ASIED: A Bayesian Adaptive Subgroup-Identification Enrichment Design (Supplementary Material)

A Split rules for random partition model

Below we describe the split rules for determining a partition of the biomarker space using the example of two rounds of splits and various types of covariates, including continuous, binary, categorical, and ordinal variables.

1. In the first split, we select biomarker k with probability ν_k, $k = 1, \ldots, K$, to split or not to split with probability ν_0, $\sum_{k=0}^{K} \nu_k = 1$. We assume $\nu_k = \frac{1}{K+1}$, indicating a uniform prior. Without loss of generality, we assume that $x_{ik} \in (-1, 1)$ if the biomarker k is continuous. Then we choose threshold c_k to split the biomarker space into two subgroups U_k and L_k. The prior of $p(c_k)$ differs depending on what type of variable the biomarker k is.

   (a) If biomarker k is binary, the split will be deterministic and we denote $U_k = \{i : x_{ik} = 0\}$ and $L_k = \{i : x_{ik} = 1\}$. Therefore $p(c_k) = 1$.

   (b) If biomarker k is continuous, denote $U_k = \{i : x_{ik} \leq c_k\}$ and $L_k = \{i : x_{ik} > c_k\}$. We assume $p(c_k) = \text{Uniform}(-1, 1) = 1/2$.

(c) If biomarker k is ordinal, let V_k denote the number of labels that biomarker k has. Let c_k denote the endpoint of the left partition, e.g., if $V_k = 5$ and $c_k = 3$, the left partition is $\{1, 2, 3\}$ and the right partition is $\{4, 5\}$. In this way we denote $U_k = \{i : x_{ik} \leq c_k\}$ and $L_k = \{i : x_{ik} > c_k\}$. Moreover, if $c_k = V_k$, it is equivalent to not splitting, which has been considered with probability ν_0. Therefore, $p(c_k) = \frac{1}{V_k - 1}$.

(d) If biomarker k is categorical, let V_k denote the number of categories corresponding to biomarker k. Let c_k denote the elements in one subset U_k. The remaining elements are stored in the other subset L_k. The c_k are elements of the powerset of $\{1, 2, \cdots, V_k\}$ without the empty-set or the full set. There are hence $2^{V_k} - 2$ options for c_k. Note that the choice of c_k is symmetric: we may flip c_k and its complement, leading to the same partition. Thus, $p(c_k) = \frac{2}{2^{V_k} - 2}$.

2. In the second split, we assume that the number of available biomarkers is $\tilde{K}$. By available, we mean that the split rules would not lead to empty subgroup. For example, if a binary biomarker was used in the first round of split, then it would no longer be available for splitting at the following rounds of split.

   In the subset U_k, we choose biomarker k_1 with probability $p(k_1)$ and threshold c_{k_1} to split the subgroup U_k into two subgroups UU_{kk_1} and UL_{kk_1}. In the subset L_k, we choose biomarker k_2 with probability $p(k_2)$ and threshold c_{k_2} to split the subgroup L_k into two subgroups LU_{kk_2} and LL_{kk_2}. Here, $p(k_1), p(k_2), p(c_{k_1})$ and $p(c_{k_2})$ differ depending on what type of variable the biomarker is and the values obtained for c_{k_1} and c_{k_2}.

   (a) If biomarker k_1 is continuous, $p(k_1) = \frac{1}{\tilde{K}+1}$. Denote $UU_{kk_1} = \{i : \boldsymbol{x}_i \in U_k \text{ and } x_{ik_1} \leq c_{k_1}\}$ and $UL_{kk_1} = \{i : \boldsymbol{x}_i \in U_k \text{ and } x_{ik_1} > c_{k_1}\}$. We have $p(c_{k_1}) = \frac{1}{2}$ if $k_1 \neq k$; $p(c_{k_1}) = \frac{1}{c_k + 1}$ if $k_1 = k$.

   If biomarker k_2 is continuous, $p(k_2) = \frac{1}{\tilde{K}+1}$. Denote $LU_{kk_2} = \{i : \boldsymbol{x}_i \in L_k \text{ and } x_{ik_2} \leq$

$c_{k_2}\}$ and $LL_{kk_2} = \{i : \boldsymbol{x}_i \in L_k \text{ and } x_{ik_2} > c_{k_2}\}$. We have $p(c_{k_2}) = \frac{1}{2}$ if $k_2 \neq k$; $p(c_{k_2}) = \frac{1}{1-c_k}$ if $k_2 = k$.

(b) If k_1 is binary and $k_1 = k$, then biomarker k_1 is not an available biomarker to choose to split U_k. If $k_1 \neq k$, denote $UU_{kk_1} = \{i : \boldsymbol{x}_i \in U_k, \text{ and } x_{ik_1} = 0\}$ and $UL_{kk_1} = \{i : \boldsymbol{x}_i \in U_k, \text{ and } x_{ik_1} = 1\}$ with $p(c_{k_1}) = 1$.

If k_2 is binary, and $k_2 = k$, then biomarker k_2 is not an available biomarker to choose to split L_k. If $k_2 \neq k$, denote $LU_{kk_2} = \{i : \boldsymbol{x}_i \in L_k, \text{ and } x_{ik_2} = 0\}$ and $LL_{kk_2} = \{i : \boldsymbol{x}_i \in L_k, \text{ and } x_{ik_2} = 1\}$ with $p(c_{k_2}) = 1$.

(c) If k_1 is ordinal, we let c_{k_1} denote the left endpoint of the second split *within* the left partition, and if c_{k_2} is ordinal, we let c_{k_2} analogously denote the left endpoint *within* the right partition. We have $p(c_{k_1}) = \frac{1}{V_{k_1}-1}$ if $k_1 \neq k$; $p(c_{k_1}) = \frac{1}{c_k}$ if $k_1 = k$ and $c_k > 1$; k_1 is not an available biomarker to choose to split U_k if $k_1 = k$ and $c_k = 1$. Also, $p(c_{k_2}) = \frac{1}{V_{k_2}-1}$ if $k_2 \neq k$; $p(c_{k_2}) = \frac{1}{V_k - c_k}$ if $k_2 = k$ and $(V_k - c_k) > 1$; k_2 is not an available biomarker to choose to split L_k if $k_2 = k$ and $V_k - c_k = 1$.

(d) If k_1 is categorical, we let c_{k_1} denote one subset of the split *within* the subset U_k. If c_{k_2} is categorical, we let c_{k_2} denote one subset of the split *within* the subset L_k. We have $p(c_{k_1}) = \frac{2}{2^{V_{k_1}}-2}$ if $k_1 \neq k$; $p(c_{k_1}) = \frac{2}{2^{|c_k|}-2}$ if $k_1 = k$ and $|c_k| > 1$; k_1 is not an available biomarker to choose to split U_k if $k_1 = k$ and $|c_k| = 1$. Also, $p(c_{k_2}) = \frac{2}{2^{V_{k_2}}-2}$ if $k_2 \neq k$; $p(c_{k_2}) = \frac{2}{2^{V_k - |c_k|}-2}$ if $k_2 = k$ and $(V_k - |c_k|) > 1$; k_2 is not an available biomarker to choose to split L_k if $k_2 = k$ and $(V_k - |c_k|) = 1$.

B Detailed MCMC derivations

Here we describe the detailed MCMC steps for binary outcomes and continuous outcomes. The MCMC derivations for categorical outcomes are similar to binary outcomes. Again, we take two rounds of splits for example. Assume we select biomarker k with probability $p(k)$

and the threshold c_k in the first round, leading to two subsets U_k and L_k if we choose to split. Then we select biomarker k_1 with threshold c_{k_1} to split U_k and biomarker k_2 with threshold c_{k_2} to split L_k.

B.1 Sampling model with priors

We first review the sampling models.

Binary outcomes. Let $y_i \in \{0, 1\}$ and $\theta_{t,m}$ be the response rate of patients in subgroup m under treatment t. In this case, $\Theta = \{\theta_{t,m}\}$. We assume

$$p(y_i = 1 \mid z_i = t, \Pi, \boldsymbol{x}_i \in S_m) = \theta_{t,m}.$$

The likelihood function is simply the product of n Bernoulli probability mass functions. We assign the prior $\theta_{t,m} \mid \Pi \overset{iid}{\sim} Beta(a, b)$, where $Beta(a, b)$ denotes a beta distribution with mean $a/(a + b)$.

Continuous outcomes. Let $y_i \in R$ and $\theta_{t,m}$ be the mean response of patients in subgroup m under treatment t. We assume

$$p(y_i \mid z_i = t, \Pi, \boldsymbol{x}_i \in S_m) = N(\theta_{t,m}, \sigma^2).$$

The likelihood can be written as follows:

$$p(\boldsymbol{Y}_n \mid \boldsymbol{X}_n, \boldsymbol{Z}_n, \Theta, \Pi) = \prod_{t=1}^{T} \prod_{m=1}^{M} \prod_{\{i: z_i = t, \boldsymbol{x}_i \in S_m\}} (2\pi\,\sigma^2)^{-1/2} \exp\{-\frac{1}{2\sigma^2}(y_i - \theta_{t,m})^2\}. \tag{B.1}$$

We assign the conjugate prior $p(\theta_{t,m}, \sigma^2) = p(\theta_{t,m}|\sigma^2)p(\sigma^2)$ with $p(\theta_{t,m}|\sigma^2) = N(\theta_0, \frac{\sigma^2}{\kappa_0})$ and $p(\sigma^2) = IG(\frac{\nu_0}{2}, \frac{SS_0^2}{2})$, where $SS_0^2 = \nu_0\sigma_0^2$.

B.2 MCMC steps

1. Update $\theta_{t,m}$ (for both binary and continuous outcomes) and σ^2 (for continuous outcome only).

   - Binary outcome:

   $$p(\theta_{t,m} \mid \cdot) \propto p(\boldsymbol{Y}_n \mid \boldsymbol{X}_n, \boldsymbol{Z}_n, \boldsymbol{\theta}, \Pi)\, p(\boldsymbol{\theta} \mid \Pi) \sim \text{beta}(n_{tm1} + a, n_{tm0} + b),$$

   where $n_{tmy} = \sum_i I(\boldsymbol{x}_i \in S_m, z_i = t, y_i = y)$, $y = 0, 1$.

   - Continuous outcome:

   $$p(\theta_{t,m}|\boldsymbol{Y}_n, \sigma^2, \boldsymbol{X}_n, \Pi) \quad \propto \quad p(\mathbf{Y}_n|\theta_{t,m}, \sigma^2, \Pi)\, p(\theta_{t,m}) \sim N(\frac{\sum y_i + \kappa_0\theta_0}{n + \kappa_0}, \frac{\sigma^2}{n + \kappa_0})$$

   $$p(\sigma^2 \mid \boldsymbol{Y}_n, \boldsymbol{\theta}, \boldsymbol{X}_n, \Pi) \propto p(\mathbf{Y}_n \mid \boldsymbol{\theta}, \sigma^2, \Pi)\, p(\sigma^2) \sim IG(\frac{n + \nu_0}{2}, \frac{SS_0^2 + \sum_{t,m} \sum_{\boldsymbol{x}_i \in S_m} (y_i - \theta_{t,m})^2}{2})$$

2. Keep the selected biomarkers k, k_1, k_2 fixed, and update the thresholds c_k, c_{k_1}, c_{k_2}.

   Assume the current value is $\beta_k = (c_k, c_{k_1}, c_{k_2})$ and the new value $\beta_k^* = (c_k^*, c_{k_1}^*, c_{k_2}^*)$ is generated from a proposal distribution $q(\beta_k^* \mid \beta_k)$. Denote the partition determined by $k, k_1, k_2, c_k^*, c_{k_1}^*, c_{k_2}^*$ to be Π^*. The acceptance ratio r is

   $$\begin{aligned} r &= \frac{p(\boldsymbol{Y}_n \mid \boldsymbol{X}_n, \boldsymbol{Z}_n, \boldsymbol{\theta}, \Pi^*)\, p(\boldsymbol{\theta} \mid \Pi^*)\, p(\Pi^* \mid k, k_1, k_2, \beta_k^*)\, p(\beta_k^*)}{p(\boldsymbol{Y}_n \mid \boldsymbol{X}_n, \boldsymbol{Z}_n, \boldsymbol{\theta}, \Pi)\, p(\boldsymbol{\theta} \mid \Pi)\, p(\Pi \mid k, k_1, k_2, \beta_k)\, p(\beta_k)} \cdot \frac{q(\beta_k \mid \beta_k^*)}{q(\beta_k^* \mid \beta_k)} \\ &= \frac{p(\boldsymbol{Y}_n \mid \boldsymbol{X}_n, \boldsymbol{Z}_n, \boldsymbol{\theta}, \Pi^*)\, p(\boldsymbol{\theta} \mid \Pi^*)\, p(c_k^* \mid k)\, p(c_{k_1}^* \mid k, c_k^*, k_1)\, p(c_{k_2}^* \mid k, c_k^*, k_2)}{p(\boldsymbol{Y}_n \mid \boldsymbol{X}_n, \boldsymbol{Z}_n, \boldsymbol{\theta}, \Pi)\, p(\boldsymbol{\theta} \mid \Pi)\, p(c_k \mid k)\, p(c_{k_1} \mid k, c_k, k_1)\, p(c_{k_2} \mid k, c_k, k_2)} \cdot \frac{q(\beta_k \mid \beta_k^*)}{q(\beta_k^* \mid \beta_k)} \end{aligned} \tag{B.2}$$

3. Update both the selected biomarkers in two rounds of splits k, k_1, k_2, and the corre-

sponding thresholds c_k, c_{k_1}, c_{k_2}. Note here we allow not to split with certain probability in each split.

Due to the possibility of not splitting, updating k, k_1, k_2 leads to potential dimension change due to the change of the number of subgroups. For computational efficiency and simplicity, in Metropolis-Hasting step we choose a specific proposal distribution that is the same as the prior, to generate the new value. That means, we sample the new k^*, k_1^*, k_2^* and the corresponding $\beta_{k^*} = (c_{k^*}, c_{k_1^*}, c_{k_2^*})$ from the prior as the proposed value, then determine if we accept it or not.

Denote the partition determined by $k, k_1, k_2, c_k^*, c_{k_1}^*, c_{k_2}^*$ to be Π^*. Then the acceptance ratio r is reduced to the following form:

$$\begin{aligned} r &= \frac{p(\boldsymbol{Y}_n \mid \boldsymbol{X}_n, \boldsymbol{Z}_n, \boldsymbol{\theta}, \Pi^*)\, p(\boldsymbol{\theta} \mid \Pi^*)\, p(\Pi^*)}{p(\boldsymbol{Y}_n \mid \boldsymbol{X}_n, \boldsymbol{Z}_n, \boldsymbol{\theta}, \Pi)\, p(\boldsymbol{\theta} \mid \Pi)\, p(\Pi)} \cdot \frac{q(\Pi \mid \Pi^*)}{q(\Pi^* \mid \Pi)} \\ &= \frac{p(\boldsymbol{Y}_n \mid \boldsymbol{X}_n, \boldsymbol{Z}_n, \boldsymbol{\theta}, \Pi^*)\, p(\boldsymbol{\theta} \mid \Pi^*)\, p(\Pi^*)}{p(\boldsymbol{Y}_n \mid \boldsymbol{X}_n, \boldsymbol{Z}_n, \boldsymbol{\theta}, \Pi)\, p(\boldsymbol{\theta} \mid \Pi)\, p(\Pi)} \cdot \frac{p(\Pi)}{p(\Pi^*)} \\ &= \frac{p(\boldsymbol{Y}_n \mid \boldsymbol{X}_n, \boldsymbol{Z}_n, \boldsymbol{\theta}, \Pi^*)\, p(\boldsymbol{\theta} \mid \Pi^*)}{p(\boldsymbol{Y}_n \mid \boldsymbol{X}_n, \boldsymbol{Z}_n, \boldsymbol{\theta}, \Pi)\, p(\boldsymbol{\theta} \mid \Pi)} \end{aligned} \tag{B.3}$$

Both (B.2) and (B.3) can be computed for binary outcome and continuous outcome by plugging the sampling model and the priors of binary outcome and continuous outcome, respectively.

C Determining ξ_1 and ξ_2

Table S1: Sensitivity analysis of the proposed ASIED design with respect to ξ_1 and ξ_2 in the first interim analysis ($n = 100$), under different β_0 values. $Pr(All)$ denotes the probability of continuing the trial with original population until the end of the trial. $Pr(Sub)$ denotes the probability of continuing the trial with an enriched subpopulation until the end of the trial. Pr(EarS) denotes the probability of stopping the trial due to futility. Pr(2All) denotes the probability of continuing the trial with original population and plan a second interim analysis. Pr(2Sub) denotes the probability of continuing the trial with an enriched subpopulation and plan a second interim analysis. All probabilities are with respect to repeated simulations.

$\xi_2 = 0.05$	$\beta_0 = 0.25$ $\beta_1 = 2$ No Effect	$\beta_0 = 0.25$ $\beta_1 = 2.55$ Sub LRV	$\beta_0 = 0.25$ $\beta_1 = 2.83$ Sub TV	$\beta_0 = 2.6$ $\beta_1 = 0$ All LRV	$\beta_0 = 3.08$ $\beta_1 = 0$ All TV
$\xi_1 = 0.80$					
$Pr(All)$	0	0	10	95	100
$Pr(Sub)$	1	96	90	0	0
Pr(EarS)	99	4	0	5	0
Pr(2All)	0	0	0	0	0
Pr(2Sub)	0	0	0	0	0
$\xi_1 = 0.85$					
$Pr(All)$	0	0	8	95	100
$Pr(Sub)$	0	92	91	0	0
Pr(EarS)	100	7	0	5	0
Pr(2All)	0	0	0	0	0
Pr(2Sub)	0	1	1	0	0
$\xi_1 = 0.90$					
$Pr(All)$	0	0	6	92	100
$Pr(Sub)$	0	88	90	0	0
Pr(EarS)	100	7	0	5	0
Pr(2All)	0	0	0	3	0
Pr(2Sub)	0	5	4	0	0

$\xi_2 = 0.10$	$\beta_0 = 0.25$	$\beta_0 = 0.25$	$\beta_0 = 0.25$	$\beta_0 = 2.6$	$\beta_0 = 3.08$
	$\beta_1 = 2$	$\beta_1 = 2.55$	$\beta_1 = 2.83$	$\beta_1 = 0$	$\beta_1 = 0$
	No Effect	Sub LRV	Sub TV	All LRV	All TV
$\xi_1 = 0.80$					
$Pr(All)$	0	0	10	95	100
$Pr(Sub)$	1	96	90	0	0
Pr(EarS)	99	4	0	5	0
Pr(2All)	0	0	0	0	0
Pr(2Sub)	0	0	0	0	0
$\xi_1 = 0.85$					
$Pr(All)$	0	0	9	95	100
$Pr(Sub)$	0	92	91	0	0
Pr(EarS)	100	8	0	5	0
Pr(2All)	0	0	0	0	0
Pr(2Sub)	0	0	0	0	0
$\xi_1 = 0.90$					
$Pr(All)$	0	0	8	95	100
$Pr(Sub)$	0	89	91	0	0
Pr(EarS)	100	11	0	5	0
Pr(2All)	0	0	0	0	0
Pr(2Sub)	0	0	1	0	0

Since we set the three risks as false stop risk=0.05, false go risk=0.1, and false enrich risk=0.15, both $\xi_1 = 0.8$, $\xi_2 = 0.05$ and $\xi_1 = 0.8$, $\xi_2 = 0.1$ achieve the desirable results. Usually pharmaceutical companies wants to be more stringent on false go in early phase drug development, we determine $\xi_2 = 0.1$.

D Additional Simulation Study

In this section, we considered one more scenario in which there is no predictive biomarker and one prognostic biomarker to evaluate the performance of the proposed Bayesian random partition (BayRP) model . We assumed all the biomarkers were continuous and generated x_{ik} from Uniform$(-1, 1)$, $i = 1, \ldots, n$ and $k = 1, \ldots, 4$. In this scenario, we assumed only the first biomarker was the prognostic biomarker and generated $y_i = 0.75 + 3.25I(z_i = 2) + 0.25I(x_{i1} > -0.4) + \epsilon_i$, where $\epsilon_i \sim N(0, 1)$. Remember that we defined the true effective subgroup as $S^o = \big\{i : [E(y_i \mid z_i = 2, \boldsymbol{x}_i) - E(y_i \mid z_i = 1, \boldsymbol{x}_i)] > \text{LRV}\big\}$, where LRV=2.37. In this scenario, the simulated true effective subgroup includes all-comers.

We simulated 100 trials for this scenario and applied the BayRP to each simulated dataset. BayRP successfully identified the all-comers as the effective group, and yielded the true positive rate being 0.96. Note that we do not report the true negative rate since the simulated true effective subgroup include all-comers, meaning that the denominator in the definition of the true negative rate is 0.