# Bayesian basket trial designs for borrowing of information across similar subpopulations

**Haiyan Zheng***, James M.S. Wason

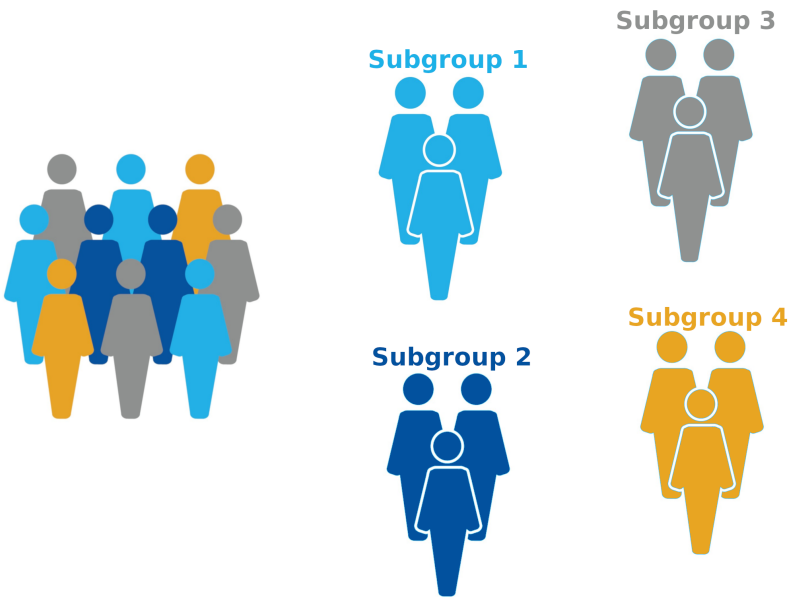
Biostatistics Research Group, Institute of Health and Society, Newcastle University

*Email: haiyan.zheng@newcastle.ac.uk

Bayes Pharma – Applied Bayesian Biostatistics Workshop
Lyon, 24th May 2019

# Clinical trials in precision medicine

⋆ **Q**: Does the drug work better than a placebo or the standard-of-care?



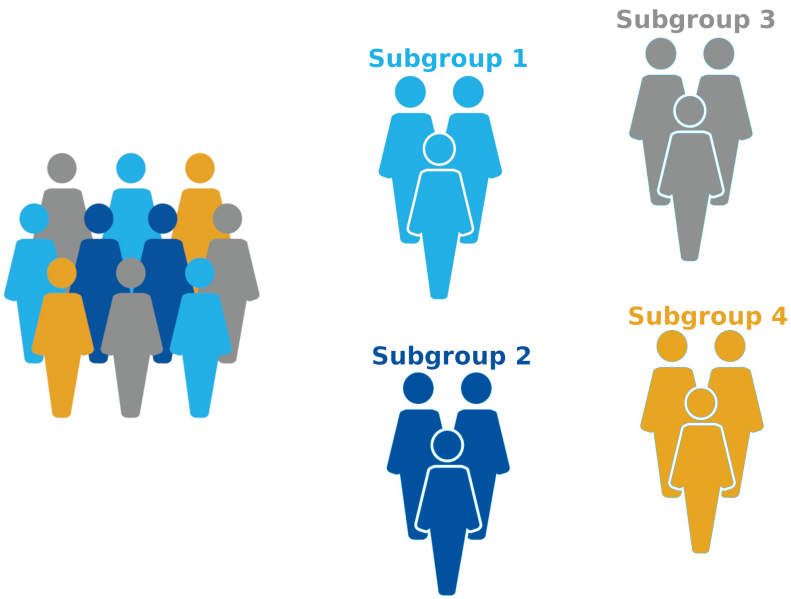

⋆ Patients may be stratified into subgroups

⋆ Population-averaged effect is less interested

⋆ Which subgroups will benefit & to what extent?

# Clinical trials in precision medicine

⋆ **Q**: Does the drug work better than a placebo or the standard-of-care?




⋆ Patients may be stratified into subgroups

⋆ Population-averaged effect is less interested

⋆ Which subgroups will benefit & to what extent?

Innovations in trial design methodology have followed.

Recent proposals include platform trials, umbrella trials, **basket trials**, etc.

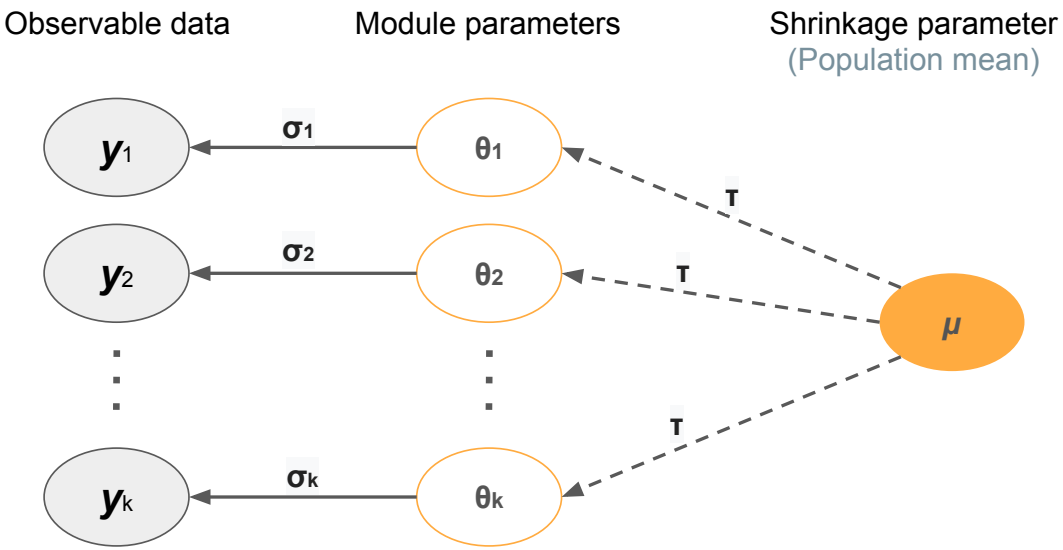
# Motivation

**Basket trials:** To assess the clinical efficacy of a new medicine to patients **harbouring a common disease trait** yet **presenting various condition subtypes** in one trial.

**Dilemma in the analysis strategy:**

- ⋆ Undesirable to pool trial data from different modules for analysis
- ⋆ Implementing subgroup-specific analysis would lead to low-powered tests
- ⋆ **Expectation**: to permit sharing of information between **similar** subgroups

# Hierarchical modelling may be limited!

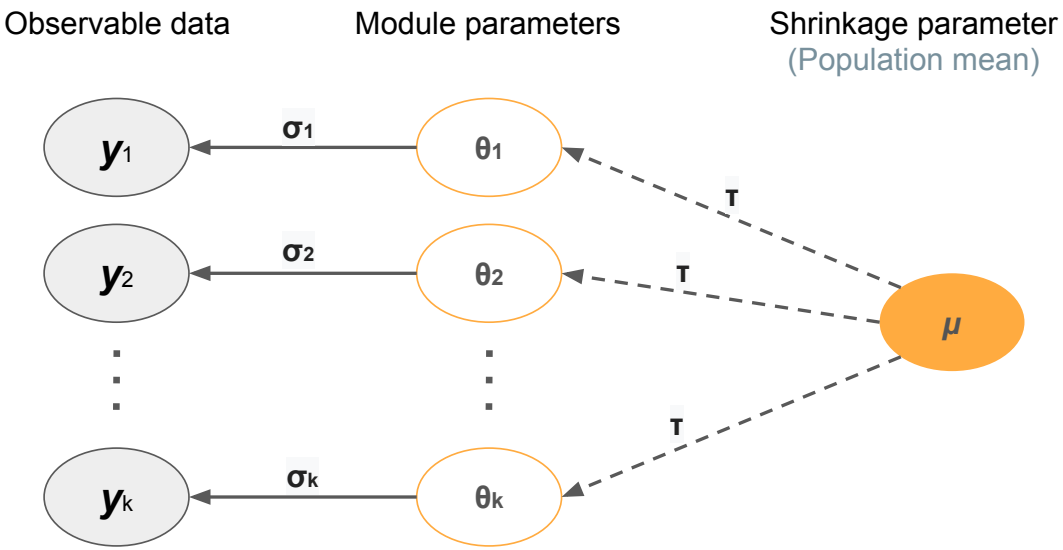



A normal-normal hierarchical model: for $i = 1, \ldots, k$,

$$y_i|\theta_i, \sigma_i \sim N(\theta_i, \sigma_i^2)$$

$$\theta_i|\mu, \tau \sim N(\mu, \tau^2)$$

# Hierarchical modelling may be limited!



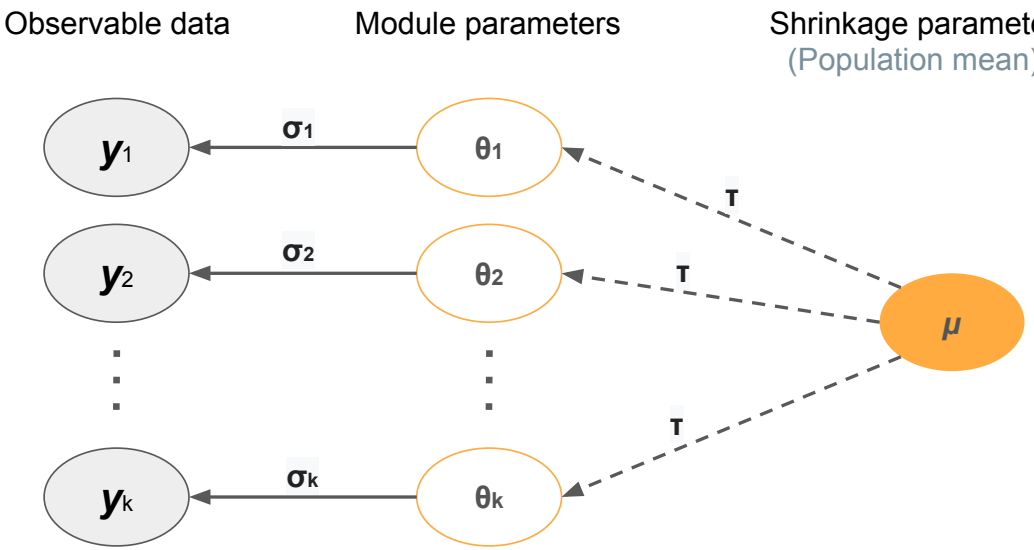

A normal-normal hierarchical model: for $i = 1, \ldots, k$,

$$y_i|\theta_i, \sigma_i \sim N(\theta_i, \sigma_i^2)$$

$$\theta_i|\mu, \tau \sim N(\mu, \tau^2)$$

- ⋆ Hierarchical modelling assumes **exchangeability** (similarity) of $\theta_i$s
- ⋆ The degree of borrowing is determined by $\tau$:
  $\tau = 0 \rightarrow$ complete pooling of data from other modules;
  $\tau = \infty \rightarrow$ no borrowing.
- ⋆ Very restrictive to suppose all $\theta_i$s will be shrunk towards one population mean

# Hierarchical modelling may be limited!




A normal-normal hierarchical model: for $i = 1, \ldots, k$,

$$y_i|\theta_i, \sigma_i \sim N(\theta_i, \sigma_i^2)$$

$$\theta_i|\mu, \tau \sim N(\mu, \tau^2)$$

- ⋆ Hierarchical modelling assumes **exchangeability** (similarity) of $\theta_i$s
- ⋆ The degree of borrowing is determined by $\tau$:
  $\tau = 0 \rightarrow$ complete pooling of data from other modules;
  $\tau = \infty \rightarrow$ no borrowing.
- ⋆ Very restrictive to suppose all $\theta_i$s will be shrunk towards one population mean
- ⋆ **More efficient analysis methodology for basket trials needed!**

# Leverage data from an external module

Let us start with a very simple scenario:

- ⋆ Incorporate **one external dataset** into the contemporary study
- ⋆ Corresponding to the setting of a basket trial, where there are $K = 2$ modules

# Leverage data from an external module

Let us start with a very simple scenario:

- ⋆ Incorporate **one external dataset** into the contemporary study
- ⋆ Corresponding to the setting of a basket trial, where there are $K = 2$ modules
- ⋆ **Idea:** represent the external information in a **prior** for the contemporary parameter

# Leverage data from an external module

Let us start with a very simple scenario:

- ⋆ Incorporate **one external dataset** into the contemporary study
- ⋆ Corresponding to the setting of a basket trial, where there are $K = 2$ modules
- ⋆ **Idea:** represent the external information in a **prior** for the contemporary parameter

*Notations*.

- Let $\theta_E$ and $\theta_C$ be the parameters that underpin the external and contemporary studies, respectively
- $\theta_E, \theta_C \in \mathbb{R}$; both are continuous location parameters

# Leverage data from an external module

Let us start with a very simple scenario:

- ⋆ Incorporate **one external dataset** into the contemporary study
- ⋆ Corresponding to the setting of a basket trial, where there are $K = 2$ modules
- ⋆ **Idea:** represent the external information in a **prior** for the contemporary parameter

*Notations*.

- Let $\theta_E$ and $\theta_C$ be the parameters that underpin the external and contemporary studies, respectively
- $\theta_E, \theta_C \in \mathbb{R}$; both are continuous location parameters

We introduce a new parameter $\nu_{EC}$ to describe the **commensurability** between $\theta_E$ and $\theta_C$, and further stipulate

- ⋆ An initial diffuse prior, $\pi_0(\theta_E)$
- ⋆ Construct a **conditional prior** opinion on $\theta_C$ as:

$$\theta_C|\theta_E, \nu_{EC} \sim N(\theta_E, 1/\nu_{EC}^2)$$

- ⋆ **Commensurate predictive prior (CPP):**

$$\pi^{\text{CPP}}(\theta_C, \nu_{EC}|\boldsymbol{y}_E, \theta_E) \propto \mathcal{L}(\boldsymbol{y}_E|\theta_E)\pi_0(\theta_E) \times \nu_{EC} \cdot \Phi((\theta_C - \theta_E) \cdot \nu_{EC})\pi(\nu_{EC})$$

# The commensurate parameter, $\nu_{EC}$

The **commensurate predictive prior (CPP)**

$$\pi^{\text{CPP}}(\theta_C, \nu_{EC} | \boldsymbol{y}_E, \theta_E) \propto \mathcal{L}(\boldsymbol{y}_E | \theta_E) \times \nu_{EC} \cdot \Phi((\theta_C - \theta_E) \cdot \nu_{EC}) \pi_0(\theta_E) \pi(\nu_{EC})$$

relies on the commensurate parameter $\nu_{EC}$.

We consider a "spike-and-slab" prior for $\nu_{EC}$ that

$$\begin{aligned} \mathbb{P}(\nu_{EC} < \mathcal{B}_1) &= 0, \\ \mathbb{P}(\nu_{EC} < u) &= \omega_{EC} \cdot \frac{u - \mathcal{B}_1}{\mathcal{B}_2 - \mathcal{B}_1}, \quad \mathcal{B}_1 \leq u \leq \mathcal{B}_2, \\ \text{and } \mathbb{P}(\nu_{EC} > \mathcal{B}_2) &= \mathbb{P}(\nu_{EC} = \mathcal{S}) = 1 - \omega_{EC}, \end{aligned}$$

where $\omega_{EC}$ is the prior probability that $\mathcal{B}_1 \leq u \leq \mathcal{B}_2$.

# The commensurate parameter, $\nu_{EC}$

The **commensurate predictive prior (CPP)**

$$\pi^{\text{CPP}}(\theta_C, \nu_{EC} | \boldsymbol{y}_E, \theta_E) \propto \mathcal{L}(\boldsymbol{y}_E | \theta_E) \times \nu_{EC} \cdot \Phi((\theta_C - \theta_E) \cdot \nu_{EC}) \pi_0(\theta_E) \pi(\nu_{EC})$$

relies on the commensurate parameter $\nu_{EC}$.

We consider a "spike-and-slab" prior for $\nu_{EC}$ that

$$\begin{aligned} \mathbb{P}(\nu_{EC} < \mathcal{B}_1) &= 0, \\ \mathbb{P}(\nu_{EC} < u) &= \omega_{EC} \cdot \frac{u - \mathcal{B}_1}{\mathcal{B}_2 - \mathcal{B}_1}, \quad \mathcal{B}_1 \leq u \leq \mathcal{B}_2, \\ \text{and } \mathbb{P}(\nu_{EC} > \mathcal{B}_2) &= \mathbb{P}(\nu_{EC} = \mathcal{S}) = 1 - \omega_{EC}, \end{aligned}$$

where $\omega_{EC}$ is the prior probability that $\mathcal{B}_1 \leq u \leq \mathcal{B}_2$.

- ⋆ The prior probability $1 - \omega_{EC} \Leftrightarrow$ borrowing of similar external information
- ⋆ Idea: link $\omega_{EC}$ with a **discrepancy measure**, for example, the Hellinger distance

# Characterising the commensurability of information

Hellinger distance:

$$d(\pi_1(\theta_C|\boldsymbol{y}_C), \pi_2(\theta_E|\boldsymbol{y}_E)) = \sqrt{\frac{1}{2}\int_{-\infty}^{\infty}\left(\sqrt{\frac{\mathrm{d}\pi_1(\theta_C|\boldsymbol{y}_C)}{\mathrm{d}\theta}} - \sqrt{\frac{\mathrm{d}\pi_2(\theta_E|\boldsymbol{y}_E)}{\mathrm{d}\theta}}\right)^2 \mathrm{d}\theta},$$

where $\pi_1(\theta_C|\boldsymbol{y}_C)$ & $\pi_2(\theta_E|\boldsymbol{y}_E)$ are updated from the initial diffuse priors, $\pi_0(\theta_C)$ & $\pi_0(\theta_E)$

Following the Cauchy–Schwarz inequality, $0 \leq d(\cdot,\cdot) \leq 1$.

Define

$$\omega_{EC} = d(\pi_1(\theta_C|\boldsymbol{y}_C), \pi_2(\theta_E|\boldsymbol{y}_E))$$

- ⋆ When $d(\pi_1(\theta_C|\boldsymbol{y}_C), \pi_2(\theta_E|\boldsymbol{y}_E)) \to 1$, it leads to effective down-weighting
- ⋆ When $d(\pi_1(\theta_C|\boldsymbol{y}_C), \pi_2(\theta_E|\boldsymbol{y}_E)) \to 0$, it implements borrowing of information

# Extending to multiple external modules

Suppose there are a total of $K \geq 3$ modules in a basket trial.

For any $\theta_k$, $k = 1, \ldots, K$, there are $(K-1)$ one-to-one commensurate predictive priors:

$$\pi^{\text{CPP}}(\theta_k, \nu_{qk} | \boldsymbol{y}_q, \theta_q),$$

for $q \neq k$.

A $K \times K$ symmetric matrix of the Hellinger distance measures can be constructed:

$$\begin{pmatrix} 0 & d_{12} & \cdots & d_{1K} \\ d_{21} & 0 & \cdots & d_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ d_{K1} & d_{K2} & \cdots & 0 \end{pmatrix}$$

- $\star$ Normalising the pairwise Hellinger distance per column $k$ as discrete **probabilities**, $p_{qk}$, that will be summed to 1.

# Synthesis of the commensurate predictive priors

⋆ With an analogue of the normal distribution likelihood, we stipulate

$$p_{qk} = \frac{\exp(-d(\pi_{\theta_q}, \pi_{\theta_k})/s_0)}{\sum \exp(-d(\pi_{\theta_q}, \pi_{\theta_k})/s_0)},$$

where $s_0$ determines the degree of rewards

- Choosing $s_0 \to \infty$, nearly same probability will be allocated to $p_{qk}$ irrespective of the value of $d(\pi_{\theta_q}, \pi_{\theta_k})$;
- Choosing $s_0 \to 0^+$, corresponding to $d(\pi_{\theta_q}, \pi_{\theta_k})$ close to 0, $p_{qk} \to 1$.

⋆ To leverage trial data from all available modules, we obtain the **marginal predictive prior** for $\theta_k$ given by

$$\pi^{\text{MPP}}(\theta_k, \nu_{qk}) = \sum_{q,k=1}^{K} p_{qk} \times \pi^{\text{CPP}}(\theta_k, \nu_{qk} | \boldsymbol{y}_q, \theta_q)$$

⋆ With inclusion of $\boldsymbol{y}_k$, decision making will be based on the posterior

$$\pi^{\text{MPP}}(\theta_k, \nu_{qk} | \boldsymbol{y}_1, \ldots, \boldsymbol{y}_K) \propto \pi^{\text{MPP}}(\theta_k, \nu_{qk}) \times \mathcal{L}(\boldsymbol{y}_k | \theta_k)$$

# A hypothetical PBC trial implementing our method

Suppose there is a basket trial in patients with Primary Biliary Cholangitis (PBC).

- ⋆ Let $i$ index patients and $k$ index modules, $i = 1, \ldots, n$; $k = 1, \ldots, K$
- ⋆ Some biomarkers are used to stratify the patients into $K = 6$ modules
- ⋆ Each module contains 5 – 10 patients per treatment arm
- ⋆ **Trial data** collected from patient $i$ in module $k$ contain
  - a pre-treatment baseline measurement, $\eta_{0ik}$
  - two covariates, $z_{1ik}$ and $z_{2ik}$
  - a binary indicator for treatment assignment, $T_{ik}$ ($T_{ik} = 1$ new treatment, and 0 for placebo)
  - a post-treatment measurement of clinical interest, $y_{ik}$
- ⋆ We consider a linear model (on the first level) given by

$$\mathbb{E}(y_{ik}|\eta_{0ik}, z_{ik}, T_{ik}) = \eta_{0ik} + z_{1ik}\gamma_{1k} + z_{2ik}\gamma_{2k} + T_{ik}\theta_k.$$

- ⋆ Set the limits of the "spike-and-slab" prior as $\mathcal{B}_0 = 0.005$, $\mathcal{B}_1 = 5$, and $\mathcal{S} = 100$
- ⋆ Interested in estimating the parameter $\theta_k$.

# A hypothetical PBC trial (I)

- ⋆ Define true values for the model parameters:

$$\gamma_{1k} = 1,\ \gamma_{2k} = 1.3,$$
$$\theta_k = c(2.1, 2.1, 4.3, 2.2, 4.2, 4.4)$$

- ⋆ Potentially two cliques with exchangeable modules are expected

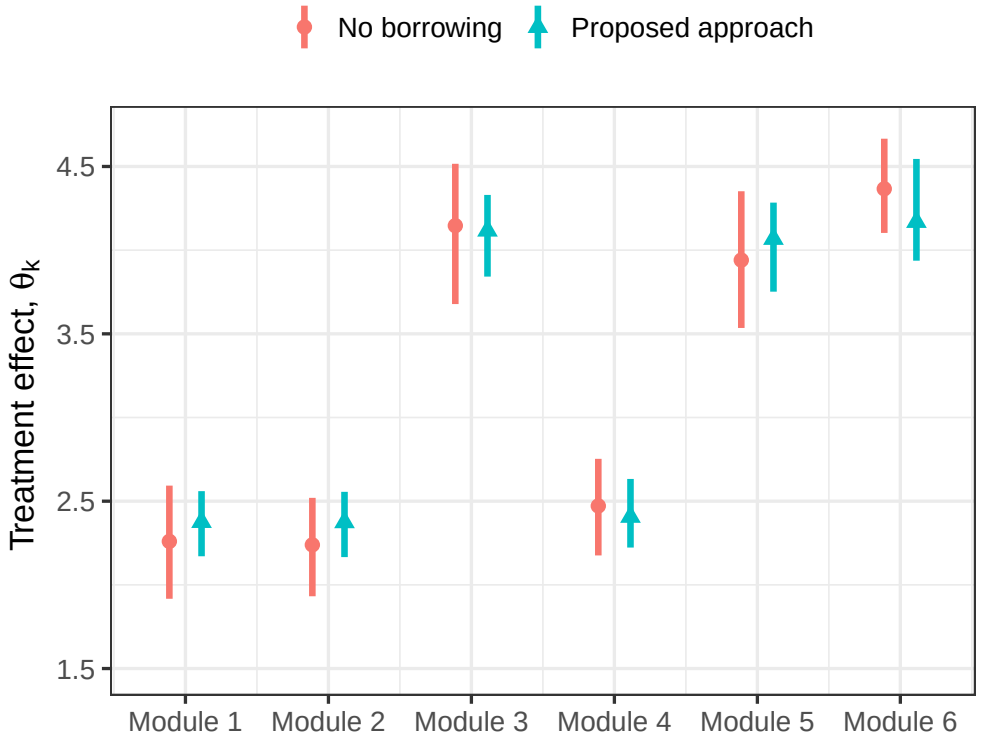



Figure : Comparing the posterior estimates of $\theta_k$, obtained using approaches of no borrowing and our proposal, respectively.

# A hypothetical PBC trial (II)

- ⋆ True values for the treatment effect:

$$\theta_k = c(2.1, 2.1, 4.3, 2.2, 4.2, 4.4)$$

- ⋆ Potentially two cliques with exchangeable modules are expected

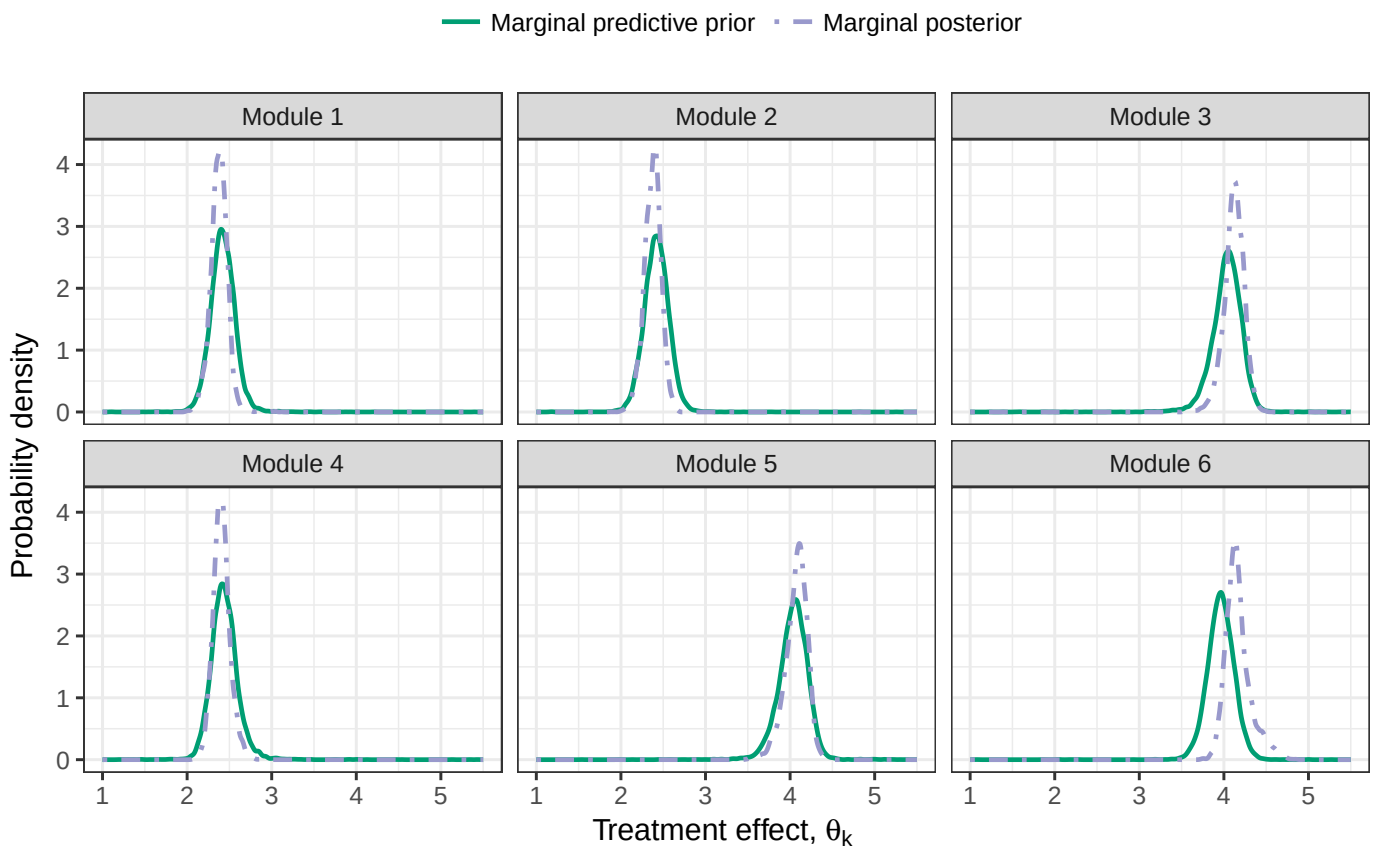



Figure : Using the proposed approach, updates of the marginal predictive prior to the marginal posterior.

# Simulation study with 1000 replicates of basket trials

⋆ We simulate 1000 replicates of the PBC trial under the following eight scenarios

- Scenario 1: $\theta_k = c(0.49, 0.54, 0.67, 0.79, 0.43, 0.35)$
- Scenario 2: $\theta_k = c(0.35, 0.80, 0.37, 1.38, 0.69, 0.40)$
- Scenario 3: $\theta_k = c(0.29, 0.68, 0.77, 0.33, 0.75, 0.30)$
- Scenario 4: $\theta_k = c(0.59, 1.02, 1.17, 0.13, 0.95, 0.75)$
- Scenario 5: $\theta_k = c(0.45, 0.45, 0.45, 0.45, 0.45, 0.45)$
- Scenario 6: $\theta_k = c(0.26, 0.26, 0.26, 0.26, 0.26, 0.26)$
- Scenario 7: $\theta_k = c(0, 0, 0, 0, 0.37, 0.37)$
- Scenario 8: $\theta_k = c(0.33, 0.82, 0, 0, 0.90, 0.83)$

⋆ Test the hypothesis of $H_0 : \theta_k \leq 0$

⋆ Decision criterion on *Go* versus *No-go*:

- *Go* if the lower bound of the **95% posterior credible interval** $\theta_{kL} > 0$, otherwise *No-go*

# Results – Erroneous decision on *Go* under null $\theta_k$

Analogue of Type I error rate

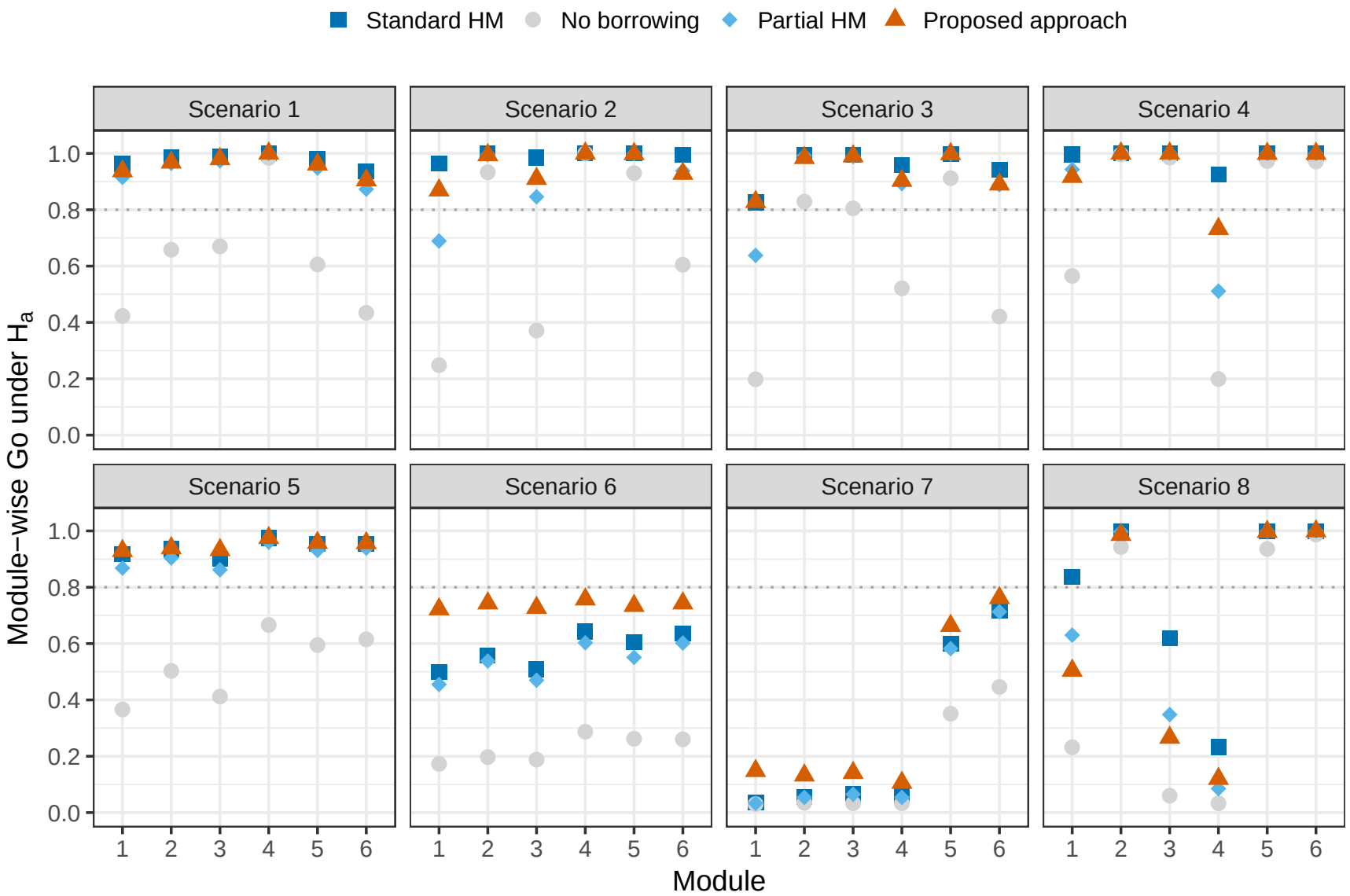
Define *Global null* as the case where all $\theta_k = 0$.

Table : Comparison of analysis models with respect to the proportion of simulated trials with an erroneous go under the specified $\theta_k$.

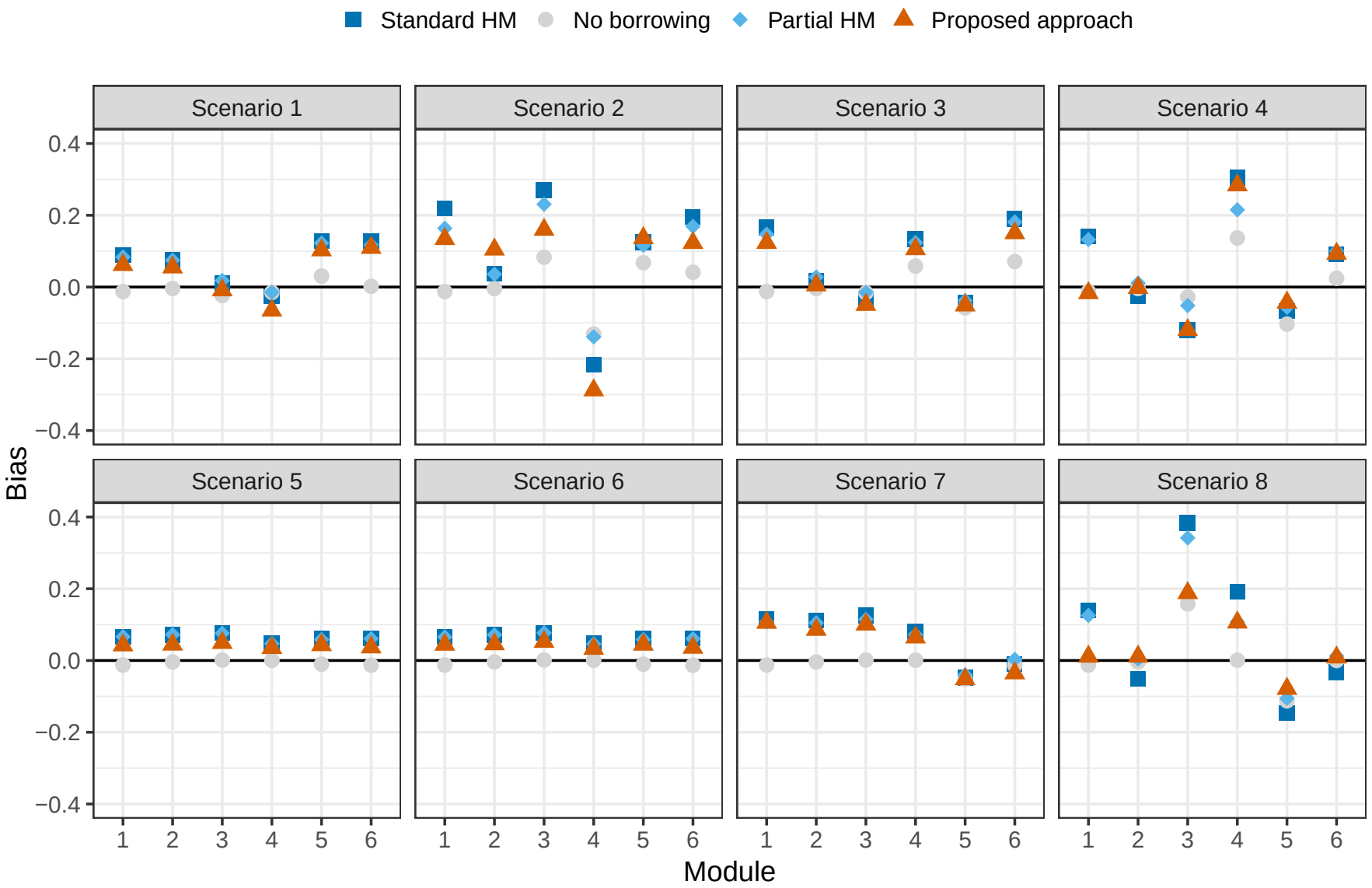
| | | Module | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| **Global null** | Standard HM | 0.018 | 0.034 | 0.029 | 0.033 | 0.022 | 0.040 |
| | No borrowing | 0.030 | 0.035 | 0.033 | 0.033 | 0.039 | 0.030 |
| | Partial HM | 0.019 | 0.035 | 0.033 | 0.032 | 0.021 | 0.040 |
| | Proposed approach | 0.059 | 0.086 | 0.080 | 0.049 | 0.054 | 0.070 |
| **Scenario 7** | Standard HM | 0.189 | 0.193 | 0.216 | 0.155 | - | - |
| | No borrowing | 0.030 | 0.035 | 0.033 | 0.033 | - | - |
| | Partial HM | 0.026 | 0.037 | 0.043 | 0.038 | - | - |
| | Proposed approach | 0.046 | 0.049 | 0.065 | 0.043 | - | - |
| **Scenario 8** | Standard HM | - | - | 0.848 | 0.364 | - | - |
| | No borrowing | - | - | 0.047 | 0.033 | - | - |
| | Partial HM | - | - | 0.374 | 0.057 | - | - |
| | Proposed approach | - | - | 0.337 | 0.150 | - | - |

# Results – Correct decision on *Go* under $H_a$

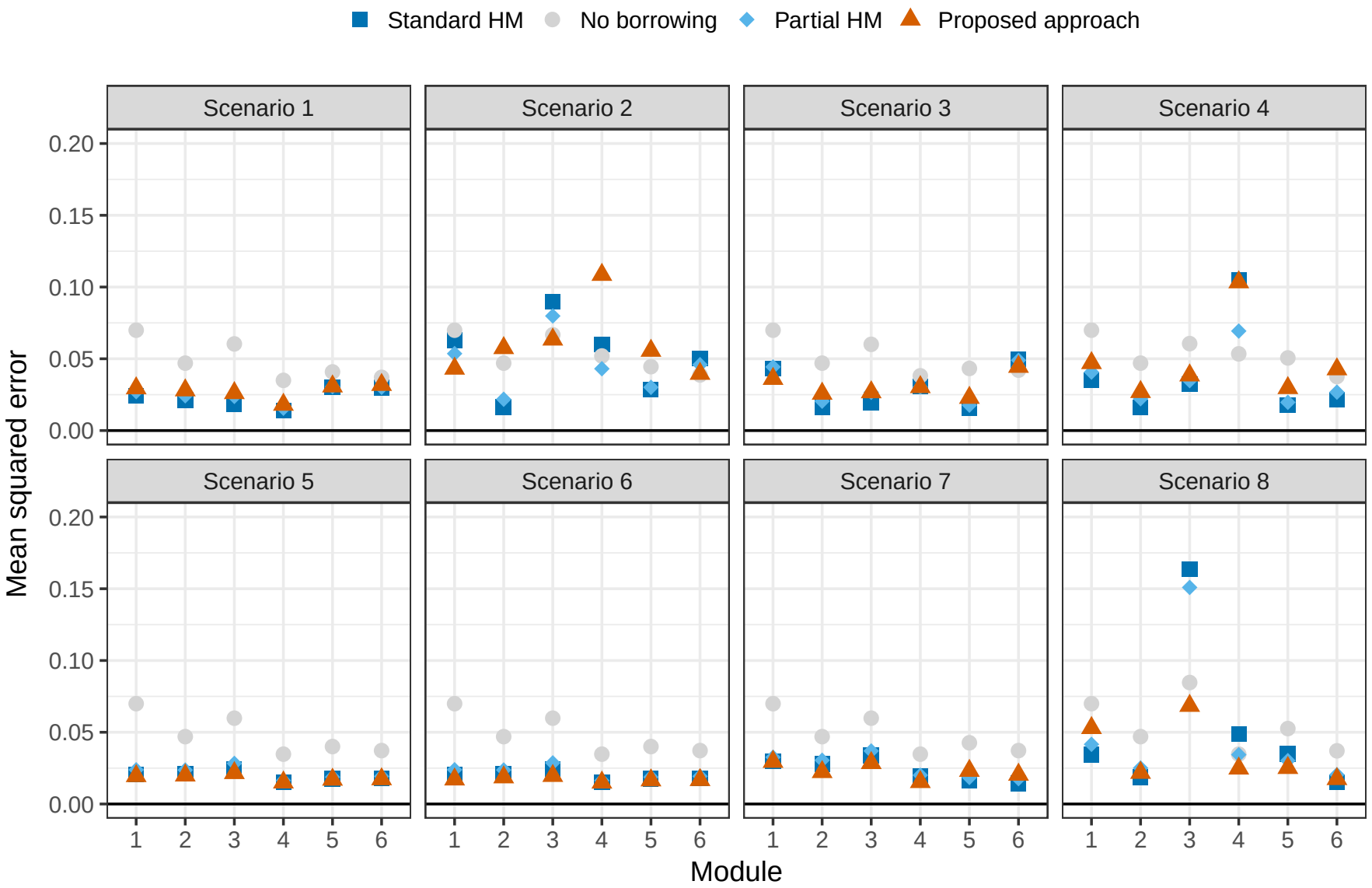
Analogue of statistical power



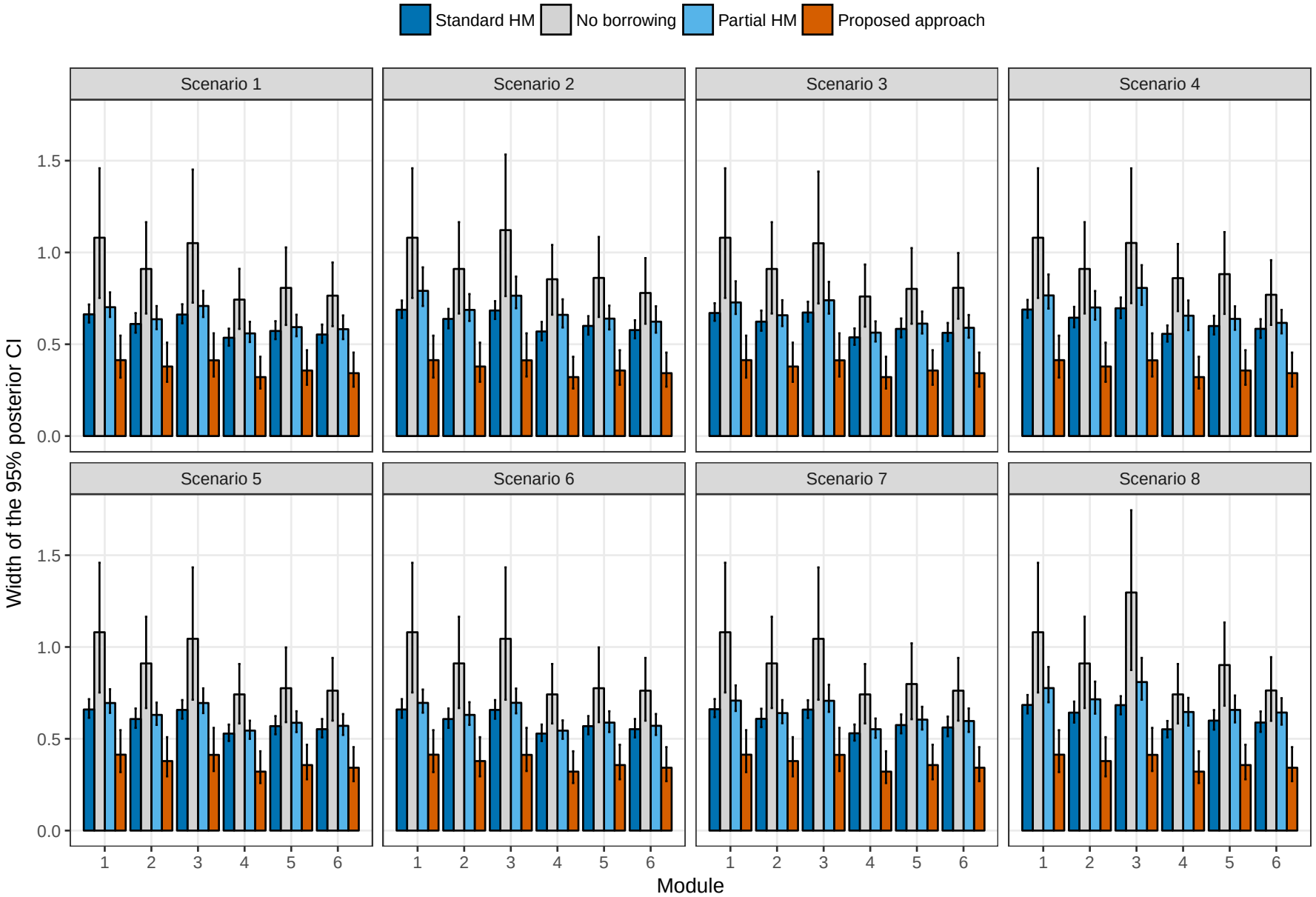
# Results – Improved precision of estimates (I)



# Results – Improved precision of estimates (II)

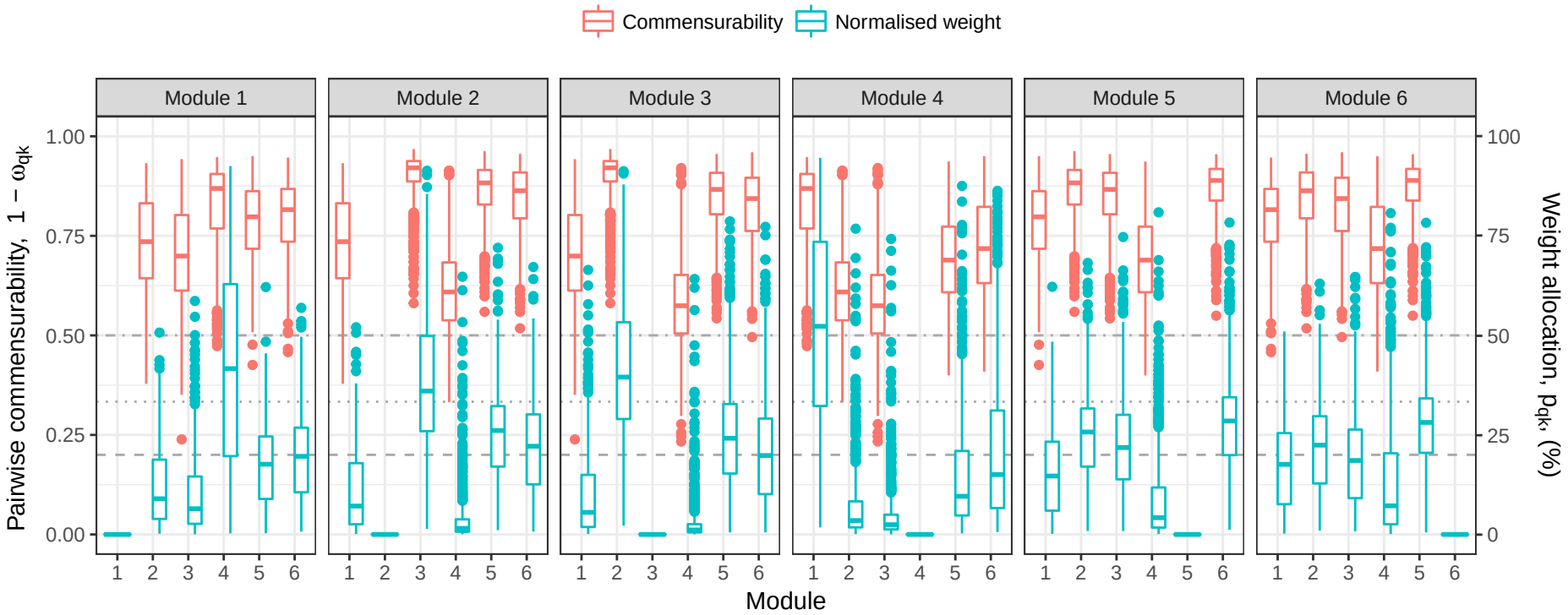


# Results – Improved precision of estimates (III)



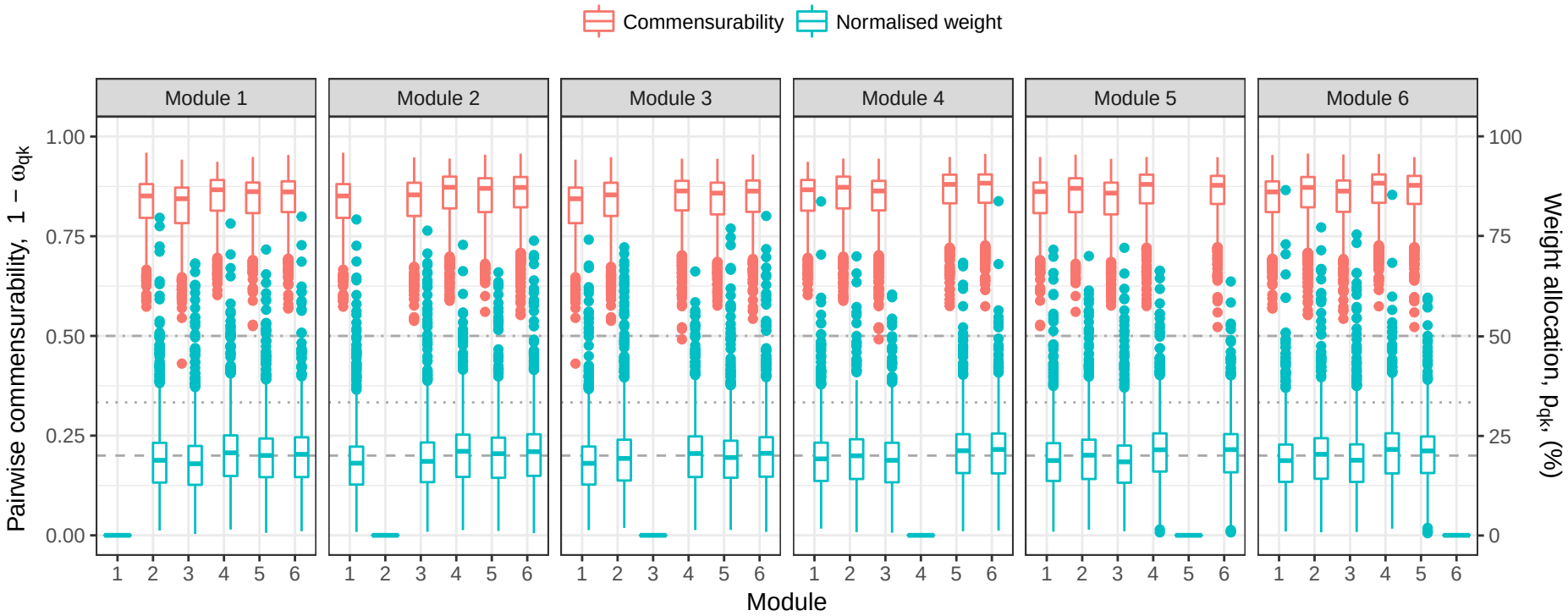
# Results – Effective borrowing and downweighting (I)

**Scenario 4**: $\theta_k = c(0.59, 1.02, 1.17, 0.13, 0.95, 0.75)$



# Results – Effective borrowing and downweighting (II)

**Scenario 5**: $\theta_k = c(0.45, 0.45, 0.45, 0.45, 0.45, 0.45)$



# Conclusions

- ⋆ We have proposed a new analysis methodology suitable for basket trials
- ⋆ Inclusion of a **distance measure** allows quantification of **commensurability** between the module parameters, and thus facilitates sensible borrowing
- ⋆ Our method **improves the precision** of estimates when there are consistent modules
- ⋆ It behaves almost like an empirical Bayes approach for using data to inform the prior

# Future work – Bayesian adaptive methods for basket trials

- ⋆ Consider more treatment arms within each module $\rightarrow$ umbrella trial designs
- ⋆ Develop sequential basket trials enabling early stopping for futility and efficacy
- ⋆ Adaptive enrichment designs to prioritise development paths in certain subgroups
- ⋆ Sample size reassessment at the interim analyses
- ⋆ Continuously add new modules in the trial $\rightarrow$ platform trial designs

# Key references


- Berry SM, Carlin BP, Lee JJ, Müller P. *Bayesian Adaptive Methods for Clinical Trials*. Chapman & Hall/CRC Biostatistics Series CRC Press; 2010.
- Viele K, Berry S, Neuenschwander B *et al*. Use of historical control data for assessing treatment effects in clinical trials. *Pharmaceutical Statistics*. 2014;13(1):41-54.
- Hobbs BP, Carlin BP, Mandrekar SJ, Sargent, DJ. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics*. 2011;67(3):1047-1056
- Schmidli H, Gsteiger S, Roychoudhury S *et al*. Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*. 2014;70(4):1023-1032.
- Woodcock J and LaVange LM. Master Protocols to Study Multiple Therapies, Multiple Diseases, or Both. *New England Journal of Medicine*. 2017;377(1):62-70.
- EMA. *Concept Paper on Extrapolation of Efficacy and Safety in Medicine Development*. European Medicines Agency: London, UK; 2013.