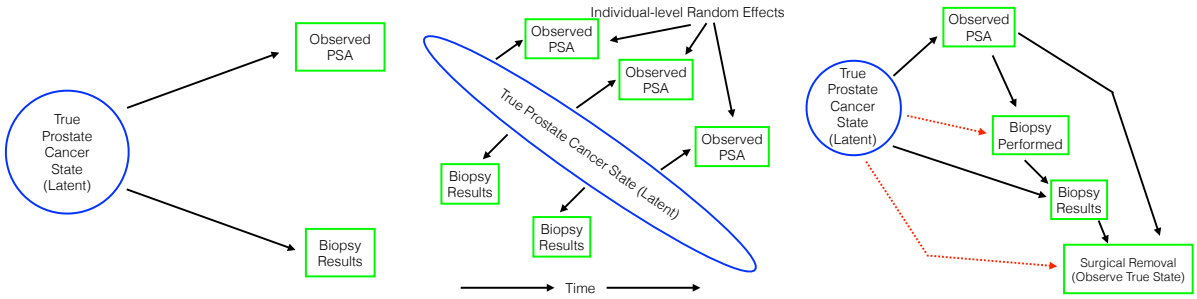
1 Introduction

Prostate cancer is the most commonly diagnosed non-skin cancer in men in the United States [16]. Upon diagnosis, early curative treatment with surgery, radiation, or androgen deprivation therapy is common [6, 31]. 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 [4, 5].

Active surveillance with curative intent offers an alternative to early treatment for individuals with lower risk disease detected [3, 18, 19, 27, 28, 29, 30]. 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[10]. 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*- with the latter defined as a true (i.e., measured without error) Gleason score of 7 or higher.



(a) (b) (c)

Figure 1

Predictions are informed by a subset of patients for whom the true state is observed, that is, patients who underwent elective prostatectomy (either before of 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 [33].

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 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 [20]. 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.)[7, 8]. 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 (LTMS), 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 [7, 8] 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 [22]. 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 [32]. 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 [11] 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[14].

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

Summary of program, use Bal's papers.

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 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

This subsection may not be necessary. 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 $i = 1, \ldots, n$. (Further description cancer state is provided below.) 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. (If no biopsy was performed, $B_{ij} = 0$, then reclassification was not observed, $R_{ij} = 0$.) Finally, S_{ij} is an indicator of surgery for individual i in year j. These variables are recorded for all individuals until any curative intervention (including surgery), death, or loss-to-follow-up, defined as two years without any PSA or biopsy measurements.

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 the 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,

$$\eta_i \sim Bern(\rho).$$

where we assume a shared underlying probability of aggressive cancer, ρ, as our data do not include any baseline predictors of cancer state, such as genetic markers or family history, other than PSA and biopsy measurements.

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, an assumption based on Author (Year), which demonstrated that the rates of de-differentiation in this population is low. 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, while tumor volume would likely differ, the same histologic category (i.e. Gleason = 6 vs. Gleason $\geq$ 7) would be observed if a full pathological 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 [20]. 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:

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

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 mth PSA measurement, $\boldsymbol{\beta}$ is a parameter vector for fixed effects, $\mathbf{b}_i$ is the subject-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, $\boldsymbol{\mu}_k$:

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

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

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, $\mathbf{Y}_i = (Y_{i1}, ..., Y_{iM_i})$, 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 active surveillance. 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. Moreover, grade determination from biopsy samples is somewhat subjective, which contributes additional error to biopsy results, 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:

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

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 [8]. 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, $j = 1, \ldots, J_i$.

3.5 Likelihood

The likelihood for the proposed model can be written as:

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

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 $\underline{\mathbf{X}_i} = [X_{i1}, \ldots, X_{iM_i}]$ and $\underline{\mathbf{Z}_i} = [Z_{i1}, \ldots, Z_{iM_i}]$.

We note that dependence structures codified in the Figure 1 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 ($\eta = 1$), 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:

$$p\Big(\rho, \boldsymbol{\beta}, (\boldsymbol{\mu}_k, \Sigma_k), \boldsymbol{\gamma}, (\eta_i, \mathbf{b}_i) \mid \big(\mathbf{Y}_i, \underline{\mathbf{X}}_i, \underline{\mathbf{Z}}_i, (B_{ij}, R_{ij}, \mathbf{V}_{ij})\big); \boldsymbol{\Theta}\Big)$$
$$\propto L\Big(\rho, \boldsymbol{\beta}, (\boldsymbol{\mu}_k, \Sigma_k), \boldsymbol{\gamma}, (\eta_i, \mathbf{b}_i) \mid \big(\mathbf{Y}_i, \underline{\mathbf{X}}_i, \underline{\mathbf{Z}}_i, (B_{ij}, R_{ij}, \mathbf{V}_{ij})\big)\Big) \times \boldsymbol{\pi}(\rho, \boldsymbol{\beta}, (\boldsymbol{\mu}_k, \Sigma_k), \boldsymbol{\gamma}, \underline{\mathbf{b}} | \boldsymbol{\Theta})$$

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

4 Informative Observation Process

In this section, we propose an addition to the joint latent class model presented above to handle 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) [22]. 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 explicity 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 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 [11].

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, $\mathbf{U}_{ij}(k)$:

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

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:

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

where $\mathbf{W}_{ij}$ is a matrix of time-varying predictors including cancer state η_i 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:

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

and

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

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 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 effect of 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 fixed 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 AS protocol is to perform a biopsy every year. A negligible 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 for the analysis are included in the model summary given in Table **??**. Posterior sampling was performed in `R2JAGS` with code available at `http://github.com/rycoley`.

5.2 Results

Among the 874 patients included in the analysis, reclassification was observed in 160 patients. 167 patients elected for surgical removal of the prostate (78 after experiencing reclassification), of which 161 had post-surgery full prostate Gleason score determination available. A total of 318 patients were censored due to curative intervention, 130 were lost to follow-up, and nineteen were censored due to death. (No patients died of prostate cancer.) 407 remained active in the program at the time of data collection.

6 Importance Sampling

Aaron writes this section.

7 Discussion

References

[1] Paul S Albert and Dean A Follmann. Shared-parameter models. *Longitudinal data analysis*, pages 433–452, 2009.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] Victor DeGruttola and Xin Ming Tu. Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics*, 50:1003–1014, 1994.

[10] 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.

[11] Dean Follmann and Margaret Wu. An approximate generalized linear model with random effects for informative missing data. *Biometrics*, pages 151–168, 1995.

[12] Alan E Gelfand, Sujit K Sahu, and Bradley P Carlin. Efficient parametrisations for normal linear mixed models. *Biometrika*, 82(3):479–488, 1995.

[13] Andrew Gelman and Jennifer Hill. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press, 2006.

[14] John Geweke. Bayesian inference in econometric models using monte carlo integration. *Econometrica: Journal of the Econometric Society*, pages 1317–1339, 1989.

[15] Robin Henderson, Peter Diggle, and Angela Dobson. Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465–480, 2000.

[16] 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.

[17] Robert E Kass, Ranjini Natarajan, et al. A default conjugate prior for variance components in generalized linear mixed models (comment on article by browne and draper). *Bayesian Analysis*, 1(3):535–542, 2006.

[18] 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.

[19] 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.

[20] Nan M Laird and James H Ware. Random-effects models for longitudinal data. *Biometrics*, pages 963–974, 1982.

[21] 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.

[22] Roderick JA Little and Donald B Rubin. *Statistical analysis with missing data*. John Wiley & Sons, 2014.

[23] 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.

[24] Jason Roy. Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. *Biometrics*, 59(4):829–836, 2003.

[25] Jason Roy. Latent class models and their application to missing-data patterns in longitudinal studies. *Statistical methods in medical research*, 2007.

[26] Mark D Schluchter. Methods for the analysis of informatively censored longitudinal data. *Statistics in medicine*, 11(14-15):1861–1870, 1992.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] 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.