Bayesian Joint Hierarchical Model for Prediction of Latent Health States with Application to Active Surveillance of Prostate Cancer

R Yates Coley, Aaron J Fisher, Mufaddal Mamawala, H Ballentine Carter, Kenneth J Pienta, Scott L Zeger

August 19, 2015

Abstract

In this article, we present a Bayesian joint hierarchical model for predicting a latent health state from longitudinal clinical measurements. Model development is motivated by an application to active surveillance of low risk prostate cancer. Existing joint latent class modeling approaches are unsuitable for this context as they do not accommodate measurement error– cancer state determinations based on biopsied tissue are prone to misclassification– nor do they allow for observations to be missing not at random. The proposed model addresses these limitations, enabling estimation of an individual's underlying prostate cancer state. We apply our model to data from a cohort of active surveillance patients at Johns Hopkins University to evaluate performance.

1 Introduction

Medicine is in a period of transition. An ever-increasing amount of information is available on patients ranging from genetic and epigentic profiles enabled by next-generation sequencing to moment-to-moment data collected by physical activity monitors. With this wealth of information comes the opportunity to provide more targeted health care including, for example, prediction of pre-clinical atherosclerosis [29], individualized cancer screening [34], sub-typing of scleroderma [35], and personalized cancer treatment[19]. In order to fully realize the promise of patient-focused medicine, principled statistical methods are needed to synthesize data from a variety of sources and provide clinicians and patients with relevant syntheses that inform decision-making. These methods must also accommodate limitations common to data generated in an observational setting including measurement error, and informative missing data patterns.

An excellent example of this challenge is low-risk prostate cancer diagnoses– tumor lethality is an aspect of an individual's latent health state that is not directly observable but is manifest in measurements. While many measurements may be available on biomarkers, histology of biopsied tissue, genetic markers, and family history of the disease, the patient's prognosis is still uncertain. As a result, individualized predictions of the disease state can guide treatment decisions. If the

tumor is potentially lethal, immediate treatment can be life-saving. Yet, some tumors are indolent and not life-threatening. In this case, treatment is not recommended due to the risk of lasting side effects including urinary incontinence and erectile dysfunction [6, 7].

Active surveillance offers an alternative to early treatment for individuals with lower risk disease detected [5, 23, 24, 38, 40, 41, 42]. Though active surveillance regimes vary, the approach generally entails regular biopsies (e.g., annually) with intervention recommended upon detection of higher risk histological features, as determined by the Gleason grading system [16, 17]. 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[11]. As a result, a biopsy does not measure the true cancer state; biopsies are instead 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 application 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 an actual Gleason score of 7 or higher. Predictions are informed by a subset of patients for whom the true state is observed– patients who, either before or after reclassification, underwent elective prostatectomy. In this sense, cancer state operates as a partially-latent class in the proposed model [44].

An individual's cancer state influences both the level and trajectory of PSA measurements as well as the outcomes from repeated biopsies. These relationships are illustrated by the directed acyclic graph (DAG) in Figure 1(a). In the model we are proposing, PSA measurements follow a mixed effects model with mean effects varying across latent classes [25]. 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.)[8, 9]. As codified in the Figure 1(a) DAG,

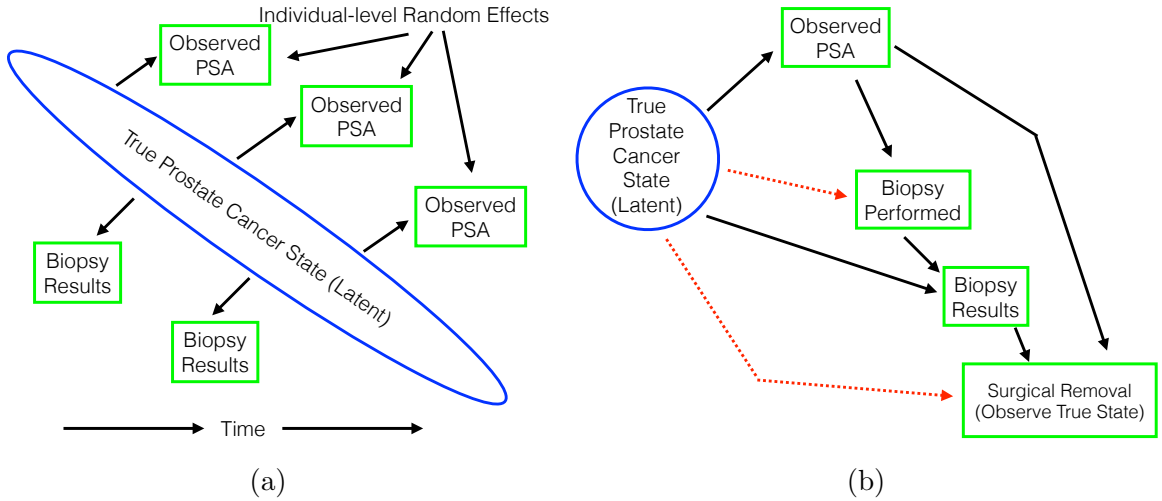

Figure 1: DAGs describing the relationships between latent class (circled) and observed measurements (squared) (a) over time and (b) in the presence of an informative observation process (at a single time point).

PSA and biopsy results are also assumed to be conditionally independent given latent class.

This first model is related to previous work by Lin et al. (2002), who proposed a joint latent class model (JLCM) to analyze 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). The 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, does not involve a priori specification of classes of interest.

Motivated by the active surveillance problem, the method proposed here addressed the limitations of JLCM and advances the development of latent variable models for multivariate outcomes in two ways. First, the proposed joint model accommodates inherent measurement error when monitoring disease progression. Second, our model is more appropriate for a clinical context where time-to-disease progression depends on non-random observation times. Among patients with low risk disease, presentation of symptoms is rare and the timing of biopsies is instead influenced by AS protocol and observed PSA results. To this latter point, the conditional independence assumption in JLCM is suspect as observation of a high PSA increases the chance of a biopsy. The proposed model addresses this limitation by modeling reclassification with a pooled

logistic regression model [8, 9] and conditioning on the presence of a biopsy.

Our approach also allows the occurrence of a biopsy and surgical removal of the prostate to depend on the latent cancer state. Absent a biopsy or surgery, we do not observe the biopsy results or true cancer state, respectively. If we treat this information as missing, our extension is to allow missing not at random (MNAR) data with missing at random (MAR) as a special case [28]. Consider the dotted lines in the DAG in Figure 1(b)– 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 [43]. 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 [13] 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.

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. Section 4 presents an application of the model to the active surveillance cohort. We close with a discussion.

2 Active Surveillance Program for Prostate Cancer at Johns Hopkins

From January 1995 to June 2014, the Johns Hopkins Active Surveillance cohort has enrolled 1,298 prostate cancer patients [39]. 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 grade reclassification, that is, when the Gleason score assigned to a biopsy exceeds 6. (Volume reclassification can also occur,

but we grade reclassification is the focus of this paper.) 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 of the underlying cancer state, measurement process, and outcomes 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 potentially non-random subset of the cohort. Predictions are also informed by the presence of some observations, which we refer to as an *informative observation process* (IOP). In this section, we introduce notation and models for the latent state, outcome measurements, and IOP and then define the likelihood in terms of these components. We then complete Bayesian specification of the model by identifying appropriate priors and defining the joint posterior distribution. Model structure is summarized in Figure 2.

Define individual i's true cancer state, η_i, as either indolent, $\eta_i = 0$, or aggressive, $\eta_i = 1$, $i = 1, \ldots, n$. We use the Gleason score that would be assigned if his entire prostate were to be removed and pathologic analysis performed to define $\eta_i = 0$ if Gleason $\leq$6 and $\eta_i = 1$ if Gleason ≥ 7. True 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, for simplicity in initial presentation. Since we observe this true cancer state on a subset of patients in active surveillance who choose surgical removal of the prostate, η_i is a partially observed latent variable.

Next, we consider PSA, which is influenced by the true cancer state as well as age, prostate volume, and other sources of inflammation. Imprecise measurements of PSA are observed at randomly selected points in time. We use a mixed effects model to model the anticipated trajectory of an individual's PSA over time [25]. Mean effects for predictors are allowed to vary across groups defined by the partially latent true cancer state as follows:

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

where Y_{im} is the observed (log-transformed) PSA, $\mathbf{X}_{im}$ and $\mathbf{Z}_{im}$ are vectors of covariates for individual i's mth PSA measurement, $\boldsymbol{\beta}$ is a parameter vector for fixed effects, 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, $\boldsymbol{\mu}_k$:

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

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.

We then consider information about the true cancer state contained in the presence and results of prostate biopsies. Biopsy data are categorized into discrete time intervals with (B_{ij}, R_{ij}) denoting binary variables for individual i in time interval j indicating whether a biopsy was performed and, when it was, if reclassification was observed, respectively. (B_{ij}, R_{ij}) is defined for $j = 1, \ldots, J_i$ where J_i is the time of reclassification or censoring for patient i. For each time interval, we use logistic regression to model the presence of a biopsy and, when a biopsy is performed, its result conditional on true cancer state:

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

$$P(R_{ij} = 1 | B_{ij} = 1, \eta_i = k, \mathbf{V}_{ij}, \boldsymbol{\gamma}) = \text{logit}^{-1}\big(\mathbf{V}_{ij}(k)\boldsymbol{\gamma}\big) \tag{2}$$

where $\mathbf{U}_{ij}(k)$ and $\mathbf{V}_{ij}(k)$ are matrices of time-varying predictors including η_i and $\boldsymbol{\nu}$ and $\boldsymbol{\gamma}$ are parameter vectors. This model specification reflects two relevant aspects of data generated in AS: first, whether or not a biopsy is performed may be informative of true cancer state, and second, reclassification on biopsies is prone to measurement error.

Lastly, to allow for the possibility that surgical removal of the prostate, and subsequent observation of the true cancer state, is informative, surgery data is included in the joint model. We define S_{ij}, a binary indicator of surgery for individual i during time interval j for $j = 1, \ldots, J_{S_i}$, where J_{S_i} is the time of surgery or other censoring for patient i. We 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 $\mathbf{W}_{ij}(k)$ is a matrix of time-varying predictors including cancer state $\eta_i = k$ and $\boldsymbol{\omega}$ is a parameter vector.

Having specified models for each information source, we define the joint probability of the

parameters given data and unobserved latent variables:

$$
\begin{aligned}
&L\Big(\rho, \boldsymbol{\beta}, (\boldsymbol{\mu}_k, \Sigma_k), \boldsymbol{\nu}, \boldsymbol{\gamma}, \boldsymbol{\omega} \mid \big(\eta_i, (\mathbf{Y}_i, \mathbf{b}_i, \underline{\mathbf{X}_i}, \underline{\mathbf{Z}_i}), (\mathbf{B}_i, \underline{\mathbf{U}_i}), (\mathbf{R}_i, \underline{\mathbf{V}_i}), (\mathbf{S}_i, \underline{\mathbf{W}_i})\big), i = 1, ..., n\Big) \\
&= \prod_{i=1}^{n} \rho^{\eta_i} (1-\rho)^{1-\eta_i} f(\mathbf{Y}_i | \eta_i, \underline{\mathbf{X}_i}, \underline{\mathbf{Z}_i}, \mathbf{b}_i, \boldsymbol{\beta}, \sigma^2)\, g(\mathbf{b}_i | \boldsymbol{\mu}_{\eta_i}, \Sigma_{\eta_i}) \\
&\qquad \prod_{j=1}^{J_i} P(B_{ij} = 1 | \eta_i, \mathbf{U}_{ij}, \boldsymbol{\nu})^{B_{ij}} P(B_{ij} = 0 | \eta_i, \mathbf{U}_{ij}, \boldsymbol{\nu})^{1-B_{ij}} \\
&\qquad\qquad \big(P(R_{ij} = 1 | \eta_i, \mathbf{V}_{ij}, \boldsymbol{\gamma})^{R_{ij}} P(R_{ij} = 0 | \eta_i, \mathbf{V}_{ij}, \boldsymbol{\gamma})^{1-R_{ij}}\big)^{B_{ij}} \\
&\qquad \prod_{j=1}^{J_{S_i}} P(S_{ij} = 1 | \eta_i, \mathbf{W}_{ij}, \boldsymbol{\omega})^{S_{ij}} P(S_{ij} = 0 | \eta_i, \mathbf{W}_{ij}, \boldsymbol{\omega})^{1-S_{ij}}. \qquad (3)
\end{aligned}
$$

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, each with mean and covariance as defined above, given covariance matrices $\underline{\mathbf{X}_i} = [X_{i1}, \ldots, X_{iM_i}]$ and $\underline{\mathbf{Z}_i} = [Z_{i1}, \ldots, Z_{iM_i}]$. $\mathbf{B}_i, \mathbf{R}_i, \mathbf{S}_i$ represent vectors of all biopsy, reclassification, and surgery observations for individual i, and $\underline{\mathbf{U}_i}, \underline{\mathbf{V}_i}, \underline{\mathbf{W}_i}$ are the associated covariance matrices. The full model likelihood is obtained by integrating over all random effects $\mathbf{b}_i$ and all unobserved cancer states η_i.

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 (ρ) 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 latent variables is given by:

$$
\begin{aligned}
&p\Big(\rho, \boldsymbol{\beta}, (\boldsymbol{\mu}_k, \Sigma_k), \boldsymbol{\nu}, \boldsymbol{\gamma}, \boldsymbol{\omega}, (\eta_i, \mathbf{b}_i) \mid \big((\mathbf{Y}_i, \underline{\mathbf{X}_i}, \underline{\mathbf{Z}_i}), (\mathbf{B}_i, \underline{\mathbf{U}_i}), (\mathbf{R}_i, \underline{\mathbf{V}_i}), (\mathbf{S}_i, \underline{\mathbf{W}_i})\big); \boldsymbol{\Theta}\Big) \\
&\quad \propto L\Big(\rho, \boldsymbol{\beta}, (\boldsymbol{\mu}_k, \Sigma_k), \boldsymbol{\nu}, \boldsymbol{\gamma}, \boldsymbol{\omega} \mid \big(\eta_i, (\mathbf{Y}_i, \mathbf{b}_i, \underline{\mathbf{X}_i}, \underline{\mathbf{Z}_i}, (\mathbf{B}_i, \underline{\mathbf{U}_i}), (\mathbf{R}_i, \underline{\mathbf{V}_i}), (\mathbf{S}_i, \underline{\mathbf{W}_i})\big)\Big) \\
&\qquad \times \boldsymbol{\pi}\big(\rho, \boldsymbol{\beta}, (\boldsymbol{\mu}_k, \Sigma_k), \boldsymbol{\nu}, \boldsymbol{\gamma}, \boldsymbol{\omega} | \boldsymbol{\Theta}\big) \qquad (4)
\end{aligned}
$$

where $\boldsymbol{\pi}(\cdot|\boldsymbol{\Theta})$ denotes the joint prior density for model parameters with hyperparameters $\boldsymbol{\Theta}$ with indexing on i, j, and k suppressed for clarity in presentation.

4 Application

4.1 Methods

We applied models with and without IOP components 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 administratively 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 above. 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' AS protocol is to perform a biopsy once per year. A small number of intervals contained two biopsies (approximately 1%). To accommodate this, we redefine the logistic regression model in Equation (1) as the probability of any biopsies during the year. Intervals with two biopsies both contributed separately to the pooled logistic regression model for time-to-reclassification.

True 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 splines were used in place of linear representations of predictors when doing so significantly 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.

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

Figure 2: Model Summary with priors used for application to AS data. D_X is the length of vector $\mathbf{X}$ and $\mathbf{I}_{D_X}$ is the identity matrix with dimension D_X.

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 [21]. Second, calibration plots were 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.

4.2 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. 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 are compared to pre-surgery reclassification 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 observed) and filled triangles indicating biopsy results; triangles at the bottom

	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.

	Reclassification		
	R=0	R=1	Total
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 ($n = 161$)

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, slightly higher than the simple 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.

Calibration plots for posterior predictions of biopsy, reclassification, and surgery are given in Figure 5. Overall, posterior probabilities reflect observed rates of each outcome, especially in ranges with more data. Posterior predictions appear less accurate for reclassification but we

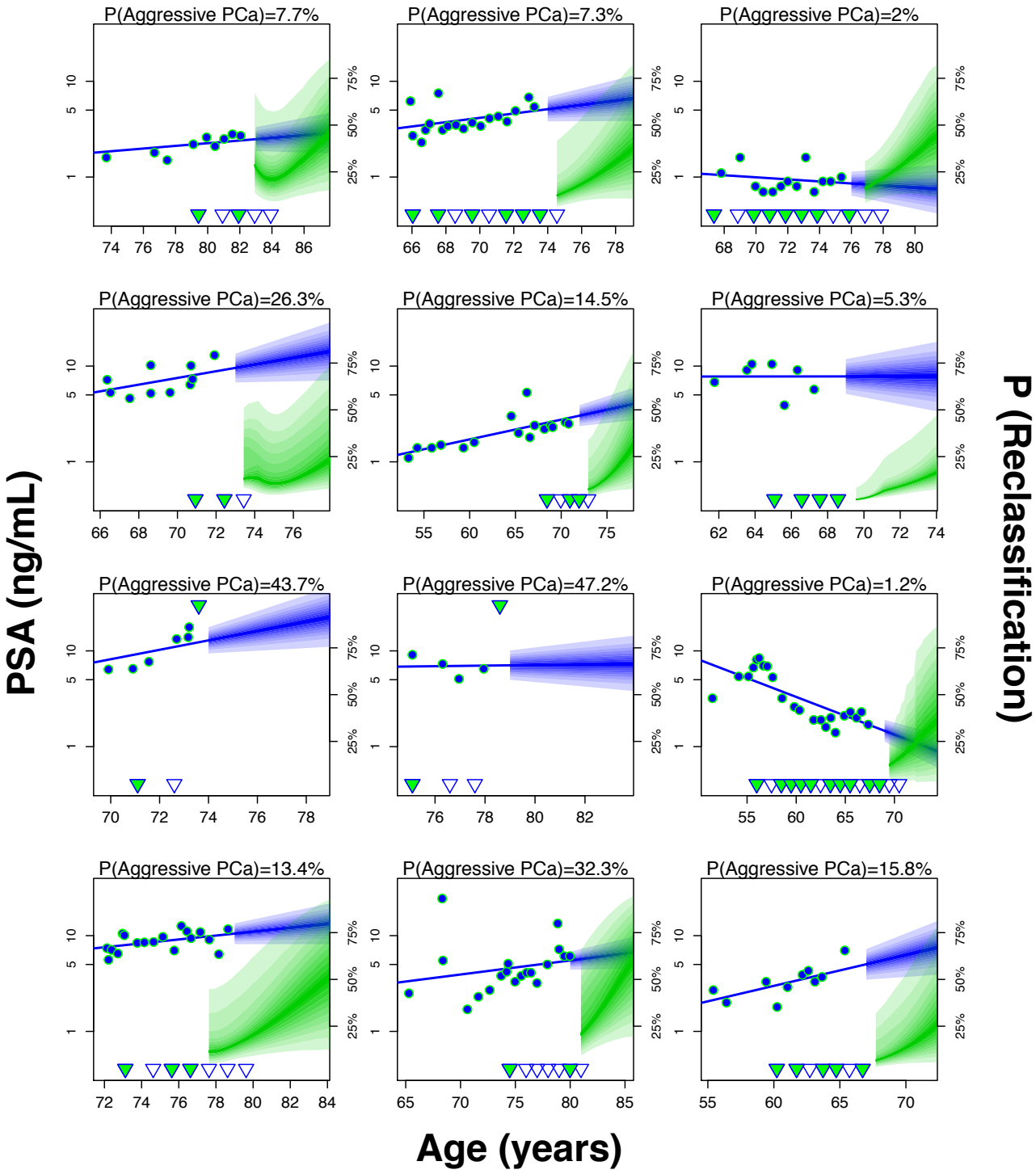

Figure 3: PSA (circles) and reclassification (triangle) data for a dozen patients. Posterior probabilities of having aggressive prostate cancer are above each patient's data. Shaded intervals show posterior credible intervals around projected PSA (blue) and reclassification (green) trajectories.

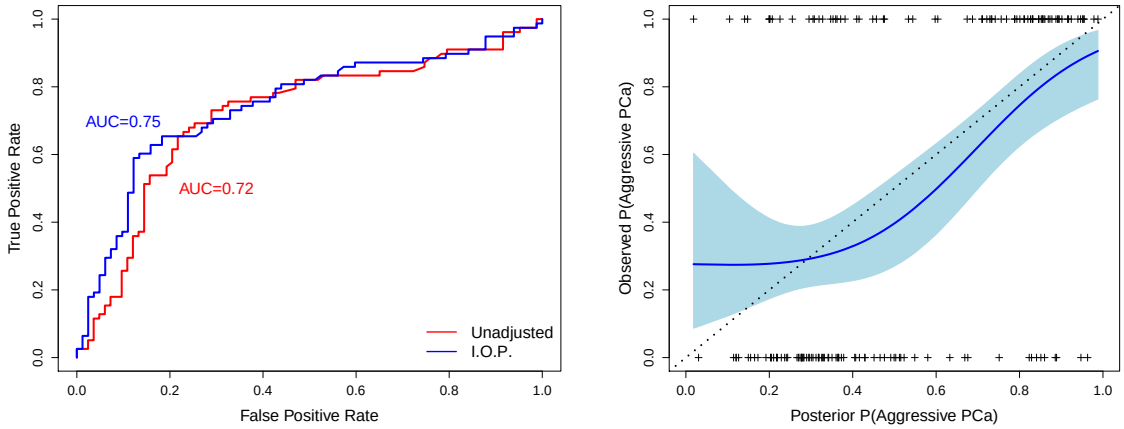

Figure 4: ROC curve (left panel) and calibration plot (right panel) for predictions of underlying cancer state, η, among those with true state observed

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.

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, indicating that identifiability of informative surgery parameters does not depend on proper priors.

5 Discussion

In this paper, we have presented a hierarchical Bayesian joint model for predicting latent cancer state among low risk prostate cancer patients from repeated biopsy and PSA results as well as information available in the observation process. While multiple models exist for providing individualized predictions of biopsy results for this population, this paper is the first to predict the outcome of chief concern- the true underlying state of an individual's prostate cancer. This approach demonstrates the potential for precision medicine approaches that provide individualized predictions of latent health states and trajectories, instead of focusing only on error-prone

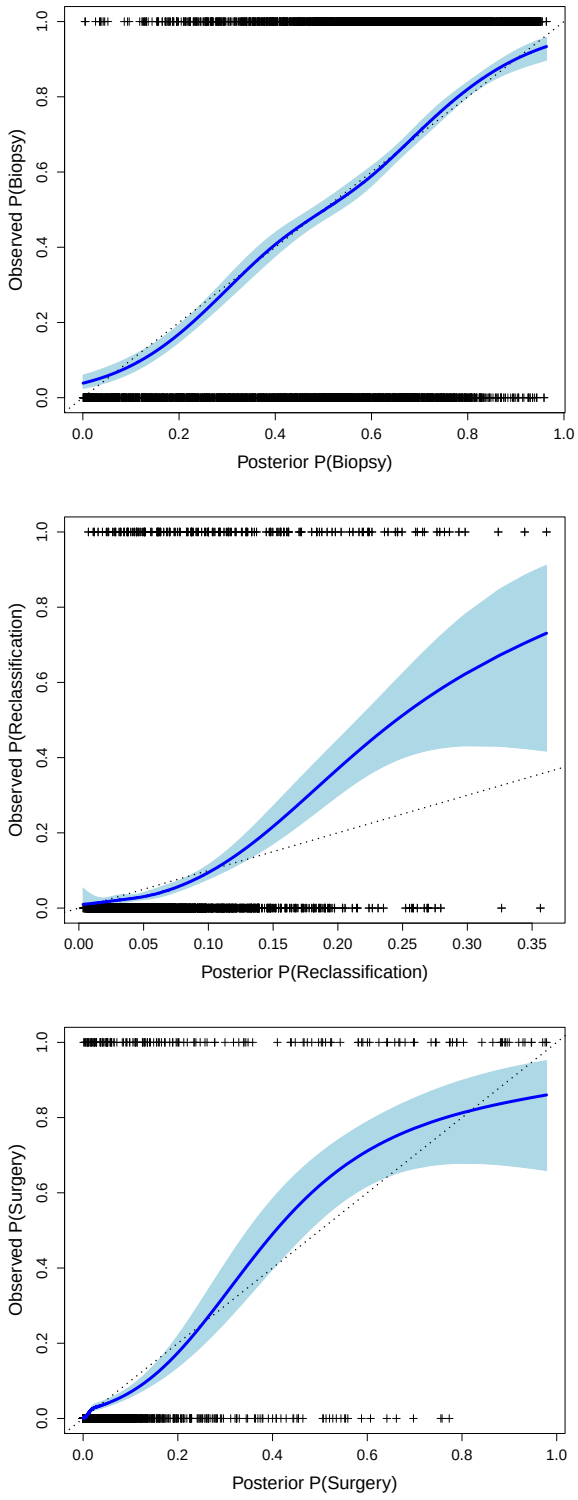

Figure 5: Calibration plots for predicting biopsies (top), reclassification (middle), and surgery (bottom) in annual intervals for all patients

measurements of the health state.

The proposed prediction model also exemplifies the statistical underpinnings of a learning health care system [18, 37], a system with the ability to continuously integrate patient data and medical knowledge to optimize patient care. 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 is likely to improve. Furthermore, online updating methods can be used to obtain real-time prediction updates based on the most current information in order to inform decision-making in a clinical setting. For example, see the accompanying technical report for an importance sampling algorithm that provides fast predictions (REF).

The proposed prediction model that targets the latent cancer state can be easily modified in response to the availability of additional data or advancement in scientific understanding. In particular, the proposed prediction model could also be extended to allow for time-varying latent states if information on the rate of disease progression in this population were available. Current model specification assumes that an individual's cancer categorization does not change over the time period under surveillance, an expectation supported by Porten et al. (2011) who showed that upgrading, particularly early on in active surveillance, is mostly due to misclassification at diagnosis rather than disease progression. This assumption is also necessary as a time-varying latent state would be nonindentifiable without additional assumptions on the rates of disease progression. For example, a recent natural history approach to modeling underlying prostate cancer state by Inoue et al. (2014), suggested a rate of disease progression in the Johns Hopkins AS cohort of 12-14% within a decade of enrollment. While these estimates were strongly dependent on prior specification, continued research on this question could inform a future extension to time-varying cancer state.

This model can also readily accommodate new predictors or outcomes. When, for example, a genetic marker for prostate cancer risk is 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 found to be predictive of true cancer state, regressions of these repeated outcome measures of latent class could also be included in the joint model. In the case that some measurements are not available for all patients, the proposed framework is also able to adjust for informative observation of predictors and outcomes. As a result, the prediction model presented here can continue to provide the most advanced decision support to physicians and patients as new knowledge and expertise is acquired.

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] 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.

[9] 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.

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

[11] 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.

[12] 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.

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

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

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

[16] D.F. Gleason. The Veteran's Administration Cooperative Urologic Research Group: Histologic grading and clinical staging of prostatic carcinoma. In M. Tannenbaum, editor, *Urologic Pathology: The Prostate*, pages 171–198. Lea and Febiger, Philadelphia, 1977.

[17] Donald F Gleason. Histologic grading of prostate cancer: a perspective. *Human pathology*, 23(3):273–279, 1992.

[18] WA Goolsby, L Olsen, and M McGinnis. Iom roundtable on value and science-driven health care. In *Clinical data as the basic staple of health learning: creating and protecting a public good: workshop summary*, pages 134–140, 2012.

[19] Erika Check Hayden. Personalized cancer therapy gets closer. *Nature News*, 458(7235):131–132, 2009.

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

[21] David W Hosmer Jr, Stanley Lemeshow, and Rodney X Sturdivant. *Applied logistic regression*, volume 398. John Wiley & Sons, 2013.

[22] Lurdes Y.T. Inoue, Bruce J. Trock, Alan W. Partin, Herbert B. Carter, and Ruth Etzioni. Modeling grade progression in an active surveillance study. *Statistics in Medicine*, 33:930–939, 2014.

[23] 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.

[24] 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.

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

[26] 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.

[27] 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.

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

[29] Michael McGeachie, Rachel L Badovinac Ramoni, Josyf C Mychaleckyj, Karen L Furie, Jonathan M Dreyfuss, Yongmei Liu, David Herrington, Xiuqing Guo, João A Lima, Wendy Post, et al. Integrative predictive model of coronary artery calcification in atherosclerosis. *Circulation*, 120(24):2448–2454, 2009.

[30] 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.

[31] 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.

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

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

[34] Sameer D Saini, Frank van Hees, and Sandeep Vijan. Smarter screening for cancer: possibilities and challenges of personalization. *JAMA*, 312(21):2211–2212, 2014.

[35] Suchi Saria and Anna Goldenberg. Subtyping: What it is and its role in precision medicine. *Intelligent Systems, IEEE*, 30(4):70–75, 2015.

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

[37] 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.

[38] 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.

[39] Jeffrey J. Tosoian, Mufaddal Mamawala, Jonathan Epstein, Patricia Landis, Sacha Wolf, Bruce J. Tock, and H. Ballentine Cater. Immediate and longer term outcomes for a prospective active surveillance program for favorable-risk prostate cancer. *Journal of Clinical Oncology*, page To appear, 2015.

[40] Jeffrey J Tosoian, Bruce J Trock, Patricia Landis, Zhaoyong Feng, Jonathan I Epstein, Alan W Partin, Patrick C Walsh, and H Ballentine Carter. Active surveillance program for prostate cancer: an update of the johns hopkins experience. *Journal of Clinical Oncology*, 29(16):2185–2190, 2011.

[41] 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. Pre-

dicting the probability of deferred radical treatment for localised prostate cancer managed by active surveillance. *European Urology*, 54:1297–1305, 2008.

[42] 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.

[43] 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.

[44] 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.