We propose a Bayesian model that predicts recovery curves based on information available before the disruptive event. A recovery curve of interest is the quantified sexual function of prostate cancer patients after prostatectomy surgery. We illustrate the utility of our model as a pre-treatment medical decision aid, producing personalized predictions that are both interpretable and accurate. We uncover covariate relationships that agree with and supplement that in existing medical literature.

1. Introduction. In the medical community, there is a pressing need for personalized predictions of how a disruptive event, such as a treatment or disease, will impact particular bodily function levels. Of particular interest is the extent to which the function is initially perturbed by the event and the ensuing pattern of recovery. In many contexts, such as mental acuity following a stroke or sexual function following prostatectomy, the post-event trajectory is expected to exhibit what we call a recovery curve shape, characterized by an initial instantaneous drop followed by a monotonic rise towards an asymptotic level not exceeding the original function level. Here, we propose a Bayesian model that can be used to predict a patient’s expected recovery curve, given information about the patient that is available before the event.

We envision our model as a decision aid for patients considering a medical treatment who want to know what adverse side effect the treatment would
have on a particular bodily function. In particular, our model will be used
to display to the patient a distribution over post-treatment function trajec-
tories, conveying the uncertainty in predictions that should be considered
in decision-making. We assume that the function level lies in some closed
interval, the pre-treatment function level is known, and the adverse effect of
the event on the function is a priori known by the medical community to be
immediate, but wearing off over time.

If a model is to be widely adopted as a medical decision aid, it is not
enough for it to merely produce predictions that are accurate; it must also
give predictions that a healthcare provider or patient can readily under-
stand, particularly when patients and providers have substantial domain
knowledge. In the applications we consider, we will predict time series that
are expected to be a recovery curves. We restrict the space of possible out-
puts to our model, therefore, to only include predictions that are recovery
curves. Thus, our model outputs a distribution of post-event trajectories
each of which is guaranteed to be a recovery curve, so that the function
level drops instantaneously downwards at event time, and rises smoothly
to approach a asymptotic level lying between the pre-treatment value and
value immediately following treatment (we will use the words “level” and
“value” interchangeably). Furthermore, our model encourages the posterior
predictive distribution over trajectories to have a well defined mode, so that
the distribution of curves, when plotted, can visually be interpreted as a
single maximum a posteriori prediction along with the uncertainty in that
single prediction.

We apply our method to data from a study that tracked the quantified
sexual function level, expressed as a number between 0 and 1, of 184 pa-
tients both before radical prostatectomy surgery and at a common set of
timepoints in the 4 years immediately following surgery. These numerical
measures of sexual function were obtained by administering the Prostate
Cancer Index, which is a multiple choice questionnaire that evaluates patient
function and bother following prostate cancer treatment, and a protocol to
convert the answers to a numerical score. Prostate cancer will affect 1 out of
6 men, and low stage patients usually have several viable treatment options,
each with different side effects. Radical prostatectomy, like most treatments
for localized prostate cancer, is known to adversely affect sexual function.
Thus it is important to be able to forecast the pattern of sexual function
level should one undergo a prostatectomy. Crucially, sexual function trajec-
tories after prostatectomy are known to follow a recovery curve, both from
past studies Potosky et al. (2004) (at least up to 5 years post-treatment,
see Resnick et al., 2013), and from our dataset. To illustrate, in Figure 1a
(a) The average function value and scaled function value over the prostatectomy dataset exhibit a recovery curve shape.

(b) Raw data of 12 randomly chosen patients who passed the filters as described in Section 6.1.

Fig 1: Plots of sexual function trajectory following prostatectomy.

we show the dataset-wide averages, for each timepoint, of sexual function level, as well as dataset-wide averages, for each timepoint, of the patients’ sexual function values scaled by their respective value immediately before prostatectomy. We also show in Figure 1 the unscaled sexual function level time series of 12 randomly selected patients, which includes their unscaled sexual function level both post-treatment and right before treatment. We hypothesize that the function levels reported by individual patients are noisy versions of a latent smooth “true” function level. In this context, we use our method to study whether there are patient covariates that correlate with post-surgery sexual function trajectory.

2. Past Work. Past work on personalized prediction of sexual function following prostate cancer treatment has attempted to predict a postoperative binary outcome, typically whether one is able to achieve an erection sufficient for sex at some single timepoint following treatment (Ayyathurai et al., 2008; Descazeaud, Debré and Flam, 2006; Eastham, Scardino and Kattan, 2008; Regan et al., 2011), or the change in (Sanda et al., 2008) or absolute level (Talcott et al., 2003) of some continuous measure of sexual function such as the IIEF-5 score. Such models have incorporated patient covariates in logistic regression models for binary outcomes, and linear regression models for continuous outcomes. One obvious shortcoming of binary modeling is that sexual function is not naturally a binary event, and it is not clear how to set a cutoff to binarize continuous measures of sexual function.
Furthermore, a serious deficiency of logistic regression and linear regression is that they are not suitable for modeling longitudinal outcomes, whereas a patient would want to know their entire post-treatment function trajectory, not just that at a single timepoint. The only longitudinal model we found in the literature was one using linear regression to relate the change in function level between two fixed timepoints after treatment (Potosky et al., 2004).

Functional data analysis (Denison, Mallick and Smith, 1998; Ramsay and Silverman, 2002) and growth curve modeling (Jung and Wickrama, 2008) are rich areas of past study, but existing models from those fields do not guarantee that the predicted time series possesses a recovery curve shape, that it drops instantaneously following the event and then monotonically approaches an asymptotic level no higher than the pre-event value. Parametric functions mentioned in Rogosa and Willett (1985) resemble the functional form we assume of recovery curves. However, they are modeling growth, not recovery, and assume the initial level of the growth curve to be known, whereas we are trying to predict the entire post-treatment trajectory, which includes the initial post-treatment value. Furthermore, they do not place an upper bound on the asymptotic level of the predicted function. Isotonic regression models (Cai and Dunson, 2007; Mammen, 1991; Neelon and Dunson, 2004; Shively, Sager and Walker, 2009), which enforce the predicted functions to be monotonic, but do not naturally output recovery curves as predictions.

Regardless of their applicability, existing models have been applied in contexts where the predicted time series is expected to exhibit a recovery curve shape. For example, in a medical context, growth curve techniques have been used to model recovery of a bodily function following a disruptive event. Warschausky, Kay and Kewman (2001) model recovery of FIM-measured function following spinal surgery as being the sum of a linear and a plateauing function. Although the FIM-score must lie within a bounded interval, they do not guarantee that the predicted scores lie within that interval. Rolfe et al. (2011) model verbal function following chemotherapy using a Bayesian latent basis model, but the model lacks incorporation of patient-correlates, and is instead specialized to infer average recovery at only two fixed timepoints. Tilling, Sterne and Wolfe (2001) model a measure of quality of life - the Barthel Index - following stroke using a multilevel model where both patient-specific and time-specific contributions are modeled as a linear combination of a fractional polynomial basis. In all these models, the predicted time series is expected to possess a recovery curve shape, but there is no explicit constraint built into the models to ensure that the predicted
trajectory actually does possess a recovery curve shape.

3. Recovery Curves. We first formally define a recovery curve and the specific parameterization of one that we use in our model. We chose this specific form of the recovery curve based on empirical evidence from the data as well as clinical expertise, which led to insights about the features of the curve that are most important to relay to patients.

3.1. Recovery Curve Definition. A recovery curve is a function $f(t)$ defined on $\mathbb{R}^+$. We will always define $f(t)$ piecewise as

$$f(t) = \begin{cases} S & \text{for } t = 0, \\ f^+(t) & \text{for } t > 0, \end{cases}$$

which is interpreted as the disruptive event occurring right after time 0, so that $S$ is the (known) pre-treatment function value, and $f^+(t)$ is the post-treatment trajectory. A recovery curve will satisfy the following:

$$f^+(t) > 0 \text{ for } t > 0, \quad (3.1)$$
$$f^+(t) \leq S \text{ for } t > 0, \quad (3.2)$$
$$f^+(t) \geq 0 \text{ for } t > 0, \quad (3.3)$$
$$S \in [0, 1]. \quad (3.4)$$

3.2. Parameterizing Recovery Curves. We will parameterize $f(t)$ scaled to the pre-treatment function value, instead of $f(t)$ itself. That is, we assume the following parameterization

$$f^+(t; S, \theta) = Sg(t; \theta).$$

Thus, the actual post-event trajectory is the shape of the post-event trajectory, $g(t; \theta)$, scaled to the pre-event function level. We choose this parameterization because to satisfy requirements of Equations 3.2 and 3.3, we just need to ensure that $g(t; \theta) \in (0, 1)$ for $t > 0$. In this work we will refer to the scaled post-event trajectory, denoted $g(t; \theta)$, as a recovery shape, and will use the term scaled function value to refer to a patient’s function value normalized by their pre-treatment value. In other words, a patient’s recovery shape is a time series over their scaled function values.

We will parameterize recovery shape $g(t; \theta)$ with 3 parameters: $A$, which represents the asymptotic drop in scaled function level after surgery, $B$, which represents the initial drop in scaled function value in excess of the
Fig 2: We parameterize a recovery curve with three parameters: $A$: the asymptotic drop in scaled function value, $B$: the initial drop in scaled function value in excess of the asymptotic drop, and $C$: the rate of recovery of the scaled function value. A recovery curve must smoothly approach an asymptotic level not exceeding the pre-treatment function level, and constraining $A,B$ to the unit interval and $C$ to be positive achieves that in the corresponding curve.

asymptotic drop, and $C$, which describes the rate of recovery of the scaled function value.

\[
\begin{align*}
\text{(3.5)} & \quad f(t; S, A, B, C) = S g(t; A, B, C) \\
\text{where} & \\
\text{(3.6)} & \quad g(t; A, B, C) = S \left(1 - A - B(1 - A) \exp \left(-\frac{t}{C} \right) \right), \\
\text{(3.7)} & \quad A \in [0, 1], \\
\text{(3.8)} & \quad B \in [0, 1], \\
\text{(3.9)} & \quad C \geq 0.
\end{align*}
\]

Figure 2 illustrates the meaning of the parameters. $f(t; S, A, B, C)$ is a recovery curve if the constraints in Equations 3.1 to 3.4 are satisfied, and we will respect those constraints in our model.

4. Model. We describe the properties we desire of our model, and the building blocks which combine to achieve those properties. We refer to the $i$-th patient’s covariate vector as $X^{(i)}$ which is of dimension $K$, and
This predictive distribution is unrealistic because some of the time series are not recovery curves, as their post-treatment function value exceeds that pre-treatment.

(b) This predictive distribution is unrealistic because the distribution is not unimodal. One can see this because there are two dark sets of curves, one for each mode of the posterior curve distribution.

Fig 3: Two unrealistic predictive distributions

their observed function value at time $t$ by $y^{(i)}(t)$. Here, $y^{(i)}(t)$ is considered a noisy measurement of their “underlying” function value at time $t$, $f(t; S^{(i)}, A^{(i)}, B^{(i)}, C^{(i)})$, where $f(\cdot; S^{(i)}, A^{(i)}, B^{(i)}, C^{(i)})$ is the parameterization of post-treatment function value described in Section 3.2. $f(t; S^{(i)}, A^{(i)}, B^{(i)}, C^{(i)})$ is a function of the patient’s pre-treatment function value $S^{(i)}$ and patient-specific random parameters $A^{(i)}, B^{(i)}, C^{(i)}$, and will be referred to as $f^{(i)}(t)$ where appropriate. Throughout this work, we assume $S^{(i)}$ to be a known, non-random quantity.

4.1. Model Requirements. As mentioned in the introduction, our model is designed to exhibit several properties. We list those properties as well as a description of our proposed solutions below:

1. **Property:** Observed within-patient function values should be dependent, and the underlying post-treatment values for patients with similar covariates should be shrunk towards each other.

   **Solution:** We adopt a hierarchical Bayesian model. As observations are naturally grouped by patient, we let $y^{(i)}(t)$ be distributed about $f(t; A^{(i)}, B^{(i)}, C^{(i)})$ according to the likelihood function we describe in 4.2.3. Shrinkage is accomplished by letting $A^{(i)}$ be drawn from a single covariate dependent distribution. In particular, $A^{(i)}$ is modelled using (a variant of) a generalized linear model. An analogous modeling approach
is used for $B^{(i)}$ and $C^{(i)}$, and we describe these three generalized linear models in more detail in Section 4.2.1.

2. **Property:** For each patient, their distribution over the underlying post-treatment function value is a recovery curve. That is, for a given patient $f(\cdot; A^{(i)}, B^{(i)}, C^{(i)})$ has support only over the space of recovery curves, namely those functions satisfying requirements 3.1 - 3.4. For example, we do not want to allow a predictive distribution like that in Figure 3a. This profile is unappealing because the distribution has support over post-treatment values that are higher than the patient’s pre-treatment values. For the treatments we consider in our application, it is unrealistic to expect that treatment will *improve* a patient’s outcome past the pre-treatment level.

**Solution:** We respect the constraints of Equations 3.7, 3.8, 3.9 in modelling $A^{(i)}, B^{(i)}, C^{(i)}$, letting their respective generalized linear models have beta, beta, and gamma response distributions, respectively, as these are canonical distributions with the desired support.

3. **Property:** We want the posterior of $f^{(i)}(t)$ to be unimodal. For example, we do not want the predictive distribution to look like that in Figure 3b. The distribution in this figure is bimodal, making it difficult to relate to physicians and patients.

**Solution:** $A^{(i)}$ depends on a shared regression parameter $b_A$, shared noise parameter $\phi_A$, as well as $X^{(i)}$, so that our model parameterizes the distribution $A^{(i)}|b_A, \phi_A; X^{(i)}$. We will constrain this conditional distribution to be unimodal, for all $X_i, b_A$, and $\phi_A$. An analogous modeling approach and constraint are used to model $B^{(i)}$ and $C^{(i)}$. Ensuring this unimodality will require special parameterizations of the beta and gamma distributions, which we describe in Section 4.2.2.

4. **Property:** $y^{(i)}(t)$ should have support on the closed unit interval, because we observed that roughly 5% of the time, patients recorded a 0 or 1 response.

**Solution:** We let $y^{(i)}(t)$ come from a mixture of a beta distribution centered at $f^{(i)}(t)$ and a Bernoulli distribution. We elaborate on this in Section 4.2.3.

5. **Property:** In the prior, a patient’s distribution over recovery shapes should be centered about some “average” shape. This is because in the absence of any data, doctors have a general idea of what will happen to a patient, on average.

**Solution:** We assume the “average” shape to be parameterized by curve parameters $\mu_A, \mu_B, \mu_C$. The GLM modeling $A^{(i)}$ depends on the hyperparameter bias term $z_A$. $z_A$ will be chosen so that in the prior, $A^{(i)}|B_A, \phi_A; X^{(i)}$
Fig 4: Our model uses separate generalized linear models to model the distribution of a patient’s three curve parameters $A^{(i)}, B^{(i)}, C^{(i)}$, which in turn determine the mode of their observed data distribution at each timepoint.

is centered at $\mu_A$. We describe the choice of $z_A$ as well as the prior distribution of $B_A$ and $\phi_A$ in Section 4.2.4.

4.2. Model Components. A diagram of the model is shown in Figure 4. The following sections give the details of the components of our model, which are, in addition to the recovery curve parameterization, described in Section 3.2:

1. Section 4.2.1: Apriori independent (variants) of generalized linear models to model $A^{(i)}, X^{(i)}, B^{(i)}, X^{(i)}, C^{(i)}, X^{(i)}$, with respective regression coefficient vectors $b_A, b_B, b_C$ and respective shared noise parameters $\phi_A, \phi_B, \phi_C$.

2. Section 4.2.2: Specialized parameterizations of the beta and gamma distributions, by their modes and a spread parameter, which are needed to satisfy Property 3.

3. Section 4.2.3: The likelihood that relates the underlying function value $f^{(i)}(t)$ to observed function value $y^{(i)}(t)$. This likelihood, which satisfies Property 4, depends on parameters $\theta, p, \phi_M$, as seen in Figure 4.

4. Section 4.2.4: A choice of the prior so that Property 5 is satisfied.

4.2.1. Generalized Linear Models for Patient Curve Parameters. In order to guarantee that $f(\cdot; A^{(i)}, B^{(i)}, C^{(i)})$ only has support over the space of recovery curves (Property 2), $A^{(i)}$ and $B^{(i)}$ will have support only on $(0, 1)$, and $C^{(i)}$ will have support only on $\mathbb{R}^+$. We choose to use model...
A^{(i)}, B^{(i)}, C^{(i)} using beta, beta, and gamma distributions, because they are canonical distributions with the desired support. In particular, we let:

\begin{align}
A^{(i)} | b_A, \phi_A; z_A, X^{(i)} &\sim \text{beta}_{m, \phi}(\text{logistic}(z_A + b_A X^{(i)}), \phi_A) \\
B^{(i)} | b_B, \phi_B; z_B, X^{(i)} &\sim \text{beta}_{m, \phi}(\text{logistic}(z_B + b_B X^{(i)}), \phi_B) \\
C^{(i)} | b_C, \phi_C; z_C, X^{(i)} &\sim \text{gamma}_{m, \phi}(\exp(z_C + b_C X^{(i)}), \phi_C)
\end{align}

with

\begin{align}
b_A, b_B, b_C &\in \mathbb{R}^K \\
\phi_A, \phi_B, \phi_C &\in (0, 1)
\end{align}

where

- \text{beta}_{m, \phi}(\cdot, \cdot) is the specialized parameterization of the beta distribution (detailed in Section 4.2.2). A \text{beta}_{m, \phi}(m', \phi') distribution has mode \( m' \) and for all \( m' \), is unimodal if and only if spread parameter \( \phi' \in (0, 1) \). Examples of such beta distributions are in Figure 16 of the Appendix.
- \text{gamma}_{m, \phi}(\cdot, \cdot) is the specialized parameterization of the gamma distribution (detailed in Section 4.2.2). A \text{gamma}_{m, \phi}(m', \phi') distribution has mode \( m' \) and for all \( m' \), is unimodal if and only if spread parameter \( \phi' \in (0, 1) \). Examples of such gamma distributions are in Figure 17 of the Appendix.
- \( z_A, z_B, z_C \) are bias terms chosen to satisfy Property 5 (details in Section 4.2.4).

The restrictions of Equation 4.5, due to the specialized parameterizations we use, ensure the unimodality of the conditional distributions of \( A^{(i)}, B^{(i)}, C^{(i)} \), satisfying Property 3.

4.2.2. Parameterizations of Unimodal Beta and Gamma Distributions. Here we describe the \text{beta}_{m, \phi}(\cdot, \cdot) and \text{gamma}_{m, \phi}(\cdot, \cdot) parameterizations of the beta and gamma distributions, respectively, and explain why the traditional parameterizations did not meet our needs.

Our \text{beta}_{m, \phi}(\cdot, \cdot) parameterization is based on the traditional \text{beta}_{\alpha, \beta}(\cdot, \cdot) parameterization. A \text{beta}_{\alpha, \beta}(\alpha', \beta') distribution is unimodal with mode \( \frac{\alpha' - 1}{\alpha' + \beta' - 1} \) if and only if \( \alpha' > 1, \beta' > 1 \). Making the substitution \( \alpha = 1 + sm \) and \( \beta = 1 + s(1 - m) \), we obtain the parameterization \text{beta}_{m, s}(m', s') = \text{beta}_{\alpha, \beta}(1 + s'm', 1 + s'(1 - m'))). It can be easily checked that a \text{beta}_{m, s}(m', s') distribution, for all \( m' \), is unimodal with mode \( m' \) if and only if \( s' > 0 \). Finally, we
A beta\(_{m,\phi}(\cdot, \cdot)\) parameterization suits our purposes because for a given \(\phi' < 1\) (a constraint we enforce), a beta\(_{m,\phi}(m', \phi')\) distribution is unimodal for all \(m'\).

(b) A beta\(_{\mu,B}(\cdot, \cdot)\) parameterization is unsuitable because for a given \(B'\), there exists \(m'\) such that a beta\(_{\mu,B}(m', B')\) distribution is not unimodal.

Fig 5: Regions of unimodality for two different beta distribution parameterizations

make the substitution \(s = \frac{1}{\phi} - 1\) to obtain the desired beta\(_{m,\phi}(\cdot, \cdot)\) parameterization, where a beta\(_{m,\phi}(m', \phi')\) distribution, for all \(m'\), is unimodal with mode \(m'\) if and only if spread parameter \(\phi' \in (0, 1)\):

\[
\text{(4.6)} \quad \text{beta}_{m,\phi}(m', \phi') = \text{beta}_{\alpha,\beta} \left( 1 + \left( \frac{1}{\phi'} - 1 \right) m', 1 + \left( \frac{1}{\phi'} - 1 \right) (1 - m') \right).
\]

Our gamma\(_{m,\phi}(\cdot, \cdot)\) parameterization is based on the traditional gamma\(_{\alpha,\beta}(\cdot, \cdot)\) parameterization. A gamma\(_{\alpha,\beta}(\alpha', \beta')\) distribution is unimodal with mode \(\frac{\alpha' - 1}{\beta'}\) if and only if \(\alpha' > 1\). Making the substitution \(\beta = \frac{\alpha - 1}{m}\) and then the substitution \(\alpha = \frac{1}{\phi}\), we obtain the desired gamma\(_{m,\phi}(\cdot, \cdot)\) parameterization, where a gamma\(_{m,\phi}(m', \phi')\) distribution is unimodal with mode \(m'\) if and only if spread parameter \(\phi' \in (0, 1)\):

\[
\text{(4.7)} \quad \text{gamma}_{m,\phi}(m', \phi') = \text{gamma}_{\alpha,\beta} \left( \frac{1}{\phi'}, \frac{1/\phi' - 1}{m'} \right).
\]

Note that both a beta\(_{m,\phi}(m', \phi')\) and gamma\(_{m,\phi}(m', \phi')\) have variance increasing in \(\phi'\). That is why we regard \(\phi'\) as a spread parameter.
The necessity for these specialized parameterizations of the beta and gamma distributions becomes clear if we consider a model that does not use them. We illustrate the necessity of our beta distribution parameterization; our gamma distribution parameterization is necessary for analogous reasons. Recall from Property 3 that for a fixed \( b_A, \phi_A \), we wanted \( A^{(i)}|b_A, \phi_A; z_A, X^{(i)} \) to be unimodal. Consider the more traditional beta \( \mu, B \) parameterization, where a beta \( \mu, \beta \) distribution has mean \( \mu' \), and \( \beta' \) is a spread parameter. Suppose we had let \( A^{(i)}|b_A, \phi_A; z_A, X^{(i)} \sim \text{beta}_{\mu, \beta}(\text{logistic}(z_A + b_A X^{(i)}), \phi_A) \). Figure 5b shows the values of \( (\mu', B') \) for which a beta \( \mu, \beta \) distribution is unimodal. Given \( \beta' \), there is some \( \mu' \) for which a beta \( \mu, \beta \) distribution is not unimodal. Thus given \( b_A \) and \( \phi_A \), there would exist some \( X^{(i)} \) for which \( A^{(i)}|b_A, \phi_A; z_A, X^{(i)} \) would not be unimodal, which violates Property 3.

4.2.3. Likelihood. Recall from Property 4 that \( y^{(i)}(t) \) should have support on the closed unit interval. Accordingly, we let \( y^{(i)}(t) \), with probability \( \theta \), come from a bernoulli(\( p \)) distribution, and come from a beta distribution centered at \( f^{(i)}(t) \) otherwise. Specifically, we let:

\[
y^{(i)}(t)|f^{(i)}(t), \theta, p, \phi_M \sim \theta \text{ bernoulli}(p) + (1 - \theta) \text{ beta}_{m, \phi}(f^{(i)}(t), \phi_M),
\]

with

\[
p, \theta, \phi_M \in (0, 1).
\]

There exist more complicated likelihoods where the proportion of unrealistic responses of 0 or 1 is covariate dependent. However, our data are noisy and limited, and does not motivate any particular such likelihood.

4.2.4. Prior. The posterior likelihood surface for the regression parameters \( b_A, b_B, b_C \) becomes exceedingly flat if the spread parameters \( \phi_A, \phi_B, \phi_C, \phi_M \) are too large. This is undesirable, as the regression coefficient parameters inform the effect of a patient’s covariates on their predicted recovery curve. Thus, we encourage the spread parameters towards 0, letting

\[
\phi_A; \lambda_A \sim \exp(\lambda_A, 1)
\]
\[
\phi_B; \lambda_B \sim \exp(\lambda_B, 1)
\]
\[
\phi_C; \lambda_C \sim \exp(\lambda_C, 1)
\]
where \( \exp(\lambda, 1) \) denotes an exponential distribution with rate parameter \( \lambda \) truncated on the right at 1 and \( \lambda_A, \lambda_B, \lambda_C \) are hyperparameters.

As mentioned in Property 5, we assume there is some “average” recovery shape \( g(\cdot; \mu_A, \mu_B, \mu_C) \) such that a patient’s expected recovery curve in the prior is centered about \( S(i) g(\cdot; \mu_A, \mu_B, \mu_C) \) (see Equation 3.5). That is, for each patient, we want the unconditional prior distributions of \( A(i), B(i), C(i) \) to be centered about \( \mu_A, \mu_B, \mu_C \), respectively. In light of Equations 4.1, 4.2, 4.3, we let

\[
(4.13) \quad z_A = \logit(\mu_A) \\
(4.14) \quad z_B = \logit(\mu_B) \\
(4.15) \quad z_C = \exp(\mu_C)
\]

and

\[
(4.16) \quad b_A \sim \text{normal}(0, s_A) \\
(4.17) \quad b_B \sim \text{normal}(0, s_B) \\
(4.18) \quad b_C \sim \text{normal}(0, s_C)
\]

where \( z_A, z_B, z_C, s_A, s_B, s_C \) are hyperparameters.

Finally, without any prior belief about the parameters \( p, \theta \) governing the likelihood, we let:

\[
(4.19) \quad p \sim \text{unif}(0, 1) \\
(4.20) \quad \theta \sim \text{unif}(0, 1).
\]

5. Simulation Studies. Here, we show the results of two simulation experiments. First, we examine the ability of our model to recover the model parameters as the amount of data simulated using those parameters grows. Second, we examine the robustness of our model to simulated data that does not fit the underlying model assumptions.

For the first experiment, we chose a single set of shared model parameters and hyperparameters \( \mu_A, \mu_B, \mu_C \). Then for various values of \( N \), we did the following: we simulated 5 datasets, where for each dataset we used that set of parameters to simulate observed function values \( y_i(t) \) for \( N \) samples at times \( t \in \{1, 2, 4, 8, 12, 18, 24, 30, 36, 42, 28\} \), which are the same times at which data were observed in the prostate cancer dataset. For each dataset, we calculated for each parameter the absolute error: the absolute difference
between the posterior median of the parameter and the true value used to simulate the dataset. For each parameter we then calculate the mean absolute error over the 5 simulated datasets. We plot in Figure 6, for each parameter, the mean absolute error over the 5 simulated datasets, as $N$ varies. Note that since the mean absolute errors are means of absolute differences, and thus must be positive.

The set of parameters we used was simply one that was not pathological. We used $b_A = 1, b_B = 2, b_C = 3, \theta = 0.1, p = 0.3, \phi_A = \phi_B = \phi_C = \phi_M = 0.01$, and $\mu_A = 0.4, \mu_B = 0.7, \mu_C = 5$. We assume only 1 feature, which for each sample is generated from a unit normal distribution. For inference, we set $s_A = s_B = s_C = 1, \lambda_A = \lambda_B = \lambda_C = 10$ and $\lambda_M = 10$. To obtain posterior samples, we used Stan (Hoffman and Gelman, 2011), obtaining 2500 samples from each of 4 chains with no thinning, and using 2500 burn-in steps for each chain. We assessed convergence both by using the Gelman statistic (Gelman and Rubin, 1992) and visual examination of the traces for each parameter. We checked that in fitting the model to each simulated dataset, the maximum Gelman statistic over parameters was less than 1.2. The meaning of errors in the regression parameter is provided by Equations 4.1-4.3, and the meaning of errors in the spread parameters $\phi_A, \phi_B, \phi_C, \phi_M$ is provided by the plots of beta and gamma distributions in Figures 16 and 17 of the Appendix.

For second experiment, for various $M$, we did the following: We simulated a single dataset by first simulating observed function values for 5000 samples using the same parameters as in the first experiment. However, then we added to the dataset an additional $M$ noise samples whose observed function values which were not generated from the model. Instead, for those $M$ noise

Fig 6: For each parameter, the mean absolute error over the simulated datasets decreases with the size of the simulated datasets.
Fig 7: For each parameter, the magnitude of the difference between posterior median parameter and true value increases with amount of noise data.

samples, at each time point for which the function value is observed, we drew the observed value uniformly from the unit interval. For inference, we used the same hyperparameters, sampling method, and convergence diagnostics as before. We then calculated, for each parameter, the error: the signed difference between its median posterior distribution and the true value used to simulate (part of) the dataset. In Figure 7, we plot for each parameter this signed difference as $M$, the number of noise data samples, varies. As expected, the magnitude of this error is increasing in $M$. Note that for each $M$, we simulate only 1 dataset, and that we calculate for each parameter the error which can be negative, and not absolute error (as done for the first simulation experiment). Also, we display in dotted lines the 25-th and 75-th percentiles of the error for each parameter.

6. Analysis of Prostate Cancer Dataset. Here we describe the dataset we study and perform an exploratory analysis of it to identify potentially useful patient covariates. Having chosen informative covariates, we use our model to uncover the relationship between those covariates and recovery trajectory, and study our model’s predictive performance through cross validation.

6.1. Dataset Description. Our data comes from a study (Gore et al., 2009, 2010) that prospectively tracked the sexual function as measured using the UCLA Prostate Cancer Index (Litwin et al., 1998) of 307 patients who underwent radical prostatectomy to treat clinically localized prostate cancer. Their sexual function levels were collected right before treatment and over a 48-month post-treatment study period. Their sexual function levels were requested via surveys sent to patients at 1,2,4,8,12,18,24,30,36,42, and 48
months after their respective treatment dates, and missing data was due to lack of survey response. The Prostate Cancer Index, derived from answers to a series of multiple-choice questions, is a numerical measure of a patient’s level of sexual function that lies between 0 and 100, which we scale to the unit interval for convenience. Various patient covariates were collected at time of treatment, including age, cancer grade/stage, physical and mental condition, urinary and bowel function, and comorbidity count.

Prior to performing any analysis, we filtered out patients using 4 criteria based on the fourth author’s urology training and experience. The filtered out patients are those that in his judgement, are not realistic and whose responses are not to be trusted. As our ultimate goal is for our model to be used by medical professionals, we opt to defer to their judgement of what constitutes realistic data, instead of attempting to model data they deem unrealistic. First, we removed 10 patients whose pre-treatment value was less than 0.1, as the recovery curve of patients whose sexual function is very low before prostatectomy is not meaningful. Then, we filtered out patients who failed at least one of the following three criteria: Firstly, we removed patients for whom function level was available at less than 6 timepoints (44 patients). Secondly, we removed patients whose representative curve, obtained by fitting a curve (under least squares error, where the asymptotic level was not constrained) to their data points, was, at 48 months post-treatment (the last time for which data was available), higher than their pre-treatment value (50 patients). We did so because prostatectomy should not have a positive effect on sexual function. Thirdly, we removed patients who at any point reported 3 or more function values of 0 in a row, as we do not trust the reporting of patients who report too many minimum scores (21 patients). Doing so reduced the number of patients from 307 to 184. (Note that some patients failed more than one of the three criteria, so that the sums of the number of patients removed for each criteria does not equal the total number of filtered out patients.)

6.2. Choosing Features. To identify potential correlates of recovery curve shapes, for every patient with a sufficient number of data points, we used curve fitting to find the $A, B, C$ parameters corresponding to their post-event recovery shapes. We made scatter plots of each of those parameters against all available covariates to identify ones that correlated with curve parameters, and identified the pre-treatment sexual function level (referred to as “init” in all figures) and patient age (at treatment time) to be the 2 covariates most strongly correlated with curve parameters. From the scatter plots shown in Figure 8, it can be seen the relationship between those
Fig 8: The three curve parameters show some dependence on a patient’s age and pre-treatment sexual function value.

2 covariates and curve parameters is likely nonlinear. Thus we decided to create 1 categorical feature based on whether a patient’s age was less than 55 years, between 55 and 65, or older than 65, and a second categorical feature based on whether a patient’s pre-treatment sexual function level was less than 0.41, between 0.41 and 0.60, between 0.60 and 0.80, or greater than 0.80. Such subdivisions matches with the fourth author’s clinical experience regarding how urologists categorize age and pre-treatment function level, and were able to account for variance in patient recovery trajectories.

Thus, in our model, patients belong to 1 of 12 classes, depending on into which of the three age groups they fall into, and which of the four intervals their pre-treatment sexual function level lies. All these features were normalized for all analysis, and we added a bias feature that was not normalized.

To visualize the effect of these 2 covariates on recovery shape from another view, we stratified the patients by age category and pre-treatment sexual function level category, and plotted in Figure 10 the average shape of the

| covariate         | bins                        |
|-------------------|-----------------------------|
| age               | (0.55), (55,65), (65,∞)     |
| pre-treatment level| (0.0.41),(0.41,0.60),(0.60,0.80),(0.80,1.0) |

Fig 9: Based on scatter plots of curve parameters against patient covariates, we decided to use categorical features based on patient age and pre-treatment sexual function level.
Fig 10: Plotting the average scaled function values over the dataset stratified by age and pre-treatment sexual function value suggests the dependence of latent recovery shapes on age and pre-treatment value.

6.3. Modeling Scaled Function Values. In general there are 2 potential approaches to modeling patients’ absolute function values - we can do so directly, or instead model their scaled function value (absolute function value divided by pre-treatment function value), and then scale it by their pre-treatment value. As we take the latter approach, we examine the dataset to justify doing so, showing that naive models that model scaled function value have superior in-sample performance to their unscaled analogue. The measure of performance we consider here and throughout the rest of the paper is loss as given by average absolute prediction error.

First, we consider a baseline model where at each timepoint, each patient is predicted to possess the average absolute function value (over the dataset) at that timepoint (labeled “average value”). We compare its in-sample predictive performance (fit) to a model where each patient is predicted to have the average scaled function value at that timepoint (labeled “average scaled value”) in Figure 11a. (These 2 models amount to predicting each patient to have the average curve and average shape, respectively, shown in Figure 1a.) Next, we plot in Figure 11b the in-sample predictive performance of two analogous models that use separate generalized linear regression models at each of the 11 common timepoints to model a patient’s absolute and scaled function value, respectively, as a function of the features shown in Figure 9. In other words, one model (labeled “regression”) regresses the absolute func-
Fig 11: In-sample performance is better when using models that model scaled function value as opposed to absolute function value.

6.4. Fitting Our Model. Now, we describe how we chose hyperparameters, and the fitting of the model. Because we use cross-validation to estimate several parameters, we must first describe how we measure the out-of-sample predictive performance of a model.

We use 5-fold cross-validation to assess out-of-sample performance, with the loss function being absolute prediction error: the absolute difference between $y_t^{(i)}$ and a point estimate of, it $\hat{y}_t^{(i)}$. With our model, we use the median of the posterior distribution of $f_t^{(i)}$ as a point estimate of $y_t^{(i)}$. The average loss over time can then be plotted.

To choose $\mu_A, \mu_B, \mu_C$, which describe the average recovery shape for the target population, we fit a recovery shape using our parametric form to the training fold-wide average scaled function value shown in Figure 1a (labeled “average shape”). For this analysis because there were not enough data to reliably estimate the variance parameters $\phi_A, \phi_B, \phi_C$, we determined those parameters by treating them as hyperparameters. Then, the hyper-
Fig 12: The prior vs posterior predictive distribution for 2 patients.

parameters of the model are $\phi_A, \phi_B, \phi_C, s_A, s_B, s_C, l_M$. The out-of-sample predictive performance we report in Section 6.5 is that for the best set of hyperparameters, which we obtained via grid search. While these hyperparameters could have been chosen via nested cross-validation, we show in Figure 18 that out-of-sample performance is not sensitive to the particular choice of those hyperparameters. The best setting of parameters $\phi_A, \phi_B, \phi_C$ was $\phi_A = 0.3, \phi_B = 0.3, \phi_C = 0.8$ and that of the hyperparameters was $s_A = s_B = s_C = 1.0$ and $l_M = 10$.

To fit the model, we used Stan (Hoffman and Gelman, 2011), for each of 4 chains, running 2500 steps with 2500 burn-in steps and no thinning, and assessed convergence using the Gelman statistic (Gelman and Rubin, 1992) (The maximum value of the Gelman statistic over all parameters was 1.11).

To assess model fit, we checked that the observed data were likely under the posterior predictive distribution. To understand the impact of the data on the posterior predictive position, we illustrate for two patients their prior and posterior distribution over recovery curves in Figure 12.

6.5. Out-of-Sample Performance. As alluded to in Section 6.4, we performed 5-fold cross-validation, obtaining for each test sample point predictions from our model as previously described, and examined the average, over the test folds, of loss as measured by absolute prediction error. (The average loss at time $t$ for a test fold consisting of patient index set $I$ is $\sum_{i \in I} |\hat{y}_i^t - y_i^t|$ where $\hat{y}_i^t$ is the point prediction of the function level of patient $i$ at time $t$.) In particular, the entire time series for patients in the testing folds are predicted given the entire time series of the patients in the training fold. Data from the early part of one patients time series is not used to predicted the same patients future values. We plot, over time, the
out-of-sample performance of our model, as well as that of several baseline models in Figure 13. To compare the improvement of our method to the status quo, in which a doctor merely tells a patient the population-wide average shape, we plotted the performance of simply predicting a patient to have the average recovery shape. This is the “scaled average value” model described in Section 6.3. We compared the performance of our model to one that, at each of the 11 common timepoints, uses a separate generalized linear regression model to relate the scaled function value at the timepoint to patient features. This is the “scaled regression” model described in Section 6.3; it uses a logistic inverse link function and assumes a normal response distribution. Finally, because of the high variance in our data, we show for comparison the in-sample performance of a model that is prone to overfitting. This model, labelled “median”, in order to make a prediction for a patient at a given time, looks at which of the 12 patient classes the patient belongs to and then calculates the median scaled function value, at the given time, of patients who belong to that patient class, over the entire dataset. The predicted value for that time is then that median scaled function value multiplied by the patient’s own pre-treatment function value. As can be seen in Figure 13, the out-of-sample performance of our model is roughly equivalent to that of the scaled regression model. Error bars show the variance in our estimates of the expected loss at each time.

6.6. Interpretability of Model. Our model achieves out-of-sample performance comparable to that of the timewise scaled regression described in Section 6.3. However, our model produces much more believable predictions, outputting a distribution over time series consisting solely of recovery curves,
so that they are smoothly increasing monotonically towards an asymptote, and do not exceed the pre-treatment value. On the other hand, the timewise scaled regression model produces a time series that is not guaranteed to be smooth or monotonically increasing. Because such an output does not look like a realistic recovery curve, these methods would not be able to be used directly in practice, whereas our curves by design would be more useful. To illustrate, in Figure 14 for several of the 12 classes of patients, we plot the scaled function values produced by scaled timewise regression, the distribution over $f_t$ from our model, and the timewise median of that distribution.

Furthermore, we have designed our model so that prediction uncertainty is easily interpretable when a patient’s posterior distribution of curves is plotted. Because we encourage the a patient’s recovery curve parameter distribution to be unimodal in the posterior, we expect the pointwise distribution of curve values, namely that of $f_t$, to be unimodal. This is why in Figure 14, the distribution of curves appears clustered about the red curve. It is important that one can visually extract from a plot of posterior distribution of curves a single most likely curve. Then, such a plot can be interpreted as giving a single curve prediction, along with the uncertainty in that prediction. On the other hand, if the posterior distribution of curves were clustered around, say, two curves, there would be no clear interpretation of the results. Such a scenario could certainly occur if we did not explicitly avoid it, because we are making predictions in a closed interval, so that the probability mass in the posterior predictive distribution could accumulate at the boundaries of that interval and produce an unrealistic second peak in the posterior.

6.7. Dependence of Recovery Shape on Covariates. Our analysis teases apart the dependence of recovery shape on age and pre-treatment value. In Figure 15, we show the predicted recovery shapes of all 12 classes of patients (the plotted time series are the median of their respective posterior predictive distributions). We stratify those recovery shapes first by pre-treatment level, to show the dependence of recovery shape on age when pre-treatment level is controlled for. We then stratify those shapes by age, to show the dependence of recovery shape on pre-treatment level when age is controlled for. We also fit our model using one categorical feature at once to see the dependence of recovery shape on each covariate, separately, without controlling for the other covariate.

In Figure 15a, we examine the effect of patient age on recovery curve shape. We find that when pre-treatment level is controlled for, patients younger than 55 years of age have a smaller asymptotic drop in sexual
Fig 14: The posterior predictive distribution over recovery curves (black) and timewise medians of it (red) convey more plausible predictions than that of timewise scaled regression (blue), whose prediction is not guaranteed to be a recovery curve.

function level, proportional to their pre-treatment level, as can be seen in Figure 15a. (To support this, we also performed a one-sided z-test that the scaled function value at 48 months for patients younger than 55 years of age was larger than those older than 55 years of age, and obtained a p-value of .034.) However, this effect is diminished for patients with pre-treatment level higher than 0.80. When pre-treatment level is not controlled for, the asymptotic proportional drop in function level for patients in the 3 different age ranges we consider are distinct, with a younger age corresponding to a smaller asymptotic proportional drop. In both cases, the initial drop in function level, proportional to pre-treatment level, appears to not depend on patient age.

In Figure 15b we examine the effect of pre-treatment sexual function level on recovery shape. We find that when age is controlled for, patients with pre-treatment level higher than 0.80 have a smaller asymptotic drop in sexual function level, proportional to their pre-treatment level. (To support this, we also performed a one-sided z-test that the scaled function value at 48 months for patients with pre-treatment value above 0.8 was higher than those with pre-treatment value below 0.8, and obtained a p-value of .0008.) However, this effect is diminished for patients younger than 55 years of age. The initial drop in function level, proportional to pre-treatment level, appears to depend very mildly on pre-treatment level. The rate of recovery shows some dependence on pre-treatment level, though not in any monotonic manner.
6.8. **Comparison to Past Prostate Cancer Studies.** Unlike past methods, which have mostly focused on modeling a continuous or binary measure of sexual function at a single fixed time, our model makes predictions of the entire post-treatment function trajectory. Regardless, we can still compare our findings to them. Past work that modeled a continuous measure of sexual function found that lower age and higher pre-treatment sexual function level are statistically linked to higher absolute levels of that measure (Talcott et al., 2003), and that lower age is linked to a smaller change in that measure of function level (Sanda et al., 2008). Likewise, when a binary indicator of satisfactory sexual function has been logistically regressed against patient covariates, lower age (Ayyathurai et al., 2008; Regan et al., 2011) and higher pre-treatment function level (Regan et al.) have been found to lead to a higher probability of having satisfactory sexual function. One can conclude from these past statistical analyses, as well as model-free data analyses (Michl et al., 2006; Rabbani et al., 2000), that lower age and higher pre-treatment sexual function level, by any measure, are linked to higher post-treatment sexual function level, which is in agreement with our findings.

We stress that unlike any previous analysis, we model the dependence of the longitudinal sexual function levels proportional to the pre-treatment level on patient features. This means, for example, that our findings that the proportional initial drop relative to the pre-treatment level does not depend on the pre-treatment level still means that the absolute initial drop does depend on pre-treatment level. Furthermore, we also emphasize that

![Diagram](image-url)
the uncertainty associated with our predictions is fairly high. We see this from the posterior distribution of the recovery shapes shown in Figure 14. However, this alludes one of the strengths of our model - that the uncertainty in predictions can be easily visually interpreted.

7. Conclusion. We have presented a natural Bayesian model that can be used to predict recovery curves, which arise in many medical contexts. Our overarching goal is to facilitate the flow of information from the data to the user, who may not be statistically inclined. Towards this end, we impart interpretability to both the model and its output, because we believe our model will be more accepted by medical practitioners if it can be easily explained to the layman, and if its output is believable. In particular, our model predicts quantities that are of natural interest, and guarantees that its output is in fact the recovery curve that we assume a domain expert to expect of a prediction. Furthermore, our model is designed for easy visualization of predictions and the uncertainty associated with them, as we encourage the posterior distribution over recovery curves to have a clear mode.

We have used our model to analyze the impact of prostatectomy on a patient’s post-treatment sexual function trajectory, and have characterized the extent of that impact on patient age and pre-treatment sexual function level, producing conclusions that both agree with and supplement past findings. Many medical treatments affect some measure of well being, whose quantified post-treatment trajectory is suitable to be modeled as a recovery curve. We have shown the promise of our model in producing insights in one specific domain, and believe it can produce similar benefits to doctors and patients in the context of other medical treatments.

Acknowledgements. This work was supported by National Science Foundation CAREER grant IIS-1053407 to C. Rudin. We thank Dr. Jim Michaelson of Massachusetts General Hospital for helpful discussions.

APPENDIX

Beta and Gamma Distribution Parameterizations. We show in Figures 16 and 17 how the distribution of the beta and gamma distributions change as the spread parameter varies.

Sensitivity of Out-Of-Sample Performance to Hyperparameters. We provide a sense in Figures 18a and 18b how out-of-sample performance changes with the values of hyperparameters $s_A, s_B, s_C, l_M$ and the fixed values of parameters $\phi_A, \phi_B, \phi_C$. The joint setting of those parameters and hyperparameter values giving optimal out-of-sample performance was with
Fig 16: Illustrations of our parameterization of the beta distribution

parameters $\phi_A = 0.3, \phi_B = 0.3, \phi_C = 0.8$ and hyperparameters. $s_A = s_B = s_C = 1.0$ and $l_M = 10$. In Figure 18a, we tie $\phi_A = \phi_B$, and illustrate how out-of-sample performance changes as the tied value of $\phi_A$ and $\phi_B$ changes, with the remaining parameters and hyperparameters set to their values at the aforementioned joint optimum. In Figure 18b, we tie $s_A = s_B = s_C$ and illustrate how out-of-sample performance changes as the tied value of $s_A, s_B, s_C$ changes, with the remaining parameters and hyperparameters set to their values at the aforementioned joint optimum.

References.

Ayyathurai, R., Manoharan, M., Nieder, A. M., Kava, B. and Soloway, M. S. (2008). Factors affecting erectile function after radical retropubic prostatectomy: results from 1620 consecutive patients. BJU international 101 833–6.

Briganti, A., Gallina, A., Suardi, N., Capitanio, U., Tutolo, M., Bianchi, M., Salonia, A., Colombo, R., Di Girolamo, V., Martinez-Salamanca, J. I., Guazzoni, G., Rigatti, P. and Montorsi, F. (2011). What is the definition of a satisfactory erectile function after bilateral nerve sparing radical prostatectomy? The Journal of Sexual Medicine 8 1210–7.

Cai, B. and Dunson, D. B. (2007). Bayesian Multivariate Isotonic Regression Splines. Journal of the American Statistical Association 102 1158–1171.

Denison, D. G. T., Mallick, B. K. and Smith, A. F. M. (1998). Automatic Bayesian Curve Fitting. Journal of Royal Statistical Society Series B.

Descazesaud, A., Debré, B. and Flam, T. A. (2006). Age Difference Between Patient and Partner is a Predictive Factor of Potency Rate Following Radical Prostatectomy. The Journal of Urology 176 2594–2598.

Eastham, J. A., Scardino, P. T. and Kattan, M. W. (2008). Predicting an Optimal Outcome After Radical Prostatectomy: The Trifecta Nomogram. The Journal of Urology 179 2207–2211.
Fig 17: Illustrations of our parameterization of the gamma distribution

Gelman, A. and Rubin, D. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science* **7** 457–511.

Gore, J. L., Kwan, L., Lee, S. P., Reiter, R. E. and Litwin, M. S. (2009). Survivorship beyond convalescence: 48-month quality-of-life outcomes after treatment for localized prostate cancer. *Journal of the National Cancer Institute* **101** 888–92.

Gore, J. L., Gollapudi, K., Bergman, J., Kwan, L., Krupski, T. L. and Litwin, M. S. (2010). Correlates of bother following treatment for clinically localized prostate cancer. *The Journal of Urology* **184** 1309–15.

Hoffman, M. D. and Gelman, A. (2011). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research* **30**.

Jung, T. and Wickrama, K. A. S. (2008). An Introduction to Latent Class Growth Analysis and Growth Mixture Modeling. *Social and Personality Psychology Compass* **2** 302–317.

Litwin, M. S., Hays, R. D., Fink, A., Ganz, P. A., Leake, B. and Brook, R. H. (1998). The UCLA Prostate Cancer Index: development, reliability, and validity of a health-related quality of life measure. *Medical Care* **36** 1002–12.

Mammen, E. (1991). Estimating A Smooth Monotone Regression Function. *The Annals of Statistics* **19** 724–740.

Michl, U. H. G., Friedrich, M. G., Graefen, M., Haese, A., Heinzer, H. and Huland, H. (2006). Prediction of postoperative sexual function after nerve sparing radical retropubic prostatectomy. *The Journal of Urology* **176** 227–31.

Neelon, B. and Dunson, D. B. (2004). Bayesian isotonic regression and trend analysis. *Biometrics* **60** 398–406.

Potosky, A. L., Davis, W. W., Hoffman, R. M., Stanford, J. L., Stephenson, R. A., Penson, D. F. and Harlan, L. C. (2004). Five-Year Outcomes After Prostatectomy or Radiotherapy for Prostate Cancer: The Prostate Cancer Outcomes Study. *Journal of the National Cancer Institute* **96** 1358–67.

Rabbani, F., Stapleton, A. M., Kattan, M. W., Wheeler, T. M. and Scardino, P. T. (2000). Factors predicting recovery of erections after radical prostatectomy. *The Journal of Urology* **164** 1929–34.
(a) Out-of-sample performance as the tied values of $\phi_A, \phi_B$ vary. The labels indicate the tied value of $\phi_A, \phi_B$.

(b) Out-of-sample performance as the tied values of $s_A, s_B, s_C$ vary. The labels indicate the tied value of $s_A, s_B, s_C$.

Fig 18: Sensitivity of out-of-sample performance to hyperparameters.

Ramsay, J. O. and Silverman, B. W., eds. (2002). *Applied Functional Data Analysis: Methods and Case Studies. Springer Series in Statistics*. Springer New York, New York, NY.

Regan, M. M., Cooperberg, M. R., Wei, J. T., Michalski, J. M., Sandler, H. M., Litwin, M. S., Klein, E., Kibel, A. S., Hamstra, D. A., Pisters, L. L., Kuban, D. A., Kaplan, I. D., Wood, D. P., Ciezki, J., Dunn, R. L., Carroll, P. R. and Sanda, M. G. (2011). Prediction of Erectile Function Following Treatment for Prostate Cancer. *Journal of the American Medical Association* 306 1205–1214.

Resnick, M. J., Koyama, T., Fan, K.-H., Albertsen, P. C., Goodman, M., Hamilton, A. S., Hoffman, R. M., Potosky, A. L., Stanford, J. L., Stroup, A. M., Van Horn, R. L. and Penson, D. F. (2013). Long-term Functional Outcomes After Treatment for Localized Prostate Cancer. *The New England Journal of Medicine* 368 436–45.

Rogosa, D. R. and Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika* 50 203–228.

Rolfe, M. I., Mengersen, K. L., Vearncombe, K. J., Andrew, B. and Beadle, G. F. (2011). Bayesian estimation of extent of recovery for aspects of verbal memory in women undergoing adjuvant chemotherapy treatment for breast cancer. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 60 655–674.

Sanda, M. G., Dunn, R. L., Michalski, J., Sandler, H. M., Northouse, L., Hembroff, L., Lin, X., Greenfield, T. K., Litwin, M. S., Saigal, C. S., Mahadevan, A., Klein, E., Kibel, A., Pisters, L. L., Kuban, D., Kaplan, I., Wood, D., Ciezki, J., Shah, N. and Wei, J. T. (2008). Quality of Life and Satisfaction with Outcome among Prostate-Cancer Survivors. *The New England Journal of Medicine* 358 1250–1261.

Shively, T. S., Sager, T. W. and Walker, S. G. (2009). A Bayesian approach to non-parametric monotone function estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71 159–175.
Talcott, J. A., Manola, J., Clark, J. A., Kaplan, I., Beard, C. J., Mitchell, S. P., Chen, R. C., O’Leary, M. P., Kantoff, P. W. and D’Amico, A. V. (2003). Time Course and Predictors of Symptoms after Primary Prostate Cancer Therapy. *Journal of Clinical Oncology* **21** 3979–86.

Tilling, K., Sterne, J. A. C. and Wolfe, C. D. A. (2001). Multilevel growth curve models with covariate effects: application to recovery after stroke. *Statistics in Medicine* **20** 685–704.

Warschausky, S., Kay, J. B. and Kewman, D. G. (2001). Hierarchical Linear Modeling of FIM Instrument Growth Curve Characteristics after Spinal Cord Injury. *Archives of Physical Medicine and Rehabilitation* **82** 329–34.