Supplement 1 for "A modular framework for early-phase seamless oncology trials"

S1 Theoretical introduction to Bayesian isotonic regression model

Let the binary variable Y_i, $i = 1, \ldots, n$ indicate dose-limiting toxicity (DLT), where $Y_i = 1$ indicates the occurrence of DLT and $Y_i = 0$ indicates no DLT. Let X_i indicate a categorical dose level taking one of K possible values. The parameter vector of interest is $\boldsymbol{\xi} = \{\xi_1, \ldots, \xi_K\}$ or, more compactly, $\{\xi_j\}_{j=1}^K$, where

$$\xi_j = \Pr(Y_i = 1 | X_i = \text{dose level } j). \tag{1}$$

A monotonic non-decreasing assumption implies that the set of increments $\{\xi_j - \xi_{j-1}\}_{j=1}^{K+1}$, where $\xi_0 \equiv 0$ and $\xi_{K+1} \equiv 1$, forms a simplex.

Now, let $\boldsymbol{\alpha} = \{\alpha_j\}_{j=1}^{K+1}$ be a set of non-negative parameters, with at least one being strictly positive, satisfying $\xi_j - \xi_{j-1} = \alpha_j / \sum_{k=1}^{K+1} \alpha_k$, $j = 1, \ldots, K$. This is equivalent to

$$\xi_j = \sum_{k=1}^{j} \alpha_k / \sum_{k=1}^{K+1} \alpha_k, \quad j = 1, \ldots, K. \tag{2}$$

The partial sum $\sum_{k=1}^{j} \alpha_k$ is non-decreasing in j, ensuring that ξ_j is similarly non-decreasing, as required.

The Bayesian isotonic regression option provided in modules 2 and 4 (`bayes_isoreg`) uses the model in (1) and (2) equipped with a prior on $\boldsymbol{\alpha}$ derived from the 'regularized horseshoe distribution' [1–5]. Specifically, let $\mathrm{N}^+(0, \sigma^2)$ be the half-normal distribution with standard deviation σ and its density function proportional to $\exp\{-x/(2\sigma^2)\} I(x \geq 0)$. Let $\mathrm{C}^+(0, 1)$

be the standard half-Cauchy distribution with the density function proportional to $(1 + x^2)^{-1} I(x \geq 0)$. Let c be a positive constant. For $j = 1, \ldots, K + 1$,

$$[\alpha_j \mid \tau, \lambda_j \,] \sim \mathrm{N}^{+}\left(0, \frac{1}{1 + 1/[c^2\tau^2\lambda_j^2]}\right), \quad \lambda_j \sim \mathrm{C}^{+}(0,1), \quad \tau \sim \mathrm{C}^{+}(0,1). \tag{3}$$

The half-horseshoe distribution serves as a hierarchical shrinkage prior, in which the global shrinkage parameter τ controls overall shrinkage to zero, and the local shrinkage parameters λ_j can be large to offset this overall shrinkage as warranted by the data. The value of c in (3) is a user-supplied hyperparameter and discussed in further detail below.

The constant value 1 that is added to $1/[c^2\tau^2\lambda_j^2]$ in the denominator of the half-normal standard deviation expression in (3) serves a two-fold purpose. First, it dominates the standard deviation expression when $c^2\tau^2\lambda_j^2$ is very large, making the standard deviation approximately equal to 1 in such cases and thinning the heavy tails that would otherwise result if no constant were added [5]. Second, adding a constant identifies the parameters c and τ. If this constant was not added, the standard deviation would reduce to $c\tau\lambda_j$, and the expression $c\tau$ would cancel out in the numerator and denominator of equation (2). Piironen and Vehtari [4] proposed a more general recipe by adding $1/d^2$ instead of a constant 1, with d either a fixed constant or a hyperparameter having a hyperprior of its own. However, in our unique extension of the horseshoe prior, d and c cannot both independently vary, again, due to relative nature of each α_k in equation (2). Thus, we set $d = 1$. Also different from previous versions of the horseshoe prior, our formulation places a **half**-normal prior on each α_j, with support only on the positive half of the real line, to match the context of the problem. To summarize, these modifications are unique to our application of the horseshoe prior to this problem and arise from the relative relationship of the scale parameters.

Our implementation of the horseshoe prior in equations (3) also has only one fixed value, namely c, that can be user-supplied or selected. The choice of c codifies an implicit assumption about the anticipated number of non-zero elements in $\boldsymbol{\alpha}$ [4], which, in our case, is the

number of non-zero jumps in the dose-toxicity curve, with larger values corresponding to anticipating more non-zero elements. The original horseshoe prior used $c = 1$, but Piironen and Vehtari [4] argue that for most applications $c \ll 1$ is a much more appropriate choice. In the simulation study, we always used $c = 10^{-7}$.

Figure 1 illustrates this Bayesian isotonic regression model fit against six different exemplar datasets. The black circles in each panel are the data. Specifically, they are $\sum_{i=1}^{n} (Y_i \times I(X_i = j)) \,/\, \sum_{i=1}^{n} I(X_i = j)$, i.e. the observed proportion of outcomes per dose level, plotted against j. Datasets in the left column have $n = 5$ observations per dose level, and datasets in the right column have $n = 25$. The types of dataset range proportions that increase at a constant rate across dose levels (top row), proportions that don't increase at all (middle row), and proportions that suddenly increase between dose levels 3 and 4 (bottom row). The solid line indicates the posterior medians of each ξ_j, which is defined in (1), and the transparent ribbon denotes the 80% pointwise credible intervals for each ξ_j.

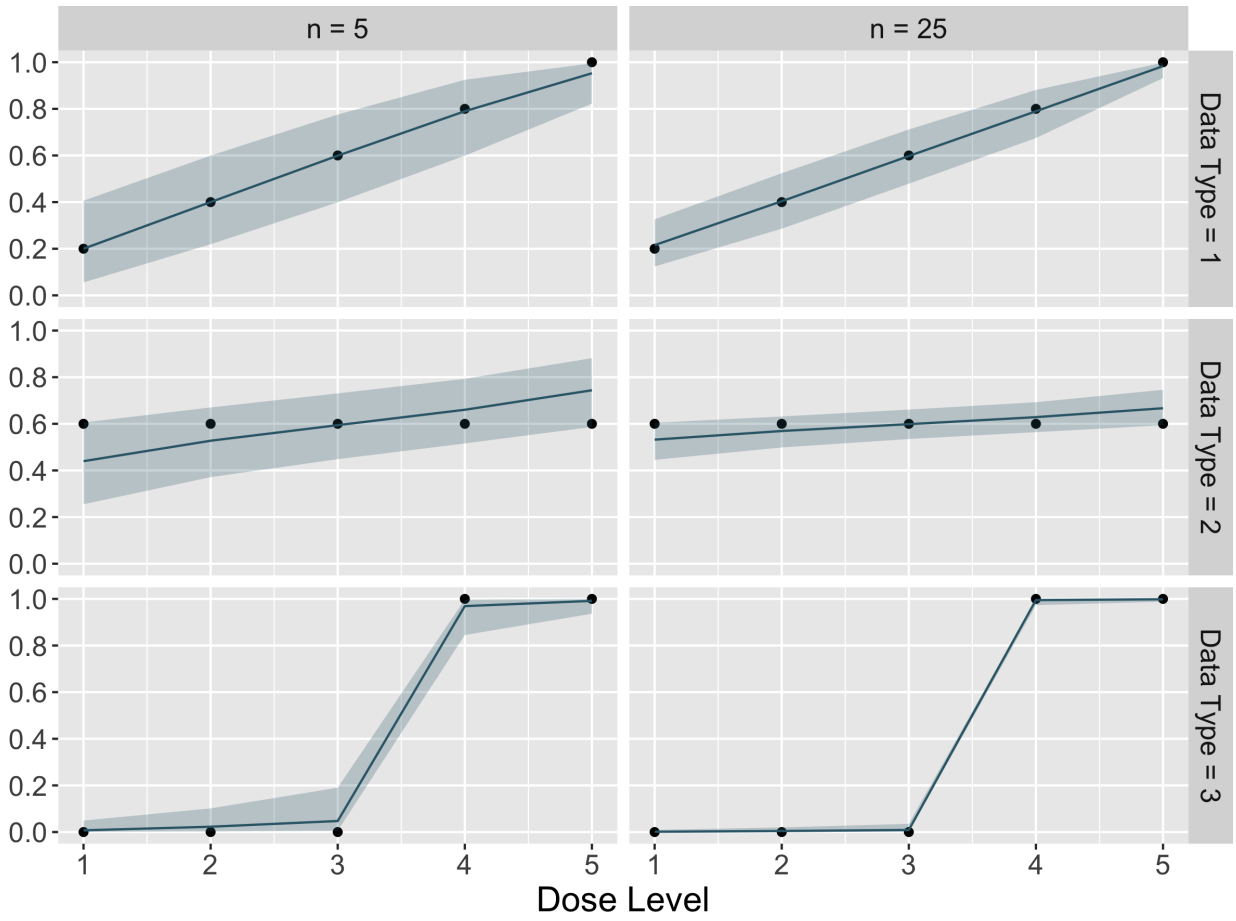

Figure S1: Comparison of fitted Bayesian isotonic regression curves and 80% posterior credible intervals (lines and shaded ribbons) and observed outcomes (single black circles) for six different exemplar datasets defined by sample size per dose level ($n = 5$ observations per dose level in left column; $n = 25$ observations per dose level in right column) and three types of data set (constant increase in top row; no increase in middle row; sudden increase in bottom row)

S2 Additional figure from trial simulator

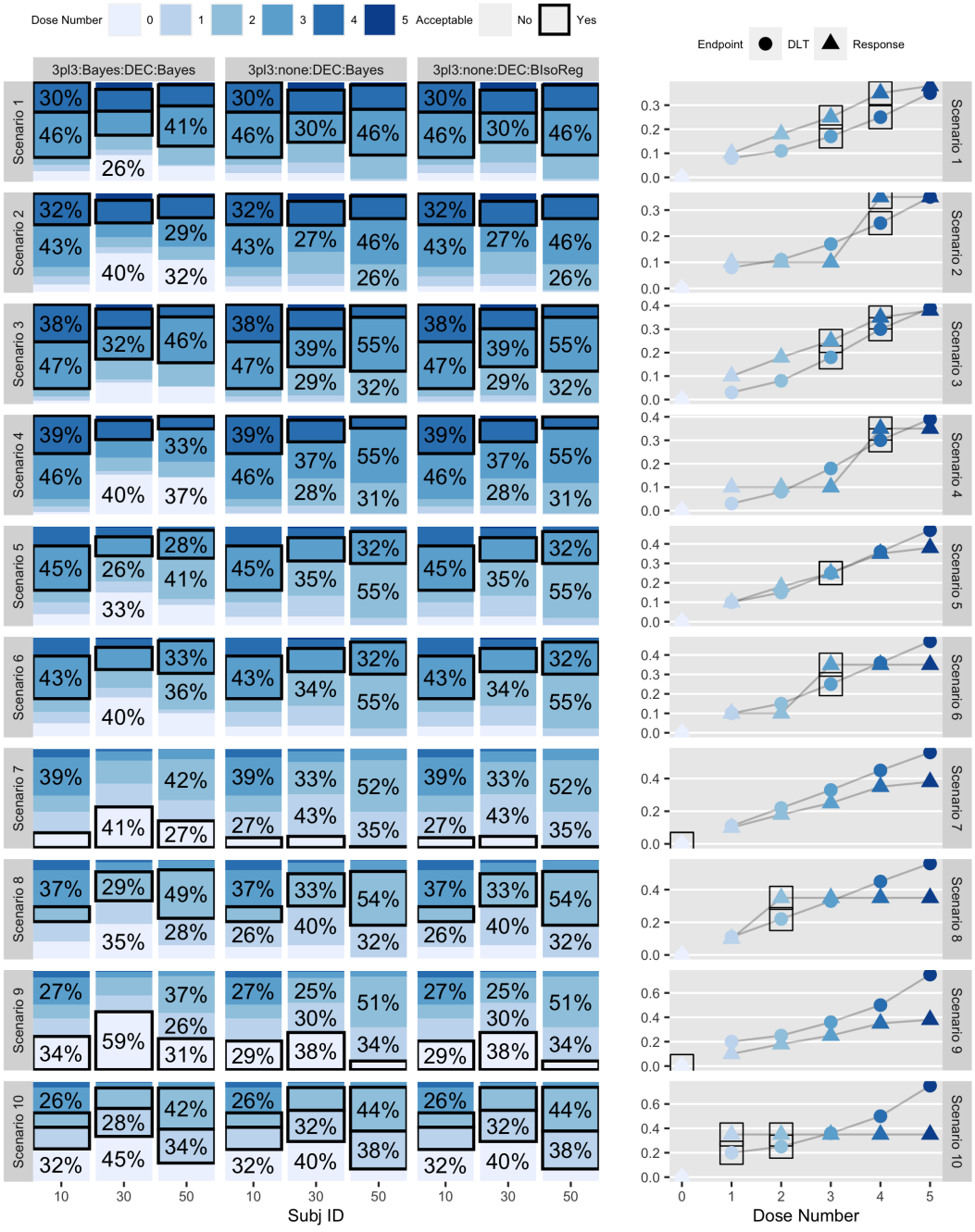

Figure S2: Distribution of dose assignments at subjects 10, 30, and 50 (based on 2000 simulations) for three 3+3-type designs (out of six total) across ten scenarios. The right-hand column gives the true generating toxicity and efficacy curves. Each row corresponds to a different scenario, and consecutive pairs of scenarios (1&2, 3&4, etc.) are linked in that they share a common dose-toxicity curve but differ in the dose-response curve. The proportion(s) corresponding to the preferred dose level are bordered by a solid box. If a trial has stopped for futility or safety, the patient was treated as having been assigned to dose level '0'. Figure 4 in the main manuscript gives the same results for the three CRM-type designs.

References

[1] Carlos M Carvalho, Nicholas G Polson, and James G Scott. Handling sparsity via the horseshoe. In David van Dyk and Max Welling, editors, Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics, volume 5 of Proceedings of Machine Learning Research, pages 73–80, Clearwater Beach, Florida USA, 2009. PMLR.

[2] CM Carvalho, NG Polson, and JG Scott. The horseshoe estimator for sparse signals. Biometrika, 97(2):465–480, 2010.

[3] J Piironen and A Vehtari. Projection predictive variable selection using Stan+R. Technical report, 2015. arXiv preprint arXiv:1508.02502.

[4] Juho Piironen and Aki Vehtari. On the hyperprior choice for the global shrinkage parameter in the horseshoe prior. In Aarti Singh and Jerry Zhu, editors, Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, volume 54 of Proceedings of Machine Learning Research, pages 905–913, Fort Lauderdale, FL, USA, 2017. PMLR.

[5] Juho Piironen and Aki Vehtari. Sparsity information and regularization in the horseshoe and other shrinkage priors. Electronic Journal of Statistics, 11(2):5018–5051, 2017.