The Annals of Applied Statistics
2015, Vol. 9, No. 3, 1549–1570
DOI: 10.1214/15-AOAS840
© Institute of Mathematical Statistics, 2015

COMBINING NONEXCHANGEABLE FUNCTIONAL OR SURVIVAL DATA SOURCES IN ONCOLOGY USING GENERALIZED MIXTURE COMMENSURATE PRIORS

BY THOMAS A. MURRAY*,[1], BRIAN P. HOBBS*,[1,2] AND
BRADLEY P. CARLIN†,[1]

University of Texas MD Anderson Cancer Center and University of Minnesota†*

Conventional approaches to statistical inference preclude structures that facilitate incorporation of supplemental information acquired from similar circumstances. For example, the analysis of data obtained using perfusion computed tomography to characterize functional imaging biomarkers in cancerous regions of the liver can benefit from partially informative data collected concurrently in noncancerous regions. This paper presents a hierarchical model structure that leverages all available information about a curve, using penalized splines, while accommodating important between-source features. Our proposed methods flexibly borrow strength from the supplemental data to a degree that reflects the commensurability of the supplemental curve with the primary curve. We investigate our method's properties for nonparametric regression via simulation, and apply it to a set of liver cancer data. We also apply our method for a semiparametric hazard model to data from a clinical trial that compares time to disease progression for three colorectal cancer treatments, while supplementing inference with information from a previous trial that tested the current standard of care.

1. Introduction. Statistical investigations begin by determining which sources(s) of information will be used to answer the motivating questions and generate hypotheses for future exploration. Conventional approaches to statistical inference preclude structures that facilitate incorporation of partially informative data, imposing polarity on the data selection process. Putatively relevant supplemental data acquired from broadly similar therapeutic interventions, patient cohorts, previous investigations or biological processes are often either excluded from statistical analysis or treated as exchangeable with the primary data. In oncology, prospective studies designed to evaluate the performance of an experimental therapy usually ignore historical information about the control therapy and limit enrollment to patients presenting lesions with a particular histological subtype, grade or performance status, or who are naïve to prior therapy. In contrast, studies

---

Received March 2014; revised March 2015.

[1] Supported in part by NCI Grant 1-R01-CA157458-01A1.

[2] Supported in part by Cancer Center Support Grant (CCSG) (P30 CA016672).

Key words and phrases. Bayesian hierarchical model, clinical trials, colorectal cancer, commensurate prior, computed tomographic imaging, evidence synthesis, mixture priors, penalized splines, proportional hazards, semiparametric methods.

with more liberal information inclusion rules, such as intention-to-treat designs, often pool information from potentially heterogeneous sources or adjust for this potential heterogeneity using simple linear regressors.

Ignoring relevant, supplemental sources of information reduces the reproducibility and scope of the primary study. On the other hand, using the supplemental information while neglecting to account for heterogeneity between the sources of information obscures understanding of the complex underlying mechanisms that produced the primary data and may lead to severely biased inference. Relaxing this dichotomy would improve the efficiency of the experimental process and enable investigators to implement statistical models that use all available information while accommodating important between-source features. Several models have been proposed for incorporating partially informative supplemental data. Pocock (1976) used generalized linear models with static, data-independent borrowing using a prespecified amount of between-source variability. Ibrahim and Chen (2000) proposed data-independent or dynamic nonhierarchical methods for partially weighting likelihoods. Bayesian [Smith, Spiegelhalter and Thomas (1995)] and frequentist [Doi, Barendregt and Mozurkewich (2011)] methods using hierarchical modeling have been developed for estimating between-source variability for univariate observables or repeated measures with generalized linear relations among covariates.

Bayesian hierarchical models facilitate dynamic partial pooling of between-source information and flexibly estimate the extent of borrowing. Here the goal is limiting bias for estimating primary effects when between-source heterogeneity occurs, while improving efficiency when approximate coherence, or *commensurability*, occurs; see Hobbs, Sargent and Carlin (2012) for generalized linear mixed models, and Hobbs, Carlin and Sargent (2013) and Murray et al. (2014) for piecewise-exponential time-to-event models. In Section 2 we discuss these methods, which can be used to borrow strength from supplemental data in parametric models for regression coefficients and other univariate parameters. We then develop a Bayesian hierarchical model structure to leverage supplemental information more generally, including for both nonparametric regression and semiparametric hazard models with penalized splines. The literature appears devoid of general methods for flexibly borrowing strength from supplemental data in semi- and nonparametric models for a group of related parameters that characterize a complex object (e.g., a curve or surface).

Two oncological applications motivate our methodological developments. The first, from diagnostic radiology, involves estimating prognostic functional imaging biomarkers acquired using perfusion computed tomography (CT) in cancerous and cancer-free liver tissue. Perfusion CT is an emerging technology that enables observation and quantification of characteristics pertaining to the passage of fluid through blood vessels. Researchers have developed physiological models to quantify a variety of perfusion characteristics derived from analysis of the distribution of contrast enhancement in tissue acquired using repeated CT scans during intravenous administration of contrast medium. Investigators have used the technology in a number of organs and tumors, including prostate, colorectal, head and neck, lung and liver. Miles and Griffiths (2003) review the clinical relevance of perfusion CT.

Our application considers three characteristics: permeability-surface area product (PS), blood volume (BV), and blood flow (BF). Each characteristic is measured at 7 to 13 acquisition times between 11 and 95 seconds following contrast injection. Data from 16 individuals comprise a total of 25 regions containing pathology-verified metastases to the liver from neuroendocrine tumors (i.e., cancerous liver tissue), and a total of 27 regions consisting of noncancerous liver tissue. PS and BF are rates measured as milliliters per minute per 100 grams of liver tissue (ml/min per 100 g), whereas BV is a volume measured as milliliters per 100 grams of liver tissue (ml per 100 g). Figure 1 displays the observed perfusion CT (CTp) curves along with Loess estimates of the average CTp curve for each characteristic in each tissue region. Characterization of the perfusion characteristics in cancerous tissue has implications for constructing biomarkers to assist in treatment monitoring, prognostication and pathophysiological understanding of metastatic tumor

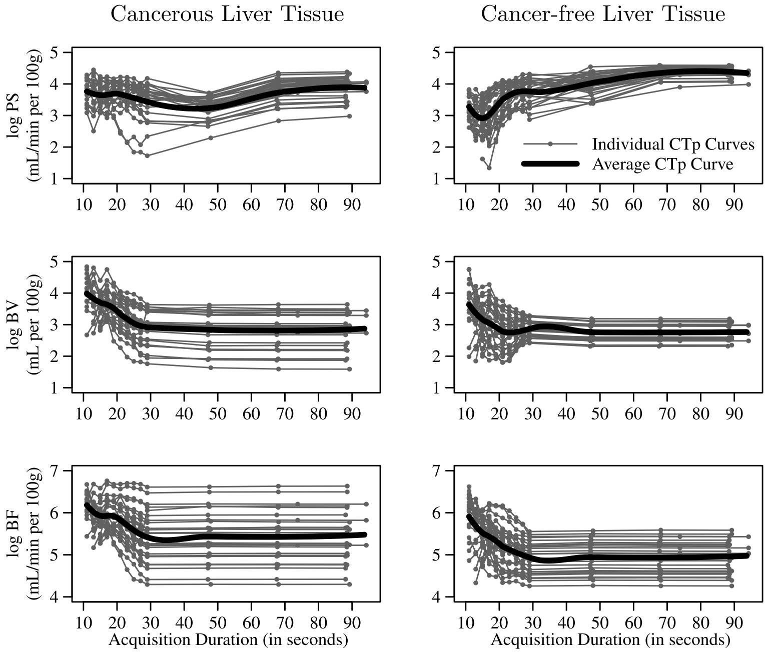

FIG. 1. *Individual-by-region perfusion CT (CTp) curves are displayed by tissue region ("Individual CTp curves") for permeability-surface area product (PS), blood volume (BV) and blood flow (BF), over acquisition durations between* 11 *and* 95 *seconds after contrast injection. The dots mark actual observations. The thick "Average CTp curve" is a Loess estimate that ignores potential clustering.*

vasculature. It is important that the acquisition duration cover the range of CTp curve instability, but once the curve has stabilized, the scan can be terminated. To limit radiation exposure and cost, the acquisition duration should be minimized [Ng et al. (2013)]. Critically, because the tissue type is unknown prior to diagnosis, any proposed acquisition period must ensure stable quantification of CTp characteristics for both types of tissue before CTp can be used for detection of metastatic sites. Though nonexchangeable, the perfusion CT data obtained in cancer-free liver tissue may inform the shape and stabilization time of the corresponding CTp curve in cancerous tissue. Figure 1 shows substantial heterogeneity in the shapes of the average CTp curves by tissue region for PS, whereas for BV and BF, the average CTp curves are more similar.

Our second application, from colorectal cancer, involves estimating progression-free survival (PFS) with data from two consecutive randomized phase III colorectal cancer trials, reported by Saltz et al. (2000) and Goldberg et al. (2004). Both trials used PFS to assess the efficacy of various treatment regimens for patients with previously untreated metastatic colorectal cancer; disease progression was defined as a 25% increase in measurable tumor size, presence of a new lesion or death. The initial trial [Saltz et al. (2000)] compared three treatment regimens: 5-Fluorouracil and Leucovorin, Irinotecan alone, and Irinotecan and bolus Fluorouracil plus Leucovorin (IFL). The results indicated that the IFL regimen was significantly more efficacious than the other two regimens, and IFL became the "standard of care" leading into the subsequent trial [Goldberg et al. (2004)], which then compared an identical IFL regimen with two novel regimens: Oxaliplatin with infused Fluorouracil plus Leucovorin (FOLFOX), and Irinotecan with Oxaliplatin (IROX).

Figure 2 shows that the PFS curves for the IFL regime are commensurate in the Saltz and Goldberg trials, with the PFS curve in the Goldberg trial tracking just above that of the Saltz trial, though within the 95% CIs for nearly all of follow-up. Moreover, Figure 2 suggests that in the Goldberg trial FOLFOX is superior to both IROX (log-rank test p-value of 0.006) and IFL (log-rank test p-value < 0.001), and that IFL and IROX perform similarly (log-rank test p-value of 0.404). The efficiency for estimating the PFS curve of the IFL regimen in the Goldberg trial may be improved by using nonexchangeable, yet relevant data on the IFL regimen in the Saltz trial.

The commonality between our two motivating examples is the availability of supplemental information about unknown curves that may aid our inference in the primary investigation (i.e., CTp curves among tissue types, and PFS curves of IFL among clinical trials). In Section 3 we investigate the borrowing properties via simulation of the proposed hierarchical model structure for nonparametric regression using penalized splines, and then we analyze the perfusion CT data. In Section 4 we analyze data from the colorectal cancer clinical trials using the proposed hierarchical structure for a semiparametric hazard model. In Section 5 we close with a discussion and propose directions for future work.

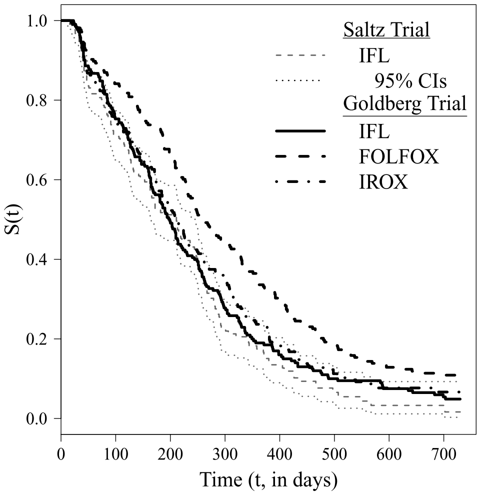

FIG. 2. *Kaplan–Meier estimates of PFS for the IFL regimen in the Saltz trial* (*along with* 95% *CIs*), *and the IFL*, *FOLFOX and IROX regimes in the Goldberg trial.*

2. Leveraging supplemental information. We restrict our attention to two Bayesian methods for leveraging supplemental information, power priors [Ibrahim and Chen (2000)] and commensurate priors [Hobbs et al. (2011)]. In general, Bayesian models consist of a likelihood for the data and prior specifications for the parameters in the likelihood [see, e.g., Carlin and Louis (2009)]. For both these modeling approaches, the *primary* and *supplemental* likelihoods are assumed to have the same structure (e.g., both Gaussian), and we denote them by $L(\boldsymbol{\theta}|\mathbf{D})$ and $L(\boldsymbol{\theta}_0|\mathbf{D}_0)$, respectively. As a result, the supplemental parameter $\boldsymbol{\theta}_0$ is analogous to the primary parameter $\boldsymbol{\theta}$, and information about $\boldsymbol{\theta}_0$ can be leveraged to aid posterior inference about $\boldsymbol{\theta}$.

2.1. *Existing methods.* Power priors assume the sources of information are exchangeable (i.e., $\boldsymbol{\theta} \equiv \boldsymbol{\theta}_0$) and downweight the supplemental likelihood by raising it to a prespecified power $a_0 \in [0, 1]$. This strategy works because raising the supplemental likelihood to the power a_0 diffuses it. In the extreme case where $a_0 = 0$, the downweighted supplemental likelihood is a noninformative constant. Formally, the posterior arises as

$$p(\boldsymbol{\theta}|\mathbf{D}, \mathbf{D}_0) \propto L(\boldsymbol{\theta}|\mathbf{D}) L(\boldsymbol{\theta}|\mathbf{D}_0)^{a_0} \pi(\boldsymbol{\theta}). \tag{1}$$

Alternatively, a_0 can be modeled as another unknown parameter in the model [Ibrahim and Chen (2000)]. Power priors are extremely general, so nothing explicitly prevents their use in any modeling context. However, they employ a single parameter (a_0) to control between-source borrowing, thereby making differential borrowing among components of $\boldsymbol{\theta}$ infeasible unless $L(\boldsymbol{\theta}|\mathbf{D})$ has a convenient factorization, as it generally does not. This is undesirable because the supplemental data may provide relevant information for only a subset of $\boldsymbol{\theta}$. Furthermore, when a_0 is assumed unknown, power priors tend to excessively downweight the supplemental likelihood, even when $\mathbf{D}_0$ and $\mathbf{D}$ are identical [Neelon and O'Malley (2010)].

Commensurate priors do not assume the sources of information are exchangeable (i.e., $\boldsymbol{\theta} \not\equiv \boldsymbol{\theta}_0$), and instead specify a hierarchical model where the posterior arises as

$$p(\boldsymbol{\theta}, \boldsymbol{\theta}_0, \boldsymbol{\eta} | \mathbf{D}, \mathbf{D}_0) \propto L(\boldsymbol{\theta}|\mathbf{D}) L(\boldsymbol{\theta}_0|\mathbf{D}_0) \pi(\boldsymbol{\theta}|\boldsymbol{\theta}_0, \boldsymbol{\eta}) \pi(\boldsymbol{\eta}) \pi(\boldsymbol{\theta}_0). \tag{2}$$

The *commensurate prior*, $\pi(\boldsymbol{\theta}|\boldsymbol{\theta}_0, \boldsymbol{\eta})$, "centers" $\boldsymbol{\theta}$ about $\boldsymbol{\theta}_0$, and $\boldsymbol{\eta}$ controls between-source borrowing to reflect the commensurability of the two parameters. In practice, $\pi(\boldsymbol{\theta}|\boldsymbol{\theta}_0, \boldsymbol{\eta})$ is defined so that $E[\boldsymbol{\theta}|\boldsymbol{\theta}_0, \boldsymbol{\eta}] = \boldsymbol{\theta}_0$ and $\mathrm{Var}[\boldsymbol{\theta}|\boldsymbol{\theta}_0, \boldsymbol{\eta}]$ is decreasing in $\boldsymbol{\eta}$.

For a unidimensional real-valued parameter θ, a useful approach takes $\theta|\theta_0, \eta \sim \mathcal{N}(\theta_0, \eta^{-1})$, a Gaussian distribution with mean θ_0 and precision η. Estimation of η is inherently difficult, but feasible by inducing sparsity over the precision domain using a "spike-and-slab" prior for η [Hobbs, Sargent and Carlin (2012)]. The spike-and-slab prior has a mixture density

$$\pi(\eta) \equiv (1 - p_0)\mathcal{U}(\eta|s_l, s_u) + p_0 \delta_{\mathcal{R}}(\eta), \tag{3}$$

where $0 \le s_l < s_u << \mathcal{R}$ and $p_0 \in [0, 1]$ are prespecified, and $\delta_{\mathcal{R}}(\eta)$ is one at $\mathcal{R}$ and zero otherwise. The distribution in (3) is locally uniform on the "slab," (s_l, s_u), with probability $(1 - p_0)$, and places probability mass p_0 at the "spike," $\mathcal{R}$, otherwise.

We prefer to modify the commensurate prior construction suggested by Hobbs, Sargent and Carlin (2012) as

$$\begin{aligned} \theta|\theta_0, \tau, \iota &\sim [\mathcal{N}(\theta_0, \tau^{-1})]^{(1-\iota)} [\mathcal{N}(\theta_0, \mathcal{R}^{-1})]^{\iota}, \\ & \qquad \iota \sim \mathrm{Bern}(p_0) \text{ and } \tau \sim \mathcal{U}(s_l, s_u), \end{aligned} \tag{4}$$

where $\mathrm{Bern}(p_0)$ denotes a Bernoulli distribution with $\Pr(\iota = 1) = p_0$. Therefore, the commensurate prior in (3) induces a two-part mixture prior distribution on θ that consists of a highly concentrated component [i.e., $\mathcal{N}(\theta_0, \mathcal{R}^{-1})$] and a relatively diffuse component [i.e., $\mathcal{N}(\theta_0, \tau^{-1})$], both centered at θ_0. The model in (4) has parallels to Bayesian variable selection methods, in which the two-part mixture prior for a regression coefficient has one component heavily concentrated about zero and the other vague [see, e.g., George and McCulloch (1997)]. Hence, the commensurate prior facilitates selective borrowing by inheriting the selective shrinkage property of the spike-and-slab distributions used in Bayesian variable selection. In particular, the commensurate prior strongly shrinks θ to θ_0 when evidence indicates that this difference is small, thereby improving efficiency, and minimally affects θ otherwise, thereby limiting bias.

We specify the spike so that if $\theta|\theta_0 \sim \mathcal{N}(\theta_0, \mathcal{R}^{-1})$, θ deviates negligibly from θ_0. On the other hand, we specify the slab to contain small values that correspond to modest shrinkage of θ toward θ_0. Hence, a $\mathcal{N}(\theta_0, s_u^{-1})$ prior for θ will be weakly informative. The tails of the prior distribution in (4) are jointly controlled by s_l and s_u: when $s_l \approx 0$, smaller values of s_u provide a prior with heavier tails. In our experience, posterior inference on θ is insensitive to modest shifts in these hyperparameters, so we suggest specifying s_l near zero, with s_u small and $\mathcal{R}$ large relative to the magnitude of a meaningful difference between θ and θ_0. Posterior inference *can* be sensitive to the fourth hyperparameter, p_0, the prior probability of effective equality between θ and θ_0. We recommend choosing the value for p_0 to deliver satisfactory operating characteristics at the anticipated sample size, such as mean-squared-error properties over a range of likely true differences between θ and θ_0; see Murray et al. (2014).

Specifying a value for s_u that is too large can result in a model that is not robust to between-source heterogeneity, whereas specifying a value for $\mathcal{R}$ that is too small will result in a model that gains little posterior precision for θ even when $\mathbf{D}_0$ and $\mathbf{D}$ are equivalent. Specifying too small a value for p_0 may also result in a model that borrows little when $\mathbf{D}_0 \equiv \mathbf{D}$, whereas specifying too large a value for p_0 may result in a model that borrows too much (resulting in unacceptable bias) in the presence of substantial between-source heterogeneity. Of course, p_0's influence on posterior inference diminishes as the primary sample size increases.

2.2. *Generalized mixture commensurate priors.* The general model structure described in (2) can accommodate a latent precision parameter (component of $\boldsymbol{\eta}$) for each component of $\boldsymbol{\theta}$. Therefore, the extent to which the supplemental data influence estimation of the primary effects can be allowed to differ among the components of $\boldsymbol{\theta}$, yielding flexibility and reducing bias. However, assuming mutual independence between the components of $\boldsymbol{\eta}$ allows strength to be borrowed from the supplemental source(s) differentially among the components of $\boldsymbol{\theta}$, reducing efficiency. Though this may be sensible, differential borrowing among the components of $\boldsymbol{\theta}$ may not always be appropriate.

Suppose $\boldsymbol{\beta} = (\beta_1, \dots, \beta_K)$ is a subset of $\boldsymbol{\theta}$ that characterizes a feature in the primary data (e.g., the shape of a CTp curve), with an analogous definition for $\boldsymbol{\beta}_0$. We may wish to borrow from the supplemental data for $\boldsymbol{\beta}_0$ to a degree that reflects its coherence with $\boldsymbol{\beta}$ in its entirety, rather than case-by-case for each β_k. Moreover, $\boldsymbol{\beta}$ (and $\boldsymbol{\beta}_0$) may conventionally be assigned a prior that induces smoothness. Ideally, (2) will inherit the smoothness properties of the conventional prior and simultaneously borrow similar amounts for related subgroups of $\boldsymbol{\theta}$ (i.e., $\boldsymbol{\beta}$). We propose to model

$$\begin{aligned}\beta_k|\iota_k, \beta_{k,0} &\sim \left[\pi^*(\beta_k|\beta_{k,0})\right]^{1-\iota_k}\left[\pi(\beta_k|\beta_{k,0})\right]^{\iota_k},\\ &\iota_k|\nu \sim \text{Bern}(\nu), \nu \sim \mathcal{B}(a_1, a_2), \beta_{0,k} \sim \pi^*(\beta_{0,k}),\end{aligned} \tag{5}$$

where $\mathcal{B}(a_1, a_2)$ denotes a beta distribution with mean $\frac{a_1}{a_1+a_2}$ and the specifications for π^* mirror a conventional analysis of these data. To facilitate borrowing of strength, we take $\pi(\beta_k|\beta_{0,k})$ in (5) to be heavily concentrated about $\beta_{0,k}$; we provide exact specifications for each application later. To borrow similar amounts of strength for each component of $\boldsymbol{\beta}$, we assume the ι_k are identically distributed with $\Pr(\iota_k = 1) = \nu$. Thus, the resulting prior is composed of K two-piece mixtures and generalizes the commensurate prior defined in (4). Hereafter, we refer to the general prior structure defined in (5) as a *generalized mixture commensurate* (GMC) prior.

3. GMC priors in nonparametric regression analysis. Suppose $\mathbf{y} = (y_1, \ldots, y_N)$ is a real-valued variable (e.g., log blood flow, as in Figure 1) that depends on a continuous covariate $\mathbf{t} = (t_1, \ldots, t_N)$ (e.g., time, as in Figure 1), and that we model $y_i \sim \mathcal{N}(\phi(t_i), \sigma^2)$. Often primary interest lies in the shape, derivatives or some other function of ϕ, and ϕ requires a smooth but flexible nonparametric specification, as opposed to a parametric linear specification $\phi(t) = \beta_0 + \beta_1 t$. Penalized splines are a practical choice for modeling ϕ; see Ruppert, Wand and Carroll (2003), Chapters 3 and 14. There are many ways to formulate a penalized spline. Low-rank thin-plate (LRTP) splines are appealing since they are defined by tractable radial basis functions and tend to exhibit fast Markov chain Monte Carlo (MCMC) convergence properties in a Bayesian context relative to truncated basis splines [Crainiceanu, Ruppert and Wand (2005)]. B-splines are another reasonable Bayesian option because they too tend to exhibit fast MCMC convergence, but they rely on a recursive algorithm to define the basis functions, making them less tractable than LRTP splines [Eilers and Marx (1996)]. Our method will work with either LRTP or B-spline formulations for ϕ, and we do not anticipate substantial differences in the behavior of our method under either formulation. Hereafter, we focus on a LRTP cubic spline specification for ϕ.

Without loss of generality, we take $t \in [0, 1]$ and model

$$\phi(t; \boldsymbol{\beta}) = \beta_0 + \beta_1 t + \sum_{k=2}^{K} \beta_k \bigl(|t - \tilde{t}_{k-1}|^3 - |\tilde{t}_{k-1}|^3\bigr), \tag{6}$$

where $\tilde{\mathbf{t}} = (0 = \tilde{t}_0 < \tilde{t}_1 < \cdots < \tilde{t}_{K-1} < \tilde{t}_K = 1)$ is a generic partition with K intervals. We discuss this choice later. Typically, a LRTP spline model is formulated as

$$\phi(t; \boldsymbol{\beta}^*) = \beta_0^* + \beta_1^* t + \sum_{k=2}^{K} \beta_k^* |t - \tilde{t}_{k-1}|^3; \tag{7}$$

however, in (7), β_0^* is not an intercept because $\phi(0; \boldsymbol{\beta}^*) = \beta_0^* + \sum_{k=2}^{K} \beta_k^* |\tilde{t}_{k-1}|^3$. We prefer the modified LRTP (mLRTP) model defined in (6), because it simplifies interpretation and prior elicitation when the regression includes an intercept, which will be the case for the CTp data, and it retains an equivalent interpretation of $(\beta_1,\dots,\beta_K)$ and $(\beta_1^*,\dots,\beta_K^*)$. In the sequel, we denote $\boldsymbol{\beta}_{(-0)}=(\beta_1,\dots,\beta_K)$, that is, the vector $\boldsymbol{\beta}$ with β_0 omitted.

A *penalized* spline requires prespecification of a fine partition $\tilde{\mathbf{t}}$, and smooths via the prior, whereas a *free-knot* spline uses complex sampling algorithms (e.g., reversible-jump MCMC) to sample over all possible partitions. We prefer penalized splines, since they are computationally much simpler and have proven to be competitive with free-knot splines when the temporal variation of ϕ is smooth [Ruppert, Wand and Carroll (2003), Section 3.16]. Wand (2000) reviews methods for free-knot spline estimation; we do not consider this approach further. To select $\tilde{\mathbf{t}}$ for a penalized spline, we consider a set of values for K with either equally-spaced or quantile-spaced $\tilde{t}_k$'s, and select the partition that has low posterior mean deviance and deviance information criterion (DIC) relative to the other partitions considered [Spiegelhalter et al. (2002)].

Following Crainiceanu, Ruppert and Wand (2005), without supplemental data, we complete the Bayesian specification of (6) by placing vague priors on β_0 and β_1, and a $\mathcal{N}(\mathbf{0},\sigma_{\boldsymbol{\beta}}^2\boldsymbol{\Omega}^-)$ prior on $(\beta_2,\dots,\beta_K)'$, where the (j,k)th entry of $\boldsymbol{\Omega}$ is defined as $|\tilde{t}_{j-1}-\tilde{t}_{k-1}|^3$, for $j,k=2,\dots,K$. In practice, we apply the transformations $b_0=\beta_0$, $b_1=\beta_1$, $(b_2,\dots,b_K)'=\boldsymbol{\Omega}^{1/2}(\beta_2,\dots,\beta_K)'$, and $\sigma_{\mathbf{b}}^2=\sigma_{\boldsymbol{\beta}}^2$, and then complete the model specification by assuming

$$\begin{aligned} \pi^*(b_k) &\equiv \mathcal{N}\bigl(b_k|0,10^4\bigr) \qquad \text{for } k=0,1,\\ \pi^*(b_k|\sigma_{\mathbf{b}}) &\equiv \mathcal{N}\bigl(b_k|0,\sigma_{\mathbf{b}}^2\bigr) \qquad \text{for } k=2,\dots,K \quad \text{and} \\ \pi^*(\sigma_{\mathbf{b}}) &\equiv \mathcal{U}(\sigma_{\mathbf{b}}|0.01,100). \end{aligned} \tag{8}$$

The prior in (8) smooths ϕ by preferring values for $(b_2,\dots,b_K)$ near zero a priori.

3.1. *GMC prior specification.*

With supplemental data $[\mathbf{y}_0=(y_{0,1},\dots,y_{0,n_0})$, $\mathbf{t}_0=(t_{0,1},\dots,t_{0,n_0})]$, we also assume $y_{0,i_0}|t_{0,i_0}\sim\mathcal{N}(\phi_0(t_{0,i_0}),\sigma_0^2)$, for $i_0=1,\dots,n_0$. To ensure that ϕ and ϕ_0 are analogous, we use the same partition $\tilde{\mathbf{t}}$ for ϕ_0, and model $\phi_0(t;\boldsymbol{\beta}_0)$ using (6) by replacing $\boldsymbol{\beta}$ with $\boldsymbol{\beta}_0$. We apply the transformation $b_{0,k}=\beta_{0,k}$, for $k=0,1$, and $(b_{0,2},\dots,b_{0,K})'=\boldsymbol{\Omega}^{1/2}(\beta_{0,2},\dots,\beta_{0,K})'$, and use the prior specifications $\pi^*(b_{0,k})\equiv\mathcal{N}(b_{0,k}|0,10^4)$, for $k=0,1$, $\pi^*(b_{0,k}|\sigma_{\mathbf{b}_0})\equiv\mathcal{N}(b_{0,k}|0,\sigma_{\mathbf{b}_0}^2)$ for $k=2,\dots,K$, and $\pi^*(\sigma_{\mathbf{b}_0})\equiv\mathcal{U}(\sigma_{\mathbf{b}_0}|0.01,100)$.

To borrow flexibly for the shape of ϕ from the supplemental data for ϕ_0, we apply the GMC prior specification defined in (5) to $\mathbf{b}_{(-0)}=(b_1,\dots,b_K)$ given $\mathbf{b}_{0,(-0)}=(b_{0,1},\dots,b_{0,K})$. We assume

$$\begin{aligned} &\pi(\mathbf{b}_{(-0)}|\mathbf{b}_{0,(-0)},\boldsymbol{\iota}_{(-0)},\sigma_{\mathbf{b}}) \\ &\quad = [\pi^*(b_1)]^{(1-\iota_1)}[\pi(b_1|b_{0,1})]^{\iota_1}\prod_{k=2}^{K}[\pi^*(b_k|\sigma_{\mathbf{b}})]^{(1-\iota_k)}[\pi(b_k|b_{0,k})]^{\iota_k}, \end{aligned}$$

where $\boldsymbol{\iota}_{(-0)} = (\iota_1, \ldots, \iota_K)$, and $\pi^*(b_1)$, $\pi^*(b_k|\sigma_{\mathbf{b}})$ and $\pi^*(\sigma_{\mathbf{b}})$ are defined in (8). We set $\pi(b_k|b_{0,k}) \equiv \mathcal{N}(b_k|b_{0,k}, \mathcal{R}_{\mathbf{b}}^{-1})$, for $k = 1, \ldots, K$, where $\mathcal{R}_{\mathbf{b}}$ is a prespecified *large* value. We then specify $\iota_k|\nu \overset{\text{iid}}{\sim} \text{Bern}(\nu)$ and $\nu \sim \mathcal{B}(a_1, a_2)$ with fixed, known hyperparameters a_1 and a_2 that reflect our prior opinion about the relevance of the supplemental data for the shape of ϕ.

To allow differential borrowing for the intercept versus the shape of ϕ, we specify an independent commensurate prior distribution for b_0 given $b_{0,0}$ following (4). We define $\pi(b_0|b_{0,0}, \iota_0, \tau) \equiv [\pi^*(b_0|b_{0,0}, \tau)]^{(1-\iota_0)}[\pi(b_0|b_{0,0})]^{\iota_0}$, where $\pi^*(b_0|b_{0,0}, \tau) \equiv \mathcal{N}(b_0|b_{0,0}, \tau^{-1})$ and $\pi(b_0|b_{0,0}) \equiv \mathcal{N}(b_0|b_{0,0}, \mathcal{R}^{-1})$. We then specify $\tau \sim \mathcal{U}(s_l, s_u)$ and $\iota_0 \sim \text{Bern}(p_0)$, where s_l, s_u, $\mathcal{R}$ and p_0 are prespecified following the guidance in Section 2. Taken together, the full posterior arises as

$$
\begin{aligned}
&p(\mathbf{b}, \mathbf{b}_0, \sigma_{\mathbf{b}}, \sigma_{\mathbf{b}_0}, \boldsymbol{\iota}, \nu, \tau|\mathbf{D}, \mathbf{D}_0) \\
&\quad \propto L(\mathbf{b}|\mathbf{D})L(\mathbf{b}_0|\mathbf{D}_0)\pi(b_0|b_{0,0}, \iota_0, \tau)\pi(\iota_0)\pi(\tau)\pi^*(b_{0,0}) \\
&\qquad \times \pi(\mathbf{b}_{(-0)}|\mathbf{b}_{0,(-0)}, \boldsymbol{\iota}_{(-0)}, \sigma_{\mathbf{b}})\pi(\boldsymbol{\iota}_{(-0)}|\nu)\pi(\nu)\pi^*(\sigma_{\mathbf{b}}) \\
&\qquad \times \pi^*(\mathbf{b}_{0,(-0)}|\sigma_{\mathbf{b}_0})\pi^*(\sigma_{\mathbf{b}_0}).
\end{aligned}
\tag{9}
$$

3.2. *Simulation assessment.* We now investigate via simulation the borrowing properties of the GMC prior model in (9) for nonparametric regression. To do so, we sample $y_i|t_i \sim \mathcal{N}\{\mu(t_i), \sigma^2\}$, $i = 1, \ldots, N$, where $\mu(t) = 5t\sin\{5t\}$, and sample $y_{0,i_0}|t_{0,i_0}, d \sim \mathcal{N}\{\mu_0(t_{0,i_0}|d), \sigma_0^2\}$, $i_0 = 1, \ldots, N_0$, where $\mu_0(t|d) = (5 + d)t\sin((5 + d)t)$ and $d \in [0, 5]$. Hence, the primary data always have the same true mean structure, and the supplemental data have a mean structure that deviates from that of the primary according to the value of d, the discordance parameter. When $d = 0$, the two curves are the same, and as d increases, the supplemental curve has increasingly greater curvature than the primary curve. Figure 3 shows curves for selected values of d.

In each run, we sample d uniformly from the set $\{0, 0.05, 0.10, 0.20, 0.35, 0.50, 0.75, 1, 1.50, 2, 3, 4, 5\}$. Given d, we then generate primary and supplemental data sets with $N = N_0 = 50$, error variances $\sigma^2 = \sigma_0^2 = 1$, and equally spaced t and t_0 values over $[0, 1]$. For each data set pair, $(\mathbf{D} = (\mathbf{y}, \mathbf{t}), \mathbf{D}_0 = (\mathbf{y}_0, \mathbf{t}_0))$, we

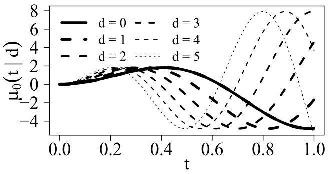

FIG. 3. *Mean curves for the supplemental data for a subset of discordance parameter (d) values. The mean curve for the primary data is denoted by the solid line (i.e., $d = 0$).*

fit the GMC prior model developed in Section 3.1. We use the mLRTP model defined in (6) with $K = 10$ and $\tilde{\mathbf{t}}$ spaced equally over [0, 1] for both $\phi_0(t_0; \boldsymbol{\beta}_0)$ and $\phi(t; \boldsymbol{\beta})$. After preliminary investigation, we set $s_l = 0$, $s_u = 2$, $\mathcal{R} = 2000$, $p_0 = 0.50$, and $\pi(\nu) \equiv \mathcal{B}(\nu|0.50, 0.50)$. The latter two choices represent an indifferent prior opinion about the relevance of the supplemental data for both the shape and the intercept of ϕ. We choose to place an extremely vague yet indifferent prior on ν to allow the primary and supplemental data to have substantial influence on whether to borrow for curve shape.

We used the `R2jags` package to call `JAGS` from `R`, and ran two chains for 20,000 iterations of burn-in, followed by 200,000 iterations for posterior estimation. We also fit the mLRTP model defined in (6) with prior specifications described in (8) to the primary data alone, as well as to the data set obtained by simply pooling the primary and supplemental data. These models feature lower MCMC autocorrelation, and thus required only two chains with 2000 iterations of burn-in, followed by 20,000 iterations for posterior estimation. These choices reflect preliminary investigations to ensure acceptable MCMC convergence and relatively small MC standard errors for the intercept and functional effect coefficients. Estimation of the GMC model took about 45 seconds, whereas estimation of each comparison model took about 5 seconds.

To evaluate the three models, we calculated four criteria at each run: mean error (ME), root-mean-square error (RMSE), mean pointwise credible interval width (CrIW), and mean pointwise coverage probability (CP). Alternatively, we could calculate simultaneous confidence bands following Krivobokova, Kneib and Claeskens (2010), Section 3. We define the four criteria as

$$\mathrm{ME}(d^{(m)}) = N^{-1}\sum_{i=1}^{N}[\hat{\phi}(t_i|d^{(m)}) - \mu(t_i|d^{(m)})],$$

$$\mathrm{RMSE}(d^{(m)}) = \left\{N^{-1}\sum_{i=1}^{N}[\hat{\phi}(t_i|d^{(m)}) - \mu(t_i|d^{(m)})]^2\right\}^{1/2},$$

$$\mathrm{CrIW}(d^{(m)}) = N^{-1}\sum_{i=1}^{N}[\hat{\phi}_{0.975}(t_i|d^{(m)}) - \hat{\phi}_{0.025}(t_i|d^{(m)})],$$

$$\mathrm{CP}(d^{(m)}) = N^{-1}\sum_{i=1}^{N} I\{\mu(t_i|d^{(m)}) \in [\hat{\phi}_{0.025}(t_i|d^{(m)}), \hat{\phi}_{0.975}(t_i|d^{(m)})]\},$$

for $m = 1, \ldots, M$, where $\hat{\phi}(t|d^{(m)})$ denotes the posterior mean estimate for $\phi(t|d^{(m)})$, and $\hat{\phi}_q(t|d^{(m)})$ denotes the qth quantile posterior estimate of $\phi(t|d^{(m)})$ in the mth run. We then compare the sampling average of each criterion over the M simulated data set pairs as a function of d, the discordance parameter. Specifically, we visually compare Loess estimates of each evaluation criteria as

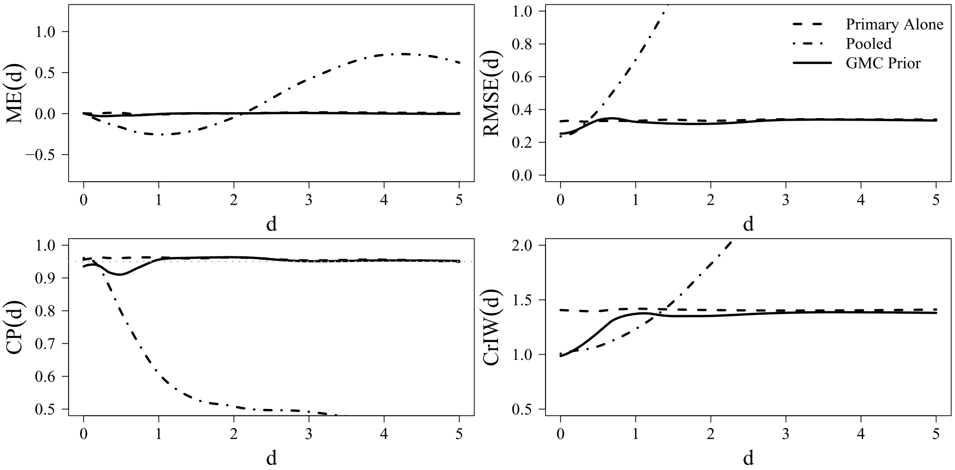

FIG. 4. *Simulation assessment of the GMC prior spline model* (*solid*) *versus a conventional spline model fit to the primary data alone* (*dashed*) *and to the data set obtained by pooling the supplemental and primary data* (*dot-dashed*). *These methods are compared on sampling averages of mean error* (*top left*), *root-mean-square error* (*top right*), *mean credible interval width* (*bottom right*) *and mean pointwise coverage probability* (*bottom left*) *as a function of* d, *the discordance parameter. All results are based on* $M = 2000$ *runs.*

a function of the discordance parameter. One simulation iteration takes about 60 seconds, which entails generating a pair of data sets, fitting the three models, and then calculating and saving the evaluation criteria, along with the sampled value of d in that run. We reduced overall computation time using the `snowfall` package for `R` to conduct the simulation runs in parallel. A `R` program to implement this simulation is available through the third author's software page http://www.biostat.umn.edu/~brad/software.html.

The results of our simulation investigation are illustrated in Figure 4. Each panel shows the Loess estimates of the sampling average of the corresponding evaluation criteria as a function of d for three models: the GMC prior model (solid), the conventional model fit to the primary data alone (dashed), and the conventional model fit to the pooled data (dot-dashed). As expected, inferential properties for the GMC prior model reflect those for the pooled approach under true concordance (i.e., $d = 0$) and approach primary data alone for increasing degrees of discordance. The primary-alone approach has the largest average CrIW and RMSE under true concordance, but its properties do not deteriorate for increasing degrees of discordance (because it does not acknowledge the supplemental data). By contrast, the pooling approach has the smallest CrIW and RMSE under true concordance, but its performance on all four evaluation criteria deteriorates substantially as discordance increases.

The ME plot (top left panel) in Figure 4 demonstrates that the GMC prior model has bias properties that interpolate the two conventional approaches for near concordance, but near $d = 1$ it learns to effectively ignore the supplemental information. Furthermore, the RMSE (top right panel) shows that the GMC prior model learns to borrow strongly from the supplemental information when the sources are commensurate (i.e., slight discordance), achieving RMSE similar to that of the pooling approach. Similarly, the CrIW plot (bottom right panel) shows for slight discordance that the credible intervals from the GMC prior model are nearly as tight as those for the pooling approach. Conversely, when the supplemental curve has greater oscillation than the primary curve, the GMC prior approach has credible interval width similar to those of the primary-alone approach. Last, the CP plot (bottom left panel) shows that the mean pointwise coverage probabilities for the GMC prior model are similar to those of the conventional spline model fit to the primary data alone.

3.3. *Application*: *Liver imaging study.* We now apply our GMC prior model structure to estimate the temporal features of each perfusion characteristic in cancerous liver tissue. Recall that similar information derived from noncancerous tissue is potentially valuable supplemental information because the shape of the average CTp curve in noncancerous regions may provide relevant information about the shape of the corresponding average CTp curve in cancerous regions. The data consist of 7 to 13 readings acquired at times between 11 and 100 seconds after contrast injection in 0 to 2 cancerous and noncancerous regions of interest (ROIs) among 16 individuals. For each perfusion characteristic, there are 687 total readings from fifteen individuals who each contribute $N_i = 11$–13 readings in $M_{T,i} = 1$–2 cancerous and $M_{N,i} = 1$–2 noncancerous ROIs, and one individual who contributes 7 readings in 1 noncancerous ROI.

Individual acquisition times $t_{i,\ell}$, $\ell = 1, \ldots, N_i$, are necessarily identical for all ROIs and all perfusion characteristics. Thus, we let $y_{r,i,j,\ell}$, for $j = 1, \ldots, M_{r,i}$, denote a reading at time $t_{i,\ell}$. We denote the average CTp curve for a given perfusion characteristic in cancerous tissue and noncancerous tissue by $\phi_T(t)$ and $\phi_N(t)$, respectively. We assume each individual's CTp curve deviates from the average CTp curve smoothly over time, and we denote these deviation curves by $\psi_{r,i}(t)$, $r = T, N$ and $i = 1, \ldots, 16$. We then model

$$y_{r,i,j,\ell} \sim \mathcal{N}\{\phi_r(t_{i,\ell}) + \psi_{r,i}(t_{i,\ell}), \sigma^2_{e,r}\}. \tag{10}$$

We use the mLRTP model defined in (6) for ϕ_r and $\psi_{r,i}$, parametrized by $\boldsymbol{\beta}_r$ and $\boldsymbol{\alpha}_{r,i}$, respectively. Therefore, the first derivative of the average CTp curve is

$$\phi'_r(t; \boldsymbol{\beta}_r) = \beta_{r,1} + \sum_{k=2}^{K} \operatorname{sign}(t - \tilde{t}_{k-1}) 3\beta_{r,k}(t - \tilde{t}_{k-1})^2, \qquad r = N, T. \tag{11}$$

To shrink the individual deviation curves $\psi_{r,i}(t; \boldsymbol{\alpha}_{r,i})$ toward the average CTp curve $\phi_r(t; \boldsymbol{\beta}_r)$, we assume $\alpha_{r,i,k} \sim \mathcal{N}(0, \sigma^2_{\mathbf{a},r,i})$, for $k = 0, \ldots, K$, $r = N, T$ and $i = 1, \ldots, 16$.

For our conventional model, we use the prior specifications on $\boldsymbol{\beta}_N$ and $\boldsymbol{\beta}_T$ described in (8), and specify vague $\mathcal{N}(0, 10^4)$ priors for $b_{N,0}$ and $b_{T,0}$. For the GMC prior model, we apply the prior developed in Section 3.1 for $\boldsymbol{\beta}_T$ given $\boldsymbol{\beta}_N$. That is, we specify a GMC prior for $\boldsymbol{\beta}_{T,(-0)} = (\boldsymbol{\beta}_{T,1}, \dots, \boldsymbol{\beta}_{T,K})$ given $\boldsymbol{\beta}_{N,(-0)} = (\boldsymbol{\beta}_{N,1}, \dots, \boldsymbol{\beta}_{N,K})$ and an independent commensurate prior for $b_{T,0}$ given $b_{N,0}$. This choice enables differential borrowing from the noncancerous tissue data for the intercept versus the curve shape parameters, and simultaneously enables borrowing similar amounts among curve shape parameters. Thus, if the average CTp curve in noncancerous regions differs from the average CTp curve in cancerous regions only for the intercept, the model still permits borrowing for the CTp curve shape. Last, we place vague $\mathcal{U}(0.01, 100)$ priors on each of the standard deviation parameters $(\sigma_{e,r}, \sigma_{\mathbf{b},r}, \sigma_{\mathbf{a},r,i})$.

We let $t \in (0, 1]$ by taking $t = \frac{t^* - t^*_{\min}}{t^*_{\max} - t^*_{\min}}$, where t^* is the original timescale. Preliminary analysis of the noncancerous regions was used to specify the hyperparameters $\mathcal{R} = 500$, $s_l = 0.01$, $s_u = 0.50$, $p_0 = 0.10$, $a_1 = 0.10$ and $a_2 = 0.90$. To select a partition, we fit our conventional model (i.e., independent spline models for the average CTp curve in each tissue type) for each perfusion parameter using $K = 5$, 10, 15 and 25 with $\tilde{\mathbf{t}}$ equally spaced over [0, 1], and using $K = 5$, 10 and 15 with $\tilde{\mathbf{t}}$ placed at equally spaced quantiles of the observed acquisition times. The model using $K = 10$ with quantile-spaced knots resulted in low deviance and low DIC relative to models with the other partitions for all three perfusion characteristics, so we chose to conduct our analysis using this partition. We estimated the conventional model using 40,000 posterior samples after 2000 burn-in samples from two MCMC chains. We estimated the GMC model posterior using 200,000 posterior samples after 20,000 burn-in samples from two MCMC chains; there was greater autocorrelation for the basis coefficient parameters than the conventional model.

Table 1 reports the posterior mean borrowing parameters (i.e., ι_k's) from the GMC model for the three perfusion characteristics. Values near 1 indicate strong borrowing for the corresponding basis parameter. The first column reports the posterior mean for ι_0, which corresponds to the intercept. PS and BF show virtually zero borrowing for the intercept, whereas BV shows little borrowing. The remaining columns illustrate borrowing for the basis parameters that control the shape

TABLE 1

Posterior borrowing parameter estimates (i.e., ι_k's). ι_0 corresponds to the CTp curve intercept, and the remaining ι_k's correspond to CTp curve shape. Values near one indicate strong borrowing

Perfusion characteristic	ι_0	ι_1	ι_2	ι_3	ι_4	ι_5	ι_6	ι_7	ι_8	ι_9	ι_{10}
PS	0.00	0.23	0.00	0.00	0.24	0.43	0.24	0.05	0.15	0.03	0.34
BV	0.20	0.92	0.95	0.92	0.82	0.86	0.90	0.82	0.88	0.96	0.96
BF	0.02	0.96	0.98	0.92	0.92	0.93	0.94	0.92	0.96	0.98	0.99

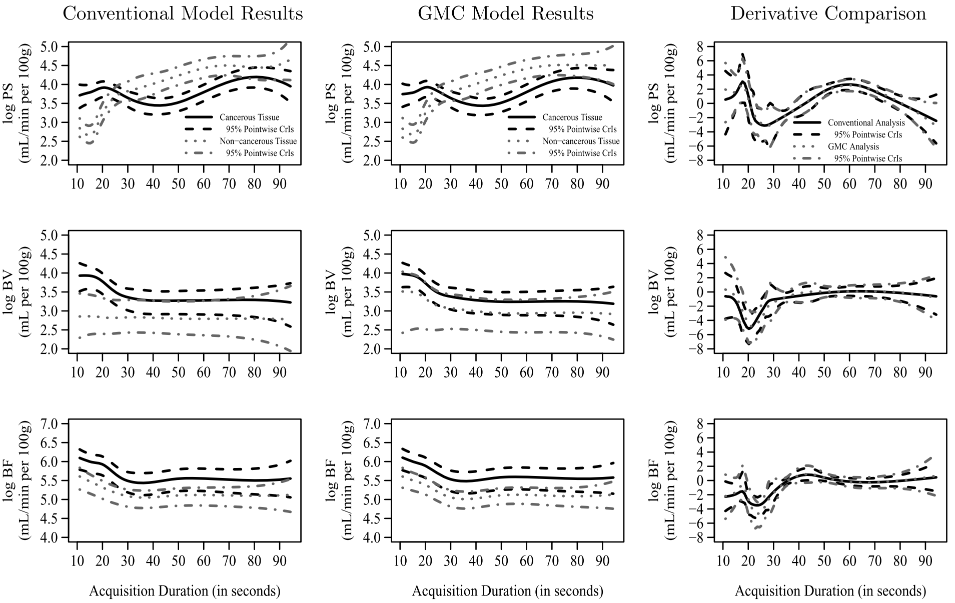

FIG. 5. *The first column displays posterior mean CTp curves from the conventional model analysis for cancerous* (*solid black*) *and noncancerous regions* (*dotted grey*), *along with* 95% *pointwise CrIs. The second column displays the same results for the GMC model analysis. The third column compares the posterior mean first derivative of the CTp curve in cancerous tissue from the conventional model* (*solid black*) *and the GMC prior model* (*dotted grey*), *along with* 95% *pointwise CrIs.*

(and thus, first derivative) of the CTp curve, which we have restricted to have similar magnitude by assuming $\iota_k|\nu \sim \text{Bern}(\nu)$, for $k = 1, \dots, 10$, a priori. For PS, these parameters all have posterior means less than 0.43, and many less than 0.25, which indicates little borrowing for CTp curve shape. In contrast, these parameters all exceed 0.82 for BV, and 0.92 for BF. Thus, for BF and BV the GMC model borrows substantially from the supplemental information in noncancerous tissue for CTp curve shape in cancerous tissue.

The first row in Figure 5 indicates that both the shape and intercept of the PS curve differ substantially by tissue type and that the results of the conventional model analysis (first column) and the GMC prior model analysis (second column) are virtually indistinguishable. Consequently, the first derivative of the CTp curve in cancerous tissue (third column) is estimated with similar precision in either the conventional or GMC model analysis. In fact, the GMC prior model results in 95% pointwise credible intervals (CrIs) that are 4% wider on average than those of the conventional model analysis. Regardless, the PS CTp curve in cancerous tissue has not stabilized after 95 seconds, so longer acquisition durations for PS appear necessary, which is consistent with the findings by Ng et al. (2013).

The second row in Figure 5 corresponds to BV. Here the GMC prior model, relative to the conventional model, results in a noncancerous CTp curve that is more similar for both the intercept and the shape to that of the cancerous tissue CTp curve. Moreover, the GMC model gains precision over the conventional model for the first derivative estimate of the cancerous tissue CTp curve via borrowing from the noncancerous tissue CTp curve shape. The average 95% pointwise CrIs for the GMC model are 18% tighter than those of the conventional model. The third row corresponds to BF and illustrates that the posterior CTp curve estimates are similar for the two modeling approaches. However, the first derivative of the cancerous tissue CTp curve is estimated more precisely for the GMC model than the conventional model, resulting in 27% tighter 95% pointwise CrIs on average. For both BF and BV, we can infer a shorter acquisition duration using the GMC prior model than using the conventional model, although we did not impose precise stability criteria here.

4. GMC priors in semiparametric survival analysis. We now describe a flexible semiparametric survival model and then apply our GMC prior technology to the colorectal cancer clinical trials data described in Section 1. For simplicity, assume each observation consists of a possibly right-censored time $t \in (0, 1]$, a binary event indicator c and a binary treatment indicator z.

4.1. *Piecewise-exponential proportional hazards model.* A flexible survival model commonly favored by Bayesians is the piecewise-exponential proportional hazards model, which is constructed by partitioning the time axis into K intervals $\tilde{\mathbf{t}} = (0 = \tilde{t}_0 < \tilde{t}_1 < \cdots < \tilde{t}_{K-1} < \tilde{t}_K = 1)$ and assuming the baseline hazard is constant in each interval [cf. Ibrahim, Chen and Sinha (2001), Section 3.1]. Under this model, the likelihood for an observation (t, c, z) is given by $h(t|z; \boldsymbol{\gamma}, \rho)^c \exp\{-\int_0^t h(s|z; \boldsymbol{\gamma}, \rho)\, ds\}$, where

$$h(t|z; \boldsymbol{\gamma}, \rho) = \exp(\gamma_k + \rho z) \qquad \text{for } t \in I_k = (\tilde{t}_{k-1}, \tilde{t}_k], k = 1, \dots, K. \tag{12}$$

Thus, $h(t|z; \boldsymbol{\gamma}, \rho)$ is assumed to be piecewise-constant, where γ_k denotes the log-hazard in the kth interval of the time-axis partition $\tilde{\mathbf{t}}$ for treatment assignment $z = 0$, and ρ denotes the log-hazard ratio for treatment assignment $z = 1$ relative to $z = 0$. Following Ibrahim, Chen and Sinha (2001), we select K using a DIC comparison over a small set of partitions. To resist overfitting, we specify a correlated prior process for $\pi^*(\boldsymbol{\gamma})$. We focus on the random-walk prior process for $\pi^*(\boldsymbol{\gamma})$ introduced by Fahrmeir and Lang (2001). Formally, we specify

$$\begin{aligned} &\pi^*(\gamma_1) \equiv \mathcal{N}(\gamma_1 | 0, 10^4), \\ &\pi^*(\gamma_k | \gamma_{k-1}, \sigma_{\boldsymbol{\gamma}}) \equiv \mathcal{N}(\gamma_k | \gamma_{k-1}, \sigma_{\boldsymbol{\gamma}}^2) \qquad \text{for } k = 2, \dots, K \quad \text{and} \\ &\pi^*(\sigma_{\boldsymbol{\gamma}}) \equiv \mathcal{U}(\sigma_{\boldsymbol{\gamma}} | 0.01, 100). \end{aligned} \tag{13}$$

The prior process in (13) smooths adjacent γ_k's toward each other by assuming their first differences are exchangeable. We also assume ρ is independent of $\boldsymbol{\gamma}$

a priori, and specify a vague $\mathcal{N}(0, 10^4)$ prior. The model defined in (12) is useful in the absence of supplemental data, and the parameter space can be decomposed into separable subsets, $\boldsymbol{\gamma}$ characterizes the baseline hazard and ρ characterizes the treatment effect.

4.2. *Application*: *Colorectal cancer trials*. Extending Hobbs, Carlin and Sargent (2013), we apply our GMC prior model structure to supplement the inference on the progression-free survival (PFS) curve among all three regimens in the trial reported by Goldberg et al. (2004) using the historical information on the IFL regimen from the trial reported by Saltz et al. (2000). We use the piecewise-exponential proportional-hazards model to estimate the PFS curves for the two trials from these data. During the first two years of follow-up, the primary (Goldberg) data set contains 197 progression events among 211 persons treated with IFL, 190 events among 216 persons treated with FOLFOX, and 189 events among 206 persons treated with IROX. For the primary data, we have two binary treatment indicators for assignment to the FOLFOX (z_F) and IROX (z_I) regimens, respectively, so following (12) we model

$$h(t|\mathbf{z}; \boldsymbol{\gamma}, \boldsymbol{\rho}) = \exp(\gamma_k + \rho_F z_F + \rho_I z_I) \qquad \text{for } t \in (\tilde{t}_{k-1}, \tilde{t}_k],$$

where $k = 1, \ldots, K$. The supplemental (Saltz) data contain 172 progression events during the first two years of follow-up among 224 persons treated with IFL. The model for the supplemental data is defined analogously, though $z_F = z_I = 0$ for all the supplemental observations. Therefore, the supplemental hazard model is completely parametrized by $\boldsymbol{\gamma}_0$, which, as required, is analogous to $\boldsymbol{\gamma}$.

We specify a GMC prior to flexibly borrow strength from the supplemental information on the IFL regimen. Namely, we specify the random-walk prior process defined by (13) for $\pi^*(\boldsymbol{\gamma}_0)$, and apply the GMC prior structure on $\boldsymbol{\gamma}$. In this setting, the prior specification follows as $\gamma_1|\gamma_{0,1}, \iota_1 \sim [\mathcal{N}(\gamma_1|0, 10^4)]^{1-\iota_1}[\mathcal{N}(\gamma_1|\gamma_{0,1}, \mathcal{R}_{\boldsymbol{\gamma}}^{-1})]^{\iota_1}$, $\gamma_k|\gamma_{k-1}, \gamma_{0,k}, \iota_k, \sigma_{\boldsymbol{\gamma}} \sim [\mathcal{N}(\gamma_k|\gamma_{k-1}, \sigma_{\boldsymbol{\gamma}}^2)]^{1-\iota_k}[\mathcal{N}(\gamma_k|\gamma_{0,k}, \mathcal{R}_{\boldsymbol{\gamma}}^{-1})]^{\iota_k}$, for $k = 2, \ldots, K$, and $\iota_k|\nu_{\boldsymbol{\gamma}} \sim \text{Bern}(\nu_{\boldsymbol{\gamma}})$, for $k = 1, \ldots, K$, where $\mathcal{R}_{\boldsymbol{\gamma}}$ is prespecified. Next, we place a $\mathcal{B}(a_1, a_2)$ prior on $\nu_{\boldsymbol{\gamma}}$, where a_1 and a_2 are also prespecified. Last, we place vague $\mathcal{U}(0.01, 100)$ priors on each of the standard deviation parameters (i.e., $\sigma_{\boldsymbol{\gamma}}$ and $\sigma_{\boldsymbol{\gamma}_0}$) and vague $\mathcal{N}(0, 10^4)$ priors on the treatment effect parameters (i.e., ρ_F and ρ_I).

We transformed the timescale so that $t \in (0, 1]$, then selected $\mathcal{R}_{\boldsymbol{\gamma}} = 10{,}000$ and specified a vague $\mathcal{B}(0.10, 0.90)$ prior for $\nu_{\boldsymbol{\gamma}}$, which represents a vague, yet skeptical prior opinion about the relevance of the supplemental data. For estimation, we used 200,000 posterior samples for estimation after 50,000 iterations of burn-in from two MCMC chains. For comparison, we also fit conventional piecewise-exponential proportional hazards models with the same time-axis partition and random-walk prior process to the primary (Goldberg) data alone, supplemental (Saltz) data alone, and the data set obtained by naively pooling the two sources

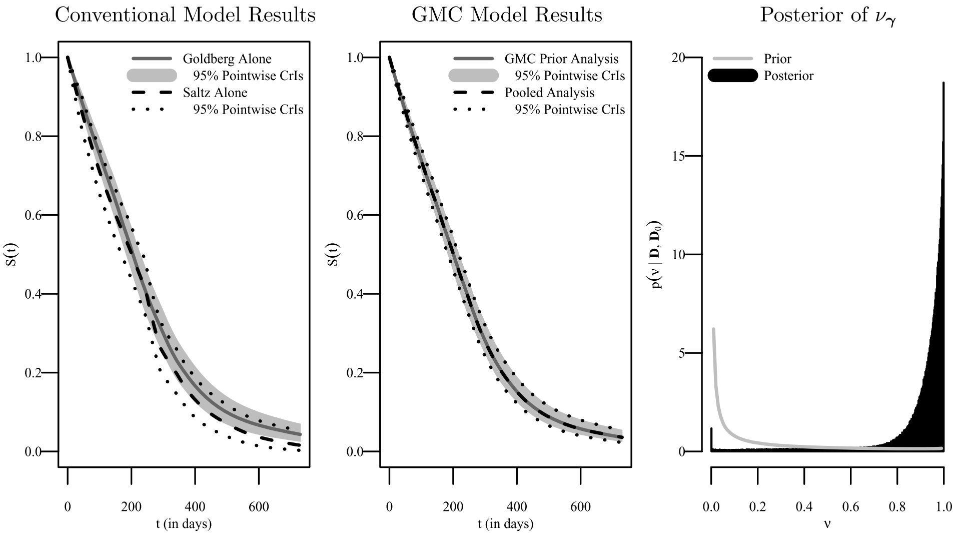

FIG. 6. *The left panel displays the posterior mean PFS curve for the IFL regimen along with* 95% *pointwise CrIs from the conventional piecewise-constant-hazard analysis of the primary* (*Goldberg*) *data alone and the supplemental* (*Saltz*) *data alone. The middle panel displays the same results from the GMC prior model analysis and the conventional model analysis of the dataset obtained by pooling the two sources of information. The third panel displays a posterior histogram for* ν_γ, *which controls borrowing for the baseline hazard among trials. For reference*, *the prior assigned to* ν_γ *is also displayed* (*grey line*).

of information. To estimate these conventional models, we used 20,000 posterior estimation draws after 2000 burn-in draws from two MCMC chains, again the shorter chain length justified by faster convergence and lower post-convergence autocorrelations.

We begin with a comparison of the estimated PFS curves for the IFL regimen from the four analyses. The left panel in Figure 6 indicates that the estimated PFS curve from the Goldberg data (solid, with shaded grey 95% pointwise CrIs) overlaps substantially with the PFS curve estimated from the Saltz data (dashed, with dotted 95% pointwise CrIs), suggesting the Saltz data provide relevant information. The middle panel in Figure 6 illustrates that the GMC analysis (solid, with shaded grey 95% pointwise CrIs) results in a PFS estimate nearly indistinguishable from that of the conventional analysis of the pooled data set (dashed, with dotted 95% pointwise CrIs), though shifted very slightly toward the primary information. The GMC prior model achieved a PFS curve estimate with noticeably greater precision than that of the conventional analysis that ignores the supplemental data without requiring bold a priori assumptions regarding the relevance of the Saltz data. Finally, the rightmost panel in Figure 6 displays a posterior histogram of ν_γ that has much of its mass shifted toward one. In fact, the ι_k's had posterior means of at least 0.87, with an average of 0.94, indicating that the GMC prior model

TABLE 2

Hazard-ratio estimates for the FOLFOX and IROX regimens relative to the IFL regimen, and estimates of median days to disease progression for each treatment regimen from the four analyses of the Goldberg data and Saltz data

Drug regimen	**IFL**	**FOLFOX**	**IROX**
Analysis		*Hazard ratios (95% CrIs)*	
Goldberg alone	–	0.70 (0.57, 0.86)	0.92 (0.75, 1.12)
Pooled analysis	–	0.66 (0.56, 0.79)	0.87 (0.73, 1.04)
GMC prior analysis	–	0.66 (0.56, 0.79)	0.87 (0.73, 1.04)
		Median days to disease progression (95% CrIs)	
Saltz alone	200 (168, 236)	–	–
Goldberg alone	205 (182, 227)	262 (235, 293)	217 (194, 242)
Pooled analysis	200 (182, 217)	263 (245, 286)	221 (201, 240)
GMC prior analysis	200 (184, 216)	265 (245, 286)	221 (203, 240)

has learned that the supplemental FOLFOX data from the Saltz trial are indeed relevant.

Turning to the comparative effectiveness of three regimens, Table 2 shows that the hazard ratios (95% CrIs) from the GMC prior model analysis for FOLFOX and IROX versus IFL are 0.66 (0.56,0.79) and 0.87 (0.73,1.04), respectively. The hazard ratios from the conventional analysis of the Goldberg data alone are slightly larger with notably wider CrI widths, 0.70 (0.57,0.86) for FOLFOX and 0.92 (0.75,1.12) for IROX. By contrast, the hazard-ratio estimates and CrI widths from the conventional analysis of the pooled data set are indistinguishable from the GMC prior model analysis. Table 2 also contains the posterior estimates from each analysis of median PFS for the IFL, FOLFOX and IROX regimens. The conventional analysis of the Saltz data alone estimates median PFS in the IFL regimen to be 200 (168, 236) days, broadly similar to the estimate of 205 (182, 227) days provided by the conventional analysis of the Goldberg data alone. The estimates of median days to disease progression for each regimen from the pooled analysis of these data are also nearly identical to the estimates from the GMC prior analysis, namely, 200, 265 and 221 days for the IFL, FOLFOX and IROX regimens, respectively. The estimates from the GMC prior model are consistent with the results reported by Goldberg et al. (2004), who found FOLFOX to be the superior regimen, and significantly better than IFL. In addition, the GMC prior analysis yields stronger evidence that IROX may be better than IFL for PFS, though the difference remains statistically insignificant.

5. Discussion and future work. Our proposed methods for prior specification in functional and survival data models with penalized splines facilitate data-dependent borrowing that is robust to biased estimation of primary effects when conflict among information sources occurs. The simulation study illustrates the beneficial flexible borrowing properties the proposed methods offer. The application in perfusion CT illustrates potential gains in CTp curve estimation from using supplemental data collected concurrently. By contrast, the colorectal cancer application illustrates the use of these methods for semiparametric survival modeling to flexibly borrow from supplemental data on a control therapy collected in a previous trial. A hierarchical model with the proposed structure allows the degree of borrowing to be estimated differentially for each feature (e.g., CTp curve intercept versus shape). The amount of strength being borrowed between sources reflects the evidence of commensurability for that feature. When substantial evidence indicates that sources differ for a feature, the proposed method will learn to effectively ignore the supplemental data for that feature, yet possibly still borrow strength for another feature. For minor discordance or concordance, the proposed method also facilitates partial (rather than full) pooling of information from the supplemental source, hence a modest yet justifiable gain in efficiency. Our general modeling strategy enables data-driven estimation of heterogeneity among information sources for complex functional relationships, and thus provides a sensible and justifiable synthesis of clinical information.

In future work, we plan to extend the commensurate prior approach to settings that have *multiple* supplemental sources of information. This setting substantially complicates the construction of a hierarchical model that facilitates flexible borrowing for each supplemental source. Ideally, the method will facilitate data-dependent differential borrowing from various sources, learning from which to borrow and which to ignore as primary data accumulate. We also hope to develop concise, interpretable summaries that quantify the amount of strength being borrowed from each information source, thereby allowing the use of these models in an adaptive trial [see Hobbs, Carlin and Sargent (2013)]. We will also consider using posterior summaries of ν in (5) to determine whether the curves are commensurate among information sources, for example, by assessing $\Pr(\nu > 0.80|\mathbf{D}, \mathbf{D}_0) > 0.90$. Furthermore, we are currently studying the properties of the piecewise-exponential model used in the colorectal cancer application, which relies on a simple, yet flexible piecewise-constant assumption for the baseline hazard. We are developing extensions that use a piecewise-linear model for the baseline hazard function, relax the proportional hazards assumption, and allow functional covariate effects with shape constraints (e.g., monotonicity).

Acknowledgments. We thank the Associate Editor and referees for their well thought-out and detailed comments during revision that substantially improved the clarity of presentation in this manuscript and the model we use in the CTp analysis.

REFERENCES

CARLIN, B. P. and LOUIS, T. A. (2009). *Bayesian Methods for Data Analysis*, 3rd ed. *Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. MR2442364

CRAINICEANU, C. M., RUPPERT, D. and WAND, M. P. (2005). Bayesian analysis for penalized spline regression using WinBUGS. *J. Stat. Softw.* **14** 1–24.

DOI, S. A. R., BARENDREGT, J. J. and MOZURKEWICH, E. L. (2011). Meta-analysis of heterogeneous clinical trials: An empirical example. *Contemp. Clin. Trials* **32** 288–298.

EILERS, P. H. C. and MARX, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statist. Sci.* **11** 89–121. MR1435485

FAHRMEIR, L. and LANG, S. (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors. *J. Roy. Statist. Soc. Ser. C* **50** 201–220. MR1833273

GEORGE, E. I. and MCCULLOCH, R. E. (1997). Approaches for Bayesian variable selection. *Statist. Sinica* **7** 339–373.

GOLDBERG, R. M., SARGENT, D. J., MORTON, R. F., FUCHS, C. S., RAMANATHAN, R. K., WILLIAMSON, S. K., FINDLAY, B. P., PITOT, H. C. and ALBERTS, S. R. (2004). A randomized controlled trial of fluorouracil plus leucovorin, irinotecan, and oxaliplatin combinations in patients with previously untreated metastatic colorectal cancer. *J. Clin. Oncol.* **22** 23–30.

HOBBS, B. P., CARLIN, B. P. and SARGENT, D. J. (2013). Adaptive adjustment of the randomization ratio using historical control data. *Clin. Trials* **10** 430–440.

HOBBS, B. P., SARGENT, D. J. and CARLIN, B. P. (2012). Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Anal.* **7** 639–673. MR2981631

HOBBS, B. P., CARLIN, B. P., MANDREKAR, S. J. and SARGENT, D. J. (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics* **67** 1047–1056. MR2829239

IBRAHIM, J. G. and CHEN, M.-H. (2000). Power prior distributions for regression models. *Statist. Sci.* **15** 46–60. MR1842236

IBRAHIM, J. G., CHEN, M.-H. and SINHA, D. (2001). *Bayesian Survival Analysis*. Springer, New York. MR1876598

KRIVOBOKOVA, T., KNEIB, T. and CLAESKENS, G. (2010). Simultaneous confidence bands for penalized spline estimators. *J. Amer. Statist. Assoc.* **105** 852–863. MR2724866

MILES, K. A. and GRIFFITHS, M. R. (2003). Perfusion CT: A worthwhile enhancement? *Br. J. Radiol.* **76** 220–231.

MURRAY, T. A., HOBBS, B. P., LYSTIG, T. C. and CARLIN, B. P. (2014). Semiparametric Bayesian commensurate survival model for post-market medical device surveillance with non-exchangeable historical data. *Biometrics* **70** 185–191. MR3251679

NEELON, B. and O'MALLEY, A. J. (2010). Bayesian analysis using power priors with application to pediatric quality of care. *J. Biometr. Biostat.* **1** 103.

NG, C. S., HOBBS, B. P., CHANDLER, A. G., ANDERSON, E. F., HERRON, D. H., CHARNSANGAVEJ, C. and YAO, J. (2013). Metastases to the liver from neuroendocrine tumors: Effect of duration of scan acquisition on CT perfusion values. *Radiology* **269** 758–767.

POCOCK, S. J. (1976). The combination of randomized and historical controls in clinical trials. *J. Chronic Dis.* **29** 175–188.

RUPPERT, D., WAND, M. P. and CARROLL, R. J. (2003). *Semiparametric Regression*. *Cambridge Series in Statistical and Probabilistic Mathematics* **12**. Cambridge Univ. Press, Cambridge. MR1998720

SALTZ, L. B., COX, J. V., BLANKE, C., ROSEN, L. S., FEHRENBACHER, L., MOORE, M. J., MAROUN, J. A., ACKLAND, S. P., LOCKER, P. K., PIROTTA, N., ELFRING, G. L. and MILLER, L. L. (2000). Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer. *N. Engl. J. Med.* **343** 905–914.

SMITH, T. C., SPIEGELHALTER, D. J. and THOMAS, A. (1995). Bayesian approaches to random-effects meta-analysis: A comparative study. *Stat. Med.* **14** 2685–2699.

SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P. and VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64** 583–639. MR1979380

WAND, M. P. (2000). A comparison of regression spline smoothing procedures. *Comput. Statist.* **15** 443–462. MR1818029

T. A. MURRAY
B. P. HOBBS
DEPARTMENT OF BIOSTATISTICS—UNIT 1411
THE UNIVERSITY OF TEXAS
MD ANDERSON CANCER CENTER
P.O. BOX 301402
HOUSTON, TEXAS 77230
USA
E-MAIL: tamurray@mdanderson.org

B. P. CARLIN
DIVISION OF BIOSTATISTICS
SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF MINNESOTA
420 DELAWARE ST SE MMC 303
MINNEAPOLIS, MINNESOTA 55455
USA
E-MAIL: carli002@umn.edu