BayesFBHborrow: An R Package for Bayesian Borrowing for Time-to-Event Data from a Flexible Baseline Hazard Function *

Darren A. V. Scott [1] Sophia Axillus [2] Alex Lewin [3]
Grant Izmirlian [1,4]

[1] AstraZeneca R&D
[2] University of Gothenburg / Chalmers University of Technology
[3] London School of Hygiene and Tropical Medicine
[4] AstraZeneca R&D, Gaithersburg

Abstract

Statistical methods that leverage external trial information to help accelerate drug development are becoming increasingly popular. Bayesian methods facilitate dynamic borrowing, where the similarity of the response guides how much information is used. We have proposed a semiparametric Bayesian borrowing model for time-to-event data, with smoothing priors that allows the baseline hazard to take any form via an ensemble average [Scott and Lewin, 2026]. By accurately modelling the baseline hazard, rather than approximating its form via fixed piecewise intervals, power is improved and bias of the estimated treatment effect reduced when the borrowing assumption of parameter exchangeability holds. A "lump-and-smear" borrowing prior makes the model robust to non-exchangeable historical data by increasing the sensitivity of borrowing to the presence of prior-data conflict, reducing the potential for type I error inflation.

We present **BayesFBHborrow**, an R package that implements our semiparametric Bayesian borrowing model with a historical control. We demonstrate how to select the optimal borrowing hyperparameters. The model supports covariate-adjusted borrowing, which can reduce prior-data conflict and improve power when differences in outcomes are attributable to changes in the covariate distribution. As the treatment effect estimator is non-collapsible, the marginal hazard ratio can be estimated via Bayesian G-computation, while still permitting an adjusted analysis to account for control group drift. We illustrate the Bayesian flexible baseline hazard model on a simulated and real dataset with a marginal estimand, for both an unadjusted and adjusted analyses.

Keywords: Bayesian borrowing, flexible baseline hazard, time-to-event, covariate adjustment, G-computation.

*Manuscript submitted to the *Journal of Statistical Software*.

1 Introduction

In early phase clinical trials, there has been a notable shift away from the traditional frequentist approach to Bayesian methods. These methods are particularly beneficial in clinical research, as they allow for the incorporation of existing knowledge with newly collected empirical data. This is relevant when conducting multiple trials for the same illness to discover effective treatments or when evaluating established therapies. Bayesian borrowing, which involves integrating information from previous trials into current research, can lead to improved efficiency through smaller and faster trials, increased statistical power, and reduced patient allocation to less effective treatments.

The implementation of incorporating historical control data into the trial design is not without challenges. Researchers must carefully consider the compatibility of historical data with new trial data to ensure that the modelling assumptions are valid. This involves evaluating the similarity of patient populations, trial designs, and standard of care Pocock [1976]. Using historical data that significantly diverge from current data can lead to issues such as an increased type I error rate and the need for longer more costly trials to offset the influence of incompatible prior data or "prior-data" conflict. These differences will be due to a data shift, either in the distribution of the prognostic covariate or the underlying relationship between predictors and outcome.

The `BayesFBHborrow` [Scott and Axillus, 2024] package, available in R [R Core Team, 2021], enables borrowing from retrospective data for time-to-event endpoints such as time-to-disease progression or time-to-death. These endpoints are the primary outcome in a variety of therapeutic areas, including oncology [Su et al., 2022] and cardiovascular diseases [Psioda et al., 2018]. The borrowing is "dynamic", as the degree of historical data integration depends on the similarity of the data between the historical and current data sets. We account for possible changes in the distribution of the prognostic covariates by incorporating covariate adjustment to minimise prior-data conflict. By accurately capturing the true underlying hazard function in our model, we improve the borrowing characteristics (power and type I error in the presence of prior-data conflict), compared with approaches which constrain the hazard function (Scott and Lewin [2026]).

Bayesian borrowing is a relatively new field and its integration with time-to-event data is still developing. Packages such as `RBesT` [Weber et al., 2021] and `psborrow2` [Gravestock, 2023] represent the leading developments in the R space. `RBesT` performs Bayesian borrowing by informing the prior for the new trial. First, the posterior predictive distribution is sampled via Monte Carlo Markov Chain (MCMC) from a hierarchical model over the historical clinical trial summary statistics. A mixture of densities, with estimated parameters that are conjugate with the likelihood of the new trial, is fitted to the posterior predictive samples of the parameter of interest. By combining this with a vague conjugate prior and a suitable weight, we have the robust meta-analytic predictive prior. This summarises the historical data, whilst accounting for the uncertainty associated with a new parameter from the model and potential prior-data conflict. However, `RBesT` is not designed to accommodate time-to-event data.

`psborrow2` incorporates external information by combining the likelihoods of the historical and the current data in one joint model for either time-to-event or binary endpoints. A simple commensurate prior [Connor Jo Lewis and Carlin, 2019], which characterises the drift between the current and historical parameter, controls the borrowing with either a Weibull or exponential likelihood. Thus, both models place a constraint on the shape of the hazard function.

Various approaches have been proposed for Bayesian borrowing in the time-to-event setting. Hobbs et al. [2013] posit a piecewise exponential model (PEM) with a commensurate prior to control the level of borrowing. They assume predetermined fixed intervals and independence across the associated baseline hazards. Han et al. [2017] use a PEM with fixed intervals and independent baseline hazards over time, to borrow the control effect across multiple studies. They incorporate patient-level covariates to enhance the efficiency of borrowing. Connor Jo Lewis and Carlin [2019] use a Weibull likelihood and commensurate prior to borrow on the baseline hazard from the one historical dataset. Bi et al. [2023] model the baseline hazard with an exponential distribution and use a Dirichlet process to cluster the historical baseline hazards. This helps to discount the historical data in the presence of prior data conflict. All of these methods impose a constraint on the shape of the baseline hazard over time, either as a horizontal line, a step function, or a monotonic function.

In the `BayesFBHborrow` package we use a joint semiparametric hierarchical model to borrow information from a historical time-to-event dataset with a smooth and flexible baseline hazard [Scott and Lewin, 2026]. We use an ensemble method, popular in tree-based approaches such as random forests, to obtain a flexible baseline hazard function. To smooth the shape of the baseline hazard function, we introduce a dependency across the time points between the log baseline hazards via a nearest-neighbour Gaussian Markov random field (NN-GMRF) prior [Besag and Kooperberg, 1995], improving the accuracy of our estimation. "Lump-and-smear" priors can control the borrowing, making the borrowing robust to both violations of the assumptions of exchangeability and prior-data conflict. The model can be used for either a marginal or conditional estimand, through either an adjusted or unadjusted model, as baseline characteristics for the historical and current trial can be included and G-computation [Robins, 1986] performed. This allows the user to target a marginal hazard ratio, whilst the borrowing model is able to account for any drift explained by the observed covariates, improving efficiency and reducing bias compared with simple marginal borrowing approaches.

Our proposed Bayesian borrowing model requires individual participant data (IPD), rather than summary data, as is the case with some other recent approaches (Roychoudhury and Neuenschwander [2020], Bi et al. [2023]). Although this is clearly more difficult to obtain, there are benefits in obtaining additional information, which is recognised by the FDA in its draft guidance (FDA [2026]). Whereas approaches which require a summary statistic constrain the shape of the baseline hazard, our model is able to borrow information whilst remaining free to accurately capture the true shape of the underlying baseline hazard. The IPD also allows the model to adjust covariates from either the current or historical trial (or both), which can help reduce posterior uncertainty within our model. Again, this is not applicable for the summary-level data models.

2 Model specification

The `BayesFBHborrow` package performs Bayesian borrowing through an ensemble method popular in tree-based regressions, such as random forests, to borrow external control information. For each iteration, we partition time and model the baseline hazard with a step function. As we sample the parameters, the partitioned region and associated baseline hazard function are recursively altered by changing the location of the split points or adding or removing split points (or both). The expected posterior baseline hazard is a smoothed function obtained via an ensemble average. Because the baseline hazard function can be

very flexible, we use our prior structure to smooth its shape. This Section will guide the user through the theoretical basis of the sampler for the Bayesian Flexible Baseline Hazard (BFBH) model. For a more detailed summary of the algorithm and the sampling methods, see Scott and Lewin [2026].

2.1 Likelihood

In order to obtain a flexible baseline hazard we begin with a piecewise exponential likelihood for the current trial of the joint model which consists of two elements, the hazard and the cumulative hazard function at the time of the event. For the current trial consisting of $i = 1, ..., n$ patients with the time-to-event outcome represented by y_i and event indicator ν_i. We divide the time axis into $J + 1$ partitions with split points $0 = s_0 < s_1 < ... < s_{J+1}$, and assume a constant baseline hazard $h_j(y_i)$ per partition for $y_i \in I_j = (s_{j-1}, s_j)$.

The vector of stratification variables or prognostic covariates $(p \times 1)$ is defined by $\mathbf{x}_i' = (x_{i1}, ..., x_{ip})$ for subject i with corresponding regression coefficients $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)$ and treatment covariate Z_i and associated effect φ. Let the interval specific baseline hazards be defined as $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_{J+1})$, the likelihood of the current data is

$$\begin{aligned}\mathcal{L}(\boldsymbol{D}|\boldsymbol{\beta}, \boldsymbol{\lambda}, \varphi, \mathbf{s}, J) = &\prod_{i=1}^{n}\prod_{j=1}^{J+1}\big(\lambda_j \exp(\mathbf{x}_i'\boldsymbol{\beta} + Z_i\varphi)\big)^{\delta_{ij}\nu_i} \exp\Big(-\delta_{ij}\big(\lambda_j(y_i - s_{j-1}) \\ &+ \sum_{g=1}^{j-1}\lambda_g(s_g - s_{g-1})\big)\exp(\mathbf{x}_i'\boldsymbol{\beta} + Z_i\varphi)\Big),\end{aligned} \tag{1}$$

where $\boldsymbol{D} = (n, \boldsymbol{y}, Z, \mathbf{X}, \boldsymbol{\nu})$ denotes the observed data with $\nu_i = 1$ if the ith subject failed (event happens) and 0 otherwise, and the variable δ_{ij} is 1 if the subject i was censored or failed in interval j, and 0 otherwise.

The likelihood for the historical data will have the same structure and split point parameters, defined as $\mathcal{L}(\boldsymbol{D}_0|\boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \mathbf{s}, J)$ with $\boldsymbol{D}_0 = (n_0, \boldsymbol{y}_0, \mathbf{X}_0, \boldsymbol{\nu}_0)$.

2.2 Priors

The total number of split points $0 = s_0 < s_1, ..., s_{J+1}$ with s_0 and s_{J+1} fixed, has a right-truncated Poisson distribution with hyperparameter ϕ and right truncation limit $J_{\max}$

$$J \sim \text{Poisson}_t(\phi) \qquad 0 \leq J \leq J_{\max}. \tag{2}$$

The truncation shrinks the mean number of split points and to a greater extent the variance below ϕ, whilst controlling the flexibility of the baseline hazard and reducing the computational burden of the sampler. The distribution of the split point locations, is the even-numbered order statistic of $2J + 1$ uniformly distributed points [Green, 1995].

To smooth the posterior baseline hazard we introduce a dependency of the historical log baseline hazards on their nearest neighbour through a GMRF prior

$$\log(\boldsymbol{\lambda}_0)|\mu, \sigma_\lambda^2 \sim \mathcal{N}_{J+1}(\mu\mathbf{1}, \sigma_\lambda^2\Sigma_s(c_\lambda)), \tag{3}$$

with means μ and the covariance matrix $\sigma_\lambda^2\Sigma_s(c_\lambda)$. The overall variability of the historical log baseline hazard is defined by σ_λ^2 and the smoothness is controlled by the hyperparameter

$c_\lambda \in (0,1)$ within Σ_s (denoted as a function of c_λ). Additionally, these parameters have the following hyperpriors

$$\sigma_\lambda^2 \sim \text{Inv-Gamma}(a_\sigma, b_\sigma), \tag{4}$$
$$\pi(\mu) \propto 1. \tag{5}$$

The priors for the treatment effect, regression coefficients corresponding to the baseline characteristics for the current and historical data are

$$\varphi \sim \mathcal{N}(0, \sigma_\beta^2), \qquad \boldsymbol{\beta} \sim \mathcal{N}(0, \sigma_\beta^2), \qquad \boldsymbol{\beta}_0 \sim \mathcal{N}(0, \sigma_{\beta_0}^2). \tag{6}$$

with σ_β^2 and $\sigma_{\beta_0}^2$ sufficiently large for a vague priors.

To account for a discrepancy between historical and current data, we use a commensurate prior to borrow on the historical log baseline hazard [Hobbs et al., 2013]

$$\log(\lambda_j)|\tau_j \sim \mathcal{N}(\log(\lambda_{0j}), \tau_j), \tag{7}$$

where τ_j is a commensurability (variance) parameter.

2.3 Borrowing method

It is the prior on τ_j that controls the borrowing, and this has two different prior parameterisation options, either the borrowing is determined per split point or across all the split points (where $\tau_j = \tau$). Both of these approaches are modelled by a mixture of inverse gamma distributions and mixture weight p_0.

$$\tau_j^{(\text{mix})} \sim p_0\text{Inv-Gamma}(a_\tau, b_\tau) + (1-p_0)\text{Inv-Gamma}(c_\tau, d_\tau), \tag{8}$$
$$\tau^{(\text{all})} \sim p_0\text{Inv-Gamma}(a_\tau, b_\tau) + (1-p_0)\text{Inv-Gamma}(c_\tau, d_\tau). \tag{9}$$

There is also the option of a simple prior (which is nested in prior (8)) with $p_0 = 1$

$$\tau_j^{(\text{uni})} \sim \text{Inv-Gamma}(a_\tau, b_\tau). \tag{10}$$

The inverse gamma prior of (10) imposes a simple assumption of exchangeability between the baseline hazards for the interval I_j. The lump-and-smear priors (8-9) allow for the possibility of large differences between the log baseline hazard of the historical and current data, which robustifies the prior against prior-data conflict.

2.4 Choice of hyperparameters

As the selection of the hyperparameters which control the borrowing are scale dependent, by default the package will standardise time in the current data so that there is an average of one event per unit of time. This standardisation constant will also be applied to the historical data simplifying the model specification, so that the principled approach to hyperparameter selection for dynamic borrowing outlined in Scott and Lewin [2026] can be followed. The MCMC posterior samples and associated plots from the package are returned in the original unit of time.

2.4.1 Smoothing parameters

When choosing the smoothing parameters of ϕ, $J_{\max}$ and c_λ, our aim is to exploit the trade-off between flexibility and variability. Although the baseline hazard can take more unusual shapes when ϕ is large, this is typically unrealistic. Scott and Lewin [2026] explore combinations of smoothing parameters for various shaped baseline hazard functions. They find that the best hyperparameter choice for their simulated historical datasets based on the fit of the model, regardless of the underlying hazard, is $\phi = 3$ and $J_{\max} = 5$.

The optimal choice of $c_\lambda \in (0, 1)$ can be made with an understanding of the shape of the baseline hazard. A regular shape hazard function requires a large c_λ, while an irregular shape requires a small value. As the true shape of the hazard function is unknown, the expected shape can be estimated by running the model, without the borrowing structure, on the external control group data.

2.4.2 Borrowing parameters

Scott and Lewin [2026] explore the impact of these hyperparameters on dynamic borrowing, through borrowing profiles which define how quickly historical information is discounted for discrepancies between the log-baseline hazard. They demonstrate that specific choices of the mixture prior hyperparameters induce favourable borrowing characteristics, close to full borrowing for tolerated differences between the log baseline hazard, whilst information is fully discounted when prior-data conflict is detected by the model. A simple approach to selecting appropriate values is proposed that is agnostic to the units of time the events are measured in and enables the user to specify their tolerance for differences between the log-baseline hazard. The approach for the mix prior parameterisation is;

- Standardise the current and historical baseline hazard so that there is an average of one event per unit of time in the current data (performed by the package). This means that the control hazard is approximately 1.
- Set $a_\tau = c_\tau = 1$ and $b_\tau = 0.001$ and $d_\tau = 5$.
- Define a limit of tolerable difference between the standardised log baseline hazards ξ. This is the point at which the posterior weight will drop below 0.5.
- Calculate the prior weight

$$p_0 = \left(1 + \frac{0.001}{d_\tau}\left(\frac{\xi^2 + 0.002}{\xi^2 + 2d_\tau}\right)^{-\frac{3}{2}}\right)^{-1}. \tag{11}$$

The posterior weight (or borrowing profile) q_0, can be plotted as a function of the potential difference between the log hazards for the choice of hyperparameters

$$q_0 = \left(1 + \frac{1 - p_0}{p_0}\frac{d_\tau}{b_\tau}\left(\frac{(\log(\lambda_j) - \log(\lambda_{0j}))^2 + 2b_\tau}{(\log(\lambda_j) - \log(\lambda_{0j}))^2 + 2d_\tau}\right)^{-\frac{3}{2}}\right)^{-1}. \tag{12}$$

Alternatively, a prior weight can be defined first, the limit of tolerable difference between the log baseline hazards calculated in (12) and the borrowing profile plotted.

If $\tau^{(\text{all})}$ (9) is used in the model, then the limit of tolerable difference ξ in (11) will be replaced by the limit of the sum of squared error. This will be harder to define, so the user may wish to plot various borrowing profiles for different choices of p_0 to decide on a suitable value.

If the simple univariate inverse gamma prior (10) is selected, then b_τ should be set to a small value (typically 0.001) to encourage borrowing. However, this prior imposes a simpler assumption of exchangeability between the baseline hazards for the interval I_j, and the lump-and-smear mixture priors, which robustifies the model against prior-data conflict, are preferable.

2.5 Design matrix for an adjusted analysis

If covariates are to be included in the model through $\mathbf{X}_0$ and $\mathbf{X}$ for an adjusted analysis, care must be taken to ensure that all historical information is utilised and the current and historical baseline hazard remain concordant. However, any type of design matrix can be used, such as reference-based or sum-to-zero, as the baseline hazard is common to all subjects, the parameterisation for the concordant design matrices does not impact the borrowing.

2.6 MCMC sampler and posterior inference

To sample from the posterior distribution, we use a Metropolis-Hastings [Hastings, 1970] within Gibbs [Geman and Geman, 1984] sampler. The sampler can be initialised with default settings, but adjustments will need to be made to the proposal variance to ensure that the MCMC is sampling the joint posterior sufficiently. We provide a brief summary of the samplers to help the user make the changes correctly and outline the arguments that control the output for the posterior inference.

For fixed J the unknown parameters in the joint model are

$$\boldsymbol{\theta}(J) = (\varphi, \boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\beta}_0, \boldsymbol{\lambda}_0, \boldsymbol{s}, \boldsymbol{\tau}, \sigma^2_\lambda, \mu), \tag{13}$$

the components of $\boldsymbol{\theta}(J)$ are updated by either exploiting conjugacies in the full conditionals or via Metropolis-Hastings steps.

The posterior treatment effect φ, current $\boldsymbol{\beta}$ and historical $\boldsymbol{\beta}_0$ regression coefficients posterior are sampled with a Metropolis-Hastings step. As the gradient and observed curvature of the log-posterior are available in closed form, we employ a Newton-Raphson, locally Gaussian proposal for both the historical and current regression coefficients. At each iteration, we form a multivariate, curvature-adjusted normal approximation around the current parameter value and propose from this distribution, using a single scalar step-size to scale the stochastic component for each coefficient vector, $(\varphi, \boldsymbol{\beta})$, and $\boldsymbol{\beta}_0$. These tuning parameters, `cprop_beta` and `cprop_beta_0` respectively, need to be varied in shorter runs so that the probability of accepting each multivariate proposal is approximately 40%, so that the posterior distribution is explored fully.

For the historical baseline hazard we use a Metropolis-Hastings where the proposal is from a conjugate posterior, after an independent vague gamma prior is combined with the historical likelihood. For the current baseline hazard, the vague gamma prior is combined with both the historical likelihood discounted by Gamma(shape $= a_\lambda$, rate $= b_\lambda$) and the current likelihood

$$\pi_{(\text{prop})}(\lambda_j | \boldsymbol{D}, \boldsymbol{D}_0, \alpha) \propto \mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{s} | \boldsymbol{D}) \mathcal{L}(\boldsymbol{\lambda}_0, \boldsymbol{\beta}_0, \boldsymbol{s} | \boldsymbol{D}_0)^\alpha \pi(\lambda_j), \tag{14}$$

where $0 \leq \alpha \leq 1$. The power parameter `alpha` is set by default at 0.4 to increase the variability of the proposal by discounting historical information. However, you may need to increase both a_λ and b_λ and (or) change α to ensure an appropriate acceptance ratio. As

the data is standardised in the sampler so that the baseline hazard is approximately 1, a_λ and b_λ can remain equal.

Our model treats the split points, both in terms of their location and the total number, as random. The total number of splits J are sampled via a reversible-jump MCMC [Green, 1995] (RJMCMC) which extends or reduces the number of split points by one. This either adds or deletes a split point, whilst adjusting the other parameters in the model. The probability of a birth move (and the respective death move) is set to 0.5 by default, but this can be changed by the user.

To obtain the smoothed posterior baseline hazard, a discrete grid of dimensions $n_{\text{iter}} \times n_{\text{grid}}$ is constructed, where n_{iter} is the number of iterations of the sampler and n_{grid} is the total number of equally spaced partitions of the time interval from 0 to $\max(y_i|\nu_i = 1)$. For each iteration, the baseline hazard associated with the corresponding time interval is saved in the grid. The smoothed expected posterior baseline hazard is an ensemble average

$$\lambda(t) = \frac{1}{n_{\text{iter}}} \sum_{\nu=1}^{n_{\text{iter}}} \lambda_j(\nu), \tag{15}$$

where $\lambda_j(\nu)$ is the interval containing t in the MCMC iteration ν and n_{iter} is the number of iterations. The smoothness of these parameters is controlled by ϕ, $J_{\max}$ (2) and c_λ (7).

The package also provides plots of the smoothed expected posterior predictive hazard function

$$\lambda(t|\mathbf{x}_i)_i = \frac{1}{n_{\text{iter}}} \sum_{\nu=1}^{n_{\text{iter}}} \lambda_j(\nu) \exp\left\{\mathbf{x}_i'\boldsymbol{\beta}(\nu) + z_i\varphi(\nu)\right\}, \tag{16}$$

and expected posterior predictive survival function

$$S(t|\mathbf{x}_i)_i = \frac{1}{n_{\text{iter}}} \sum_{\nu=1}^{n_{\text{iter}}} \exp\left\{-\int_0^t \lambda_j(\nu) \exp\left\{\mathbf{x}_i'\boldsymbol{\beta}(\nu) + z_i\varphi(\nu)\right\} d\nu\right\}, \tag{17}$$

where $\boldsymbol{\beta}(\nu)$ and $\varphi(\nu)$ are the posterior samples of the regression coefficients and the treatment coefficient for an individual with a covariate set $\mathbf{x}_i$ in the treatment group z_i. These are plotted along with the associated credible intervals.

3 Estimand and adjusted models

In Section 4, we demonstrate how to use the package under different population-level summaries, a key element of the estimand framework [Gogtay et al., 2021]. We adopt the terminology from [Daniel et al., 2021], where we use conditional/marginal to refer to the estimand (or parameter that defines our target treatment effect) and adjusted/unadjusted to refer to the analysis model. In our setting, as the estimand of interest in our time-to-event setting is non-collapsible, we cannot simply use an adjusted model (which is typically more robust to prior-data conflict) if we wish to target a marginal estimand. In the first example, rather than omitting covariates from the model to estimate our marginal treatment effect, we account for any potential drift of the historical control by first performing a covariate adjusted analysis and then marginalising over the adjusted probability measure using an approach called standardisation or Bayesian G-computation [Keil et al., 2018]. This approach sits naturally within the Bayesian framework, where both marginalisation and inference are simple steps that are performed with the MCMC sampled posterior distribution. In the

second example, our focus is on a conditional estimand. After a discussion on covariate adjustment and the implications in our model, we provide a short summary of G-computation for our choice of estimand with our time-to-event data, before illustrating the application of our model.

3.1 Covariate adjustment

In many of the Bayesian dynamic methods that borrow control information, the focus is on borrowing with a marginal sufficient statistic. However, the control outcome is likely to be affected by variables, for example, in the oncology setting the tumour stage is predictive of survival. If the distribution of these key prognostic covariates varies across the current and control trial, either due to differences in the study's target population or "drift" as the standard of care improves over time or chance, then the possibility of prior-data conflict increases. By including prognostic covariates in an adjusted model, we can account for this "imbalance", reducing the risk of prior-data conflict.

When we adjust for covariates in our model through $(\mathbf{X}, \mathbf{X}_0)$, the assumption of exchangeability of the current and historical log baseline hazard in (7) is now conditional on the covariates. For borrowing to be meaningful, the true log-hazard function of the covariates must be the same in both the historical and control datasets. We recommend examining the current and historical covariate effect posterior summaries to assess how reasonable this assumption is in your application. The assumption that both datasets are from the same source population; under similar measurement, inclusion criteria, clinical practice, and follow-up processes is even more important [Pocock, 1976] when working with an adjusted model.

Unfortunately, the log link function in our joint hazard ratio model leads to a non-collapsible treatment estimand, if the treatment is effective (all effect measures are collapsible for null treatment effects) (Daniel et al. [2021]). The treatment estimand from a model with covariate adjustment (even with an omission of any treatment covariate interaction) does not have a marginal interpretation, the conditioning changes the nature of the treatment effect we are estimating. In the following subsection we describe a procedure which allows us to adjust for covariates and increase precision, whilst still targeting a marginal estimand. This highlights another aspect of time-to-event modelling, the effect of individual heterogeneity or frailty and drop-out on the estimated treatment effect (in our case the hazard ratio). This phenomenon has also been widely discussed in the literature, particularly in the context of a causal interpretation (Fay and Li [2024], Hernán [2010], Aalen et al. [2015]).

3.2 G-computation and marginal estimands from adjusted models

The marginalisation procedure ensures we exploit both covariate adjustment and available external data borrowed in a robust fashion, to optimise the power for testing a marginal treatment estimand. Intuitively, as we are interested in population-level effects, we work with the survival distribution rather than the hazard function (which is conditional on surviving up to the argument of the function, time t) directly. The marginalised survival function for treatment Z, from expressing the survival probability in terms of the cumulative hazard

function is

$$S(t|Z=z,\boldsymbol{\theta}) = \mathbb{E}[S(t|Z=z,\mathbf{X},\boldsymbol{\theta})]$$
$$= \int_{\mathbf{X}} \exp(-H_0(t))^{\exp(\mathbf{x}'\boldsymbol{\beta}+z\varphi)} p(\mathbf{X}) d\mathbf{X}, \tag{18}$$

where the expectation is over the covariate distribution and $H_0(t)$ is the cumulative baseline hazard function given the baseline hazards λ, and time partition parameters $\mathbf{s}$ and J,

$$H_0(t|\boldsymbol{\theta}) = \prod_{j=1}^{J+1} \delta_{ij}\big(\lambda_j(t - s_{j-1}) + \sum_{g=1}^{j-1} \lambda_g(s_g - s_{g-1})\big)$$

with $\delta_{ij} = 1$ when t is in the jth interval and 0 otherwise.

Rather than performing a full Bayesian analysis, in which a joint prior distribution is specified for the covariates (discrete, continuous, or both) (Keil et al. [2018]), we adopt an approximation procedure based on the Bayesian bootstrap (Willard et al. [2024]). For each unique covariate pattern $\mathbf{x}_i = \mathbf{x}_{k_i}$, $k = 1, ..., K$, we specify a Dirichlet prior on the pattern weights

$$\begin{aligned}
\boldsymbol{\pi} &\sim \text{Dirichlet}(\alpha_1, ..., \alpha_K), \\
k_i &\sim \text{categorical}(\boldsymbol{\pi}), \\
\boldsymbol{\pi}|\boldsymbol{k} &\sim \text{Dirichlet}(\alpha_1 + k_1, ..., \alpha_K + k_K)
\end{aligned}$$

where $k = (k_1, ..., k_K)$ are the counts of observations in each unique covariate pattern. As is standard practise Rubin [1981], we set $\alpha_i = 0$ for all $i = 1, ..., K$. The conjugate posterior weights $\boldsymbol{\pi}|\boldsymbol{k}$ define a random discrete distribution over the observed covariate patterns and are used to approximate the integral over $p(\mathbf{X})$ via a Bayesian bootstrap. The marginalised survival distribution can then be transformed to obtain posterior samples of the log hazard ratio $\gamma(\boldsymbol{\theta})$

$$\gamma(\theta)_m = \log(-\log(S(t|Z=1,\boldsymbol{\theta}_m)) - \log(-\log(S(t|Z=0,\boldsymbol{\theta}_m)).$$

The full procedure in the package is described in the Appendix A. Our routine returns the marginalised hazard for each treatment group over a grid of time. This allows the user to specify a time point, plot the posterior expectation and variance over time, or report a single measure across time.

The interpretation of hazards and hazard ratios for comparing treatment and exposure groups is complicated by the possible effects of unmeasured frailty on event times [Fay and Li, 2024]. Under the model structure in (1), as we assume hazards conditional on the split point and covariates

$$\lambda_j(t)\exp(\boldsymbol{x}_i'\boldsymbol{\beta} + Z_i\varphi) \tag{19}$$

are correctly specified, the model implied by the G-computation step will, in general, be misspecified. This is because the composition of the risk sets under each arm at any given time generally differs from that at the start of the trial, because under a non-null treatment effect, patients with higher hazards tend to experience events earlier. This dynamic selection alters the balance of covariates and unmeasured risk factors in the risk set over time, despite being exchangeable at baseline due to randomisation. Consequently, the true marginal hazard ratio typically varies with time, an effect that is not captured when fitting a marginal

proportional hazards model that assumes a constant hazard ratio. This emphasises the importance of interpreting a hazard as, the instantaneous event rate in the group of survivors at each time, rather than a common hazard applicable to each individual in the study.

This limitation reflects the behaviour of marginal proportional hazards models when an underlying conditional relationship is omitted, which is well documented in the literature (Hernán [2010], Stensrud et al. [2017]). In randomised controlled trials, the marginal estimand is often targeted, and a simplifying assumption of time-constant treatment effect (a constant marginal hazard ratio) is made, despite the likely presence of frailty. In this case, our covariate-adjusted estimator from the g-computation procedure will have a very similar expectation to the unadjusted estimator but with increased precision.

Under a non-null treatment effect, the effect of time varying risk sets in time-to-event data due to unmeasured frailty will affect the marginal posterior hazard ratio of our model, even without additional prognostic covariate adjustment. This is because we obtain a time-varying hazard, by assuming a constant hazard (and constant hazard ratio) for each time partition, and then average over unmeasured frailty and time partitions rather than conditioning on them. The balance of the treated and control group will change dynamically as higher-risk or "sicker" patients fail earlier, leading to a slight attenuation toward the null of the marginal hazard ratio, which is no longer constant over time.

4 Package Functionality through Examples

The design philosophy is based upon the premise that this tool is intended for the analysis of survival data. Accordingly, the canonical R methods for providing a summary, extracting coefficients, and generating plots are primarily aimed at presenting the results of survival analysis, with a secondary focus on producing diagnostics for the MCMC sampler. In this Section we explain the essentials of the package functionality, e.g. calling sequence, meaning of arguments, and most importantly, the connection between the methodology presented in the previous sections and consequences of various options provided by the package, as well as proper interpretation of the results. This will be presented within the context of a series of related examples using data generated via an included simulation function, `genBFBHBdat`, to conclude with analysis of the German Breast Cancer Study (GBCS) of Tamoxifen which is publicly available in the R package, `condSURV`. Note that this document can be processed from the package vignette directory using the command

```
echo 'Sweave("Using-BayesFBHborrow.Rnw")' | R --no-save --quiet &
```

This will process the document and create all the figures and tables using pre-cooked sampler runs stored in R datasets included with this vignette. If you wish to rerun a collection or all of the sampler examples from scratch, locate the `do_examples_option` line at the top of the first R code block in the vignette source file and set it appropriately.

4.1 Examples using simulated data

The simulated current trial and historical control dataset are obtained via the included data generation function `genBFBHBdat()`. We begin by specifying the sample sizes and true values of the parameters, where the variable names are assigned to their corresponding argument names. First, `n_cc_1` and `n_cc_0` are the sizes of the current trial treated and control arms, and `n_hst` is the size of the historical controls data set.

```
R> n_cc_1 <- 200
R> n_cc_0 <- 100
R> n_hst <- 100
```

The data generating mechanism is a proportional hazards Weibull model. The true values of the shape parameter and the effect of treatment in the current trial are `shape` and `B_trt`.

```
R> shape <- 2
R> B_trt <- log(0.55)
```

In this example, we will have 3 pretreatment covariates consisting of two continuous variables and a factor variable. This is specified via the coefficient vector arguments, `B_x_cc`, `B_x_hst`, and the argument `X_fact_levs` in the following way. The argument `X_fact_levs` gives the number (its length) and levels (its components) of factor variables, so for one factor variable of three levels, we set `X_fact_levs=3`. This will require two non-referent effects in both of the coefficient vectors. Continuous covariates are specified by filling the coefficient vectors out further on the front end. Notice below that the coefficient vectors are of length 4 corresponding to a single factor variable with two non-referent levels and two continuous covariates. Here we use equal pretreatment coefficient vectors for the current trial and historical trial but this need not be the case in general. Lastly we specify the intercepts, `int_cc` and `int_hst`, in the current trial and historical controls linear predictors. They are here also taken to be equal but this need not be the case.

```
R> X_fact_levs <- 3
R> B_x_cc <- B_x_hst <- c(-0.3,0.5,0.25,-0.5)
R> int_cc <- int_hst <- -log(3)
```

We call our data generation function, which returns a list of `data.frame`s, named `DAT_cc` and `DAT_hst`, and assign them to the corresponding objects.

```
R> dat_lst <-
   +     genBFBHBdat(n_cc_1=n_cc_1, n_cc_0=n_cc_0, n_hst=n_hst,
   +                 B_trt=B_trt,
   +                 B_x_cc=B_x_cc,
   +                 B_x_hst=B_x_hst,
   +                 int_cc=int_cc, int_hst=int_hst, shape=shape,
   +                 t_er=0.5, t_fin=1.5, X_fact_levs=X_fact_levs)
R> DAT_cc <- dat_lst$DAT_cc
R> DAT_hst <- dat_lst$DAT_hst
```

4.1.1 Selection of Hyperparameters

We outline two approaches to model hyperparameter selection. In our first set of examples based upon simulated data, rather than following Section 2.4 in full to select our borrowing hyperparameters, we outline an abbreviated approach according to an effective sample size (ESS). The borrowing is controlled by the hyperparameters in the τ_j hyperprior. To robustify the model, a lump-and-smear mixture of inverse gamma distributions (8) is selected via `model_choice = "mix"`. We fix `a_tau` = `c_tau` = 1, define the lump to encourage borrowing with `b_tau` = 0.001 and choose the vague smear component with `d_tau` = 5 as in the principled approach outlined in Section 2.4. Our choice of borrowing hyperparameters ensures a sufficiently diffuse density for the vague component for the standardised data,

so that a sample of the commensurate parameter is likely to fully discount the external information. Our prior weight of $p_0 = 0.8$ can therefore be interpreted to have an approximate effective sample size of 80% of the historical sample size. Solving (11) for our choice of p_0, means that our commensurate mixture prior can be interpreted as defining a limit of tolerable difference between current and historical log hazards (ξ) on a standardised scale of 0.29 (the call function is below) or a hazard ratio of 1.35. This is the threshold for data conflict, as the posterior weight associated with the component which encourages borrowing drops below 0.5.

```
R> xifinder(b_tau = 0.001, d_tau = 5, p_0 = 0.8)

[1] 0.2913806
```

The smoothness of the baseline hazard is controlled by the truncated Poisson prior on the split points and the NN-GMRF prior hyperparameter c_λ. To manage the trade off between flexibility and variability, we follow the example of Scott and Lewin [2026] who demonstrate the importance of restricting the split points to prevent the model from over fitting the posterior baseline hazard to the data, with `phi`=3 and `Jmax`=5. This constrains the flexibility of the smoothed posterior baseline hazard, the wider time intervals within the PEM ensure that there is more data to estimate the baseline hazard at later time-to-events, where there are fewer events, reducing the overall variability of the baseline hazard estimate. As we know the underlying baseline hazard is monotonic, we choose a `clam_smooth` of 0.8. The default setting of a vague prior choice of (1,1) for `a_sigma` and `b_sigma` (4) controls the overall variability of the historical baseline. Finally, the variance of the vague prior on the treatment parameter is set by `beta_prior`=10^2.

```
R>     hyperparameters_sim <-
   +         list(beta_prior = 10^2,
   +              beta_0_prior = 10^2,
   +              a_tau = 1,
   +              b_tau = 0.001,
   +              c_tau = 1,
   +              d_tau = 5,
   +              p_0 = 0.8,
   +              a_sigma = 1,
   +              b_sigma = 1,
   +              clam_smooth = 0.8,
   +              phi = 3,
   +              Jmax = 5)
```

The MCMC sampler will select the initial parameter values. The historical baseline hazard is sampled from a gamma distribution with shape and scale equal to the number of events and the total exposure time within the historical data. The estimated mean and standard deviation of the log transformed data is used for the initial values of μ and σ^2_λ in the NN-GMRF prior (3). The commensurate parameters, τ_j are sampled from the prior.

4.1.2 MCMC sampling options

The `tuning_parameters` control the performance of the MCMC sampler, and in particular the variance of the proposal distributions. The parameter that scales the Metropolis Hastings

proposal for the regression coefficients `cprop_beta`, is tuned to achieve an acceptance ratio in the region of $30\% \sim 40\%$. The baseline hazard parameters for the proposal prior `a_lambda` and `b_lambda` are set to 0.01 and `alpha` = 0.3. The probability of a birth move is set to the default value of 0.5.

```
R> tuning_parameters_sim <-
   +     list(cprop_beta = 1.35,
   +          cprop_beta_0 = 1.35,
   +          a_lambda = 0.01,
   +          b_lambda = 0.01,
   +          pi_b = 0.5,
   +          alpha = 0.4)
```

In the next four examples, we explore how to estimate; (i) the baseline hazard using data on the historical controls alone, (ii) the baseline hazard and treatment effect using data from the current trial alone (e.g. without borrowing) and (iii) the baseline hazard and treatment effect using data from the current trial and from the historical controls (e.g. with borrowing). We perform the estimation for both marginal and conditional treatment effect estimands. In the context of a marginal treatment estimand, we demonstrate how an adjusted model can be used via G-computation.

4.1.3 Current trial / historical controls inference only

The model can be fitted to the historical controls alone for an estimate of the posterior baseline hazard with a call similar to the one shown below. The first argument, `formula`, names the time to event and event indicator variables, `tte` and `event`, wrapped in the `Surv` function as is typically done in survival models in R. On the right hand side we specify either an intercept only model, `1`, or any pre-treatment variables related to randomisation. These can be continuous or factor variables, or character intended for coercion to factor. The reference level can be set in the usual way using an internal function R such as `relevel`. The arguments `data` and `CntlOnly` are set to `data=DAT_hst`, and `CntlOnly=TRUE`. The model `hyperparameters` and `tuning_parameters`, discussed above, are specified. We select 6000 iterations for the sampler after 2000 warmup iterations.

```
R> fit_sim_hst_cndl <-
   +     BayesFBHborrow(formula=Surv(tte, event)~X_01+X_02+X_03,
   +                    data = DAT_hst, CntlOnly = TRUE,
   +                    hyperparameters = hyperparameters_sim,
   +                    tuning_parameters = tuning_parameters_sim,
   +                    warmup_iter = 2000, iter = 6000,
   +                    refresh = 2000, verbose = TRUE)
```

A plot of the baseline hazard can be used to judge the adequacy of the choice of the smoothing parameters, (`J_max`, `phi`, `c_lambda`). Use of the plot method for the class is described below.

The model can also be fitted to the current trial alone using a call similar to the above, but with the following changes. The `data` argument is set to the current trial dataset, `DAT_cc`, and the treatment indicator variable is listed first among the regressors on the right side of the formula argument. There is no need to set the `CntlOnly` argument to `FALSE`, as this is its default value.

```
R> fit_sim_nb_cndl <-
   +     BayesFBHborrow(formula=Surv(tte, event)~X_trt+X_01+X_02+X_03,
   +                    data = DAT_cc,
   +                    hyperparameters = hyperparameters_sim,
   +                    tuning_parameters = tuning_parameters_sim,
   +                    warmup_iter = 2000, iter = 6000,
   +                    refresh = 2000, verbose = TRUE)
```

4.1.4 Inference on current trial with borrowing, with covariates

To undertake inference on the current trial, whilst borrowing from historical controls, the following changes are made to the previous call. In addition to specification of the argument `data` with data on the current trial, we must now specify the argument `data_hist` with a `data.frame` containing data on historical controls. The borrowing model, which determines the parameterisation for the variance of the baseline hazard prior, is selected using the argument `model_choice`. With `model_choice="all"`, the variance of the log baseline hazard is aggregated across all split points. Alternatively, it can be defined separately for each split point using `model_choice="mix"` (the default). There is also an option for the standard univariate inverse gamma for each split via `model_choice="uni"`, but a "lump-and-smear" mixture is recommended. Here we use the default option, `model_choice="mix"`.

```
R> fit_sim_wb_cndl <-
   +     BayesFBHborrow(formula=Surv(tte, event)~X_trt+X_01+X_02+X_03,
   +                    data = DAT_cc, data_hist = DAT_hst,
   +                    model_choice="mix",
   +                    hyperparameters = hyperparameters_sim,
   +                    tuning_parameters = tuning_parameters_sim,
   +                    warmup_iter = 2000, iter = 6000,
   +                    refresh = 2000, verbose = TRUE)
```

4.1.5 Output: Conditional Estimand versus Marginal Contrast

The print method shows two tables: treatment/pretreatment effects and survival probabilities, shown here formatted as booktabs tables. The treatment/pretreatment effects correspond to the model log hazard ratios for each MCMC sample and are summarised by the posterior median together with the 95% credible interval, as shown in Table 1.

Table 1: Conditional Treatment Effect and Pre-treatment Covariates Effects in model borrowing from historical controls.

	logHR	exp(logHR)	CrIn.L	CrIn.H
X_trt	-0.9305	0.3944	-1.2085	-0.5895
X_01	-0.2880	0.7497	-0.4464	-0.1456
X_02	0.5221	1.6855	0.3807	0.6799
X_03b	0.6525	1.9203	0.4703	0.8621
X_03c	-0.3892	0.6776	-0.6369	-0.1781
X_01_0	-0.0971	0.9075	-0.3996	0.1960
X_02_0	0.5397	1.7155	0.3527	0.7336
X_03b_0	0.2238	1.2509	-0.0696	0.5153
X_03c_0	-0.2564	0.7738	-0.5665	0.0346

Table 1 reveals a statistically significant treatment effect, with the conditional treatment effect indicating a risk reduction of 61%, and the credible interval fully on one side of zero.

Displayed survival probabilities are by default conditional per arm survival probabilities per MCMC sample e.g. $S(t_f)$, in the control arm and $S(t_f)^{\mathrm{e}^{\beta_{\mathrm{trt}}}}$ in the intervention arm at $f = 25\%, 50\%, 75\%$ and 100% information time. Here, t_f denotes the time at which the indicated portion, f of total events is reached. These quantities are again summarised by the posterior median together with the 95% credible interval.

Notice that here "conditional" means that each of these estimates represents survival in a hypothetical group of patients with pretreatment covariates equal to their referent values. For continuous variables, these are their means over the dataset supplied in the argument `data`. For factor variables, they are also set at their means, but under a hypothetical design balanced across the levels of each factor. These results are shown in Table 2.

Table 2: Conditional per arm survival probabilities at landmark information fraction in model with borrowing from historical controls, simulated data.

Arm	InfFrac	Time	Survival	Surv_L	Surv_U
C	0.25	0.5243	0.9747	0.9631	0.9828
C	0.50	0.8177	0.9346	0.9157	0.9530
C	0.75	1.0378	0.8983	0.8746	0.9231
C	1.00	1.4966	0.8127	0.7557	0.8608
I	0.25	0.5243	0.9899	0.9854	0.9935
I	0.50	0.8177	0.9737	0.9655	0.9805
I	0.75	1.0378	0.9586	0.9473	0.9683
I	1.00	1.4966	0.9212	0.8963	0.9409

4.1.6 G-computation

When the model is fit to a trial, either with or without borrowing, and pre-treatment covariates are included, then after the sampler runs and conditional estimands are derived, G-computation is performed to obtain the marginal per arm survival probabilities and

marginal treatment effect (MTE). As the MTE is the difference of log-logged marginal per arm survival function averaged over pretreatment variables, it is a time-varying quantity and is therefore shown at landmark information fractions in Table 3.

The print method is coerced into displaying marginal contrasts by specifying `G_compute=TRUE` either in the original call sequence, or in an application of the `update` method. The latter is done without any need to rerun the sampler.

```
R> mgnl_cntrst_sim_wb <- update(fit_sim_wb_cndl, G_compute=TRUE)
```

Table 3: Marginal treatment effect at landmark information fraction in model with borrowing from historical controls, simulated data.

InfFrac	Time	MTE	exp(MTE)	MTE_L	MTE_U
0.25	0.5243	-0.2320	0.7929	-0.3117	-0.1406
0.50	0.8177	-0.2197	0.8027	-0.2963	-0.1333
0.75	1.0378	-0.2133	0.8079	-0.2886	-0.1292
1.00	1.4966	-0.1998	0.8189	-0.2734	-0.1217

Some final comments regarding the calling sequence. Notice when this formula/data interface to the function is used, variable names are entirely immaterial, with the only unusual requirement that the treatment indicator, when included, must be listed first, in addition to consistency of variable names among all datasets when there is more than one supplied. This provision is recommended in most cases. There is also a data only interface which offers additional flexibility, e.g. allows interactions et cetera. This option is used by omitting the `formula` argument, and instead, adhering to the following strict variable naming conventions. The time to event and event indicator variables must be named `tte` and `event` respectively, and variable names must begin with the prefix `X_`. Again, a treatment indicator, when used, must be the first of these so named variables in all datasets. Following this approach, factor variables can still be left as is or as character with intention to coerce as factor, and the function will recode them to dumy indcator variables by default, but without any possibility of entering interaction terms. If the user wishes to include interactions then in addition to using the data only interface, the argument `preprocess` must be set to `FALSE` and all recoding must be done manually.

For publication purposes, the user can extract the tables shown above for downstream wrapping in a table formatter of their own choosing. For example, if `my_bfbhb` is the result of a call with `G_compute=FALSE` then conditional estimand tables are extracted via the following commands:

- Coefficients Table: `coef(my_bfbhb)`.
- Survival Probabilities Table: `summary(my_bfbhb)$surv_summary`.

with marginal contrast versions obtained by wrapping the object first in a call to the update method `update(my_bfbhb, G_compute=TRUE)`. Note as well that the survival probabilities at landmark information fraction are internally extracted from the more detailed version which contains corresponding survival probabilities at every time in the fine meshed internal grid of length `max_grid` (default 2000). These are in the component `surv_dat` of the returned object, representing conditional estimand or marginal contrast depending upon the value of `G_compute`.

4.1.7 Output–MCMC Sampling details

The complete sampler history is recorded and accessible, allowing derivation of any posterior functional. Specifically, the `samples` component of the returned object is a data.frame with `iter` rows, containing all scalar-valued components such as:

- Model coefficients (for current and historical data likelihood portions)
- Number and locations of split points (J, $s_0, s_1, \ldots, s_J, s_{J+1}$)
- Sampled hyperparameters, such as (μ, σ^2_λ) from NN-GMRF prior on the baseline hazard values.
- Piecewise constant baseline hazards ($\lambda_1, \cdots, \lambda_J$ and $\lambda_{0,1}, \cdots, \lambda_{0,J}$), for both current and historical data (maximum split points: J)

While `samples` provides a comprehensive sampler output representation, the most practical form for baseline hazard samples is their snapped-to-common-grid version (each sample of length `max_grid` per sample). This large matrix is written as a binary file to facilitate efficient storage, with a unique file name that is generated and stored in the returned object.

If the full snapped-to-common-grid matrix of samples (dimensions: `max_grid` $\times$ `iter`) for the conditional baseline hazard is needed, use the function `read_haz_mcmc_smpls` as shown below.

```
R> surv_dat_full <- read_haz_mcmc_smpls(fit_sim_wb_cndl)
```

The result is a list with 3 components, `cnd_haz_0`, `mgnl_haz_0` and `mgnl_haz_1`. The first is the conditional baseline hazard while the last two are the log-logged arm specific survival functions from which the marginal treatment effect is derived as the difference.

4.1.8 Plot method

The plot method for the `BayesFBHborrow` class will create upon the user's choice, a plot of control and intervention arm hazards or survival functions in the form of a median coupled with 95% pointwise credible interval ribbons, or density plot of the treatment effect (conditional or marginal at specified landmark information fraction) or a trace plot of any one of the state variables in the output component `samples`. These will have a differing interpretation depending on whether `G_compute` was set to `FALSE` or `TRUE` as described above, namely that the relevant quantities are conditional or marginal estimates from the G-computation. A plot of the hazard function estimated from historical controls is created by

```
R> p_cndl_haz_hst <- plot(fit_sim_hst_cndl, type="hazard")
```

This is used to check the adequacy of the smoothing parameters, (`J_max`, `phi`, `c_lambda`), as mentioned above, and as displayed in the Appendix (Figure 6). The next two lines create plots of the per arm conditional hazard functions from the fit to the current trial without borrowing, and fit to the current trial with borrowing from historical controls, respectively.

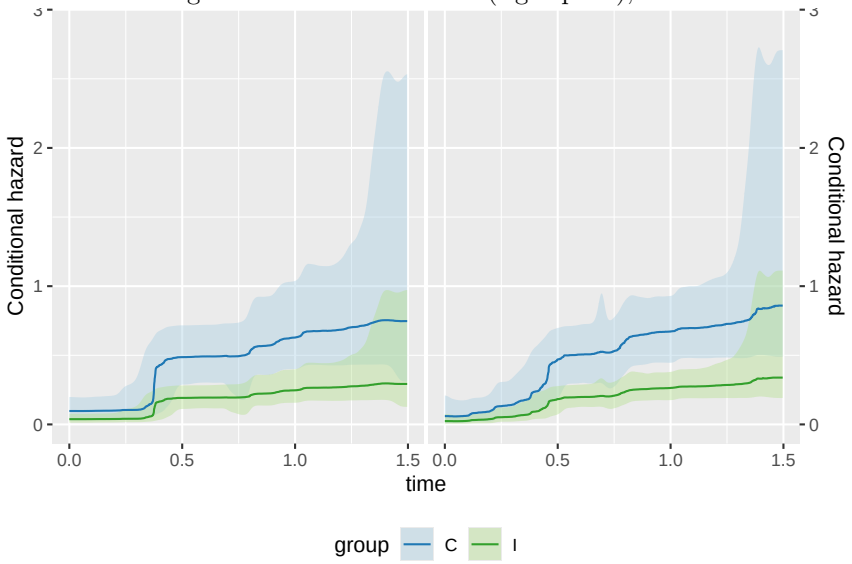
```
R> p_cndl_haz_nb <- plot(fit_sim_nb_cndl, type="hazard")
R> p_cndl_haz_wb <- plot(fit_sim_wb_cndl, type="hazard")
```

The `Combine()` function is used to neatly combine two of the plots as shown in Figure 1.

```
R> p_cndl_haz_cmb <- Combine(p_cndl_haz_nb, p_cndl_haz_wb)
```

Corresponding plots of conditional per arm survival curves are obtained by specifying `type="survival"` instead.

Figure 1: Estimated conditional per arm hazard functions in fit to current trial alone (left pane) and with borrowing from historical controls (right pane), simulated data



4.2 Analysis of German Breast Cancer Study Data

For the second series of examples, the model is applied to the German Breast Cancer Study Data (gbcsCS) [Sauerbrei and Royston, 1999], available from the R package `condSURV` [Meira-Machado and Sestelo, 2023]. The original dataset comprises 686 observations across 16 variables, for patients with node-positive breast cancer, recruited between July 1984 and December 1989. Additional patient-level information in the dataset includes the treatment (hormone), time to recurrence or censoring (rectime), recurrence or censoring indicator (censrec), tumour grade (grade), and size (size).

We are interested in investigating the effect of treatment with the hormone Tamoxifen on the hazard ratio after adjusting for pretreatment variables, and therefore target a conditional estimand. We split the control arm into two using the median time of enrollment, to create a historical set and a current set. Thus, there are 193 and 247 patients in the historical and current control groups, and 96 treated patients who enrolled after the median for the current treated group.

After loading the dataset, we perform some data wrangling. We use `diagdateb` to distinguish the current and historical datasets, and define the treatment variable, `tamoxifen`, as one less than the {1,2} valued original variable, `hormone`. We do the same for the variable `menopause` and convert the original variable, `grade` into a factor with levels corresponding to the original ordinal values 1, 2, and 3. We define a variable `hst_flg` to help us artificially create our current trial and historical datasets, based on whether a patient enrolled before or after the median enrollment time. Then, we create a factor variable, `group`, which we'll use to construct Kaplan-Meier plots. Finally, we split the dataset into "current" and "historical" portions according to the previously mentioned chronology of enrollement.

```
R> data(gbcsCS, package="condSURV")
```

```
R> gbcs_full <- gbcsCS
R> gbcs_full$diagdateb <- as.Date(format(as.character(gbcs_full$diagdateb),
   +                                     format="%d-%m-%Y"), format="%d-%m-%Y")
R> gbcs_full$tamoxifen <- gbcs_full$hormone - 1
R> gbcs_full$menopause <- gbcs_full$menopause - 1
R> gbcs_full$grade <- factor(gbcs_full$grade, levels=as.character(1:3))
R> gbcs_full$hst_flg <- 1*with(gbcs_full, diagdateb < median(diagdateb))
R> grp_lvls <- c("Hist Cntl","Curr Cntl","Curr Trt")
R> gbcs_full$group <-
   +   factor(with(gbcs_full,
   +                grp_lvls[tamoxifen + (1 - hst_flg) + 1]),
   +          levels=grp_lvls)
R> gbcs_curr <- gbcs_full[gbcs_full$hst_flg==0,]
R> gbcs_hist <- gbcs_full[with(gbcs_full, (hst_flg==1) & (tamoxifen==0)),]
```

Here are the first six rows of the current dataset.

Table 4: German Breast Cancer Study Data– Enrollment after the Median – Browsing 1st 6 lines

id	diagdateb	rectime	censrec	tamoxifen	menopause	estrg_recp	size	grade
17	1986-10-14	518	1	0	1	9	20	2
22	1986-10-14	1722	0	1	0	20	21	2
84	1987-10-16	535	1	0	0	8	55	3
85	1987-04-02	1653	0	0	1	14	30	2
194	1987-02-05	1182	0	0	0	11	25	2
195	1987-03-20	71	0	0	0	41	18	2

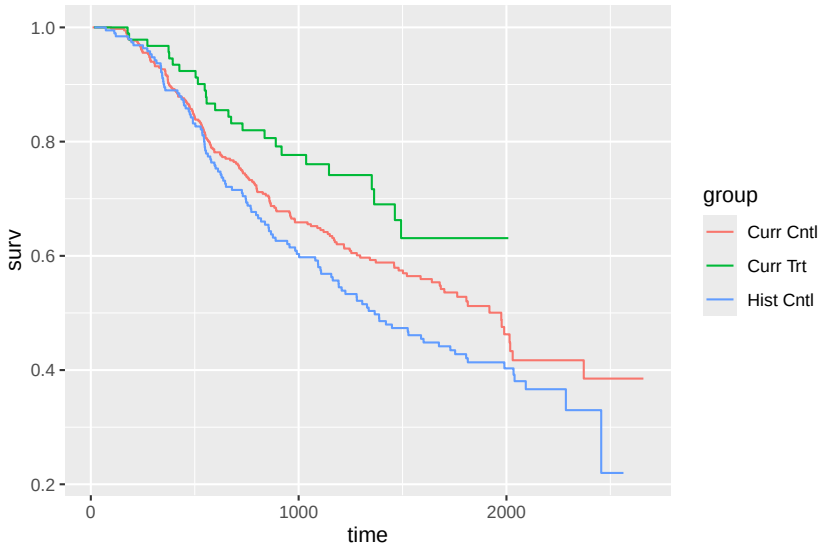
We take a look at Kaplan-Meier plots blocked on chronological dataset and treatment group.

```
R> grp_lvls <- c("Hist Cntl","Curr Cntl","Curr Trt")
R> fit_sf <- survfit(Surv(rectime, censrec)~group, data=gbcs_full)
R> KM_dat <- data.frame(time=fit_sf$time, surv=fit_sf$surv)
R> n_strat <- fit_sf$strata
R> KM_dat$group <-
   +     c(rep(grp_lvls[1], n_strat[1]), rep(grp_lvls[2], n_strat[2]),
   +       rep(grp_lvls[3], n_strat[3]))
R> p_KM <- ggplot(data=KM_dat) + geom_step(aes(time, surv, group=group, color=group))
```

The Kaplan-Meier plots (Figure 2) appear to exhibit a drift in the marginal probability of survival over time. This may be caused by various factors such as improvements in the underlying standard of care or changes in the patient demographics. This motivates the importance of covariate adjusted borrowing, which will account for drift if it is driven by the pretreatment variables.

We adopt a conservative approach to borrowing, through the principled approach to selection of borrowing hyperparameters outlined in Scott and Lewin [2026], in order to

Figure 2: Kaplan-Meier Plot for German Breast Cancer Data Stratified on Enrollment before or after median and Intervention Arm



address the possible drift between the control groups. We set $a_\tau = c_\tau = 1$, $b_\tau = 0.001$ and $d_\tau = 5$ (as in our previous example) and define a limit of tolerable difference for the current and historical standardised log baseline hazard of approximately $\xi = 0.18$, which leads to a prior weight of 0.5 (the borrowing profile is in Figure 5).

```
R> xifinder(b_tau = 0.001, d_tau = 5, p_0 = 0.5)

[1] 0.1797401

R> hyperparameters_gbcs <- list(beta_prior = 10^2,
   +                            beta_0_prior = 10^2,
   +                            a_tau = 1,
   +                            b_tau = 0.001,
   +                            c_tau = 1,
   +                            d_tau = 5,
   +                            p_0 = 0.5,
   +                            a_sigma = 1,
   +                            b_sigma = 1,
   +                            clam_smooth = 0.8,
   +                            phi = 3,
   +                            Jmax = 5)
```

We specify the variance on our vague prior on the regression coefficients of the historical study with `beta_0_prior`=10^2, and both the treatment effect and regression coefficients in the control group with `beta_prior`=10^2. The MCMC sampler proposal scalars, `cprop_beta` and `cprop_beta_0`, are tuned by running subsequent models, with much fewer total iterations,

so that we reach an acceptance ratio of 30% $\sim$ 40%. For example, we specify the following tuning parameters;

```
R> tuning_parameters_gbcs <- list(cprop_beta = 1.17,
   +                              cprop_beta_0 = 1.21,
   +                              a_lambda = 0.5,
   +                              b_lambda = 0.5,
   +                              pi_b = 0.5,
   +                              alpha = 0.4)
```

and conduct a test run with 500 iterations to check the acceptance ratios,

```
R> test <-
   + BayesFBHborrow(Surv(rectime, censrec) ~ tamoxifen + menopause + size + grade,
   +                data=gbcs_curr, data_hist=gbcs_hist,
   +                model_choice="mix",
   +                tuning_parameters=tuning_parameters_gbcs,
   +                hyperparameters=hyperparameters_gbcs,
   +                iter=500, warmup_iter=100, refresh=0)
```

leading to acceptable acceptance ratios of `beta_acc_ratio`=0.306 and `beta_0_acc_ratio`=0.344, respectively.

Our first model fit, for illustration purposes, is without borrowing. Our current dataset is `gbcs_curr` and the treatment variable is `tamoxifen`. We adjust for the pretreatment variables `menopase` (0/1) `size` (tumour size, continuous) and `grade` (histologic grade of tumour, factor).

```
R> fit_gbcs_nb_cndl <-
   +     BayesFBHborrow(Surv(rectime, censrec)~tamoxifen + menopause + size + grade,
   +                    data=gbcs_curr,
   +                    tuning_parameters=tuning_parameters_gbcs,
   +                    hyperparameters=hyperparameters_gbcs,
   +                    iter=6000, warmup_iter=2000, refresh=2000,
   +                    verbose=TRUE)
```

Next, we fit the model with borrowing from historical controls.

```
R> fit_gbcs_wb_cndl <-
   +   BayesFBHborrow(Surv(rectime, censrec)~ tamoxifen + menopause + size + grade,
   +                  data=gbcs_curr, data_hist=gbcs_hist,
   +                  model_choice="mix",
   +                  tuning_parameters=tuning_parameters_gbcs,
   +                  hyperparameters=hyperparameters_gbcs,
   +                  iter=6000, warmup_iter=2000, refresh=2000,
   +                  verbose=TRUE,
   +                  max_grid=2000,
   +                  standardise=TRUE)
R>
R>
```

resulting in the treatment effect estimate in Table 5. The convergence of the sampler is checked by inspecting the trace plot of the parameters, such as the conditional treatment effect `tamoxifen` in Figure 7.

```
R> p_gbcs_wb_trace <- plot(fit_gbcs_wb_cndl, type="trace", col="tamoxifen")
```

Table 5: Conditional Treatment Effect and Pre-treatment Covariates Effects in model borrowing from historical controls.

	logHR	exp(logHR)	CrIn.L	CrIn.H
tamoxifen	-0.4549	0.6345	-0.7756	-0.1303
menopause	-0.0776	0.9254	-0.3464	0.1768
size	0.0113	1.0114	0.0035	0.0192
grade2	0.2134	1.2379	0.0175	0.4190
grade3	0.2804	1.3236	0.0450	0.5130
menopause_0	0.3382	1.4025	0.0628	0.6016
size_0	0.0191	1.0193	0.0107	0.0275
grade2_0	0.3341	1.3967	0.1077	0.5623
grade3_0	0.4723	1.6036	0.2271	0.7389

The conditional treatment effect of tamoxifen is a reduction in recurrence of 36.5%, with credible interval ranging from 12.2% to 54%.

As previously mentioned one can look at the predictive curves with calls to the `plot()` methods.

```
R> p_gbcs_cnd_nb_haz <- plot(fit_gbcs_nb_cndl, type="hazard", ylim=c(0,0.0010))
R> p_gbcs_cnd_wb_haz <- plot(fit_gbcs_wb_cndl, type="hazard", ylim=c(0,0.0010))
R> p_gbcs_cnd_haz_cmb <- Combine(p_gbcs_cnd_nb_haz, p_gbcs_cnd_wb_haz)
R> p_gbcs_cnd_nb_TrtEff <- plot(fit_gbcs_nb_cndl, type="TrtEff")
R> p_gbcs_cnd_wb_TrtEff <- plot(fit_gbcs_wb_cndl, type="TrtEff")
R> p_gbcs_cnd_TrtEff_cmb <- Combine(p_gbcs_cnd_nb_TrtEff, p_gbcs_cnd_wb_TrtEff,
   +                                lgnd=c("NB","WB"))
```

5 Summary

The `BayesFBHborrow` package enables Bayesian borrowing within a joint hierarchical model from historical control data in a time-to-event setting. By integrating over the uncertainty of the location and number of split points in our ensemble method and incorporating smoothing priors, we induce a flexible baseline hazard in our joint semiparametric model. This leads to improved borrowing characteristics compared to standard dynamic approaches, which constrain the shape of the baseline hazard. The borrowing is controlled by a commensurability parameter τ_j, where a prior mixture ensures that the borrowing is more robust to prior-data conflict. The model allows for different types of estimands through the addition of prognostic covariates, which may also improve the precision of the estimate.

Possible extensions of the `BayesFBHborrow` package include the implementation of borrowing across multiple historical datasets and a more varied joint prior structure to capture the between trial heterogeneity. With a similar structure to our commensurate prior (8), Neuenschwander et al. [2016], proposed an extension to allow for non-exchangeability. Alternatively, to account for departures from the exchangeability assumption, a Dirichlet process

Figure 3: Conditional Estimate of baseline hazard function from model fit to current trial alone and fit with borrowing from historical controls for the German Breast Cancer Study

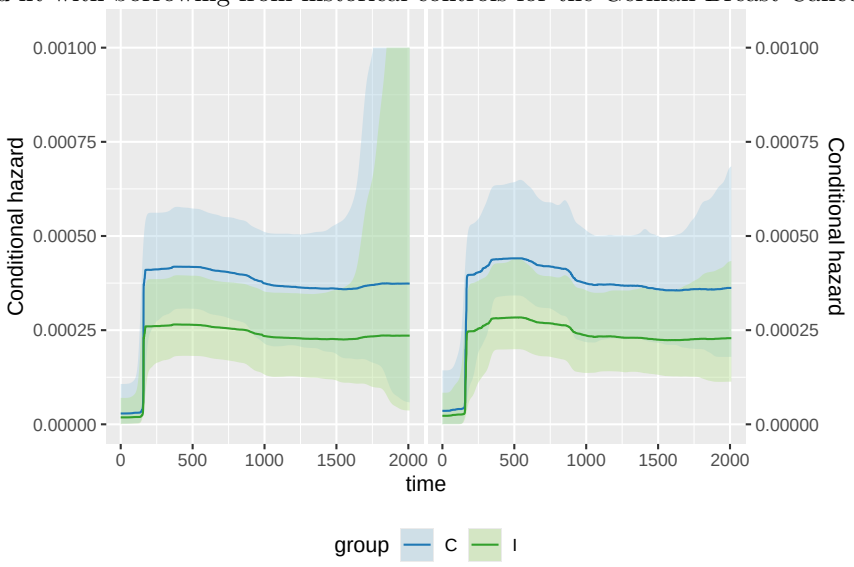
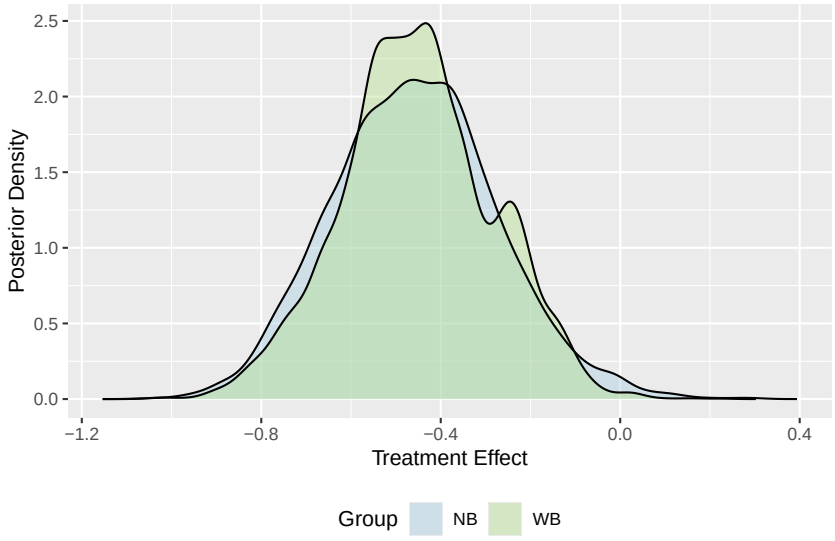


Figure 4: Posterior density of conditional treatment effect from model fit to current trial alone and fit with borrowing from historical controls for the German Breast Cancer Study



mixture prior [Escobar and West, 1995] could be used to create a data-dependent clustering mechanism. This approach has been adopted by Hupf et al. [2021] in the context of a binary end point and in a time-to-event setting [Bi et al., 2023]. As the number of historical trials is usually small, this type of clustering can also be achieved using a RJMCMC approach.

Currently there is an option to run an analysis with and without borrowing to quantify the amount of information within the prior in terms of an historical ESS. This retrospective measure can only be performed once the current data has been realised. When there is conjugacy, the Fisher information can be used to quantify the prior information centred at the likelihood Neuenschwander et al. [2020]. Further research is required to be able to quantify the prospective information in the prior for more complicated models.

Disclosure

Darren Scott and Grant Izmirlian are employed full-time and possess equity holdings in AstraZeneca.

References

Odd O. Aalen, Richard J. Cook, and Kjetil Røysland. Does cox analysis of a randomized survival study yield a causal treatment effect? *Lifetime Data Analysis*, 21:579–593, 10 2015. ISSN 1380-7870. doi: 10.1007/s10985-015-9335-y.

Julian Besag and Charles Kooperberg. On conditional and intrinsic autoregressions. *Biometrika*, 82:733–779, 1995.

Dehua Bi, Meizi Liu, Jianchang Lin, and Rachael Liu. Beats: Bayesian hybrid design with flexible sample size adaptation for time-to-event endpoints. *Statistics in Medicine*, 2023. ISSN 10970258. doi: 10.1002/sim.9936.

Jiawen Zhu Connor Jo Lewis, Somnath Sarkar and Bradley P. Carlin. Borrowing from historical control data in cancer drug development: A cautionary tale and practical guidelines. *Statistics in Biopharmaceutical Research*, 11(1):67–78, 2019. doi: 10.1080/19466315.2018.1497533. URL `https://doi.org/10.1080/19466315.2018.1497533`. PMID: 31435458.

Rhian Daniel, Jingjing Zhang, and Daniel Farewell. Making apples from oranges: Comparing noncollapsible effect estimators and their standard errors after adjustment for different covariate sets. *Biometrical Journal*, 63:528–557, 3 2021. ISSN 15214036. doi: 10.1002/bimj.201900297.

Michael D. Escobar and Mike West. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90:577–588, 1995. ISSN 1537274X. doi: 10.1080/01621459.1995.10476550.

Michael P Fay and Fan Li. Causal interpretation of the hazard ratio in randomized clinical trials. *Clinical Trials*, 21:623–635, 10 2024. ISSN 1740-7745. doi: 10.1177/17407745241243308. URL `https://journals.sagepub.com/doi/10.1177/17407745241243308`.

FDA. Use of bayesian methodology in clinical trials of drug and biological products. Technical report, 2026.

Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, 1984. doi: 10.1109/TPAMI.1984.4767596.

Nithya Jaideep Gogtay, Priya Ranganathan, and Rakesh Aggarwal. Understanding estimands. *Perspectives in Clinical Research*, 12:106–112, 4 2021. ISSN 22295488. doi: 10.4103/picr.picr_384_20.

Isaac Gravestock. *psborrow2: An R Package for Bayesian Dynamic Borrowing Simulation Study and Analysis*, 2023. R package version 0.0.2.0.

Peter J. Green. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732, 1995. ISSN 00063444. URL `http://www.jstor.org/stable/2337340`.

Baoguang Han, Jia Zhan, Z. John Zhong, Dawei Liu, and Stacy Lindborg. Covariate-adjusted borrowing of historical control data in randomized clinical trials. *Pharmaceutical Statistics*, 16:296–308, 7 2017. ISSN 15391612. doi: 10.1002/pst.1815.

W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 04 1970. ISSN 0006-3444. doi: 10.1093/biomet/57.1.97. URL `https://doi.org/10.1093/biomet/57.1.97`.

Miguel A. Hernán. The hazards of hazard ratios. *Epidemiology*, 21:13–15, 1 2010. ISSN 10443983. doi: 10.1097/EDE.0b013e3181c1ea43.

Brian P Hobbs, Bradley P Carlin, and Daniel J Sargent. Adaptive adjustment of the randomization ratio using historical control data. *Clinical Trials*, 10(3):430–440, 2013. doi: 10.1177/1740774513483934. URL `https://doi.org/10.1177/1740774513483934`. PMID: 23690095.

Bradley Hupf, Veronica Bunn, Jianchang Lin, and Cheng Dong. Bayesian semiparametric meta-analytic-predictive prior for historical control borrowing in clinical trials. *Statistics in Medicine*, 40:3385–3399, 2021. ISSN 10970258. doi: 10.1002/sim.8970.

Alexander P. Keil, Eric J. Daza, Stephanie M. Engel, Jessie P. Buckley, and Jessie K. Edwards. A bayesian approach to the g-formula. *Statistical Methods in Medical Research*, 27:3183–3204, 10 2018. ISSN 14770334. doi: 10.1177/0962280217694665.

Luis Meira-Machado and Marta Sestelo. *condSURV: Estimation of the Conditional Survival Function for Ordered Multivariate Failure Time Data*, 2023. URL `https://CRAN.R-project.org/package=condSURV`. R package version 2.0.4.

Beat Neuenschwander, Simon Wandel, Satrajit Roychoudhury, and Stuart Bailey. Robust exchangeability designs for early phase clinical trials with multiple strata. *Pharmaceutical Statistics*, 15:123–134, 3 2016. ISSN 15391612. doi: 10.1002/pst.1730.

Beat Neuenschwander, Sebastian Weber, Heinz Schmidli, and Anthony O'Hagan. Predictively consistent prior effective sample sizes. *Biometrics*, 76:578–587, 6 2020. ISSN 15410420. doi: 10.1111/biom.13252.

Stuart J. Pocock. The combination of randomized and historical controls in clinical trials. *Journal of Chronic Diseases*, 29(3):175–188, 1976. ISSN 0021-9681. doi: https://doi.org/10.1016/0021-9681(76)90044-8. URL https://www.sciencedirect.com/science/article/pii/0021968176900448.

Matthew A Psioda, Mat Soukup, and Joseph G Ibrahim. A practical bayesian adaptive design incorporating data from historical controls. *Stat. Med.*, 37(27):4054–4070, November 2018.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL https://www.R-project.org/.

James Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7:1393–1512, 1986. ISSN 02700255. doi: 10.1016/0270-0255(86)90088-6.

Satrajit Roychoudhury and Beat Neuenschwander. Bayesian leveraging of historical control data for a clinical trial with time-to-event endpoint. *Statistics in Medicine*, 39:984–995, 3 2020. ISSN 10970258. doi: 10.1002/sim.8456.

Donald B. Rubin. The bayesian bootstrap. *The Annals of Statistics*, 9, 1 1981. ISSN 0090-5364. doi: 10.1214/aos/1176345338.

W. Sauerbrei and P. Royston. Building multivariable prognostic and diagnostic models: Transformation of the predictors by using fractional polynomials. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 162(1):71–94, 1999. ISSN 09641998, 1467985X. URL http://www.jstor.org/stable/2680468.

Darren Scott and Sophia Axillus. *BayesFBHborrow: Bayesian Dynamic Borrowing with Flexible Baseline Hazard Function*, 2024. URL https://CRAN.R-project.org/package=BayesFBHborrow. R package version 2.0.1.

Darren A. V. Scott and Alex Lewin. Borrowing from historical control data in a bayesian time-to-event model with flexible baseline hazard function. *Biostatistics*, 27(1):1–19, 2026. URL https://doi.org/10.1093/biostatistics/kxag006.

Mats Julius Stensrud, Morten Valberg, Kjetil Røysland, and Odd O. Aalen. Exploring selection bias by causal frailty models the magnitude matters. *Epidemiology*, 28:379–386, 5 2017. ISSN 15315487. doi: 10.1097/EDE.0000000000000621.

Liwen Su, Xin Chen, Jingyi Zhang, and Fangrong Yan. Comparative study of bayesian information borrowing methods in oncology clinical trials. *JCO Precision Oncology*, 6 (6):e2100394, 2022. doi: 10.1200/PO.21.00394. URL https://doi.org/10.1200/PO.21.00394. PMID: 35263169.

Sebastian Weber, Yue Li, John W. Seaman, Tomoyuki Kakizume, and Heinz Schmidli. Applying meta-analytic-predictive priors with the R Bayesian evidence synthesis tools. *Journal of Statistical Software*, 100(19):1–32, 2021. doi: 10.18637/jss.v100.i19.

James Willard, Shirin Golchi, and Erica E.M. Moodie. Covariate adjustment in bayesian adaptive randomized controlled trials. *Statistical Methods in Medical Research*, 33:480–497, 3 2024. ISSN 14770334. doi: 10.1177/09622802241227957.

A G-computation procedure

The following G-computation procedure, via a Bayesian bootstrap, is performed in the `BayesFBHborrow` package to obtain the marginal log-hazard ratio treatment estimate. After sampling $m = 1, ..., M$ Monte Carlo samples from the joint posterior distribution of the model parameters $\boldsymbol{\theta}|\boldsymbol{D}, \boldsymbol{D}_0$, two copies of the sampled covariate data $\boldsymbol{X}$ are created, where $Z_i = z$ defines the treatment allocation for all $i = 1, ..., n$. For each $\boldsymbol{\theta}_s$ and treatment allocation we calculate the posterior marginal hazard;

1. For each $i = 1, ..., n$ the conditional survival probabilities are calculated at time t corresponding to the time from the start of the trial to the current analysis.
2. Sample $\boldsymbol{\pi}_m = (\pi_{m1}, ..., \pi_{mn})$ from the posterior distribution Dirichlet(1_n) for each retained iteration of the MCMC sampler m.
3. Average over these n values to marginalise with respect to the observed covariates,

$$S(t|Z = z, \boldsymbol{\theta}_m) \approx \sum_{i=1}^{n} \pi_{mi} \exp(-H_0(t))^{\exp\{\mathbf{x}_i'\boldsymbol{\beta}_m + z\varphi_m\}}$$

   where i in $\mathbf{x}_i'$ identifies the covariate row. This is on the survival probability not the hazard (as the hazard is not a probability density but a conditional rate).
4. Apply a $\log(-\log(\cdot))$ transformation to yield a single sample from the posterior distribution of $\log(h(t|Z = z))$.

M samples of the posterior distribution of the log-hazard ratio $\gamma(\theta)$ at time t are obtained from

$$\gamma(\theta)_m = \log(-\log(S(t|Z = 1, \boldsymbol{\theta}_m)) - \log(-\log(S(t|Z = 0, \boldsymbol{\theta}_m)).$$

Posterior summaries can be easily for the hazard ratio at any time point. Alternatively a posterior hazard ratio marginalised over time can be also estimated.

B Borrowing profiles

Following Section 2.4, the choice of the prior weight can be aided by using the borrowing profiles (12). This is a series of curves mapping the posterior weight as a function of the difference between the standardised log baseline hazards of the control and historical data, for various prior weights. They illustrate how sensitive the borrowing is to changes in these differences. As the data is standardised in the algorithm, we choose our commensurate prior hyperparameters as $a_\tau = c_\tau = 1$, $b_\tau = 0.001$ and $d_\tau = 5$. Our choice of 0.8 for the prior weight p_0 in the simulated data application, means our tolerable difference between the current and historical standardised log hazards of 0.29 (or a hazard ratio of 1.34). Beyond this value, the posterior weight is below 0.5 and smear mixture element is more likely. To be less conservative with borrowing, we can increase the prior weight. For our applied data example, we choose a tolerable difference between the current and historical standardised log hazards of 0.18 (or a hazard ratio of 1.20). This yields a prior weight, given our hyperparameter selection of $p_0 = 0.5$. Both choices our illustrated, on the set of borrowing profiles (for our hyperparameters).

C Additional Plots

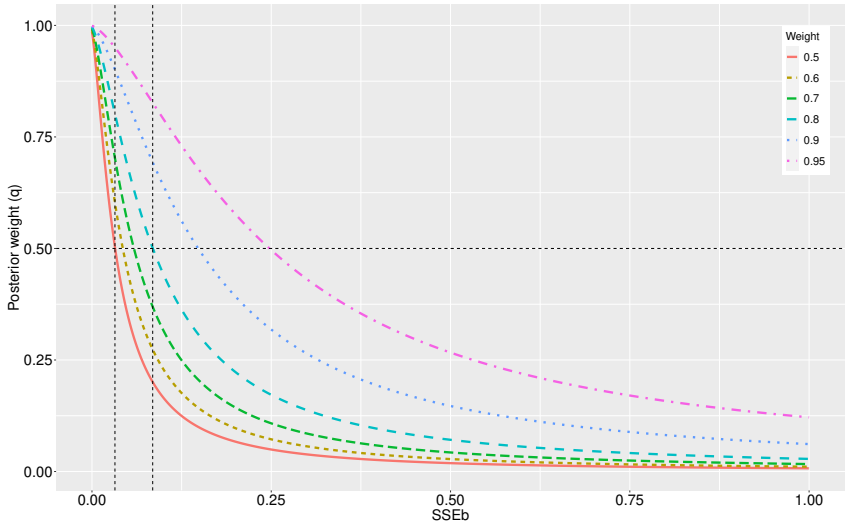


Figure 5: Borrowing profiles (posterior weight as a function of the sum of squared error (SSEb) of the standardised log baseline hazards) from $b_\tau = 0.001$ and $d_\tau = 5$, for various prior weights (p_0). From left to right, the first vertical dashed line at $\xi = 0.18^2$ represents our threshold of tolerable difference between the log hazards (SSEb, or ξ^2), leading to a prior weight of $p_0 = 0.5$. The second vertical dashed line of $\xi = 0.29^2$ is the threshold of tolerable difference from a prior weight of $p_0 = 0.8$. The horizontal dashed line shows where the posterior weight begins to fall below the tipping point of 0.5.

Figure 6: Estimated hazard from historical controls (left pane) and per arm hazard from current trial, no borrowing (right pane), simulated data

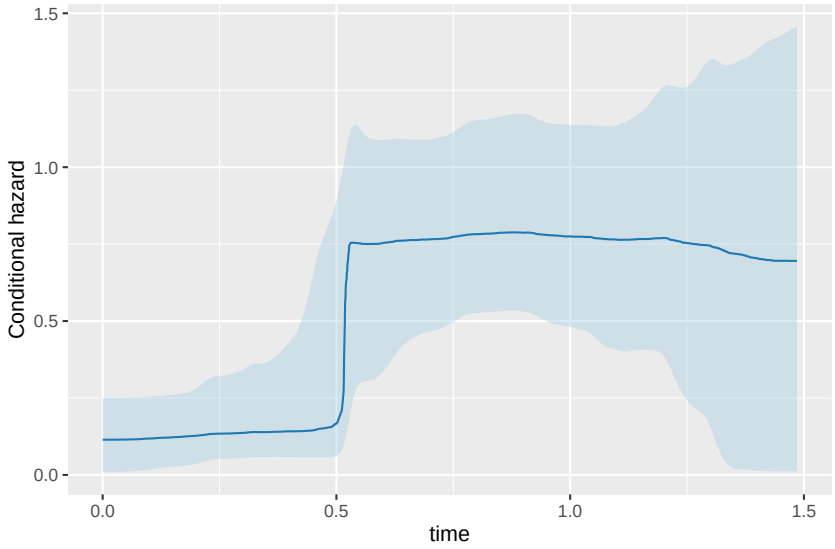
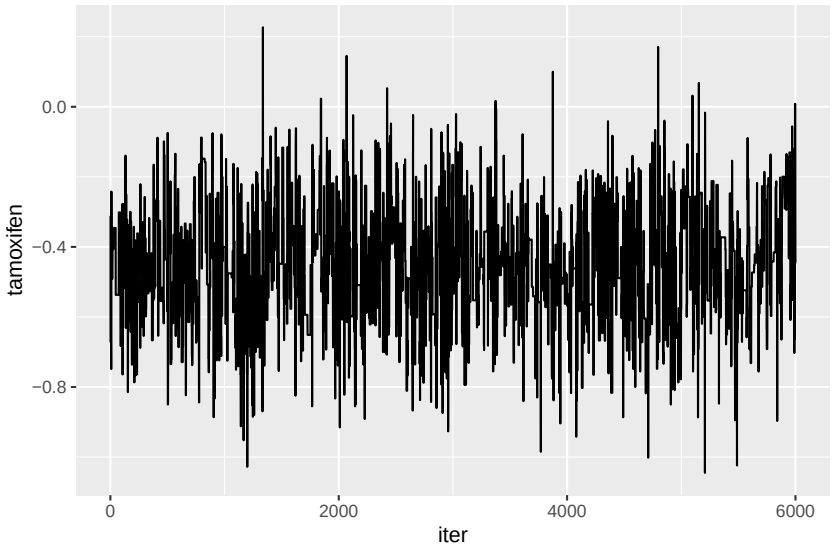


Figure 7: Trace plot for MCMC samples of the Conditional Treatment Effect of Tamoxifen, in model fit to current trial with borrowing on historical controls, German Breast Cancer Study



Affiliation:
Darren A. V. Scott
R&D
AstraZeneca
1 Francis Crick Avenue, Cambridge, Cambridgeshire, UK, CB2 0AA
E-mail: darren.scott@astrazeneca.com

Alex Lewin
Department of Medical Statistics
London School of Hygiene and Tropical Medicine
Keppel Street, London, UK, WC1E 7HT
E-mail: alex.lewin@lshtm.ac.uk

Sophia Axillus
Department of Mathematical Sciences
Chalmers University of Technology and University of Gothenburg
S-412 96 Gothenburg, Sweden
E-mail: axillus@chalmers.se

Grant Izmirlian
R&D
AstraZeneca
101 Orchard Ridge Drive, Gaithersburg, Maryland 20874, USA
E-mail: grant.izmirlian@astrazeneca.com