AUGMENTING CLINICAL TRIAL DATA WITH EXTERNAL CONTROLS THROUGH ENERGY BALANCING WEIGHTED POWER PRIOR

Shaoming Yin

25OCT2024

REAL WORLD DATA (RWD)

- FDA:
  - Real-World Data (RWD): “data relating to patient health status and/or delivery of health care routinely collected from a variety of sources.”
  - Real-World Evidence (RWE): “the clinical evidence regarding the usage, the potential benefits or risks, of a medical product derived from analysis of RWD.”

- Sources of RWD include historical data from previous clinical trials, procedure or disease registry, electronic health records (EHRs), medical claims and billing data, patient-reported outcomes.

- Use of historical controls in rare disease and oncology has become common in the regulatory setting (via 21st Century Cures Act).

- Recent years have seen the release of multiple guidance documents on the use of RWD by FDA, EMA, PMDA, and ICH.

REGULATORY GUIDANCE ON RWD

Guidance for Industry

E 10 Choice of Control Group and Related Issues in Clinical Trials

Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products

Guidance for Industry

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to https://www.regulations.gov. Submit written comments to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products

Rare Diseases: Natural History Studies for Drug Development

Guidance for Industry

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to https://www.regulations.gov. Submit written comments to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

MOTIVATION

Incorporate information from RWD to augment control arm data in early phase/proof-of-concept studies or rare disease studies

Benefits

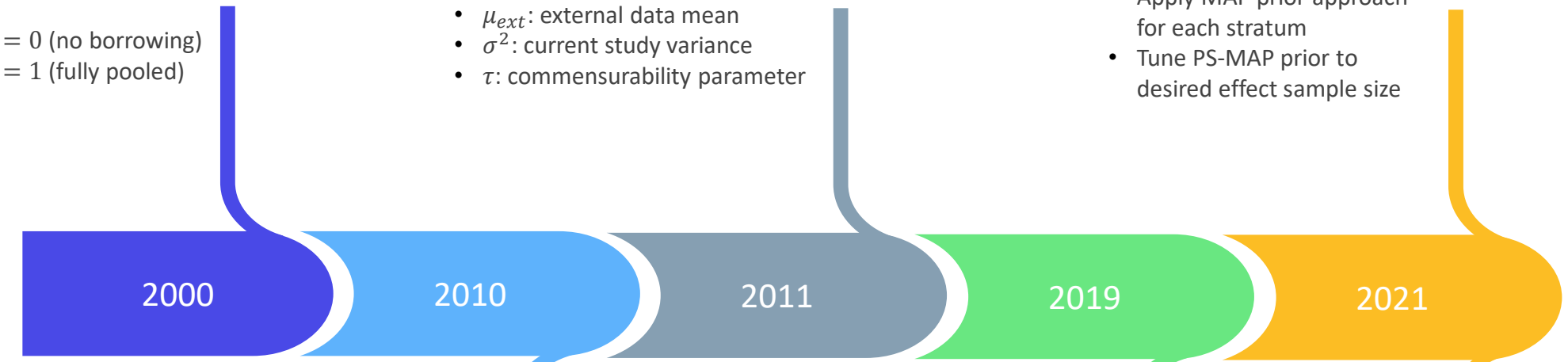
- Smaller control arms
  - More patients receive treatment
  - Cost effective
- Better estimation of control response
  - Increased power
  - Decreased type I error rate

Potential Risks

- Historical data conflicts with the observed control data
  - Bias
  - Decreased Power
  - Increased type I error

BAYESIAN DYNAMIC BORROWING METHODOLOGY

EXTERNAL DATA INFORMS PRIOR: $p(\theta|\mathrm{D}) \propto p_0(\theta|D_{ext},\alpha)L(\theta|D)$



Power Prior

(Ibrahim and Chen)

- $p_0(\theta|D_{ext},\alpha) \propto p_0(\theta)L(\theta|D_{ext})^{\alpha}$
- $0 < \alpha < 1$
  - $\alpha = 0$ (no borrowing)
  - $\alpha = 1$ (fully pooled)

Commensurate Prior

(Hobbs et al.)

- $p_0(\theta|D_{ext},\tau) \sim N\left(\mu_{ext},\frac{\sigma^2}{\tau^2}\right)$
  - μ_{ext}: external data mean
  - σ^2: current study variance
  - τ: commensurability parameter

PS + MAP Prior

(Liu et al.)

- Stratify subjects using PS
- Apply MAP prior approach for each stratum
- Tune PS-MAP prior to desired effect sample size

MAP Prior

(Neuenschwander et al.)

- $y_j|\, D_{ext}, D, \theta \sim Bin(n_j, p_j)$
- $\theta_j = \log\left(\frac{p_j}{1-p_j}\right) = \mu_0 + \delta_j$
  - $\mu_0 \sim \mathrm{p}(\mu_0)$; $\delta_j \sim \mathrm{N}(0,\sigma_\theta^2)$; $\sigma_\theta^2 \sim \mathrm{p}(\sigma_\theta^2)$
- Accounts for heterogeneity through σ_θ^2

PS + Power Prior

(Wang et al.)

- Stratify subjects using PS
- Use stratum-specific power prior:
  - $p_0(\theta|D_{ext,s},\alpha_s) \propto p_0(\theta)L(\theta|D_{ext,s})^{\alpha_s}$
- Combine stratum-specific prior

Individual Weights Prior

(Golchi)

- $p_0(\theta|D_{ext},\alpha) \propto p_0(\theta)\prod_i^{N_{ext}} p(y_{ext,i};\theta)^{\alpha_i}$
  - α_i: based on mahalanobis distance of covariates between external subject and current study

MAP: Meta-Analytic Predictive; PS: Propensity Score

LIMITATIONS OF CURRENT METHODS

- Traditional Bayesian dynamic borrowing methods down-weights external data in the presence of heterogeneity.
- However, they mainly adjust for differences in outcomes, not covariates.
- Propensity score (PS) based methods have also been used to augment clinical trial data with external controls, by adjusting for differences in covariates, not outcomes.
- Recent methods combined PS approaches with either Power Prior (Wang et al. 2019) or MAP Prior (Liu et al. 2021) to adjust for differences in both outcomes and covariates.
- However, the use of PS relies on the modeling of an unknown treatment-assignment mechanism (with respect to both covariates and functional form).
- Individual weights (IW) prior (Golchi 2021) avoids PS but relies on covariate mahalanobis distance that may be computationally intensive, with relative inferior performance as the number of covariates increase.

MORE ON INDIVIDUAL WEIGHTS (IW) PRIOR

- Shown to **outperform previous Bayesian methods** when external data are **partially exchangeable** with current data (Golchi 2021) – based on 1 baseline covariate and continuous outcomes.

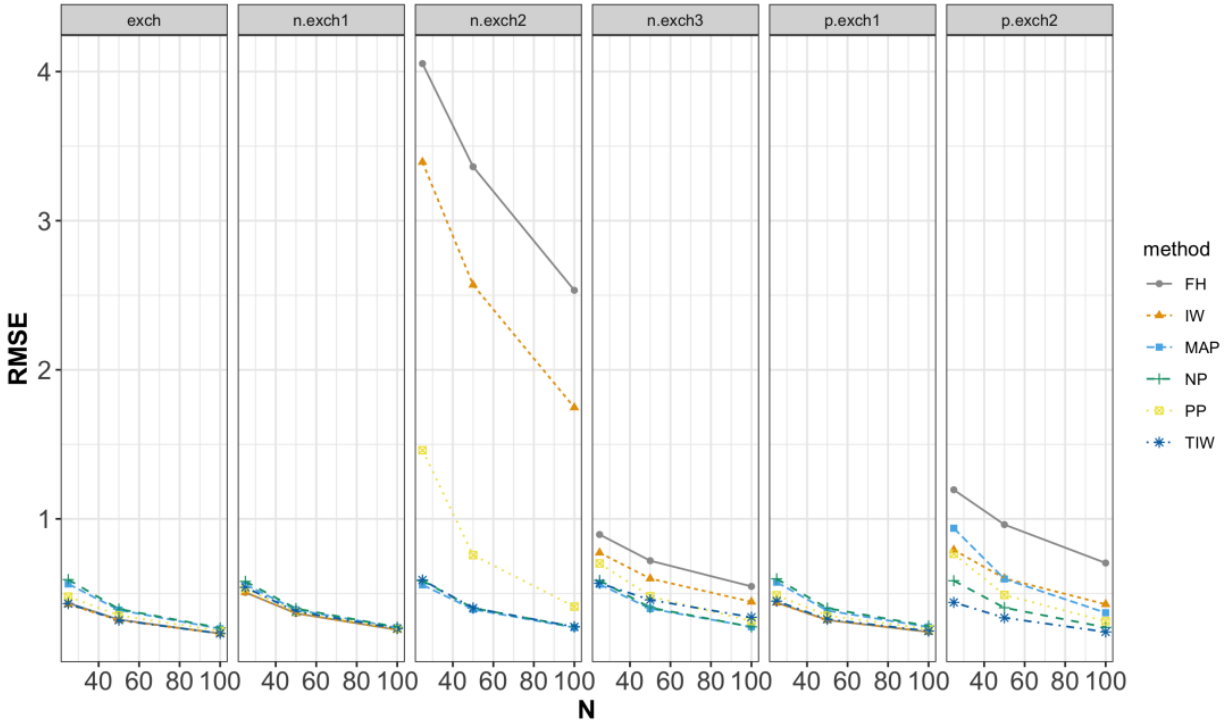


Figure 4: RMSE for estimating the treatment effect averaged over 500 simulation iterations for six simulation scenarios (column panels), six methods incorporating various amounts of external control data (legend) and increasing sample sizes (X axis).

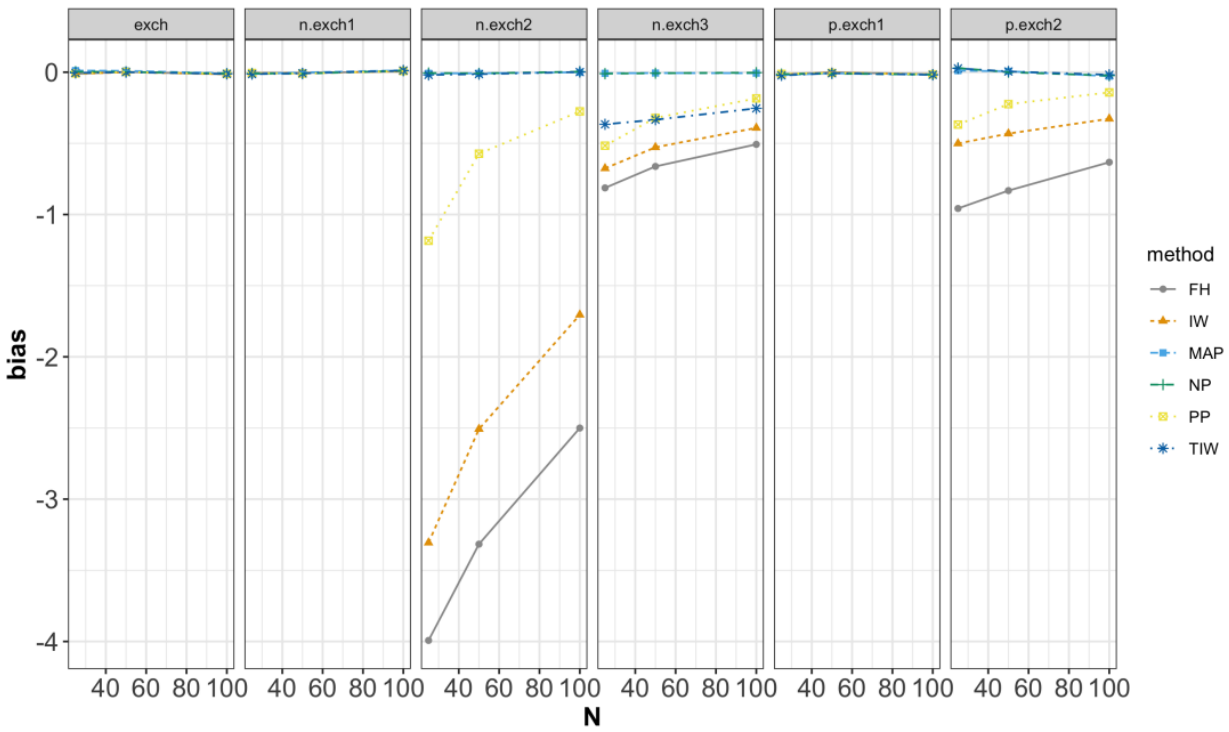


Figure 5: Bias for estimating the treatment effect averaged over 500 simulation iterations for six simulation scenarios (column panels), six methods incorporating various amounts of external control data (legend) and increasing sample sizes (X axis).

FH: Full History Prior; IW: Individually Weighted Prior; MAP: Meta-Analytic Predictive Prior; NP: No Prior; PP: Power Prior; TIW: Truncated Individually Weighted Prior

PROPOSAL: ENERGY BALANCING WEIGHTED POWER PRIOR

Individual power weights α_i(Golchi 2021) can be chosen such that it balances energy distance (Huling and Mak, 2024) of all covariates:

$$p_0(\theta|D_{ext},\alpha) \propto p_0(\theta)\prod_{i}^{N_{ext}} p\left(y_{ext,i};\theta\right)^{\alpha_i}$$

Intuition: want to find a similarity measure of external individual to current trial based on the entire distribution of covariates

- **Step 1**: Obtain individual w_i by minimizing the energy distance between weighted ECDFs, via quadratic program (with linear inequality constraints) optimization.

$$\mathrm{w} \in \underset{w=(w_1,\dots w_n)}{argmin}\{\varepsilon\left(F_{n,1,w},F_n\right)+\varepsilon\left(F_{n,0,w},F_n\right)+\varepsilon\left(F_{n,0,w},F_{n,1,w}\right)\}$$

$$s.t.\ \sum_{i=1}^{n} w_i(1-A_i)=n_0, w_i=1\ \forall A_i=1, w_i\geq 0\ for\ i=1,\dots n$$

Where:

$A_i \in \{0,1\},\ n_1=\sum_{i=1}^{n} A_i,\ n_0=n-n_1,\ F_{n,a,w}(X)=\sum_{i=1}^{n}\frac{w_i I(X_i\leq X, A_i=a)}{n_a},\ a\in\{0,1\},\ ||.||_2$ is the Euclidean norm

$$\varepsilon\left(F_{n,a,w},F_n\right)=\frac{2}{n_a n}\sum_{i=1}^{n}\sum_{j=1}^{n} w_i I(A_i=a)\,||X_i-X_j||_2-\frac{1}{n_a^2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j I\left(A_i=A_j=a\right)||X_i-X_j||_2-\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}||X_i-X_j||_2$$

(weighted energy distance: distance between weighted covariate ECDFs for treated/control and combined covariate ECDF)

PROPOSAL: ENERGY BALANCING WEIGHTED POWER PRIOR

$$\varepsilon\left(F_{n,0,w}, F_{n,1,w}\right) = \frac{2}{n_1 n_0}\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j A_i\left(1 - A_j\right)||X_i - X_j||_2 \; - \frac{1}{n_1^2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \, A_i A_j||X_i - X_j||_2$$
$$- \frac{1}{n_0^2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j (1 - A_i)\left(1 - A_j\right)||X_i - X_j||_2$$

(weighted energy distance: distance between weighted covariate ECDFs for treated and control)

- Note: examples of optimization algorithms that could be applied include interior point methods (via 'cccp' R package), augmented Lagrangian techniques, extensions of the simplex algorithm, operator splitting approach in Stellato et al.

- **Step 2**: Apply min-max scaling to individual w_i obtained from step 1, such that the individual power weights α_i is between 0 and 1.

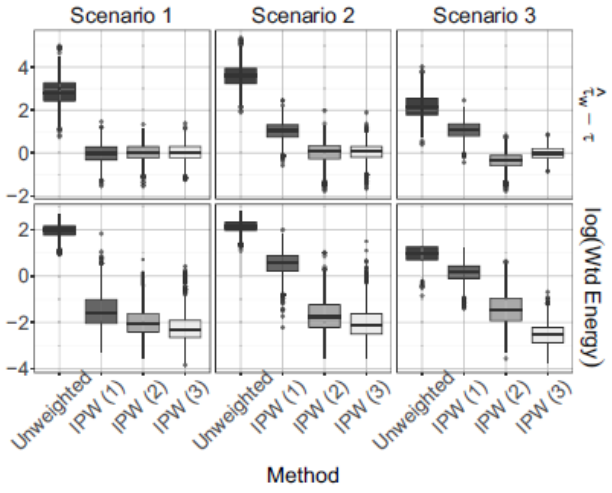
$$\alpha_i = \frac{w_i \; - \; min_n w_i}{max_n w_i \; - min_n w_i}$$

- **Step 3**: Apply truncation of individual power weights to prevent potential estimation bias due to many small weights.

$$p_0(\theta|D_{ext}, \alpha) \propto \; p_0(\theta)\prod_i^{N_{ext}} p\left(y_{ext,i}; \theta\right)^{\alpha_i \mathbb{1}(\alpha_i > \rho)}$$, where $\rho = \; q_{0.05} = \inf\{\alpha\colon F_{\alpha_i}(\alpha) \geq 0.05\}$

ENERGY BALANCING

(HULING AND MAK, 2024)



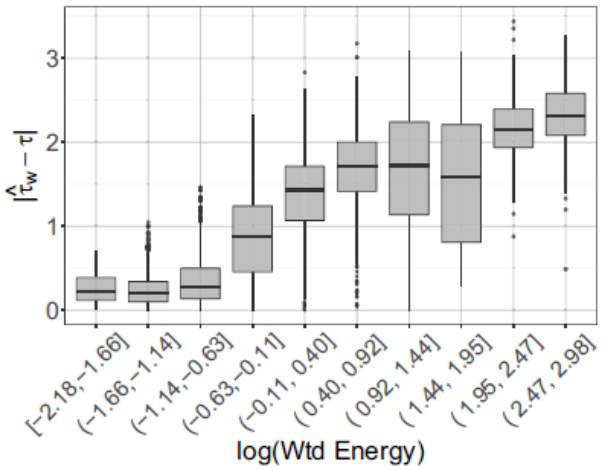


Figure 1: (a, left) Energy distances and biases for IPW estimates based on weights from the three fitted logistic regression models; (b, right) Boxplots of the biases for IPW estimates versus weighted energy distance based on weights estimated by several methods, each with different combinations of moments included for balancing or estimation.

Table 5: Displayed are the median, mean, standard deviation, and maximum RMSEs for each method across the 100 simulation settings using the IAC data. The bold values indicate the best performance across all methods for a given setting

	Unweighted	CBPS	IPW	Cal	EBW	iEBW
	Constant treatment effect					
Median RMSE	8.0151	3.2899	7.9435	3.4895	3.0276	**1.8319**
Mean RMSE	9.1296	3.8542	9.5545	3.7609	3.5113	**2.2087**
SD RMSE	6.7091	2.5588	7.5263	2.5293	2.5028	**1.5785**
Max RMSE	32.3715	11.0479	39.1095	12.0521	12.4231	**7.2066**
	Heterogeneous treatment effect					
Median RMSE	11.9037	3.8587	15.1933	3.7128	3.0732	**1.7355**
Mean RMSE	13.5594	4.4330	18.4590	4.1689	3.9079	**1.8688**
SD RMSE	9.9665	3.1225	14.6490	2.8963	2.8718	**1.3415**
Max RMSE	48.0820	12.3672	75.3899	12.8563	13.8571	**5.6715**

CBPS: Covariate Balancing PS Weights; IPW: Inverse PS Weights; Cal: Empirical Calibration Balancing Weights; EBW: Energy Balancing Weights; iEBW: Improved EBW

SIMULATION SETUP

- Similar setup to Golchi 2021 and Li 2022.
- Concurrent trial with 1:1 randomization (N_c = 50, 100, 200, 300), and external RWD (N_e = 200).
- 4 baseline covariates ($X = [X_1,\ldots,X_4]$) generated from multivariate Normal distribution (MVN):
  - For concurrent trial: $X_c \sim MVN(1, \Sigma)$, with $\Sigma_{ij} = 1*\delta_{ij} + 0.5*(1 - \delta_{ij})$, where $\delta_{ij} = 1$ if i = j and 0 o.w.
  - For external RWD: $X_e \sim MVN(\mu_e, \Sigma)$
- Outcome is continuous:
  - For concurrent trial: $Y_c \sim MVN(X\beta + Z\theta, 1)$, where θ is the treatment effect and Z is treatment assignment (0 for control and 1 for treatment). We set true θ to 1.
  - For external RWD: $Y_e \sim MVN(X\beta + \delta_e, 1)$, where δ_e is a mean shift of external data not explained by X.
- 6 simulation scenarios were evaluated:

Scenario	μ_e	δ_e	Exchangeable X	Exchangeable Y
1	1	0	Yes	Yes
2	1 or 2	0	Partially Yes	Yes
3	1	0 or 1.5	Yes	Partially Yes
4	2	0	No	Yes
5	1	1.5	Yes	No
6	2	1.5	No	No

COMPARED METHODS

Classic propensity score-based methods:

- Propensity score 1:1 matching (no replacement; caliper of 0.2*SD of logit of PS) - PSM
- Inverse probability of treatment weighting with ATT weights - IPTW

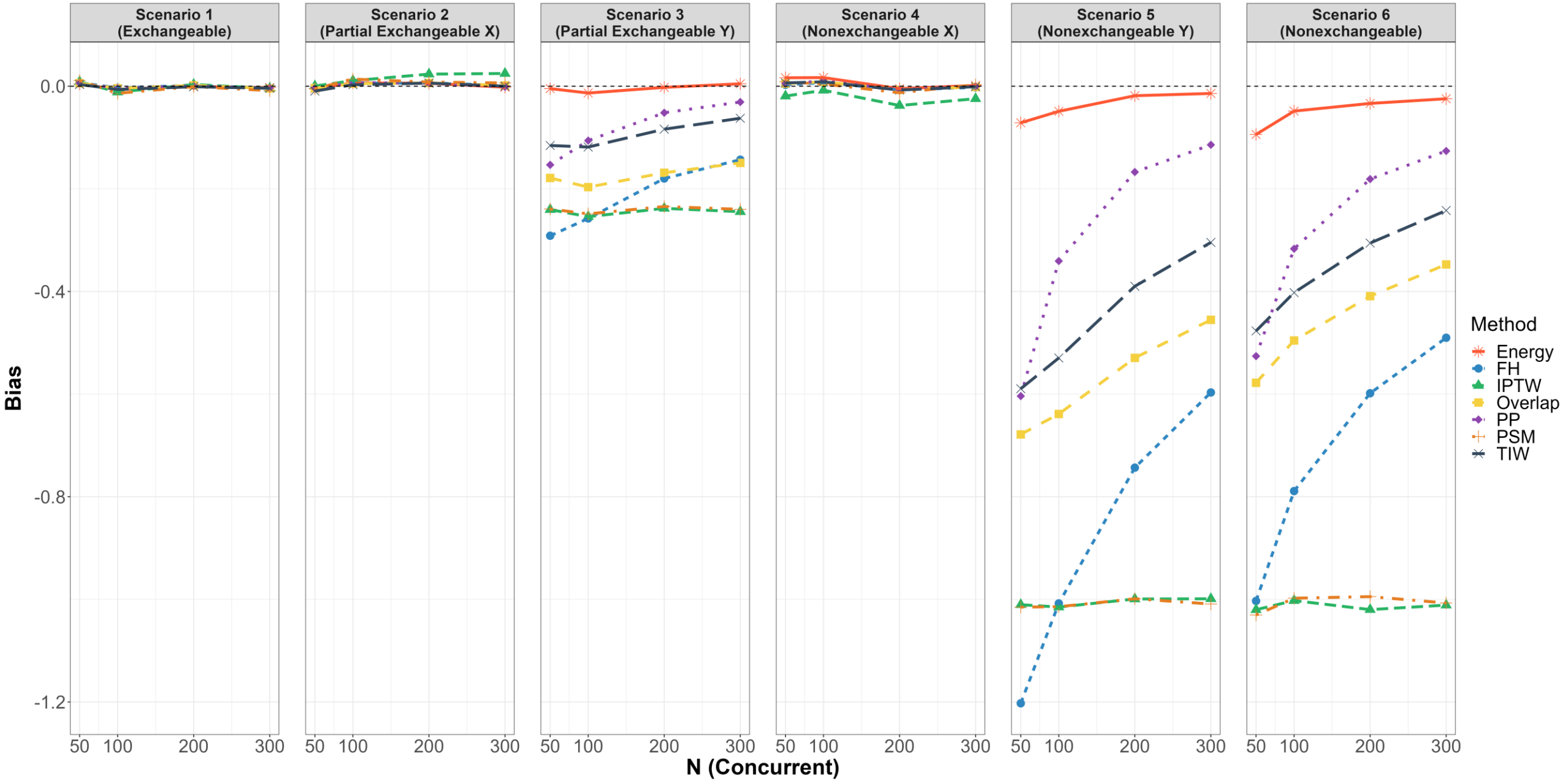
Bayesian borrowing methods:

- Power prior with full history borrowing ($\alpha = 1$) - FH
- Power prior with study weight obtained as penalized likelihood-type criterion (Ibrahim et al. 2003) - PP

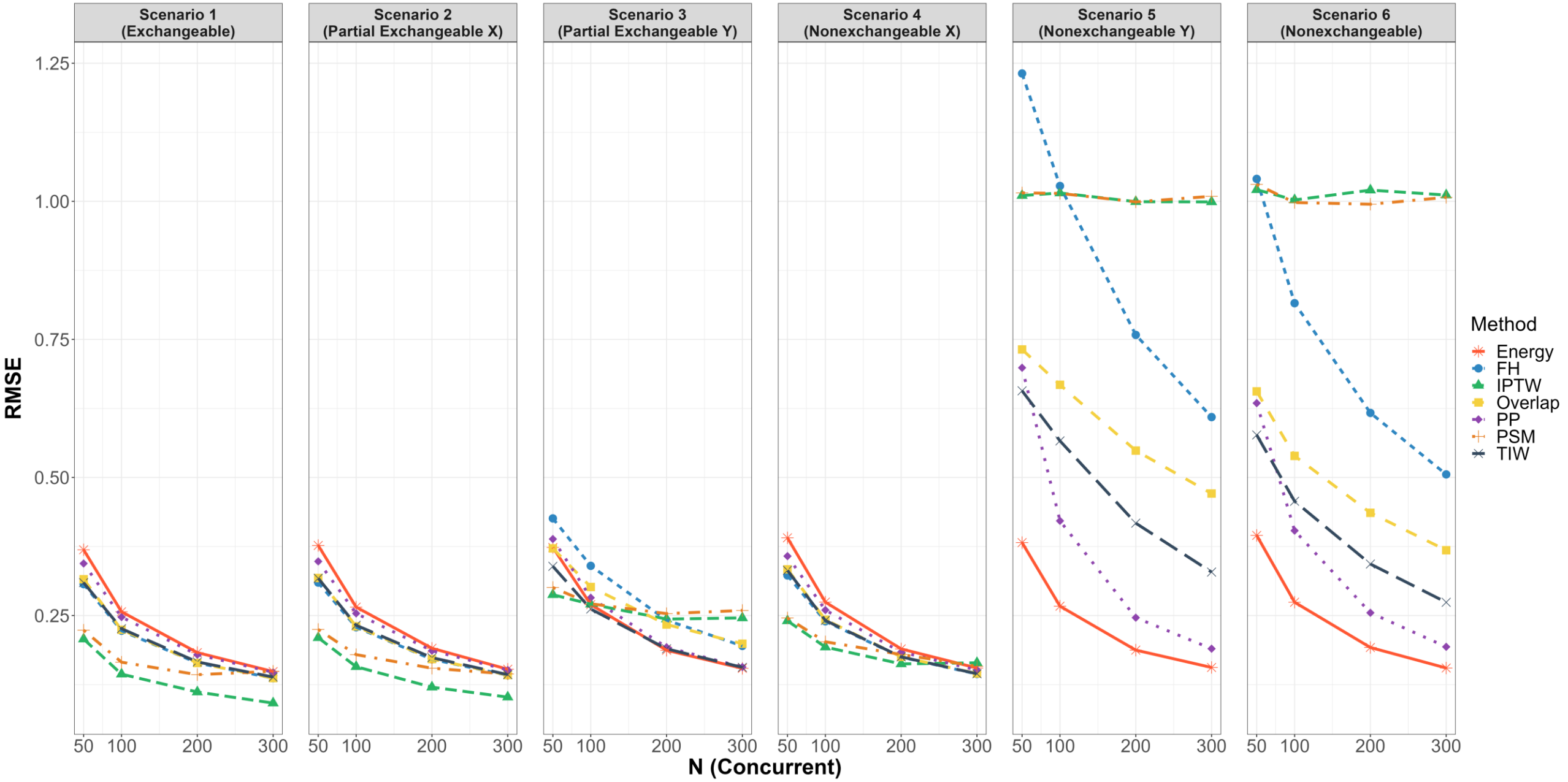
Hybrid approaches

- Individual weights prior with truncation (Golchi 2021) - TIW
- Individual overlap weights prior (Li 2022) - Overlap
- Proposed energy-weighted power prior - Energy

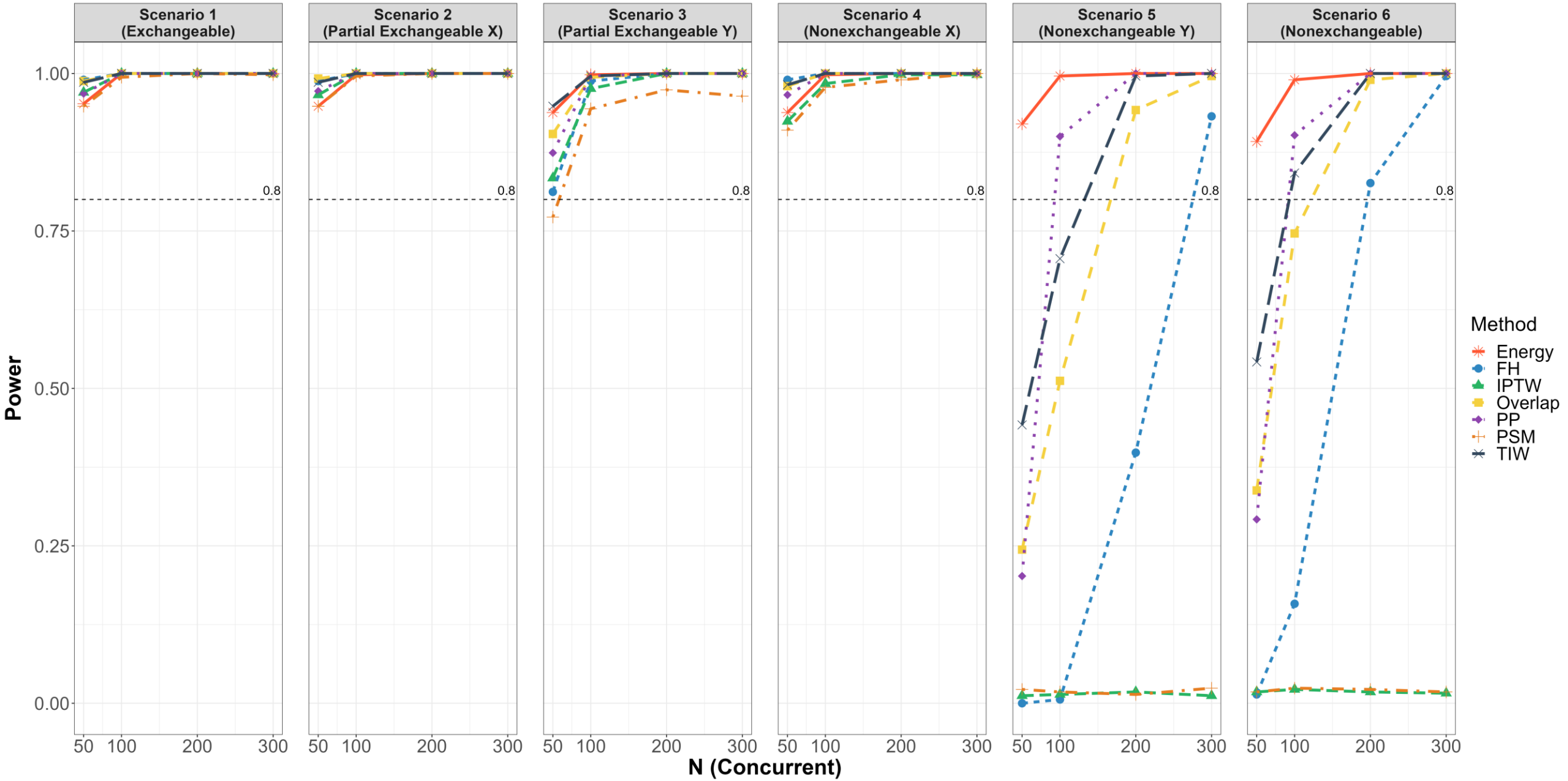
SIMULATION RESULTS - BIAS



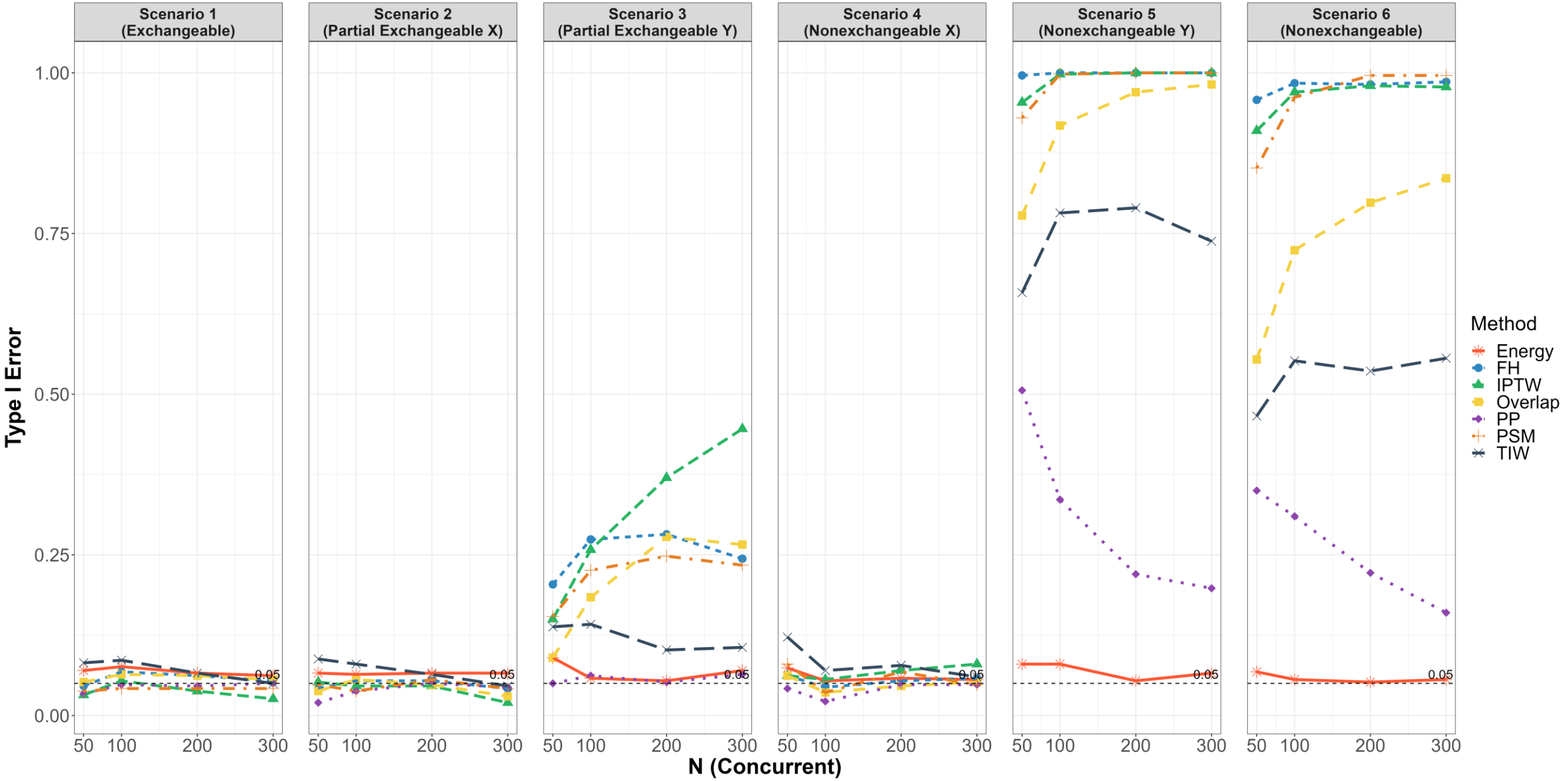
SIMULATION RESULTS - RMSE



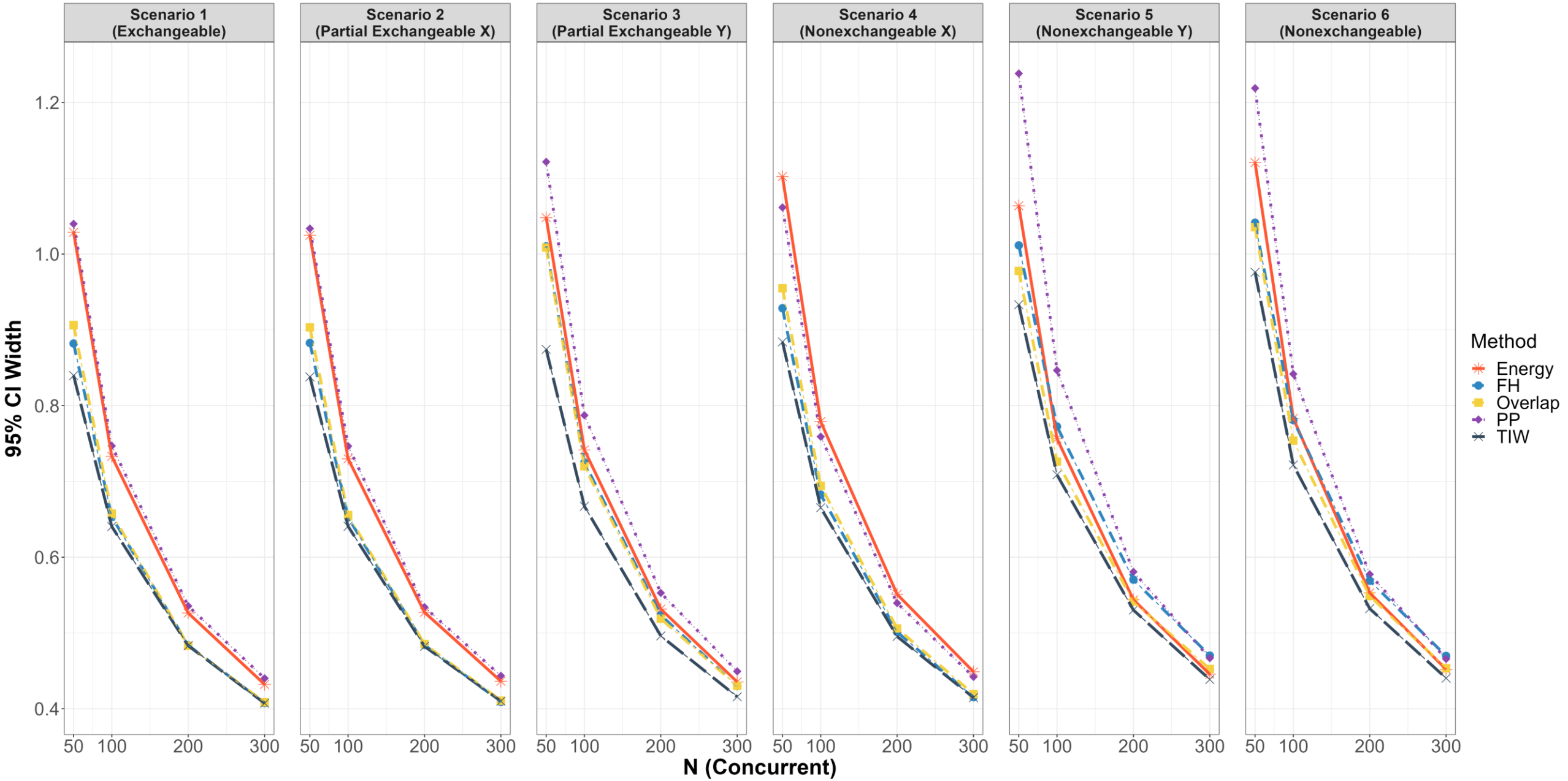
SIMULATION RESULTS - POWER



SIMULATION RESULTS – TYPE I ERROR



SIMULATION RESULTS – 95% CI WIDTH



PROS AND CONS OF PROPOSED METHOD

- **Advantages:**
  - Hybrid approach adjusts for differences in both outcomes and covariates.
  - Model-free and robust, does not require tuning parameters.
  - Directly targets distributional imbalance of covariates between treatment groups, and avoidance of covariate moment(s) specification.
  - Takes advantage of external individual patient data (IPD), instead of only relying on external study-level/aggregate information.
  - Computationally efficient.
  - Individual power weights for the external control will always be between 0 to 1, hence less susceptible for overestimation of precision, and potentially more acceptable to regulators.
  - Has among the best performance in the investigated scenarios.

- **Limitations/Considerations:**
  - Need to specify and justify a list of shared confounders/prognostic covariates for the optimization algorithm.
  - Treatment effect estimates may be biased if external data comes from a different distribution and yet receives large weights in the analysis.
  - Requires patient-level external data.
  - Capping the weights to 1 may not capture all measured confounding.
  - Performance has not been characterized for other endpoint types and scenarios. Further work is needed.

REGULATORY CONSIDERATIONS IN USING APPROACH/RWD

Fundamental question: without randomization, is it possible for the study design to generate evidence capable of distinguishing effect of drug from outcomes attributable to other factors?

Key considerations:

- Is treatment effect size anticipated to be large?
- Selection of 'fit-for-use' external data:
  - Similarity of populations.
  - How many patients could be extracted from RWD after I/E criteria?
  - Extent of bias/confounding and bias mitigation plan.
  - Availability of important prognostic characteristics.
  - Comparability/reliability of outcome assessment.
  - Access to regulatory grade patient-level data.
- Early engagement with regulatory agencies.
- Selection of endpoints and assess their comparability and reliability:
  - Outcome assessment blinded to treatment status?
  - Outcome consistently assessed across arms (with respect to timing, frequency, method, rigor)?
- Selection of index date.
- Determination/justification of analysis methods, assumptions and evaluation of operating characteristics.

THANK YOU!!