Bayesian Analysis (2025) **TBA**, Number TBA, pp. 1–31

The Scale Transformed Power Prior for Time-To-Event Data

Ethan M. Alt[*,‖], Brady Nifong[†,‖], Xinxin Chen[‡], Matthew A. Psioda[§], and Joseph G. Ibrahim[¶]

Abstract. In survival analysis, borrowing information from historical data can increase precision and power. However existing methods often assume that both current and historical data are of the same type. This assumption becomes problematic when data types differ, such as in cancer trials where phase 2 studies may use binary endpoints (e.g., response rates) while phase 3 studies typically use time-to-event endpoints. To address this limitation, we propose the partial-borrowing scale-transformed power prior (straPP), specifically designed for survival models with heterogeneous historical data. By using a functional rescaling based on the Fisher information matrices, the straPP aligns parameter vectors across differing data types, enabling partial borrowing of historical information while mitigating biases associated with borrowing from mismatched endpoints. Additionally, we introduce the generalized scale-transformed power prior (Gen-straPP) to further guard against biases in circumstances where scaling alone is insufficient. Through simulations and analyses of real cancer trial data from the Eastern Cooperative Oncology Group, we demonstrate that the (Gen-) straPP can outperform traditional priors in controlling type I error, power, coverage probabilities, and model fit, making it a robust choice for time-to-event analysis in these contexts.

Keywords: Bayesian analysis, historical data, heterogeneous endpoints, scale transformation, time-to-event data.

1 Introduction

Over the past few decades, the use of historical data in the design and analysis of new studies has become increasingly common. The availability of historical data is quite common in a variety of settings, including clinical trials, carcinogenicity studies, and environmental studies, where large data sets are available from similar, previously completed studies. One common way to incorporate historical data is through the power prior developed by Ibrahim and Chen (2000). Since then, extensive research has been conducted on the power prior and its modifications for a substantial number of statistical models, including generalized linear models, survival models, models for longitudinal data and semi-parametric models. Some, but certainly not all, of these papers

[*]Department of Biostatistics, University of North Carolina at Chapel Hill, NC, U.S.A.
[†]GlaxoSmithKline
[‡]Department of Biostatistics, University of North Carolina at Chapel Hill, NC, U.S.A.
[§]GlaxoSmithKline
[¶]Department of Biostatistics, University of North Carolina at Chapel Hill, NC, U.S.A., ibrahim@bios.unc.edu
[‖]E. M. Alt and B. Nifong are co-first authors and contributed equally to this work. Their names are listed in alphabetical order.

 https://doi.org/10.1214/24-BA1504

include Ibrahim and Chen (1998), Chen et al. (1999b, 2002), Ibrahim et al. (1999, 2003), Chen and Ibrahim (2006), De Santis (2006), Neuenschwander et al. (2009), Neelon and O'Malley (2010), Rietbergen et al. (2011), Fouskakis et al. (2015), Gravestock et al. (2017), Banbeta et al. (2019, 2022), and Pawel et al. (2023). Various extensions of the power prior include the *normalized power prior*, initially discussed by Ibrahim and Chen (2000) and further developed by Duan et al. (2006), Shen et al. (2024) and the references therein, Pawel et al. (2024), the *partial-borrowing power prior* proposed by Ibrahim et al. (2012b) and Chen et al. (1998, 2011). More recently, propensity score methods have be weaved into the power prior, including papers by Wang et al. (2019, 2022, 2024), and Lu et al. (2022). Several other important variations of the power prior have been proposed in the literature which are too numerous to list here.

Some alternative methods for borrowing information from historical data, distinct from the power prior and its variations, have been proposed in the literature. These include the commensurate prior (Hobbs et al., 2011), the meta-analytic-predictive prior (MAP) and robust MAP of Schmidli et al. (2014), and the Bayesian hierarchical model (BHM), where the normal linear model was explored and discussed in detail by Lindley and Smith (1972). Although each approach allows for discounting of the historical data in some sense, each borrowing approach has its own pros and cons, and a direct comparison between them is out of the scope of this paper. However, much research has been done comparing the performance of these approaches in practice, particularly in the context of clinical trials (Viele et al., 2014; Ibrahim et al., 2015; Lewis et al., 2019; Su et al., 2022; Lesaffre et al., 2024). Still, to our knowledge, none of these priors are suitable when the current and historical data outcomes are of different types. Recently, Suder et al. (2023) introduced the general concept of *Bayesian transfer learning*. Transfer learning is a recent concept in statistical machine learning that seeks to improve inference and/or predictive accuracy on a domain of interest by leveraging data from related domains. Suder et al. (2023) show how the power prior, normalized power prior, robust MAP, and BHM can emerge as special cases of the general approach.

We emphasize here, however, that *all* of the aforementioned papers assume the fundamental notation that the historical data is of the *same data type* as the current data. For example, the historical data may be discrete, such as binary or count data, while the current data is continuous, such as, time-to-event, normally distributed, or gamma distributed, or vice versa–where the current data is discrete but the historical data is continuous. To the best of our knowledge, the *only* paper that handles historical and current data of different types using power prior ideas is that of Alt et al. (2023a). It is not at all clear if the paper by Suder et al. (2023) can be applied or extended to these different data type settings. In this paper, we extend and develop the framework by Alt et al. (2023a) for right censored time-to-event data, which requires a totally new development, theory, and machinery for handling censoring in an appropriate way.

As noted above, the use of the power prior has been demonstrated in time-to-event settings in the context of cancer clinical trials (Chen et al., 1999a, 2002; Ibrahim et al., 2003, 2012a, 2015). These papers address analysis with the power prior for many survival models, including the proportional hazards (PH) model with constant baseline hazard, parametric cure rate model, semi-parametric cure rate model and promotion time cure

rate model. Additional papers by Psioda and Ibrahim (2018, 2019) consider use of the power prior for the design of clinical trials with time-to-event endpoints using the cure rate model and Cox model, respectively. Other historical data borrowing methods, aside from the power prior and its variations, have also been developed for time-to-event settings, including the meta-analytic methods of Roychoudhury and Neuenschwander (2020) and the references therein. Their approach is based on a robust hierarchical model for piecewise exponential data, accommodating varying degrees of between-trial heterogeneity and leveraging both individual and aggregate data.

However, as mentioned earlier, these papers fail to address the scenario in which the historical and current data have different data types and therefore require different models. Such differences in data type (i.e., binary vs time-to-event) may lead to parameters from the historical and current data that have non-comparable magnitudes. Recently, Psioda et al. (2020a,b) developed approaches for design and analysis of concurrent clinical trials with information borrowing across different outcome models, but the methods proposed do not address the problem of incorporating prior information from existing data, which is the primary focus in this paper.

More recently, Alt et al. (2023a) proposed a methodology for incorporating information from a historical data set where the outcome has a different data type compared to the outcome measured in a current data. They consider the case where the outcomes can be reasonably assumed to follow two generalized linear models (GLMs). To account for differences in parameter magnitudes for the two data models, those authors develop the scale transformed power prior (straPP). Specifically, the straPP is constructed via a scale transformation that equates the historical and current model parameters are equivalent. The rescaling is achieved by multiplying each parameter by the square root of the Fisher information matrix for the respective model. However, Alt et al. (2023a) do not consider censored time-to-event data in their development, which necessitates substantive new developments in order to formulate a straPP. In particular, for settings with time-to-event outcomes, the straPP transformation must accommodate nuisance parameters such as the baseline hazard, right censored data, and the construction of an appropriate one-to-one transformation in which the regression coefficient vectors in the historical and current data models are of necessarily different dimensions. Our contribution in this work addresses all of these challenges.

In this paper, we develop the partial-borrowing straPP for two commonly used survival models: a PH model with piecewise constant baseline hazard and the mixture cure rate model that assumes a common probability of cure for all subjects paired with a PH model with piecewise constant baseline hazard for the non-cured population. In addition, we develop a generalized version of the partial-borrowing straPP, which we call the generalized scale transformed power prior (Gen-straPP), that allows for random deviations that occur between the components of the rescaled parameter vectors through a hierarchical modeling framework. We explore in simulations and real data analyses the relative merits of the partial-borrowing straPP and Gen-straPP compared to other commonly used priors.

The rest of the paper proceeds as follows. Section 2 motivates the straPP transformation using two pairs of clinical data sets. Section 3 generalizes the straPP and

Gen-straPP of Alt et al. (2023a) to accommodate time-to-event data. In Section 4, we derive the Fisher information matrices for the piecewise constant PH model and mixture cure rate model with piecewise constant PH for the non-cured fraction, which are needed for the straPP and Gen-Strapp transformations. We present results from simulation studies in Section 5. In Sections 6 and 7, we analyze two pairs of data sets, where the historical data outcomes are respectively assumed to be binary and normal and the current data outcomes are time-to-event. In Section 8, we discuss the strengths and limitations of our approach.

2 Motivation

We consider two case studies based on melanoma clinical trials in which the current data sets both had relapse-free survival (RFS) as a key outcome and the historical data sets had outcomes believed to be related to (but not identical to) RFS. The first case study, described in Section 2.1 discusses a case study based on the Eastern Cooperative Oncology Group (ECOG) trials E1684 and E1690 trials. This first case study suffices to illustrate the performance of the proposed method in cases where it may be reasonable to apply. The second case study, described in Section 2.2, involves the E2696 and E1694 ECOG studies. This second case study was chosen to illustrate the robustness the proposed method when the scale transformation motivating using the straPP is clearly violated. Summary statistics for the baseline characteristics considered in each study are presented in Table 1.

Variable	**Study**			
	E1684, N = 262[1]	**E1690**, N = 426[1]	**E1694**, N = 200[1]	**E2696**, N = 98[1]
Treatment				
Control	128 (49%)	211 (50%)	99 (50%)	67 (68%)
Treatment	134 (51%)	215 (50%)	101 (50%)	31 (32%)
Sex				
Male	158 (60%)	267 (63%)	129 (64%)	56 (57%)
Female	104 (40%)	159 (37%)	71 (36%)	42 (43%)
Performance				
Fully active	233 (89%)	372 (87%)	172 (86%)	71 (72%)
Ambulatory	29 (11%)	54 (13%)	28 (14%)	27 (28%)
Age	47.0 (13.1)	47.9 (13.2)	53.4 (13.0)	48.1 (12.8)

[1]Mean (Standard Deviation) or Frequency (%).

Table 1: Summary statistics for the ECOG data sets.

2.1 The ECOG E1684 and E1690 Trials

The E1690 trial (Kirkwood et al., 2000) was a randomized three-arm trial conducted to investigate the effectiveness of interferon alfa-2b (IFN) on RFS for melanoma patients.

RFS is defined to be the duration from treatment until the patient experiences a relapse or experiences death from any cause. The three arms in the trial were high-dose IFN, low-dose IFN, and observation. However, we only consider the high-dose IFN arm and the observation arm for analysis. We consider an indicator of relapse within two-years from the E1684 trial as a mature, binary outcome in the historical data. The E1684 trial evaluated the effectiveness of high-dose IFN on RFS (Kirkwood et al., 1996) compared to observation. Two-year RFS is considered an important threshold and milestone for melanoma clinical trials. Moreover, many cancer clinical trials consider milestone survival as the primary endpoint, including, but not limited to, Buunen et al. (2009), Warren et al. (2012), Kim et al. (2012), Nilsson et al. (2013), and Joshua et al. (2015). As both the E1684 and E1690 trials investigate the effect of IFN on related outcomes, they provide an example where the partial-borrowing straPP may provide improved performance compared to other commonly used priors (and compared to not using the seemingly relevant prior information in analysis).

2.2 The ECOG E2696 and E1694 Trials

The E1694 trial (Kirkwood et al., 2001b) was a randomized trial to evaluate the effectiveness of the GM2-KLH/QS-21 (GMK) vaccine in comparison to high-dose IFN adjuvant therapy on RFS for patients with resectable high-risk or very high-risk melanoma. An earlier trial, E2696 (Kirkwood et al., 2001a), was conducted to evaluate the clinical and immunologic effects of the combination of the GMK vaccine and IFN compared to GMK alone. For this historical data, we consider immunoglobulin (IgM) levels at day 28 as an important immunological outcome. There was no treatment effect on IgM levels in the E2696 study (by design) because all treatment arms contained GMK as a component (and GMK was the driver of immunological response). In fact, one of the objectives of the study was to confirm that IFN did not impede immunological response. Thus, this case study corresponds to a scenario where treatment (i.e., IFN) has no effect in the historical data outcome and where treatment is shown to have an effect in the current data. We use this analysis as a opportunity to explore the robustness of the proposed methodology when its underlying assumptions are violated.

3 The Partial-Borrowing StraPP and Its Generalizations

We develop the partial-borrowing straPP of Alt et al. (2023a) for situations where the historical data outcome is related to, but different than, the current data outcome, which we assume to be time-to-event. For example, the historical data outcome may be an indicator of relapse within two years and the current data outcome might be time to relapse. Developing the partial-borrowing straPP for time-to-event outcomes presents unique challenges. In this setting, there are often nuisance parameters, such as the parameters corresponding to a baseline hazard, upon which one does not wish to borrow information.

To fix ideas, let $\boldsymbol{\theta}$ denote the parameters for the current (time-to-event) data and let $\boldsymbol{\eta}$ denote those for the historical data. Suppose that we can decompose $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$ and

$\boldsymbol{\eta} = (\boldsymbol{\eta}_1', \boldsymbol{\eta}_2')'$, where $\boldsymbol{\theta}_1$ and $\boldsymbol{\eta}_1$ are p-dimensional vectors (e.g., regression coefficients from a PH model and a GLM, respectively), $\boldsymbol{\theta}_2$ is an r-dimensional vector of nuisance parameters (e.g., parameters for the baseline hazard), and $\boldsymbol{\eta}_2$ is an s-dimensional vector of nuisance parameters (e.g., an intercept and a dispersion parameter).

We wish to construct an informative prior for $\boldsymbol{\theta}_1$ on the basis of $\boldsymbol{\eta}_1$ without borrowing any information on the nuisance parameters. We begin by assuming a power prior for $\boldsymbol{\eta}$, i.e.,

$$\pi_{\text{PP}}(\boldsymbol{\eta}|D_0, a_0) \propto \mathcal{L}_0(\boldsymbol{\eta}|D_0)^{a_0}\pi_0(\boldsymbol{\eta}), \tag{1}$$

where $\mathcal{L}_0(\cdot|D_0)$ denotes the likelihood function for the historical data, $a_0 \in [0, 1]$ is a scalar discounting parameter, and π_0 is an initial prior. The power prior in (1) is an informative prior for the parameters for the historical data, but it would be inappropriate to assume $\boldsymbol{\eta}_1 = \boldsymbol{\theta}_1$ as the outcomes are measured on *different scales* (e.g., the outcomes are of two different data types). To remedy this, Alt et al. (2023a) apply a scale transformation for the parameters in the power prior in (1). However, their approach was developed in the context where the historical and current data models are GLMs. The incorporation of time-to-event data introduces more nuisance parameters, and a generalization of their approach is needed.

Let the Fisher information matrices for the historical and current data sets be represented in block form as

$$\mathcal{I}_{\boldsymbol{\eta}}(\boldsymbol{\eta}|\boldsymbol{X}_0) = \begin{pmatrix} \mathcal{I}_{\boldsymbol{\eta}_1\boldsymbol{\eta}_1}(\boldsymbol{\eta}|\boldsymbol{X}_0) & \mathcal{I}_{\boldsymbol{\eta}_1\boldsymbol{\eta}_2}(\boldsymbol{\eta}|\boldsymbol{X}_0) \\ \mathcal{I}_{\boldsymbol{\eta}_1\boldsymbol{\eta}_2}(\boldsymbol{\eta}|\boldsymbol{X}_0)' & \mathcal{I}_{\boldsymbol{\eta}_2\boldsymbol{\eta}_2}(\boldsymbol{\eta}|\boldsymbol{X}_0) \end{pmatrix}$$

and

$$\mathcal{I}_{\boldsymbol{\theta}}(\boldsymbol{\theta}|\boldsymbol{X}_0) = \begin{pmatrix} \mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\theta}|\boldsymbol{X}_0) & \mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2}(\boldsymbol{\theta}|\boldsymbol{X}_0) \\ \mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2}(\boldsymbol{\theta}|\boldsymbol{X}_0)' & \mathcal{I}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}(\boldsymbol{\theta}|\boldsymbol{X}_0) \end{pmatrix},$$

respectively, where we evaluate both Fisher information matrices using the historical data design matrix $\boldsymbol{X}_0$. The purpose of evaluating the Fisher information matrix using the historical data is to formulate a prior for $\boldsymbol{\theta}$ that depends on the historical data alone. Consider the $(p + r + s)$-dimensional "augmented" vectors $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3)$ and $\tilde{\boldsymbol{\eta}} = (\boldsymbol{\eta}_1, \boldsymbol{\eta}_3, \boldsymbol{\eta}_2)$ and $(p + r + s) \times (p + r + s)$ matrices given by

$$\tilde{\mathcal{I}}_{\boldsymbol{\eta}}(\tilde{\boldsymbol{\eta}}|\boldsymbol{X}_0) = \begin{pmatrix} \mathcal{I}_{\boldsymbol{\eta}_1\boldsymbol{\eta}_1}(\boldsymbol{\eta}|\boldsymbol{X}_0) & \mathbf{0}_{p\times r} & \mathbf{0}_{p\times s} \\ \mathbf{0}_{r\times p} & \boldsymbol{I}_r & \mathbf{0}_{r\times s} \\ \mathbf{0}_{s\times p} & \mathbf{0}_{s\times r} & \boldsymbol{I}_s \end{pmatrix}$$

and

$$\tilde{\mathcal{I}}_{\boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}|\boldsymbol{X}_0) = \begin{pmatrix} \mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\theta}|\boldsymbol{X}_0) & \mathbf{0}_{p\times r} & \mathbf{0}_{p\times s} \\ \mathbf{0}_{r\times p} & \boldsymbol{I}_r & \mathbf{0}_{r\times s} \\ \mathbf{0}_{s\times p} & \mathbf{0}_{s\times r} & \boldsymbol{I}_s \end{pmatrix},$$

where $\mathbf{0}_{a\times b}$ is an $a \times b$ matrix of zeros and $\boldsymbol{I}_q$ is the q-dimensional identity matrix. The partial borrowing straPP transformation in this setting is then given by solving

$$\tilde{\mathcal{I}}_{\boldsymbol{\eta}}(\tilde{\boldsymbol{\eta}}|\boldsymbol{X}_0)^{1/2}\tilde{\boldsymbol{\eta}} = \tilde{\mathcal{I}}_{\boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}|\boldsymbol{X}_0)^{1/2}\tilde{\boldsymbol{\theta}}. \tag{2}$$

Note that (2) implies $\boldsymbol{\theta}_3 = \boldsymbol{\eta}_2$ and $\boldsymbol{\eta}_3 = \boldsymbol{\theta}_2$, so that one only needs to solve the equation

$$\mathcal{I}_{\boldsymbol{\eta}_1\boldsymbol{\eta}_1}(\boldsymbol{\eta}|\boldsymbol{X}_0)^{1/2}\boldsymbol{\eta}_1 = \mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\theta}|\boldsymbol{X}_0)^{1/2}\boldsymbol{\theta}_1. \tag{3}$$

If one considers the nuisance parameters as fixed and known, note that $\mathcal{I}^{-1}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\cdot)$ and $\mathcal{I}^{-1}_{\boldsymbol{\eta}_1\boldsymbol{\eta}_1}(\cdot)$ are the asymptotic covariance matrices for $\boldsymbol{\theta}_1$ and $\boldsymbol{\eta}_1$, respectively. Thus, the parameters $\mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\theta}|\boldsymbol{X}_0)^{1/2}\boldsymbol{\theta}_1$ and $\mathcal{I}_{\boldsymbol{\eta}_1\boldsymbol{\eta}_1}(\boldsymbol{\eta}|\boldsymbol{X}_0)^{1/2}\boldsymbol{\eta}_1$ can be viewed as standardized (or unitless). We construct the asymptotic covariance matrix for the current data likelihood using the historical covariate matrix to allow the partial-borrowing straPP to be constructed using information entirely from the historical data (e.g., outcome and covariates).

The transformation in (3) implies the existence of a function g_1 such that we may solve $\boldsymbol{\eta}_1 = g_1^{-1}(\boldsymbol{\theta}, \boldsymbol{\eta}_2)$. Hence, the partial borrowing straPP for time-to-event data is given by applying a Jacobian adjustment to the power prior in (1), i.e.,

$$\pi(\boldsymbol{\theta}|D_0, a_0) = \int \pi_{\text{PP}}(g_1^{-1}(\boldsymbol{\theta}, \boldsymbol{\eta}_2), \boldsymbol{\eta}_2|D_0, a_0) \left|\frac{\partial g_1^{-1}}{\partial \boldsymbol{\theta}_1}\right| \pi_0(\boldsymbol{\theta}_2) d\boldsymbol{\eta}_2, \tag{4}$$

where $\pi_0(\boldsymbol{\theta}_2)$ is an initial prior for $\boldsymbol{\theta}_2$. We stress that the straPP in (4) does not assume that the transformation in (3) holds for the data model. The transformation is only used as a mechanism to formulate a prior for $\boldsymbol{\theta}_1$ that is compatible with the current data outcome.

Under the special case $\boldsymbol{\theta}_1 = \boldsymbol{\eta}_1$, the transformation in (2) results in an identity transformation. As a result, the straPP in (4) reduces to the partial borrowing power prior (Ibrahim et al., 2012b). Partial discounting (e.g., in models with latent variables, it is often desirable to only discount the non-latent parameters), can be easily handled in the approach by effectively including them in the vector of nuisance parameters $\boldsymbol{\eta}_1$. The block-matrix approach giving the transformation in (2) effectively unifies the concepts of partial borrowing and partial discounting, while offering a generalization in that transformations of parameters may be implemented.

3.1 The Generalized StraPP

As a more flexible alternative to the straPP, Alt et al. (2023a) developed the generalized straPP (Gen-straPP). Using the transformation in (2), we generalize their approach. Specifically, the Gen-straPP imposes the transformation

$$\mathcal{I}_{\boldsymbol{\eta}_1\boldsymbol{\eta}_1}(\boldsymbol{\eta})^{1/2}\boldsymbol{\eta}_1 = \mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\boldsymbol{\theta})^{1/2}\boldsymbol{\theta}_1 + \boldsymbol{c}_0, \tag{5}$$

where $\mathcal{I}_{\boldsymbol{\eta}_1\boldsymbol{\eta}_1}(\cdot)$ and $\mathcal{I}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}(\cdot)$ are as defined in (3) and $\boldsymbol{c}_0$ is an $r \times 1$ vector that allows component-specific deviations from the standardized parameter values for $\boldsymbol{\eta}_1$ and $\boldsymbol{\theta}_1$. We may write the transformation induced by (5) as $\boldsymbol{\eta}_1 = g_{1,\boldsymbol{c}_0}(\boldsymbol{\theta}, \boldsymbol{\eta}_2)$, and thus the joint partial-borrowing Gen-straPP is

$$\pi(\boldsymbol{\theta}, \boldsymbol{c}_0, \omega_0|D_0, a_0) = \pi_{\text{PP}}\left(g_{1,\boldsymbol{c}_0}(\boldsymbol{\theta}, \boldsymbol{\eta}_2), \boldsymbol{\eta}_2|D_0, a_0\right) \left|\frac{\partial g_{1,\boldsymbol{c}_0}^{-1}}{\partial \boldsymbol{\theta}_1}\right| \pi_0(\boldsymbol{c}_0|\omega_0)\pi_0(\omega_0)\pi_0(\boldsymbol{\theta}_2) d\boldsymbol{\eta}_2, \tag{6}$$

where here we specify $\boldsymbol{c_0} \sim N_p(\boldsymbol{0}, \omega_0^2 \boldsymbol{I}_p)$ and a half standard normal prior on ω_0, i.e., $\pi_0(\omega_0) \propto \phi(\omega_0) 1\{\omega_0 > 0\}$, where $\phi(\cdot)$ is the standard normal density function.

4 The Partial-Borrowing StraPP Transformation for Time-to-Event Data

We now develop the Fisher information matrices needed for implementation of the partial-borrowing straPP for two different survival models: (1) a PH model with piecewise constant baseline hazard and (2) a mixture cure rate model that assumes a common probability of cure for all subjects paired with a PH model with piecewise constant baseline hazard for the non-cured population. These models are discussed in Sections 4.2 and 4.3, respectively. While traditional analyses of time-to-event data use the observed information matrix, this is not possible for the straPP priors since the historical data are not time-to-event. In principle, one could use the observed information matrix of the current data, but this would result in a prior that depends on outcome of the current data set. Thus, to develop straPP priors when the historical data are not time-to-event, we must make assumptions about censoring. In general, we will assume non-informative censoring, meaning that the probability of censoring does not depend on the underlying event time.

For the remainder of this paper, we treat the historical data distribution as arising from a GLM and the current data as following a time-to-event model. Let $D_0 = (\boldsymbol{y}_0, \boldsymbol{X}_0, n_0)$ denote the historical data of size n_0, where $\boldsymbol{y}_0$ is a n_0-dimensional vector of responses and $\boldsymbol{X}_0$ is a $n_0 \times p$ design matrix, which does *not* include an intercept term. Thus, using the notation in Section 3, $\boldsymbol{\eta} = (\boldsymbol{\beta}_{01}', \beta_{00}, \phi_0)'$, where $\boldsymbol{\beta}_{01}$ is a p-dimensional vector of regression coefficients pertaining to covariates, β_{00} is an intercept, and ϕ_0 is a dispersion parameter, which may be known for some models. The Fisher information for the historical data may be expressed as

$$I_{\boldsymbol{\eta}_1 \boldsymbol{\eta}_1}(\boldsymbol{\eta} | \boldsymbol{X}_0) = \frac{1}{\phi_0} \boldsymbol{X}_0' \boldsymbol{W}(\beta_{00}, \boldsymbol{\beta}_{01}) \boldsymbol{X}_0, \tag{7}$$

where $\boldsymbol{W}(\beta_{00}, \boldsymbol{\beta}_{01})$ is a diagonal matrix of variance functions that, in general, depends on the regression coefficients.

In this paper we focus on the piecewise constant PH model and mixture cure rate model with piecewise constant PH for the non-cured individuals but note that the straPP can be applied to other time-to-event models. In Supplementary Material (Alt et al., 2023b) Section 2, we analytically derive the Fisher information matrix for applying the straPP and Gen-straPP to the promotion time cure rate model.

4.1 The StraPP for PH Models under Right Censoring

We now develop the straPP for PH models in full generality. We begin by deriving the Fisher information matrix for PH models. We then discuss why making assumptions regarding censoring is necessary in order to conduct a straPP transformation.

Let $D_1 = (\boldsymbol{y}_1, \boldsymbol{X}_1, \boldsymbol{\nu}_1, n_1)$ denote the current data, in which $\boldsymbol{y}_1$ is the vector of observation times, $\boldsymbol{X}_1 = (\boldsymbol{x}_{11}, \ldots, \boldsymbol{x}_{1,n_1})'$ is the $n_1 \times p$ design matrix, $\boldsymbol{\nu} = (\nu_{11}, \ldots, \nu_{1,n_1})$ is a vector of indicators for whether an individual had an event observed, and n_1 is the sample size. The likelihood for parametric PH models may be expressed as

$$\mathcal{L}_1(\boldsymbol{\beta}, \boldsymbol{\lambda} | D_1) = \prod_{i=1}^{n_1} \left[h_0(y_{1i}|\boldsymbol{\lambda}) e^{\boldsymbol{x}_{1i}'\boldsymbol{\beta}} \right]^{\nu_{1i}} \exp\left\{ -H_0(y_{1i}|\boldsymbol{\lambda}) e^{\boldsymbol{x}_{1i}'\boldsymbol{\beta}} \right\} \tag{8}$$

where $\boldsymbol{\lambda}$ is a vector of parameters pertaining to the baseline hazard, $\boldsymbol{\beta}$ is a vector of regression coefficients associated with covariates $(\boldsymbol{x}_{1i}, i = 1, \ldots, n)$, $y_{1i} = \min\{t_{1i}, c_{1i}\}$ is the observed time (i.e., the minimum of the event time t_{1i} and the censoring time c_{1i}), and $\nu_{1i} = I(t_{1i} \leq c_{1i})$ is the event indicator.

Suppose we possess historical data $D_0 = \{(y_{0i}, \boldsymbol{x}_{0i}), i = 1, \ldots, n_0\}$ with likelihood $\mathcal{L}_0(\boldsymbol{\beta}_0, \boldsymbol{\eta}_0)$. We assume that the historical data outcomes are related to (but possibly of different types than) the current data outcomes. Our goal is to apply the (Gen-)straPP to construct an informative prior for $\boldsymbol{\beta}$. To do this, we require the Fisher information for the likelihoods pertaining to the current and historical data sets. The Fisher information for the current data based on the likelihood in (8) is given by

$$\mathcal{I}^1(\boldsymbol{\beta}, \boldsymbol{\lambda} | D_1) = \sum_{i=1}^{n_1} E\left[\begin{pmatrix} H_0(y_{1i}|\boldsymbol{\lambda}) e^{\boldsymbol{x}_{1i}'\boldsymbol{\beta}} \boldsymbol{x}_{1i} \boldsymbol{x}_{1i}' & e^{\boldsymbol{x}_{1i}'\boldsymbol{\beta}} \left[\frac{\partial H_0(y_{1i}|\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} \right]' \boldsymbol{x}_{1i} \\ e^{\boldsymbol{x}_{1i}'\boldsymbol{\beta}} \boldsymbol{x}_{1i}' \left[\frac{\partial H_0(y_{1i}|\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} \right] & e^{\boldsymbol{x}_{1i}'\boldsymbol{\beta}} \frac{\partial^2 H_0(y_{1i}|\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} - \nu_{1i} \frac{\partial^2 \log h_0(y_{1i}|\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} \end{pmatrix} \right]. \tag{9}$$

Let $\mathcal{I}^0(\boldsymbol{\beta}_0, \boldsymbol{\eta}_0 | D_0)$ denote the Fisher information for the historical data likelihood evaluated at the historical data D_0. To conduct the straPP transformation so the prior depends only on the historical data D_0, we need to solve

$$\left[\mathcal{I}^1_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\beta}, \boldsymbol{\lambda} | D_0) \right]^{1/2} \boldsymbol{\beta} = \left[\mathcal{I}^0_{\boldsymbol{\beta}_0\boldsymbol{\beta}_0}(\boldsymbol{\beta}_0, \boldsymbol{\eta}_0 | D_0) \right]^{1/2} \boldsymbol{\beta}_0 \tag{10}$$

for $\boldsymbol{\beta}$, giving $\boldsymbol{\beta}_0 = g(\boldsymbol{\beta})$. The posterior density for the current data parameters is then given by

$$p(\boldsymbol{\beta}, \boldsymbol{\lambda} | D_1, D_0) \propto \mathcal{L}_1(\boldsymbol{\beta}, \boldsymbol{\lambda} | D_1) \mathcal{L}_0(g(\boldsymbol{\beta}), \boldsymbol{\eta}_0)^{a_0} \pi_0(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\eta}_0) d\boldsymbol{\eta}_0. \tag{11}$$

However, the Fisher information in (9) depends on expectations involving the joint distribution of the time-to-event and censoring times. To avoid making assumptions about this joint distribution, we instead compute the information under no censoring for the transformation in (10), which allows for analytically tractable expectations for the Fisher information matrix $\mathcal{I}^1_{\boldsymbol{\beta}\boldsymbol{\beta}}(\cdot | D_0)$ regardless of censoring distribution. Note that, in principle, an observed Fisher information matrix based on the historical data could be used, but this would require the current and historical data outcomes to be directly compatible, obviating the need for a transformation. Thus, while the assumption seems somewhat restrictive it is necessary in general to convert from the historical parameter space to the current parameter space. We emphasize that this assumption only applies for the transformation to solve (10); the current data likelihood in (8) still incorporates censoring in the posterior density provided in (11).

4.2 The Fisher Information Matrix for the PH Model

We assume a PH regression model with piecewise constant baseline hazard having J components corresponding to time axis partition $s_0 = 0 < s_1 < \cdots < s_{J-1} < s_J = \infty$. Using the notation of Section 3, the parameters for the PH model may be expressed as $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\lambda}')$, where $\boldsymbol{\beta}$ is a p-dimensional vector of regression coefficients pertaining to covariates and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_J)'$ is a vector of baseline hazards. The current data likelihood may be written as

$$\mathcal{L}^{\text{PH}}(\boldsymbol{\theta}|D_1) = \prod_{i=1}^{n_1}\prod_{j=1}^{J}\left\{\left(\lambda_j e^{\boldsymbol{x}_{1i}^T\boldsymbol{\beta}}\right)^{\nu_i} \times \exp\left[-\left\{\lambda_j(y_{1i} - s_{j-1}) + \sum_{g=1}^{j-1}\lambda_g(s_g - s_{g-1})\right\}e^{\boldsymbol{x}_{1i}^T\boldsymbol{\beta}}\right]\right\}^{\delta_{ij}}, \tag{12}$$

where $\boldsymbol{\delta}$ is a matrix in which $\delta_{ij} = 1\{s_{j-1} < y_{1i} \leq s_j\}$ denotes whether the i^{th} individual had an event or was censored in the j^{th} interval, $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_{n_1})$ is a vector of indicators for whether an individual had an event observed, and n_1 is the sample size.

Let $\mathcal{L}_i^{\text{PH}}$ denote the i^{th} individual's contribution to the PH likelihood in (12). As shown by Gelfand and Mallick (1995), we may write

$$\mathcal{L}_i^{\text{PH}} = \left[h_0(y_{1i})e^{\boldsymbol{x}_{1i}'\boldsymbol{\beta}_1}\right]^{\nu_i}\exp\left\{-H_0(y_{1i})e^{\boldsymbol{x}_{1i}'\boldsymbol{\beta}_1}\right\},$$

where $h_0(y_{1i}) = \prod_{j=1}^{J}\lambda_j^{\delta_{ij}}$ and

$$\begin{aligned} H_0(y_{1i}) &= \prod_{j=1}^{J}\left\{\lambda_j(y_{1i} - s_{j-1}) + \sum_{g=1}^{j-1}\lambda_g(s_g - s_{g-1})\right\}^{\delta_{ij}} \\ &= \prod_{j=1}^{J}\left\{\lambda_j(y_{1i} - s_{j-1}) + H_0(s_{j-1})\right\}^{\delta_{ij}}. \end{aligned}$$

Let $\ell^{\text{PH}}(\boldsymbol{\beta}, \boldsymbol{\lambda}|D_1)$ denote the log likelihood based on (12). Assuming no censoring (i.e., $\nu_i = 1$ for all $i \in \{1, \ldots, n_1\}$), then

$$\frac{\partial^2\ell^{\text{PH}}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} = -\sum_{i=1}^{n_1}H_0(y_{1i})e^{\boldsymbol{x}_{1i}'\boldsymbol{\beta}_1}\boldsymbol{x}_{1i}\boldsymbol{x}_{1i}'. \tag{13}$$

Now,

$$\begin{aligned} E\left[H_0(y_{1i})\right] &= \sum_{j=1}^{J}E\left[\delta_{ij}\left(\lambda_j(y_{1i} - s_{j-1}) + H_0(s_{j-1})\right)\right], \\ &= \sum_{j=1}^{J}\int_{s_{j-1}}^{s_j}\left[\lambda_j(y_{1i} - s_{j-1}) + H_0(s_{j-1})\right]f(y_{1i}|\boldsymbol{\beta}, \boldsymbol{\lambda})dy_{1i}. \end{aligned} \tag{14}$$

For each of the J integrands in the sum in (14), let $u_i = e^{\boldsymbol{x}'_{1i}\boldsymbol{\beta}}[\lambda_j(y_{1i} - s_{j-1}) + H_0(s_{j-1})]$ so that $y_{1i} = s_{j-1} + \frac{u_i - H_0(s_{j-1})e^{\boldsymbol{x}'_{1i}\boldsymbol{\beta}}}{\lambda_j e^{\boldsymbol{x}'_{1i}\boldsymbol{\beta}}}$ and $dy_{1i} = [\lambda_j e^{\boldsymbol{x}'_{1i}\boldsymbol{\beta}}]^{-1} du_i$. Then

$$\begin{aligned} E\left[H_0(y_{1i})\right] &= e^{-\boldsymbol{x}'_{1i}\boldsymbol{\beta}} \sum_{j=1}^{J} \int_{H_0(s_{j-1})e^{\boldsymbol{x}'_{1i}\boldsymbol{\beta}}}^{H_0(s_j)e^{\boldsymbol{x}'_{1i}\boldsymbol{\beta}}} u_i e^{-u_i} du_i \\ &= e^{-\boldsymbol{x}'_{1i}\boldsymbol{\beta}} \sum_{j=1}^{J} \left\{ H(s_{j-1})e^{-H(s_{j-1})} - H(s_j)e^{-H(s_j)} + e^{-H(s_{j-1})} - e^{-H(s_j)} \right\} \\ &= e^{-\boldsymbol{x}'_i\boldsymbol{\beta}} \left\{ \left[H(s_0)e^{-H(s_0)} - H(s_J)e^{-H(s_J)} \right] + [S(s_0) - S(s_J)] \right\}. \end{aligned}$$

Note that since $s_0 = 0$, $H(s_0) = 0$ and $S(s_0) = 1$. Moreover, since $s_J = \infty$, $H(s_J)e^{-H(s_J)} = \lim_{x\to\infty} xe^{-x} = 0$ and $S(s_J) = 0$. It follows that $E[H_0(y_{1i})] = e^{-\boldsymbol{x}'_{1i}\boldsymbol{\beta}}$. Substituting this result into the expectation of (13), we get that the Fisher information matrix for the regression coefficients for the PH model is given by

$$\mathcal{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\beta}, \boldsymbol{\lambda} | \boldsymbol{X}_1) = \boldsymbol{X}_1' \boldsymbol{X}_1. \tag{15}$$

We now describe how the Gen-straPP works in the Markov chain Monte Carlo (MCMC) scheme for the PH model with piecewise constant baseline hazards. Suppose the t^{th} iteration of the MCMC scheme yields $(\boldsymbol{\eta}^{(t)}, \boldsymbol{\lambda}^{(t)}, \boldsymbol{c}_0^{(t)})$. Substituting the Fisher information matrix for the historical data in (7) and that for the current data in (15) evaluated at $\boldsymbol{X}_0$ into the Gen-straPP transformation in (5) and solving, we may obtain $\boldsymbol{\beta}^{(t)}$ as

$$\boldsymbol{\beta}^{(t)} = [\boldsymbol{X}_0'\boldsymbol{X}_0]^{-1/2} \left\{ \left[\phi_0^{(t)} \right]^{-1/2} \left[\boldsymbol{X}_0' \boldsymbol{W} \left(\beta_{00}^{(t)}, \boldsymbol{\beta}_{01}^{(t)} \right) \boldsymbol{X}_0 \right]^{1/2} \boldsymbol{\beta}_{01}^{(t)} - \boldsymbol{c}_0^{(t)} \right\} \tag{16}$$

where we may substitute $\boldsymbol{c}_0^{(t)} = \boldsymbol{0}$ in (16) to obtain the analogous straPP transformation in (3). We may thus use "complementary sampling" techniques to sample from the posterior density (Alt et al., 2023a). We provide details regarding the complementary sampling technique in Supplementary Material (Alt et al., 2023b) Section 1.

4.3 The Fisher Information Matrix for the Mixture Cure Rate Model

For the mixture cure rate model, we let $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\lambda}', \boldsymbol{p}_c')'$ denote the current data model parameters, where $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ are defined in Section 4.2 and $\boldsymbol{p}_c = (p_{c1}, \ldots, p_{cn_1})'$ is a n_1-dimensional vector of cure proportions. We can write the *complete* data likelihood (i.e., the likelihood including the latent indicator of cure) as

$$\mathcal{L}^c(\boldsymbol{\theta} | D_1^c) \propto \prod_{i=1}^{n_1} p_{ci}^{\Delta_i(1-\nu_i)} \left[(1 - p_{ci}) \tilde{\mathcal{L}}_i(\boldsymbol{\beta}, \boldsymbol{\lambda}) \right]^{1-\Delta_i}, \tag{17}$$

where $D_1^c = (\boldsymbol{y}_1, \boldsymbol{X}_1, \boldsymbol{\Delta}, \boldsymbol{\nu}, n_1)$ denotes the current data, in which $\boldsymbol{\Delta} = (\Delta_1, \ldots, \Delta_{n_1})$, where $\Delta_i = 1\{y_{1i} = \infty\}$ is an indicator for whether the i^{th} individual is cured, p_{ci}

denotes probability of cure for subject i, which may depend on covariates, and

$$\tilde{\mathcal{L}}_i(\boldsymbol{\beta}, \boldsymbol{\lambda}) = S(y_{1i}|\boldsymbol{\beta}, \boldsymbol{\lambda})^{1-\nu_i} f(y_{1i}|\boldsymbol{\beta}, \boldsymbol{\lambda})^{\nu_i}$$

is the density function for the observed failure times and the survival time for censored individuals for some model of event times. Note that though we can infer an individual is not cured if we observe that the individual has an event, we cannot determine whether an individual is cured, thus $\boldsymbol{\Delta}$ is a partially latent variable.

We digress momentarily to further clarify our use of term complete data likelihood in this context. Several technical points are relevant here. First, our use of the term complete data likelihood should not be construed to imply that each individual's event time is observed (indeed this is impossible if one is cured), but rather that the data D_1^c include the latent indicators of cure. Moreover, we acknowledge that under dependent censoring mechanisms, the *full* likelihood would include a model for the random censorship mechanism, which we do not include here. This is based on an implicit assumption of independent censoring or a dependent censoring mechanism such as administrative censoring in an event driven trial. As illustrated by the recent work of (Rühl et al., 2023), the impact of ignoring such dependence is negligible for inference unless the number of events is smaller than is typical in a reasonably powered clinical trial.

Here we specify a PH model with piecewise constant baseline hazard with J intervals for the event times in the non-cured population. Let $\tilde{\mathcal{L}}_i^{\text{PH}}(\boldsymbol{\beta}, \boldsymbol{\lambda})$ denote the survival function if individual i's time to event is right-censored and the density function otherwise. Using (17) and (12), we can write the complete data log-likelihood as

$$\ell^c(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{p}_c|D_1^c) = \sum_{i=1}^{n_1} \left\{ \Delta_i(1-\nu_i) \log p_{ci} + (1-\Delta_i) \left[\log(1-p_{ci}) + \log \tilde{\mathcal{L}}_i^{\text{PH}}(\boldsymbol{\beta}, \boldsymbol{\lambda}) \right] \right\}.$$

In Section 4.1, we discussed the necessity of assuming no censoring for the purposes of computing the Fisher information matrix. For a cure rate model, such an assumption, even for a transformation, would be inappropriate since there would be no cure fraction if everyone experienced the event. Thus, to compute the transformation, we compute the Fisher information for the cure rate model assuming all censored individuals are cured.

To that end, let $D = (\boldsymbol{y}_1, \boldsymbol{X}_1, \boldsymbol{\delta}, \boldsymbol{\nu}, n_1)$ denote the observed data. Assuming all censored individuals are cured, we have $\Delta_i = 1\{\nu_i = 0\} = 1 - \nu_i$. Thus, the complete data log likelihood is free of latent variables and hence the observed data log likelihood is

$$\ell(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{p}_c|D_1^c) = \sum_{i=1}^{n} \left\{ (1-\nu_i) \log p_{ci} + \nu_i \left[\log(1-p_{ci}) + \log \tilde{\mathcal{L}}_i^{\text{PH}}(\boldsymbol{\beta}, \boldsymbol{\lambda}) \right] \right\},$$

which leads to the second derivative

$$\frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \sum_{i=1}^{n} \nu_i \frac{\partial^2 \log \tilde{\mathcal{L}}_i^{\text{PH}}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}.$$

The Fisher information matrix is thus given by

$$\begin{aligned}\mathcal{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta}) &= \sum_{i=1}^{n_1} E_{(y_{1i},\nu_i)}\left[\nu_i \frac{\partial^2 \log \tilde{L}_i^{\text{PH}}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right], \\ &= \sum_{i=1}^{n_1}(1-p_{ci})\int \frac{\partial^2 \log \tilde{L}_i^{\text{PH}}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} f(y_{1i}|\nu_i=1)dy_{1i}, \\ &= \boldsymbol{X}'\left(\boldsymbol{I}_n - \boldsymbol{P}_c\right)\boldsymbol{X}, \end{aligned} \tag{18}$$

where $\boldsymbol{P}_c = \text{diag}\{p_{c1},\dots p_{cn}\}$ and the expectation of the second derivative with respect to the regression coefficients was evaluated in Section 4.2. Note that the assumption that all censored individuals are cured leads to a *discounted* Fisher information matrix, e.g., we may write the Fisher information matrix as $I_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\theta}) = \sum_{i=1}^{n}(1-p_{ci})\boldsymbol{x}_i\boldsymbol{x}_i'$. Thus, the contribution of each individual to the Fisher information is a fraction of what it would be without such an assumption. If we conversely assumed no censoring, there would be no discounting. As a result, the assumption "all censored individuals are cured" is a conservative one, protecting against too much borrowing.

Note that the Fisher information for the mixture cure rate model in (18), like that for the PH model, is free of the regression coefficients. Thus, similar to the piecewise constant PH model in Section 4.2, we may use complementary sampling techniques (see Supplementary Material (Alt et al., 2023b) Section 1 for details). If the historical data is a GLM, substituting the Fisher information matrices (18) and (7) into the Gen-straPP transformation (5), we have for iteration t of the MCMC sampling scheme that

$$\boldsymbol{\beta}^{(t)} = \left[\boldsymbol{X}_0'\left(\boldsymbol{I}_{n_0} - \boldsymbol{P}_{c0}^{(t)}\right)\boldsymbol{X}_0\right]^{-1/2}\left\{\left[\phi_0^{(t)}\right]^{-1/2}\left[\boldsymbol{X}_0'\boldsymbol{W}\left(\beta_{00}^{(t)},\boldsymbol{\beta}_{01}^{(t)}\right)\boldsymbol{X}_0\right]^{1/2}\boldsymbol{\beta}_{01}^{(t)} - \boldsymbol{c}_0^{(t)}\right\}, \tag{19}$$

where $\boldsymbol{P}_{c0}^{(t)} = \text{diag}\{p_{c01}^{(t)},\dots,p_{c0,n_0}^{(t)}\}$ is a diagonal matrix and $p_{c0i}^{(t)}$ is the probability that subject i of the historical data set is cured based on parameters at the t^{th} MCMC sampling iteration, which may depend on the subject's covariates.

For our simulations and data analysis, we assume that the proportion cured does not depend on covariates. The reason for this is that the outcome for the historical data (relapse within two years) is more closely related to the time-to-event endpoint for the current data (i.e., the hazard ratio for progression-free survival). Thus, it may be inappropriate to borrow information from the historical data to inform the cure fraction. While out of the scope of this paper, it is straightforward to incorporate covariates in the proportion cured use non-informative priors for the regression coefficients associated with the cure fraction.

For model fitting with the current data, we consider the observed data likelihood so that the posterior density is given by

$$p(\boldsymbol{\beta},\boldsymbol{\lambda}|D_1,D_0) \propto \pi(\boldsymbol{\beta},\boldsymbol{\lambda}|D_0)\prod_{i=1}^{n_1}\left\{p_c(1-\nu_i) + (1-p_c)\tilde{\mathcal{L}}_i^{\text{PH}}(\boldsymbol{\beta},\boldsymbol{\lambda})\right\},$$

where $\pi(\boldsymbol{\beta}, \boldsymbol{\lambda}|D_0)$ is the (Gen-)straPP and $D_1 = (\boldsymbol{y}_1, \boldsymbol{X}_1, \boldsymbol{\nu}, n_1)$ is the observed current data. Thus, while we assume all censored subjects are cured for the purposes of the information matrix calculation to specify the prior, we make no such assumption in the analysis of the data.

5 Simulation Study

In this section, we present a simulation study designed to evaluate the performance of the partial-borrowing straPP and partial-borrowing Gen-straPP compared to each other and to the partial-borrowing power prior, described in Section 2, and a non-informative reference prior. We note that the authors investigated the commensurate prior in an earlier study and found that it underperformed in comparison to the partial-borrowing straPP and partial-borrowing Gen-straPP (Alt et al., 2023a), thus it was excluded from simulations in this paper.

5.1 Simulation Setup

To perform the simulation study, we considered the data from the E1684 trial as the historical data (sample size $n_0 = 262$ after removing patients with missing values for number of cancerous lymph nodes) and data from the E1690 trial as the current data (sample size $n_1 = 426$). We discuss the current and historical data sets in more detail in Section 6. In order to permit a more comprehensive evaluation of the priors considered, we investigated multiple values of the sample size in the current study. To obtain current data sets with reduced sample size, we took bootstrap samples from the E1690 data. This approach preserves the empirical relationships between the covariates and event times while providing current data sets of varying sample sizes for investigation. To evaluate the performance of the priors, we computed several metrics: the average posterior variance, bias of the posterior mean, mean square error (MSE) of the posterior mean, coverage probabilities associated with 95% credible intervals, and average interval scores (Gneiting and Raftery, 2007) which account for both credible interval width and coverage probabilities. For these evaluations, we treated the posterior means from analysis of the full E1690 data set, from which bootstrap samples were taken, as the "true" values of the parameters. In this sense, the full E1690 data set defines a super population from which we drew random samples. Thus, the posterior means (i.e., approximate maximum likelihood estimates (MLEs)) of the full data set may be appropriately viewed as true parameter values for this super population. In Section 8 of the Supplementary Material (Alt et al., 2023b), we additionally provide simulation results using a parametric model to generate the current data, showing the performance of our proposed method across varying values of a_0.

We assumed that the outcomes from the historical data (indicators of a relapse within two years) were independently distributed according to a logistic regression model, i.e., the likelihood for the historical data is given by

$$\mathcal{L}(\beta_{00}, \boldsymbol{\beta}_{01}|D_0) \propto \prod_{i=1}^{n_0} \gamma_{0i}^{y_{0i}} (1 - \gamma_{0i})^{1-y_{0i}},$$

where $\gamma_{0i} = \text{logit}^{-1}(\beta_{00} + \boldsymbol{x}_{0i}'\boldsymbol{\beta}_{01}) = \Pr(y_{0i} = 1|\boldsymbol{x}_{0i})$ is the probability that subject i of the historical data set experiences a relapse within two years, β_{00} is an intercept term, $\boldsymbol{x}_{0i} = (z_{0i}, \boldsymbol{x}_{01i})$ is a p-dimensional vector, where z_{0i} is the treatment indicator and $\boldsymbol{x}_{01i}$ is a vector of covariates. The Fisher information matrix for the logistic regression model may be expressed as

$$\mathcal{I}_{\boldsymbol{\beta}_{01}\boldsymbol{\beta}_{01}}(\beta_{00}, \boldsymbol{\beta}_{01}|\boldsymbol{X}_0) = \boldsymbol{X}_0'\boldsymbol{W}(\beta_{00}, \boldsymbol{\beta}_{01})\boldsymbol{X}_0,$$

where $\boldsymbol{W}(\beta_{00}, \boldsymbol{\beta}_{01}) = \text{diag}\{\gamma_{0i}(1 - \gamma_{0i}), i = 1, \ldots, n_0\}$.

For the current data, we assumed that the time-to-relapse outcomes were independently distributed according to a mixture cure rate model, as described in Section 4.3. For our analysis, the covariates of interest are standardized age, an indicator for gender (0 = male; 1 = female), an indicator for ECOG performance status (0 = fully active; 1 = ambulatory), and an indicator for receiving high-dose IFN (1 = high-dose IFN; 0 = observation). Age was centered and scaled by the sample mean and standard deviation, respectively. Henceforth we refer to standardized age simply as age for ease of exposition. We included an intercept term in the historical data model but not in the current data model.

As the historical data were binary, we solved the partial-borrowing straPP transformation for the current regression coefficients and performed analysis with the partial-borrowing straPP and partial-borrowing Gen-straPP via sampling from the complementary posterior distribution outlined in Supplementary Material (Alt et al., 2023b) Section 1. Using (19), solving for the current data regression coefficients yields, at iteration t of the MCMC sampling scheme,

$$\boldsymbol{\beta}^{(t)} = \left(1 - p_c^{(t)}\right)^{-1/2} (\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1/2} \left\{ \left[\boldsymbol{X}_0'\boldsymbol{W}\left(\beta_{00}^{(t)}, \boldsymbol{\beta}_{01}^{(t)}\right)\boldsymbol{X}_0\right]^{1/2} \boldsymbol{\beta}_{01}^{(t)} - \boldsymbol{c}_0^{(t)} \right\}, \quad (20)$$

where $\boldsymbol{\beta}^{(t)}$ is p-dimensional vector of regression coefficients for the PH component of the mixture cure rate model and where $\boldsymbol{c}_0^{(t)} = \boldsymbol{0}$ for the straPP transformation.

Additionally, we determined the value for a_0 via a grid search with the optimal value being the one that minimized the deviance information criterion (Spiegelhalter et al., 2002) (DIC). The identified values were $a_0 = 0.50$ for the partial-borrowing straPP and partial-borrowing Gen-straPP and $a_0 = 1.00$ for the partial-borrowing power prior. More details on the procedure for identifying a_0 are provided in Section 6. To make a comprehensive comparison, we investigated the performance of all partial-borrowing priors with $a_0 \in \{0.5, 1\}$.

For the simulation, we evaluated current data sample sizes, n_1, in the set $n_1 \in \{100, 125, 150, \ldots, 325\}$. When $n_1 = 100$, the average number of events per generated current data set was approximately 56. We assumed 5 intervals for the piecewise constant baseline hazard for the non-cured population across all simulation studies for consistency and to ensure an acceptable amount of events per interval were observed. When computing estimates of bias, MSE, coverage probability, and interval score, we calculated the posterior means of the regression coefficients based on the entire current data

set using a non-informative reference prior. For the reference prior, we specified a non-informative $N(0, 10^2)$ prior for all current regression coefficients and independent and identically distributed (i.i.d.) gamma priors for each baseline hazard component with shape and inverse scale parameters equal to 0.1. A Uniform$(0, 1)$ prior was also specified on the probability of cure. This led to the following "true" values of the current regression parameters: $\beta_{age} = 0.1105$, $\beta_{trt} = -0.1791$, $\beta_{gend} = -0.1604$, and $\beta_{perf} = 0.1094$, corresponding to age, treatment, gender, and performance status, respectively.

For analysis, we specified a non-informative $N(0, 10^2)$ prior for all regression coefficients, including the historical data model intercept, for the partial-borrowing priors. For all priors considered, we specified a Uniform$(0, 1)$ prior on the probability of cure and i.i.d. gamma priors for each baseline hazard component with shape and inverse scale parameters equal to 0.1. For the partial-borrowing Gen-straPP, we considered a hierarchical prior $\boldsymbol{c}_0 \sim N(\mathbf{0}, \omega_0^2 I_p)$ and $\omega_0 \sim N^+(0, 1)$, where N^+ denotes the half-normal distribution.

To simulate current data, we utilized a bootstrap procedure to generate 10,000 data sets, where we sampled from the current data from the real trial with replacement. For each prior evaluated, we obtained 25,000 samples after a burn-in period of 2,000 iterations from the posterior distribution.

5.2 Simulation Results

Key Findings

Figure 1 displays the average log posterior variance, bias, log MSE, and average interval score for the partial-borrowing straPP, partial-borrowing Gen-straPP, partial-borrowing power prior, and reference prior. In addition, the coverage probabilities associated with 95% credible intervals are shown in Section 3 of the Supplementary Material (Alt et al., 2023b). The first row in Figure 1 (panels (a)–(d)) shows that both the straPP and Gen-straPP consistently achieved lower average log posterior variance compared to the reference prior and power prior, indicating more stable estimates. Although the historical data were discounted more for the straPP and Gen-straPP fitted with $a_0 = 0.5$, the average posterior variance was slightly lower under these two priors compared to the power prior with $a_0 = 1$. This suggests that the straPP and Gen-straPP may be more informative priors than the power prior, despite the difference in the amount of discounting.

Panels (e)–(h) show that none of the priors performed uniformly better than the others in terms of bias. The log mean squared error (MSE) (depicted in the third row, panels (i)–(l)), which is a function of both bias of the posterior mean point estimator and its sampling variability, is lowest for the straPP with $a_0 = 1$ across all regression coefficients. For analyses using the same a_0 value, the log MSE under the Gen-straPP was often between the log MSE under the power prior and the straPP, illustrating the bias-variance trade-off for including the location parameter $\boldsymbol{c}_0$ in the Gen-straPP.

In the final row of Figure 1 (panels (m)–(p)), the average interval scores, as proposed by Gneiting and Raftery (2007), were computed using the width of 95% equal-tailed

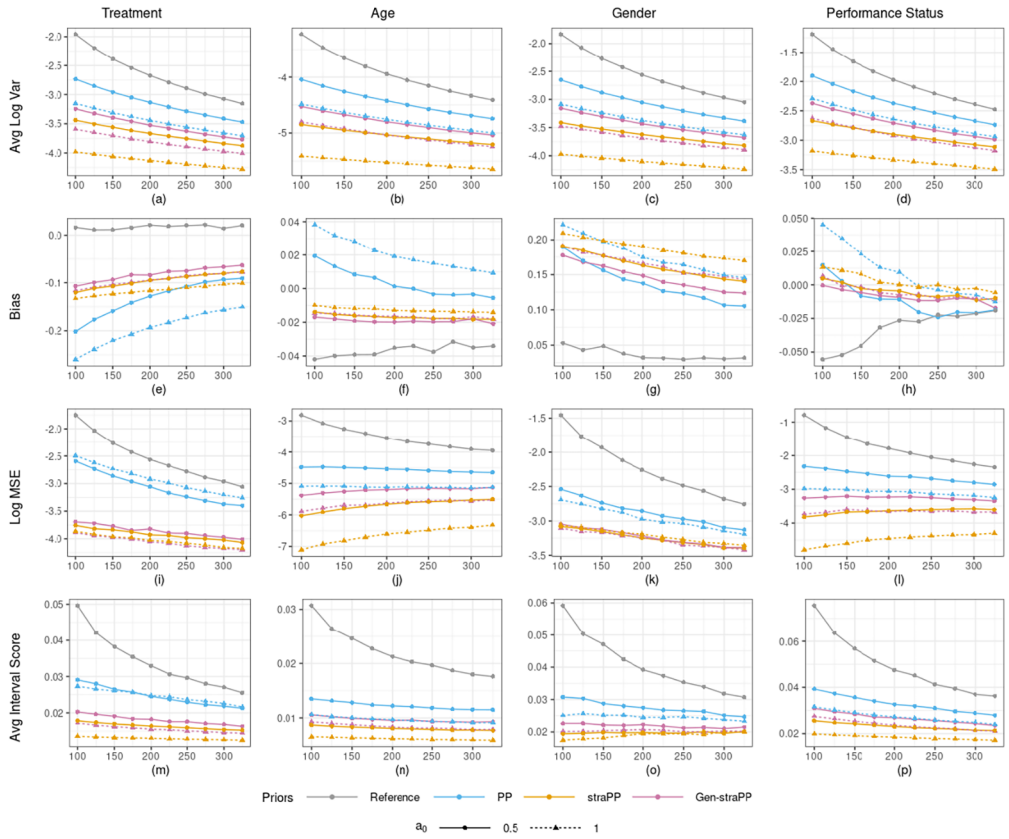

Figure 1: Panels (a)–(p) present the average log posterior variance, bias of the posterior mean, log MSE of the posterior mean, and average interval score for each current data model regression coefficient as a function of the current data bootstrap sample size. Each color represents a distinct prior. For each partial-borrowing prior, we fitted the model with $a_0 = 0.5$ (shown in solid lines) and $a_0 = 1$ (shown in dashed lines). straPP, scale transformed power prior; PP, power prior; Gen-straPP, generalized scale transformed power prior.

credible intervals, with a penalty term proportional to the distance between the estimates (that fall outside these intervals) and the nearest boundary point of the intervals. Lower interval scores indicates better performance. Both the straPP and Gen-straPP maintained lower average interval scores compared to the reference prior and power prior across all sample sizes, indicating better interval estimation. Analyses based on the straPP achieved similar or slightly lower interval scores than those based on the Gen-straPP, regardless of the a_0 value.

Overall, the straPP was the best performer among all the information borrowing priors. Of note, when fitted with the same a_0 value, the straPP exhibited lower posterior

variance, lower mean squared error, and better interval score than the Gen-straPP, with relatively small difference in bias. Thus, although the Gen-straPP is more flexible than the straPP, it does not necessarily translate to better performance in terms of operating characteristics. In practice, it is advisable to fit the model using both the straPP and Gen-straPP to see which prior results in a better fit.

We see from the simulation studies that it is in fact possible to obtain higher than 95% coverage probabilities under an informative prior when the informative prior has high mass around the truth. To see this point, for example, consider an *i.i.d.* setting where $y_i \mid \mu \sim N(\mu, 1)$ and suppose $\mu \sim N(\mu_0, \sigma_0^2)$. Suppose further that the "truth" is $\mu = \mu_0$. As $\sigma_0^2 \to 0$, the prior is converging to a point mass at the truth, in which case we will have 100% coverage probability a posteriori. For small values of $\sigma_0^2 > 0$, we will have slightly lower than 100% coverage, but we will only have 95% coverage for large values of σ_0^2. However, if the prior has mass far from the truth, then the coverage can of course be substantially less than the nominal value.

Investigation of Marginal Prior Densities

To further investigate the information borrowing priors, we examined the marginal prior densities for each regression coefficient in the current data model, as illustrated in Figure 2. To highlight differences among these priors, we set the x-axis limits of each density plot to the 0.5th and 99.5th percentiles of the estimates. By Figure 2, the power prior appears to have heavier tails than the straPP and Gen-straPP across all regression coefficients, which is consistent with the difference in average log posterior variance shown in Figure 1 panels (a)–(d). The prior modes from using the straPP and Gen-straPP are similar for each regression coefficient. In contrast, the prior modes from using the power prior differ from those of straPP and Gen-straPP for the regression coefficients associated with treatment and age but are relatively close for the regression coefficients associated with gender and performance status.

Note that in the straPP paper focusing on generalized linear models by Alt et al. (2023a), we investigated the normalized straPP, comparing its performance and operating characteristics in detail to the straPP and various other priors through extensive simulations and real data analyses. The results demonstrated that the normalized straPP performs very similarly to the straPP. Therefore, we do not carry out a simulation study for the normalized straPP for time-to-event data here for brevity.

Choice of Intervals and Spacing

Finally, we emphasize that choosing the number of intervals and their spacing is a crucial aspect in fitting a piecewise constant hazard model. As discussed in detail in the book by Ibrahim et al. (2001), there are four main strategies for choosing the number of intervals and their spacing. (i) The most common approach is to pick the number of intervals such that one has an equal number of events (failures) in each interval. This strategy provides the best guarantee for stability in parameter estimation and MCMC convergence and constitutes the most efficient construction. In most applications, 5–10

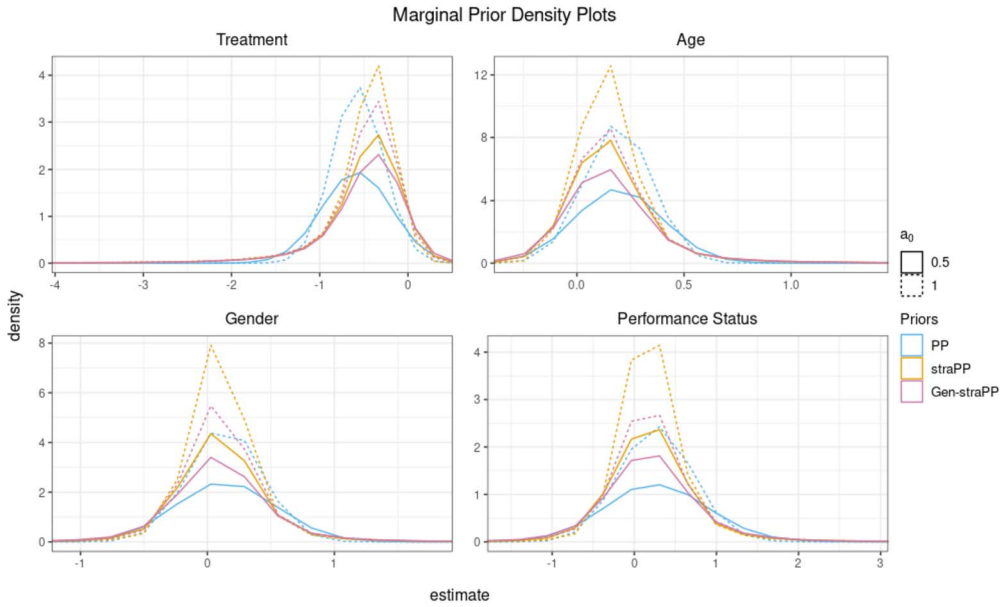

Figure 2: Marginal prior densities for each current data model regression coefficient. For each subplot, the x-axis limits are set to the 0.5th and 99.5th percentiles of the estimates. Each color represents a distinct prior. For each prior, we fitted the model with $a_0 = 0.5$ (shown in solid lines) and $a_0 = 1$ (shown in dashed lines). straPP, scale transformed power prior; PP, power prior; Gen-straPP, generalized scale transformed power prior.

intervals suffice and one can use Bayesian model of fit statistics, such as DIC, to find the optimal number of intervals. (ii) Another strategy is to break up the time axis into quantiles (say 5–10). This approach gives very similar results as (i). (iii) One may also pick 5–10 equally-spaced intervals. This is typically not a good strategy since it may result in intervals with very few failures, resulting in estimation instability. Models with no failures in an interval are not identifiable. (iv) Lastly, one can pick each interval to contain exactly one failure. This strategy offers the most flexible construction of the baseline hazard, but again, it suffers from instability in parameter estimation due to only one event in each interval. Moreover, it may result in an unnecessarily large number of intervals, leading to further instability in estimation, a high computational burden, and an inefficient interval construction which may not fit the data well. Consequently, most authors use strategy (i) or (ii) as we have done here for the reasons stated.

6 Analysis of the ECOG E1684 and E1690 Trials

In this section, we consider the E1684 trial as the historical data and the E1690 trial as the current data. See Section 2 for more details on these studies. For the historical

data, the outcome of interest was an indicator of relapse within two years, which we modeled as a logistic regression model. For the current data, the outcome of interest was taken to be time-to-relapse which was modeled using a mixture cure rate model with a piecewise constant PH regression model for the event times in the non-cured population. For covariates, we consider an indicator for high-dose IFN (0 = observation; 1 = high-dose), standardized age, an indicator for sex (0 = male; 1 = female), and an indicator for ECOG performance status (0 = fully active; 1 = ambulatory). In the historical and current data, 262 and 426 patients, respectively, had non-missing values for these covariates and were thus included in our analyses. For the historical data, there were only 3 subjects who were censored within 2 years, and these subjects were removed from the final analysis so that the binary response variable contains no censored subjects. As in Section 5, we included an intercept term in the historical data model but not in the current data model.

As the historical data were binary in this case, one cannot solve for the historical data model regression parameters in (3) and (5). However, as stated in Section 4.3, the Fisher information matrix for the mixture cure rate model does not involve the regression parameters. Thus, we solved for the current data model regression coefficients using the transformation formula in (20) and performed the analysis with the partial-borrowing straPP and partial-borrowing Gen-straPP via sampling from the complementary posterior distribution, as described in Section 3. We also investigated the partial-borrowing power prior and a non-informative reference prior for comparison purposes.

To determine the interval size, we analyzed the current data using the reference prior for $J \in \{2, \ldots, 10\}$ and computed the DIC values. The DIC for $J = 5$ (992.48) was marginally higher than that for $J = 10$ (991.35), both much lower than the DIC values from using other interval sizes. A sharp increase in DIC was observed after $J = 5$ (Figure 2 in the Supplementary Material (Alt et al., 2023b)). Thus, we selected $J = 5$ to obtain a simpler, more parsimonious model. The time axis was discretized so that an approximately equal number of events was observed in each of the $J = 5$ intervals.

For the partial-borrowing priors, we specified a non-informative $N(0, 10^2)$ prior for all regression coefficients pertaining to covariates of interest and a $N(0, 10^2)$ prior for the historical data model intercept. We specified i.i.d. gamma priors for each baseline hazard component with shape and inverse scale parameters equal to 0.1. Via sensitivity analysis shown in Supplementary Material (Alt et al., 2023b) Section 7, we found that the results are relatively robust to changes in the gamma priors for the baseline hazards. We took $J = 5$ hazard components and chose the time axis partition so that there was an equal number of events observed in each interval. For the partial-borrowing Gen-straPP, we considered a hierarchical prior $\boldsymbol{c}_0 \sim N(\mathbf{0}, \omega_0^2 I_p)$ and $\omega_0 \sim N^+(0, 1)$, where N^+ denotes the half-normal distribution.

For each prior, we obtained 25,000 posterior samples after a burn-in period of 2,000 iterations. No thinning was performed. Posterior sampling was conducted via a Hamiltonian Monte Carlo (HMC) algorithm as implemented in the Stan software (Carpenter et al., 2017).

			Variable							
			Treatment		Age		Sex		Performance	
Priors	a_0	DIC	Mean	95% CI	Mean	95% CI	Mean	95% CI	Mean	95% CI
Reference	0.00	993.31	−0.18	(−0.52, 0.17)	0.11	(−0.09, 0.29)	−0.16	(−0.52, 0.21)	0.11	(−0.39, 0.58)
straPP	0.25	989.77	−0.24	(−0.52, 0.05)	0.11	(−0.04, 0.26)	−0.09	(−0.39, 0.21)	0.12	(−0.33, 0.53)
	0.50	989.40	−0.26	(−0.51, −0.01)	0.11	(−0.02, 0.24)	−0.06	(−0.32, 0.20)	0.11	(−0.28, 0.47)
	0.75	989.93	−0.27	(−0.50, −0.04)	0.11	(−0.01, 0.22)	−0.04	(−0.27, 0.20)	0.11	(−0.25, 0.44)
	1.00	990.40	−0.28	(−0.49, −0.07)	0.11	(−0.00, 0.21)	−0.02	(−0.24, 0.20)	0.11	(−0.22, 0.42)
Gen-straPP	0.25	990.12	−0.23	(−0.52, 0.06)	0.11	(−0.05, 0.26)	−0.10	(−0.41, 0.21)	0.12	(−0.33, 0.54)
	0.50	988.77	−0.25	(−0.50, 0.01)	0.11	(−0.03, 0.24)	−0.07	(−0.34, 0.20)	0.11	(−0.29, 0.49)
	0.75	991.23	−0.26	(−0.50, −0.01)	0.11	(−0.02, 0.23)	−0.05	(−0.31, 0.20)	0.11	(−0.26, 0.46)
	1.00	990.68	−0.26	(−0.49, −0.03)	0.11	(−0.02, 0.23)	−0.04	(−0.29, 0.20)	0.11	(−0.25, 0.44)
PP	0.25	987.25	−0.23	(−0.54, 0.10)	0.12	(−0.06, 0.28)	−0.12	(−0.46, 0.22)	0.10	(−0.38, 0.55)
	0.50	983.56	−0.26	(−0.56, 0.04)	0.12	(−0.04, 0.28)	−0.10	(−0.41, 0.23)	0.10	(−0.35, 0.52)
	0.75	981.30	−0.29	(−0.58, −0.01)	0.13	(−0.03, 0.27)	−0.08	(−0.37, 0.22)	0.10	(−0.33, 0.51)
	1.00	980.03	−0.31	(−0.58, −0.03)	0.13	(−0.01, 0.27)	−0.06	(−0.34, 0.23)	0.10	(−0.31, 0.49)

Gen-straPP, generalized scale transformed power prior; straPP, scaled transformed power prior; PP, power prior.

Table 2: Posterior Summaries for the E1684 and E1690 data sets.

6.1 Analysis Results

Table 2 presents the DIC, value for a_0, posterior mean, and 95% equal-tailed credible interval (CI) for each regression coefficient and for each prior distribution. Posterior estimates for the cure fraction and baseline hazards are provided in Table 1 of the Supplementary Material (Alt et al., 2023b). Note that the row corresponding to the lowest DIC for each prior family is highlighted in gray (i.e., the value of a_0 that minimizes the DIC in the grid search).

The power prior with $a_0 = 1.00$ minimizes the DIC overall, suggesting that a scale transformation is not necessary and that the data sets are compatible. Of course, this is illogical, since the historical data outcome is binary and the current data outcome is time-to-event. Conversely, for the straPP and Gen-straPP families, the DIC is minimized by taking $a_0 = 0.50$, resulting in more discounting of the historical data compared to the power prior. As mentioned in Section 5, a larger amount of discounting for the straPP and Gen-straPP does not necessarily translate to a less informative prior. Indeed, the 95% credible intervals under the power prior with $a_0 = 1.00$ are wider than those for the straPP and Gen-straPP with $a_0 = 0.50$.

In this example, the posterior means and 95% credible intervals using the optimal straPP and Gen-straPP are virtually identical, with the optimal Gen-straPP offering a lower DIC. This suggests that the bias introduced by the straPP transformation is sufficiently large compared to the increased complexity required by the Gen-straPP. However, the posterior means and 95% credible intervals are virtually identical between these two priors. Moreover, the posterior mean for the treatment effect under the power prior was more sensitive to the choice of a_0 than compared to the straPP or Gen-straPP, which may be attributable to the fact that the historical data and current data outcomes are of different types.

7 Analysis of the ECOG E2696 and E1694 Trials

For the analysis presented in this section, we took continuous log(IgM) measured 28 days after vaccination in the E2696 trial as the outcome for the historical data and time-to-relapse from the E1694 trial as the outcome for the current data. See Section 2 for more details on these studies and recall that the data sets were chosen to reflect a scenario where the straPP assumption is violated (i.e., there is no effect of treatment in the historical data and an apparent treatment effect in the current data). Hence, our objective is to explore robustness of the partial-borrowing straPP and Gen-straPP in settings where the partial-borrowing straPP assumption is violated.

For analysis, we assumed that the outcomes from the historical data were independently distributed according to a linear regression model. For the current data, we compared two different models to illustrate the proposed method: a PH regression model with piecewise constant baseline hazard and a mixture cure rate model with a piecewise constant PH regression model for the event times in the non-cured population. The covariates of interest for each regression model were standardized age, an indicator for treatment (0 = GMK; 1 = IFN and GMK for E2696, and IFN for E1694), an indicator for gender (0 = male; 1 = female), and an indicator for ECOG performance status (0 = fully active; 1 = ambulatory). We included an intercept term in the historical data model but not in the two current data models.

We analyzed the E2696 and E1694 data sets using the partial-borrowing straPP, partial-borrowing Gen-straPP, partial-borrowing power prior, and a non-informative (reference) prior. For the reference prior, we specified a $N(0, 10^2)$ prior for all current regression coefficients, including the current data model intercept (if it is in the model). For the information borrowing priors, we specified a $N(0, 10^2)$ prior for all regression coefficients, including the historical data model intercept and the current data model intercept (if it is in the model). We specified a gamma prior for the historical data model dispersion parameter with shape and inverse scale parameters equal to 0.1. For all models considered, we specified i.i.d. gamma priors for all hazard components with shape and inverse scale parameters equal to 0.1. For the partial-borrowing Gen-straPP, we considered the hierarchical prior $\boldsymbol{c}_0 \sim N(\boldsymbol{0}, \omega_0^2 \boldsymbol{I}_p)$ and $\omega_0 \sim N^+(0, 1)$.

In this case, the Fisher information matrix for the historical data is simply

$$I_{\boldsymbol{\beta}_{01}\boldsymbol{\beta}_{01}}(\beta_{00}, \boldsymbol{\beta}_{01}, \sigma_0^2 | \boldsymbol{X}_0) = \sigma_0^{-2} \boldsymbol{X}_0' \boldsymbol{X}_0.$$

Thus, using the Gen-straPP solutions (16) and (19), we may solve for the current data regression coefficients at iteration t of the MCMC sampling scheme via

$$\boldsymbol{\beta}^{(t)} = \left(1 - p_c^{(t)}\right)^{-1/2} \left(\frac{\boldsymbol{\beta}_{01}^{(t)}}{\sigma_0^{(t)}} - \left[\boldsymbol{X}_0' \boldsymbol{X}_0\right]^{-1/2} \boldsymbol{c}_0^{(t)} \right),$$

where we may substitute $p_c^{(t)} = 0$ for the PH model and $\boldsymbol{c}_0^{(t)} = \boldsymbol{0}$ when using the straPP. We use the complementary sampling approach described in Supplementary Material (Alt et al., 2023b) Section 1.

To determine the interval size, we analyzed the current data using the PH model with piecewise constant baseline hazard and the mixture cure rate model under reference priors for $J \in \{2, \ldots, 10\}$, computing the DIC values. For the PH regression model, the DIC for $J = 5$ (771.98) was marginally higher than that for $J = 8$ (771.42), both much lower than the DIC values from using other interval sizes. A sharp increase in DIC was observed after $J = 5$. A similar trend in DIC was observed for the mixture cure rate model (Figure 3 in the Supplementary Material (Alt et al., 2023b)). Thus, we selected $J = 5$ to obtain a simpler, more parsimonious model. The time axis was then selected so that an approximately equal number of events was observed in each interval.

The sample size of the E2696 data set is more than eight times smaller than that of the E1694 data set ($n_0 = 98$ vs $n_1 = 873$). Due to this difference in total information content, the historical data would contribute comparatively little information regardless of the prior used if no other adjustment was made. Thus, for the sake of illustration in this section, we restricted the current data to only include patients without nodal metastases, which led to a current data sample size of 200. We note that the E2696 data set only contains patients without nodal metastases, thus consistent with the restricted population in the current data. Several patients in the E1694 data set had failure times of zero, which we coded as 0.50 (i.e., half of a day) for analysis.

7.1 Analysis Results

For analyses using each information borrowing prior, we computed the DIC for each $a_0 \in \{0.25, 0.50, 0.75, 1.00\}$. Table 3 presents the DIC, value for a_0, posterior mean, and 95% equal-tailed credible interval (CI) for each parameter and for each model and prior distribution combination. The posterior estimates were calculated based on 25,000 posterior samples obtained via HMC after a burn-in period of 2,000 iterations. No thinning was performed. Note that the row corresponding to the lowest DIC for each model and prior combination is highlighted in gray. Posterior estimates for the cure fraction and baseline hazards are presented in Table 2 of the Supplementary Material (Alt et al., 2023b).

For the mixture cure rate model, one can see that the power prior with $a_0 = 0.25$ results in the lowest DIC among the families of priors considered. Focusing on the optimal prior within each family (optimal over the a_0 values considered), the optimal straPP prior results in the highest DIC, with the comparatively poor performance likely attributable to the fact that the straPP assumption is violated for these data. The fact that the DIC for the Gen-straPP is somewhat lower than that for the straPP provides support for this conjecture.

For the PH model, the optimal straPP and Gen-straPP priors result in essentially identical values for the DIC, which are lower than those for the reference prior and the power prior. This may indicate that, for the PH model, a scale transformation of the power prior was necessary, and the complexity introduced by the Gen-straPP transformation did not result in sufficiently better model fit compared to the straPP transformation.

				Variable							
				Treatment		Age		Sex		Performance	
Model	Prior	a_0	DIC	Mean	95% CI	Mean	95% CI	Mean	95% CI	Mean	95% CI
Mixture cure	Reference	0.00	731.29	−0.49	(−1.08, 0.12)	0.03	(−0.35, 0.35)	−0.15	(−0.78, 0.50)	−0.32	(−1.26, 0.59)
	straPP	0.25	733.12	−0.37	(−0.87, 0.17)	−0.06	(−0.41, 0.22)	−0.17	(−0.73, 0.37)	−0.32	(−1.03, 0.39)
		0.50	734.02	−0.29	(−0.77, 0.21)	−0.15	(−0.47, 0.12)	−0.21	(−0.74, 0.28)	−0.32	(−0.96, 0.31)
		0.75	735.16	−0.23	(−0.69, 0.24)	−0.21	(−0.50, 0.05)	−0.25	(−0.75, 0.21)	−0.34	(−0.91, 0.22)
		1.00	736.26	−0.19	(−0.62, 0.24)	−0.25	(−0.52, −0.00)	−0.27	(−0.74, 0.16)	−0.35	(−0.88, 0.18)
	Gen-straPP	0.25	732.62	−0.38	(−0.90, 0.15)	−0.05	(−0.39, 0.23)	−0.17	(−0.72, 0.37)	−0.31	(−1.05, 0.40)
		0.50	733.52	−0.32	(−0.82, 0.19)	−0.11	(−0.44, 0.17)	−0.20	(−0.71, 0.32)	−0.31	(−0.99, 0.40)
		0.75	733.57	−0.29	(−0.78, 0.21)	−0.15	(−0.47, 0.15)	−0.22	(−0.73, 0.28)	−0.33	(−0.95, 0.28)
		1.00	733.98	−0.27	(−0.75, 0.23)	−0.17	(−0.49, 0.13)	−0.23	(−0.73, 0.26)	−0.33	(−0.93, 0.28)
	PP	0.25	730.51	−0.41	(−0.97, 0.18)	−0.06	(−0.44, 0.25)	−0.19	(−0.79, 0.42)	−0.37	(−1.18, 0.44)
		0.50	731.11	−0.34	(−0.88, 0.23)	−0.15	(−0.50, 0.16)	−0.23	(−0.81, 0.34)	−0.39	(−1.12, 0.33)
		0.75	732.13	−0.27	(−0.79, 0.26)	−0.22	(−0.54, 0.09)	−0.27	(−0.83, 0.26)	−0.41	(−1.08, 0.26)
		1.00	733.45	−0.23	(−0.72, 0.27)	−0.28	(−0.56, 0.02)	−0.31	(−0.82, 0.20)	−0.41	(−1.03, 0.20)
PH	Reference	0.00	771.95	−0.48	(−0.95, −0.02)	0.11	(−0.13, 0.35)	−0.18	(−0.69, 0.31)	−0.37	(−1.19, 0.34)
	straPP	0.25	770.59	−0.36	(−0.77, 0.04)	0.01	(−0.19, 0.22)	−0.16	(−0.59, 0.26)	−0.30	(−0.88, 0.25)
		0.50	770.83	−0.29	(−0.65, 0.07)	−0.05	(−0.23, 0.14)	−0.17	(−0.55, 0.20)	−0.29	(−0.78, 0.19)
		0.75	771.29	−0.24	(−0.58, 0.09)	−0.09	(−0.26, 0.08)	−0.18	(−0.53, 0.17)	−0.29	(−0.72, 0.13)
		1.00	771.98	−0.21	(−0.52, 0.10)	−0.12	(−0.28, 0.04)	−0.19	(−0.50, 0.13)	−0.29	(−0.68, 0.09)
	Gen-straPP	0.25	770.47	−0.37	(−0.79, 0.02)	0.02	(−0.19, 0.24)	−0.16	(−0.59, 0.26)	−0.31	(−0.91, 0.27)
		0.50	770.59	−0.33	(−0.72, 0.06)	−0.02	(−0.21, 0.19)	−0.16	(−0.56, 0.23)	−0.29	(−0.85, 0.22)
		0.75	771.00	−0.30	(−0.70, 0.06)	−0.04	(−0.24, 0.18)	−0.16	(−0.54, 0.21)	−0.29	(−0.80, 0.19)
		1.00	771.01	−0.29	(−0.68, 0.07)	−0.05	(−0.24, 0.17)	−0.16	(−0.53, 0.20)	−0.30	(−0.79, 0.18)
	PP	0.25	771.47	−0.43	(−0.88, 0.00)	0.05	(−0.18, 0.28)	−0.19	(−0.66, 0.28)	−0.39	(−1.12, 0.26)
		0.50	771.20	−0.40	(−0.83, 0.03)	0.00	(−0.22, 0.21)	−0.19	(−0.64, 0.24)	−0.41	(−1.07, 0.20)
		0.75	771.65	−0.36	(−0.78, 0.04)	−0.04	(−0.25, 0.17)	−0.20	(−0.63, 0.22)	−0.42	(−1.02, 0.14)
		1.00	772.05	−0.34	(−0.73, 0.05)	−0.08	(−0.28, 0.12)	−0.21	(−0.62, 0.19)	−0.43	(−0.99, 0.10)

Gen-straPP, generalized scale transformed power prior; straPP, scaled transformed power prior; PP, power prior; PH, proportional hazards.

Table 3: Posterior Summaries for the E2696 and E1694 data sets.

For both the PH and the mixture cure rate models, the posterior density of $\boldsymbol{c}_0$ from using the Gen-straPP with $a_0 = 1$ (Figure 4 in the Supplementary Material (Alt et al., 2023b)) shows some support for the Gen-straPP transformation. The value $a_0 = 1$ was selected because the effects of $\boldsymbol{c}_0$ play a bigger role in the estimation when more historical data are borrowed (i.e., when the a_0 value is larger).

8 Discussion

In this paper, we have developed the partial-borrowing scale transformed power prior for applications with time-to-event data. Applications with time-to-event data require the use of a partial-borrowing version of the straPP as time-to-event data involve nuisance parameters, such as the parameters in a piecewise constant hazard, that have no analog in non-time-to-event models. Though the simulations and case studies discussed in the paper involve historical data following generalized linear models and current data following time-to-event models, it is straightforward to extend the methods presented to settings in which the opposite is true. The case studies presented in the paper consider commonly used time-to-event models and collectively suggest that the partial-borrowing straPP and the partial-borrowing Gen-straPP may effectively account for the scale differences in outcomes in these complex settings.

The straPP and Gen-straPP are designed for settings where historical data are available but where the outcomes assessed in the historical data are not the same as in the current data (or perhaps even when the outcomes measured in the current data

are measured in the historical data as well, but are much less mature in the historical data). There are several contexts in which this could occur. First, in early phase trials, surrogate outcomes that can be measured quickly are often used (e.g., response) with later phase trials frequently using outcomes that more clearly indicate clinical benefit (e.g., overall survival). In this setting, the straPP and Gen-straPP could be used to translate effects on binary response outcomes to time-to-event outcomes. Thus, although our examples involved surrogate endpoints for historical and current data sets, our approach is broadly applicable. In other settings, the types of outcomes measured (e.g., patient reported outcomes) evolve over time and this could provide a context where the straPP and Gen-straPP could be useful.

Another useful application of our proposed method could be in postmarketing safety studies. Since adverse event (AE) data are typically reported in clinical publications as binary, one could use a straPP transformation to elicit an informative prior for the hazard ratio based on summary statistics for the binary endpoint.

Regardless of context, it is not in dispute that application of the straPP and Gen-straPP will be subject to more scrutiny than, for example, applications of a simple power prior for contexts where the power prior is appropriate (exchangeable populations with the same outcomes measured). Thus, careful planned sensitivity analyses will be important for users of the straPP and Gen-straPP. The straPP and Gen-straPP are not designed to replace commonly used priors such as the power prior in contexts where commonly used priors are appropriate.

For the analyses with the Gen-straPP, we considered a hierarchical prior $\boldsymbol{c}_0 \sim N(\mathbf{0}, \omega_0^2 I_p)$ and $\omega_0 \sim N^+(0,1)$. As an alternative strategy, one may consider a tipping point analysis whereby an initial value of ω_0 is increased until such time as the evidence in favor of some claim ceases to be substantial. Such an approach can facilitate a discussion of whether the value of ω_0 that corresponds to the tipping point is sufficiently small to be of concern. In fact, a similar strategy can be used with the straPP for the a_0 parameter value. This type of analysis may be particularly advantageous as a prespecified sensitivity analysis in cases where it is desired to fully specify a single formulation of the straPP or Gen-straPP as the primary prior for analysis (e.g., as one would need to do for a clinical trial protocol).

Note that the Gen-straPP transformation in (5) allows for the partial-borrowing straPP assumption to be violated in a specific way. Accordingly, the Gen-straPP will provide added robustness compared to the straPP when the violation is well captured by the additional additive term (i.e., $\boldsymbol{c}_0$). If the straPP transformation is violated in other ways, it is not clear the degree to which the Gen-straPP can counteract such violations. Quality of performance will likely be case-specific. Future work will compare the Gen-straPP robustification strategy to other strategies, such as constructing a robust mixture prior where one component of the mixture is a straPP and another component is a non-informative prior.

For future work, the authors plan to develop Bayesian sample size determination methods using the straPP for clinical trial applications with time-to-event outcomes, including trials with group sequential designs. While this work primarily focuses on

estimation quality (e.g., bias, MSE) as a means to understand the potential value of the straPP and Gen-straPP, applications to clinical trial design will necessarily need to address how one can use the straPP and Gen-straPP while balancing important operating characteristics such as type I error control and power which are of paramount importance in that setting. Finally, it is worth mentioning that, in situations where the current data measures the outcome in the historical data (but the primary endpoint is different), a joint modeling approach could potentially be used. This is an area for future research.

Acknowledgments

The first two authors contributed equally to the production of this article.

Supplementary Material

Supplementary Material to "The Scale Transformed Power Prior for Time-To-Event Data" (DOI: 10.1214/24-BA1504SUPP; .pdf). The Supplementary Material presents the details of the complementary sampling approach, the derivation of the Fisher information matrix for applying the straPP and Gen-straPP to the promotion time cure rate model, supplementary simulation and data analysis results, and an additional simulation study. Software and data used for the simulations and data analyses in the development of this paper are publicly available at `https://github.com/ethan-alt/strapp_survival`.

References

Alt, E. M., Nifong, B., Chen, X., Psioda, M. A., and Ibrahim, J. G. (2023a). "The scale transformed power prior for use with historical data from a different outcome model." *Statistics in Medicine*, 42(1): 1–14. MR4527710. doi: `https://doi.org/10.1002/sim.9598`. 2, 3, 4, 5, 6, 7, 11, 14, 18

Alt, E. M., Nifong, B., Chen, X., Psioda, M. A., and Ibrahim, J. G. (2023b). "Supplement to "The scale transformed power prior for time-to-event data"." *Bayesian Analysis*. doi: `https://doi.org/10.1214/24-BA1504SUPP`. 8, 11, 13, 14, 15, 16, 20, 21, 22, 23, 24

Banbeta, A., Lesaffre, E., and van Rosmalen, J. (2022). "The power prior with multiple historical controls for the linear regression model." *Pharmaceutical Statistics*, 21(2): 418–438. 2

Banbeta, A., van Rosmalen, J., Dejardin, D., and Lesaffre, E. (2019). "Modified power prior with multiple historical trials for binary endpoints." *Statistics in Medicine*, 38(7): 1147–1169. MR3920603. doi: `https://doi.org/10.1002/sim.8019`. 2

Buunen, M., Veldkamp, R., Hop, W. C., Kuhry, E., Jeekel, J., et al. (2009). "Survival after laparoscopic surgery versus open surgery for colon cancer: long-term outcome of a randomised clinical trial." *The Lancet Oncology*, 10(1): 44–52. 5

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). "Stan: A probabilistic programming language." *Journal of Statistical Software*, 76(1). 20

Chen, M.-H., Harrington, D. P., and Ibrahim, J. G. (2002). "Bayesian cure rate models for malignant melanoma: a case-study of Eastern Cooperative Oncology Group trial E1690." *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 51(2): 135–150. MR1900162. doi: https://doi.org/10.1111/1467-9876.00259. 2

Chen, M.-H. and Ibrahim, J. G. (2006). "The relationship between the power prior and hierarchical models." *Bayesian Analysis*, 1(3): 551–574. MR2221288. doi: https://doi.org/10.1214/06-BA118. 2

Chen, M.-H., Ibrahim, J. G., Lam, P., Yu, A., and Zhang, Y. (2011). "Bayesian design of noninferiority trials for medical devices using historical data." *Biometrics*, 67(3): 1163–1170. MR2829252. doi: https://doi.org/10.1111/j.1541-0420.2011.01561.x. 2

Chen, M.-H., Ibrahim, J. G., and Sinha, D. (1999a). "A new Bayesian model for survival data with a surviving fraction." *Journal of the American Statistical Association*, 94(447): 909–919. MR1723307. doi: https://doi.org/10.2307/2670006. 2

Chen, M.-H., Ibrahim, J. G., and Yiannoutsos, C. (1999b). "Prior elicitation, variable selection and Bayesian computation for logistic regression models." *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 61(1): 223–242. MR1664057. doi: https://doi.org/10.1111/1467-9868.00173. 2

Chen, M.-H., Manatunga, A. K., and Williams, C. J. (1998). "Heritability estimates from human twin data by incorporating historical prior information." *Biometrics*, 54(4): 1348–1362. 2

De Santis, F. (2006). "Power priors and their use in clinical trials." *The American Statistician*, 60(2): 122–129. MR2224209. doi: https://doi.org/10.1198/000313006X109269. 2

Duan, Y., Ye, K., and Smith, E. P. (2006). "Evaluating water quality using power priors to incorporate historical information." *Environmetrics*, 17(1): 95–106. MR2222036. doi: https://doi.org/10.1002/env.752. 2

Fouskakis, D., Ntzoufras, I., and Draper, D. (2015). "Power-expected-posterior priors for variable selection in Gaussian linear models." *Bayesian Analysis*, 10(1): 75–107. MR3420898. doi: https://doi.org/10.1214/14-BA887. 2

Gelfand, A. E. and Mallick, B. K. (1995). "Bayesian analysis of proportional hazards models built from monotone functions." *Biometrics*, 51(3): 843–852. 10

Gneiting, T. and Raftery, A. E. (2007). "Strictly proper scoring rules, prediction, and estimation." *Journal of the American Statistical Association*, 102(477): 359–378. MR2345548. doi: https://doi.org/10.1198/016214506000001437. 14, 16

Gravestock, I., Held, L., and Consortium, O. b. o. t. C.-N. (2017). "Adaptive power pri-

ors with empirical Bayes for clinical trials." *Pharmaceutical Statistics*, 16(5): 349–360. 2

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(3): 1047–1056. MR2829239. doi: https://doi.org/10.1111/j.1541-0420.2011.01564.x. 2

Ibrahim, J. G. and Chen, M.-H. (1998). "Prior distributions and Bayesian computation for proportional hazards models." *Sankhyā: The Indian Journal of Statistics, Series B (1960–2002)*, 60(1): 48–64. MR1717075. 2

Ibrahim, J. G. and Chen, M.-H. (2000). "Power prior distributions for regression models." *Statistical Science*, 15(1): 46–60. MR1842236. doi: https://doi.org/10.1214/ss/1009212673. 1, 2

Ibrahim, J. G., Chen, M.-H., and Chu, H. (2012a). "Bayesian methods in clinical trials: a Bayesian analysis of ECOG trials E1684 and E1690." *BMC Medical Research Methodology*, 12: 183. 2

Ibrahim, J. G., Chen, M.-H., Gwon, Y., and Chen, F. (2015). "The power prior: theory and applications." *Statistics in Medicine*, 34(28): 3724–3749. MR3422144. doi: https://doi.org/10.1002/sim.6728. 2

Ibrahim, J. G., Chen, M.-H., and MacEachern, S. N. (1999). "Bayesian variable selection for proportional hazards models." *Canadian Journal of Statistics*, 27(4): 701–717. MR1767142. doi: https://doi.org/10.2307/3316126. 2

Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2001). *Bayesian Survival Analysis*, chapter 3, 47–86. New York: Springer. MR1876598. doi: https://doi.org/10.1007/978-1-4757-3447-8. 18

Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2003). "On optimality properties of the power prior." *Journal of the American Statistical Association*, 98(461): 204–213. MR1965686. doi: https://doi.org/10.1198/016214503388619229. 2

Ibrahim, J. G., Chen, M.-H., Xia, H. A., and Liu, T. (2012b). "Bayesian meta-experimental design: Evaluating cardiovascular risk in new antidiabetic therapies to treat type 2 diabetes." *Biometrics*, 68(2): 578–586. MR2959625. doi: https://doi.org/10.1111/j.1541-0420.2011.01679.x. 2, 7

Joshua, A. M., Monzon, J. G., Mihalcioiu, C., Hogg, D., Smylie, M., and Cheng, T. (2015). "A phase 2 study of tremelimumab in patients with advanced uveal melanoma." *Melanoma Research*, 24(4): 342–347. 5

Kim, K. B., Prieto, V., Joseph, R. W., Diwan, A. H., Gallick, G. E., et al. (2012). "A randomized phase II study of cilengitide (EMD 121974) in patients with metastatic melanoma." *Melanoma Research*, 22(4): 294–301. 5

Kirkwood, J. M., Ibrahim, J., Lawson, D. H., Atkins, M. B., Agarwala, S. S., et al. (2001a). "High-dose interferon alfa-2b does not diminish antibody response to GM2 vaccination in patients with resected melanoma: Results of the multicenter eastern

cooperative oncology group phase II trial E2696." *Journal of Clinical Oncology*, 19(5): 1430–1436. 5

Kirkwood, J. M., Ibrahim, J. G., Sondak, V. K., Richards, J., Flaherty, L. E., et al. (2000). "High- and low-dose interferon alfa-2b in high-risk melanoma: First analysis of intergroup trial E1690/S9111/C9190." *Journal of Clinical Oncology*, 18(12): 2444–2458. 4

Kirkwood, J. M., Ibrahim, J. G., Sosman, J. A., Sondak, V. K., Agarwala, S. S., et al. (2001b). "High-dose interferon alfa-2b signifficantly prolongs relapse-free and overall survival compared with the GM2-KLH/QS-21 vaccine in patients with resected stage IIB-III melanoma: Results of intergroup trial E1694/S9512/C509801." *Journal of Oncology*, 19(9): 2370–2380. 5

Kirkwood, J. M., Strawderman, M. H., Ernstoff, M. S., Smith, T. J., Borden, E. C., and Blum, R. H. (1996). "Interferon alfa-2b adjuvant therapy of high-risk resected cutaneous melanoma: The Eastern Cooperative Oncology Group Trial EST 1684." *Journal of Clinical Oncology*, 14(1): 7–17. 5

Lesaffre, E., Qi, H., Banbeta, A., and van Rosmalen, J. (2024). "A review of dynamic borrowing methods with applications in pharmaceutical research." *Brazilian Journal of Probability and Statistics*, 38(1): 1–31. MR4718423. doi: https://doi.org/10.1214/24-bjps598. 2

Lewis, C. J., Sarkar, S., Zhu, J., and Carlin, B. P. (2019). "Borrowing from historical control data in cancer drug development: A cautionary tale and practical guidelines." *Statistics in biopharmaceutical research*, 11(1): 67–78. 2

Lindley, D. V. and Smith, A. F. M. (1972). "Bayes estimates for the linear model." *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(1): 1–18. MR0415861. 2

Lu, N., Wang, C., Chen, W.-C., Li, H., Song, C., Tiwari, R., Xu, Y., and Yue, L. Q. (2022). "Propensity score-integrated power prior approach for augmenting the control arm of a randomized controlled trial by incorporating multiple external data sources." *Journal of Biopharmaceutical Statistics*, 32(1): 158–169. 2

Neelon, B. and O'Malley, A. J. (2010). "Bayesian analysis using power priors with application to pediatric quality of care." *Journal of Biometrics & Biostatistics*, 01(01). 2

Neuenschwander, B., Branson, M., and Spiegelhalter, D. J. (2009). "A note on the power prior." *Statistics in Medicine*, 28(28): 3562–3566. MR2744383. doi: https://doi.org/10.1002/sim.3722. 2

Nilsson, S., Franzén, L., Parker, C., Tyrrell, C., Blom, R., et al. (2013). "Two-year survival follow-up of the randomized, double-blind, placebo-controlled phase II study of radium-223 chloride in patients with castration-resistant prostate cancer and bone metastases." *Clinical Genitourinary Cancer*, 11(1): 20–26. 5

Pawel, S., Aust, F., Held, L., and Wagenmakers, E.-J. (2023). "Normalized power priors

always discount historical data." *Stat*, 12(1): e591. MR4641212. doi: https://doi.org/10.1002/sta4.591. 2

Pawel, S., Aust, F., Held, L., and Wagenmakers, E.-J. (2024). "Power priors for replication studies." *TEST*, 33(1): 127–154. MR4727784. doi: https://doi.org/10.1007/s11749-023-00888-5. 2

Psioda, M. A., Hu, K., Zhang, Y., Pan, J., and Ibrahim, J. G. (2020a). "Bayesian design of biosimilars clinical programs involving multiple therapeutic indications." *Biometrics*, 76: 630–642. MR4125283. doi: https://doi.org/10.1111/biom.13163. 3

Psioda, M. A. and Ibrahim, J. G. (2018). "Bayesian design of a survival trial with a cured fraction using historical data." *Statistics in Medicine*, 37(26): 3814–3831. MR3869156. doi: https://doi.org/10.1002/sim.7846. 3

Psioda, M. A. and Ibrahim, J. G. (2019). "Bayesian clinical trial design using historical data that inform the treatment effect." *Biostatistics*, 20(3): 400–415. MR3973117. doi: https://doi.org/10.1093/biostatistics/kxy009. 3

Psioda, M. A., Xia, H. A., Jiang, X., Xu, J., and Ibrahim, J. G. (2020b). "Bayesian adaptive design for concurrent trials involving biologically related diseases." *Biostatistics*. kxab008. MR4454967. doi: https://doi.org/10.1093/biostatistics/kxab008. 3

Rietbergen, C., Klugkist, I., Janssen, K. J. M., Moons, K. G. M., and Hoijtink, H. J. A. (2011). "Incorporation of historical data in the analysis of randomized therapeutic trials." *Contemporary Clinical Trials*, 32(6): 848–855. 2

Roychoudhury, S. and Neuenschwander, B. (2020). "Bayesian leveraging of historical control data for a clinical trial with time-to-event endpoint." *Statistics in Medicine*, 39(7): 984–995. MR4075841. doi: https://doi.org/10.1002/sim.8456. 3

Rühl, J., Beyersmann, J., and Friedrich, S. (2023). "General independent censoring in event-driven trials with staggered entry." *Biometrics*, 79(3): 1737–1748. MR4643952. doi: https://doi.org/10.1111/biom.13710. 12

Schmidli, H., Gsteiger, S., Roychoudhury, S., O'Hagan, A., Spiegelhalter, D., and Neuenschwander, B. (2014). "Robust meta-analytic-predictive priors in clinical trials with historical control information." *Biometrics*, 70(4): 1023–1032. MR3295763. doi: https://doi.org/10.1111/biom.12242. 2

Shen, Y., Carvalho, L. M., Psioda, M. A., and Ibrahim, J. G. (2024). "Optimal priors for the discounting parameter of the normalized power prior." *Statistica Sinica; in press*. MR4330550. doi: https://doi.org/10.1002/sim.9124. 2

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). "Bayesian measures of model complexity and fit." *Journal of Royal Statistical Society: Series B (Statistical Methodology)*, 64(4): 583–639. MR1979380. doi: https://doi.org/10.1111/1467-9868.00353. 15

Su, L., Chen, X., Zhang, J., and Yan, F. (2022). "Comparative study of bayesian

information borrowing methods in oncology clinical trials." *JCO Precision Oncology*, 6: e2100394. 2

Suder, P. M., Xu, J., and Dunson, D. B. (2023). "Bayesian transfer learning." *arXiv preprint at* arXiv:2312.13484. 2

Viele, K., Berry, S., Neuenschwander, B., Amzal, B., Chen, F., Enas, N., Hobbs, B., Ibrahim, J. G., Kinnersley, N., Lindborg, S., Micallef, S., Roychoudhury, S., and Thompson, L. (2014). "Use of historical control data for assessing treatment effects in clinical trials." *Pharmaceutical Statistics*, 13(1): 41–54. 2

Wang, C., Li, H., Chen, W.-C., Lu, N., Tiwari, R., Xu, Y., and Yue, L. Q. (2019). "Propensity score-integrated power prior approach for incorporating real-world evidence in single-arm clinical studies." *Journal of Biopharmaceutical Statistics*, 29(5): 731–748. 2

Wang, J., Zhang, H., and Tiwari, R. (2024). "A propensity-score integrated approach to Bayesian dynamic power prior borrowing." *Statistics in Biopharmaceutical Research*, 16(2): 182–191. 2

Wang, X., Suttner, L., Jemielita, T., and Li, X. (2022). "Propensity score-integrated Bayesian prior approaches for augmented control designs: A simulation study." *Journal of Biopharmaceutical Statistics*, 32(1): 170–190. 2

Warren, K., Bent, R., Wolters, P. L., Prager, A., Hanson, R., et al. (2012). "A phase 2 study of pegylated interferon α-2b (PEG-Intron(®)) in children with diffuse intrinsic pontine glioma." *Cancer*, 118(14): 3607–3613. 5