A Sequential Basket Trial Design Based on Multi-Source Exchangeability With Predictive Probability Monitoring

Alex Kaizer

Assistant Professor
Department of Biostatistics and Informatics
Colorado School of Public Health

BAYES 2022
October 13, 2022

Graduate School
UNIVERSITY OF COLORADO
DENVER | ANSCHUTZ MEDICAL CAMPUS

Acknowledgments

PLOS ONE

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Bayesian and frequentist approaches to sequential monitoring for futility in oncology basket trials: A comparison of Simon's two-stage design and Bayesian predictive probability monitoring with information sharing across baskets

Alexander Kaizer, Emily Zabor, Lei Nie, Brian Hobbs

Published: August 2, 2022 • https://doi.org/10.1371/journal.pone.0272367

Acknowledgements: Nan Chen

Funding Source: K01-HL151754

A Move Towards Precision Medicine in Oncology

- In oncology, studies traditionally are designed for a particular histology or “indication” (e.g., lung cancer)

A Move Towards Precision Medicine in Oncology

- In oncology, studies traditionally are designed for a particular histology or “indication” (e.g., lung cancer)
- We can now partition cancers into many small molecular subtypes and have developed therapies to target these genetic alterations

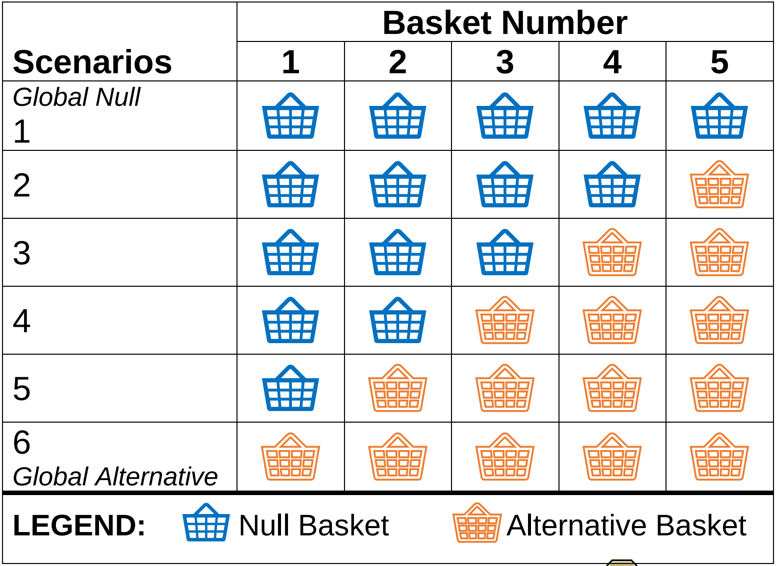
A Move Towards Precision Medicine in Oncology

- In oncology, studies traditionally are designed for a particular histology or “indication” (e.g., lung cancer)
- We can now partition cancers into many small molecular subtypes and have developed therapies to target these genetic alterations
- However, there may be potential heterogeneity in treatment benefit by indication

A Move Towards Precision Medicine in Oncology

- In oncology, studies traditionally are designed for a particular histology or “indication” (e.g., lung cancer)
- We can now partition cancers into many small molecular subtypes and have developed therapies to target these genetic alterations
- However, there may be potential heterogeneity in treatment benefit by indication
- Basket trials address this scientific context, but have their own challenges, including interim monitoring for futility (Woodcock and LaVange, 2017; Hobbs et al., 2018)

Example Basket Trial with 5 Baskets



Sequential Monitoring for Futility

1. Simon's two-stage designs

2. Bayesian predictive probability monitoring

3. Addition of information sharing across baskets with multi-source exchangeability models (MEMs)

Sequential Monitoring for Futility

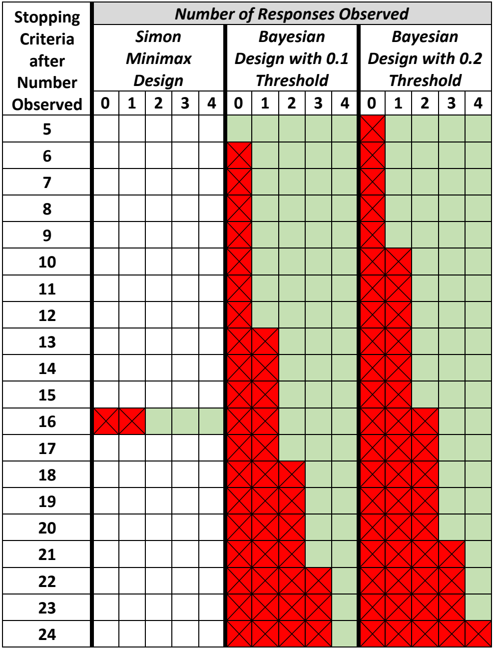
1. Simon's two-stage designs
   - Pros: Simple and easy to implement
   - Cons: Only one interim evaluation, no sharing information across baskets
2. Bayesian predictive probability monitoring
3. Addition of information sharing across baskets with multi-source exchangeability models (MEMs)

Sequential Monitoring for Futility

1. Simon's two-stage designs
   - Pros: Simple and easy to implement
   - Cons: Only one interim evaluation, no sharing information across baskets
2. Bayesian predictive probability monitoring
   - Pros: Can design with more than one interim analysis, interpretation makes sense for interim monitoring
   - Cons: No sharing information across baskets
3. Addition of information sharing across baskets with multi-source exchangeability models (MEMs)

Sequential Monitoring for Futility

1. Simon's two-stage designs
   - Pros: Simple and easy to implement
   - Cons: Only one interim evaluation, no sharing information across baskets
2. Bayesian predictive probability monitoring
   - Pros: Can design with more than one interim analysis, interpretation makes sense for interim monitoring
   - Cons: No sharing information across baskets
3. Addition of information sharing across baskets with multi-source exchangeability models (MEMs)
   - Pros: Can share exchangeable information across baskets
   - Cons: More complex, concerns about heterogeneity
   - We will use an empirically Bayesian prior with a single hyperparameter B, where $0 \leq B \leq 1$



Designs assuming a 10% null response and 30% alternative response *without* information sharing calibrated for α=0.1 and 90% power (Simon and Bayesian designs without interim monitoring).

Legend	
	Not evaluated
	Terminate for futility
	Continue enrollment

Simulation Set-up

Assume a 10 basket trial where $p_0 = 0.1$ and our target response is $p_1 = 0.3$

Generate 1,000 trials with $N = 25$ per basket under two scenarios:

1. Global scenario (all null or all alternative baskets)
2. Mixed scenario (8 null and 2 alternative baskets)

Compare three designs:

1. Simon minimax two-stage design
2. Bayesian design with predictive probability monitoring *without* information sharing (i.e., $B = 0$)
3. Bayesian design with predictive probability monitoring *with* information sharing via MEMs set at $B = 0.1$

Bayesian Calibration and Monitoring

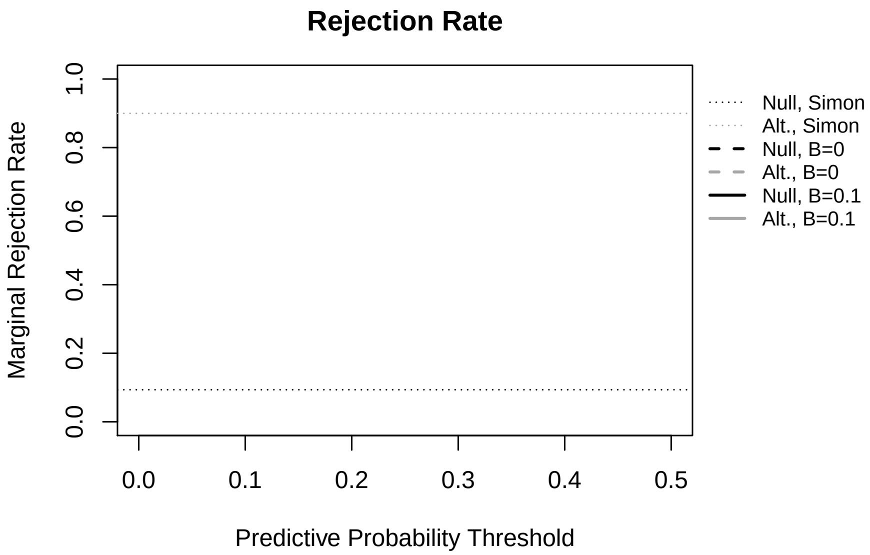
Assume a Beta(0.5,0.5) prior on treatment response.

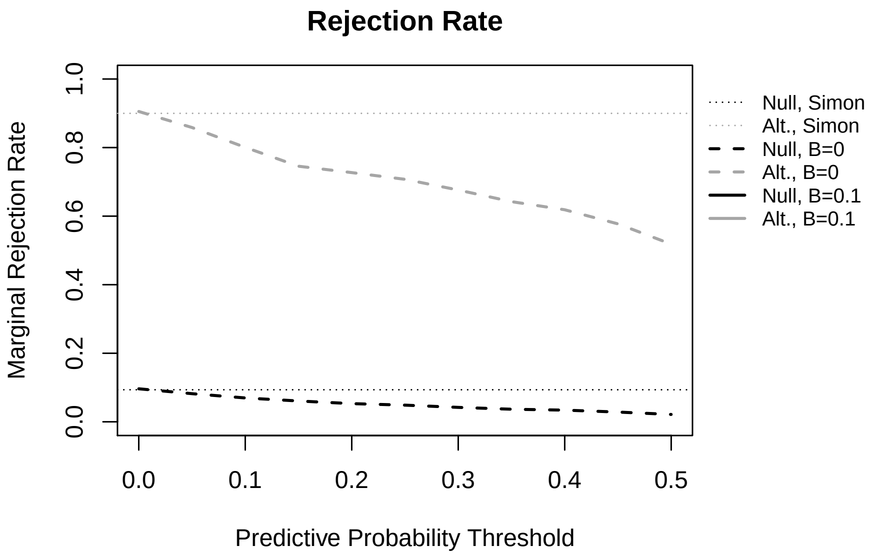
Posterior probability thresholds calibrated to achieve a 10% type I error rate under the **global null scenario** with $N = 25$ and **no interim monitoring**:

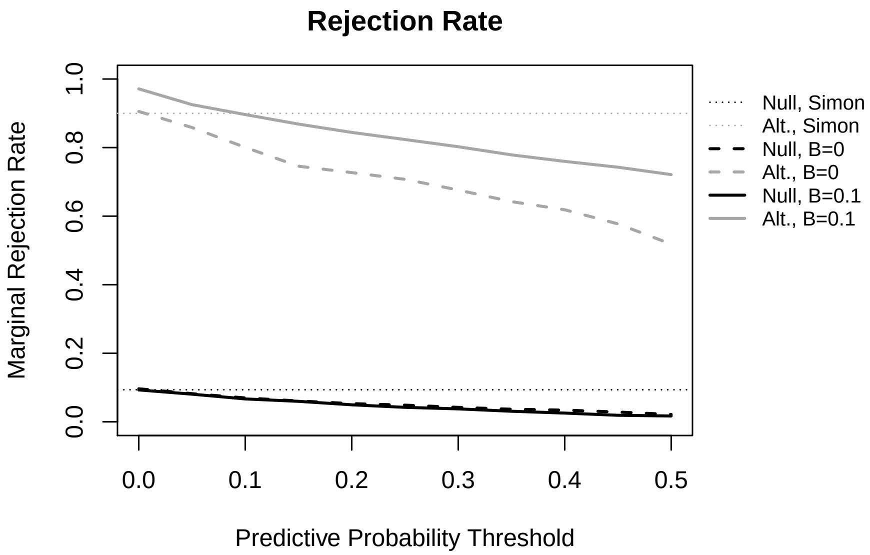
- When $B = 0$, conclude "success" if above 0.900.
- When $B = 0.1$, conclude "success" if above 0.848.

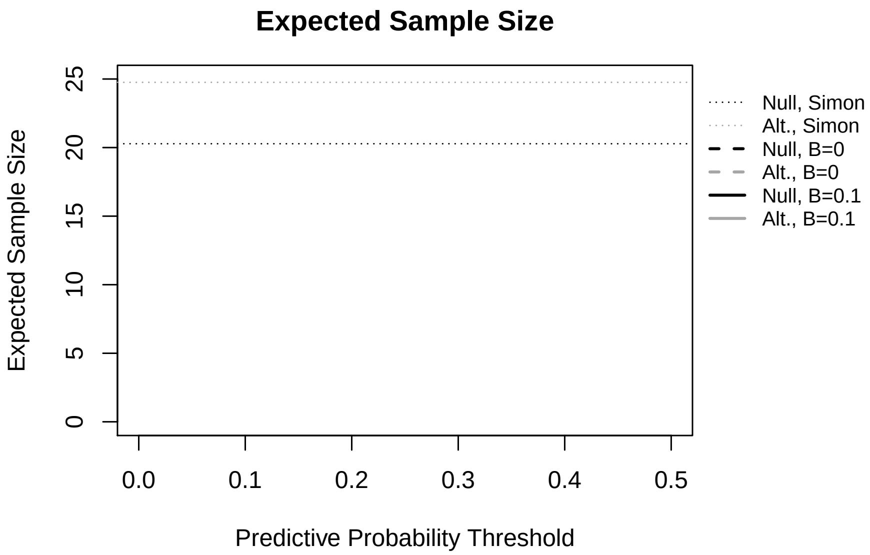
Predictive probability monitoring for futility is implemented continually after the 5th participant in each basket.

Explore thresholds across a grid of values from 0 to 0.5 in increments of 0.05.

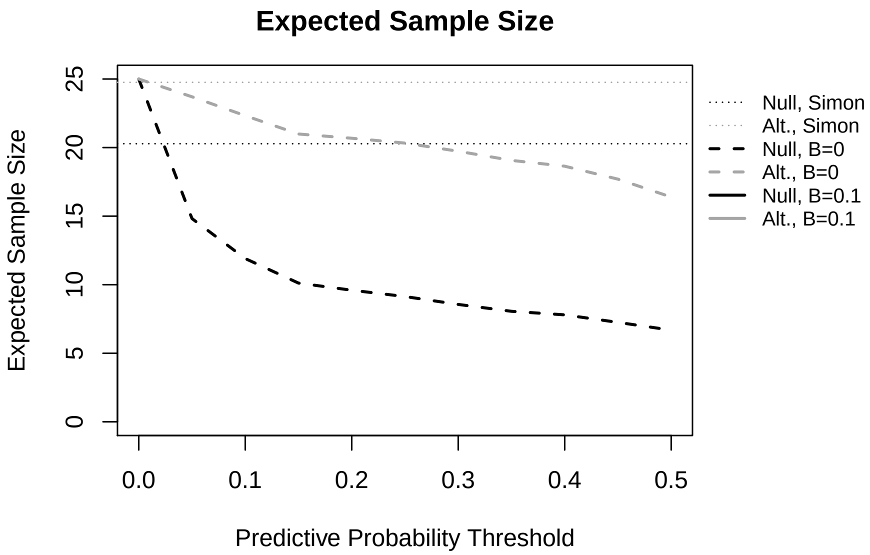

Rejection Rate
Marginal Rejection Rate
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4
0.5
Predictive Probability Threshold
Null, Simon
Alt., Simon
Null, B=0
Alt., B=0
Null, B=0.1
Alt., B=0.1


Rejection Rate
Marginal Rejection Rate
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4
0.5
Predictive Probability Threshold
Null, Simon
Alt., Simon
Null, B=0
Alt., B=0
Null, B=0.1
Alt., B=0.1

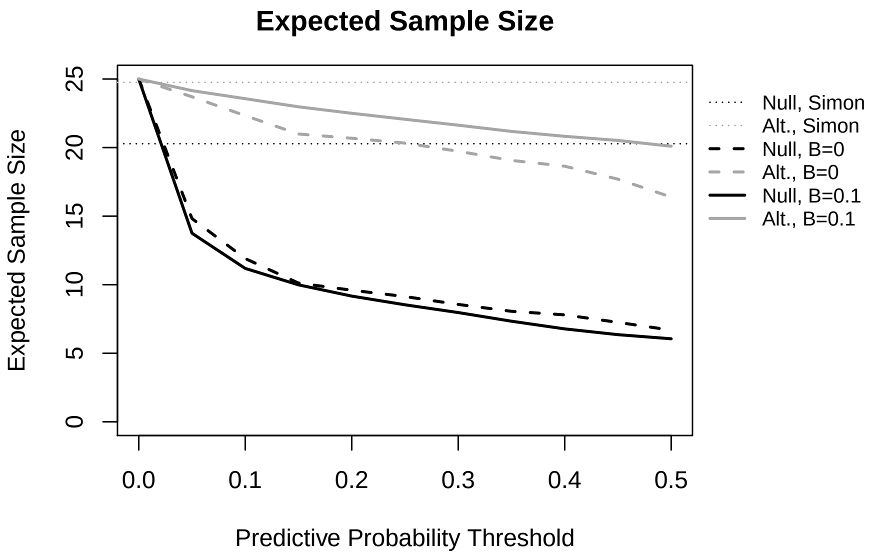

Rejection Rate
Marginal Rejection Rate
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4
0.5
Predictive Probability Threshold
Null, Simon
Alt., Simon
Null, B=0
Alt., B=0
Null, B=0.1
Alt., B=0.1



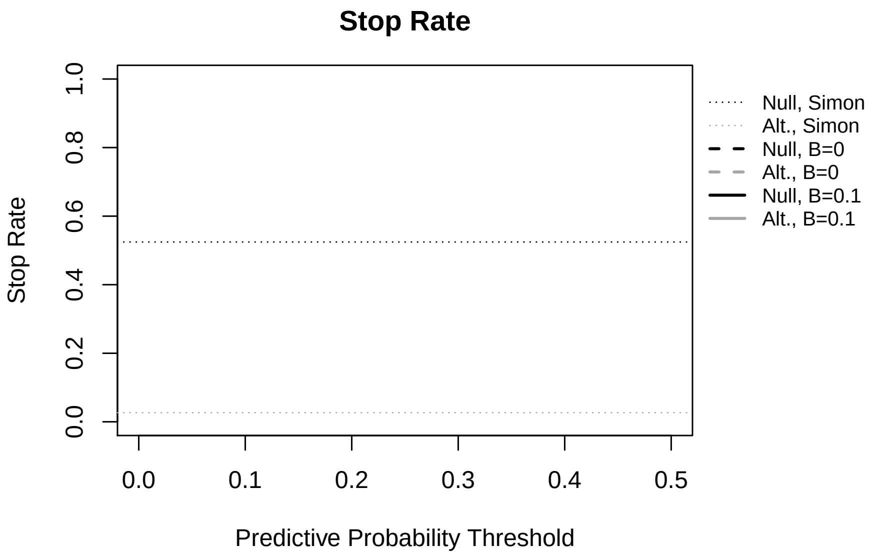
Expected Sample Size
Expected Sample Size
0
5
10
15
20
25
0.0
0.1
0.2
0.3
0.4
0.5
Predictive Probability Threshold
Null, Simon
Alt., Simon
Null, B=0
Alt., B=0
Null, B=0.1
Alt., B=0.1

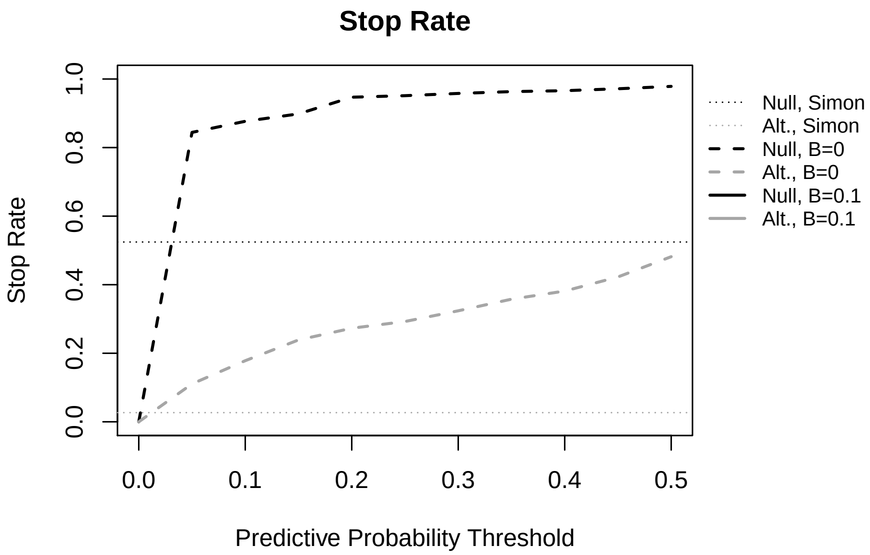


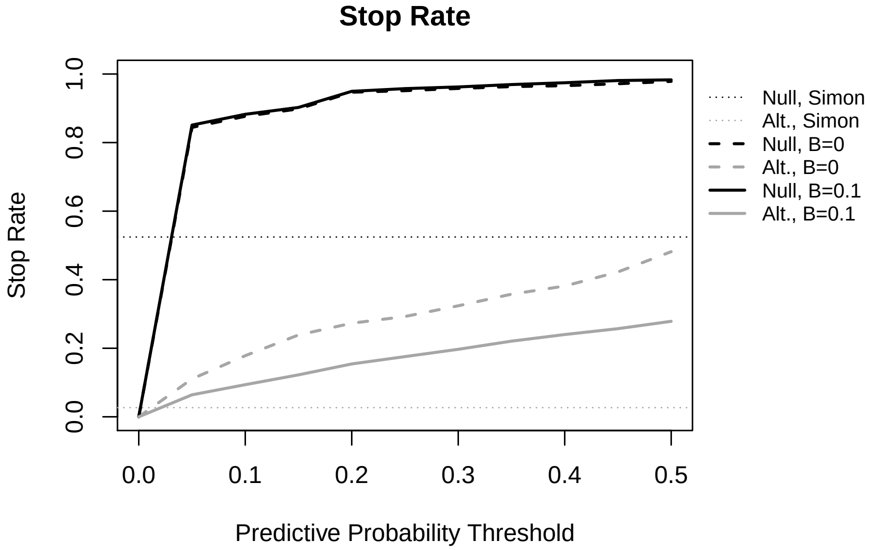
Expected Sample Size
Expected Sample Size
Predictive Probability Threshold
0
5
10
15
20
25
0.0
0.1
0.2
0.3
0.4
0.5
Null, Simon
Alt., Simon
Null, B=0
Alt., B=0
Null, B=0.1
Alt., B=0.1



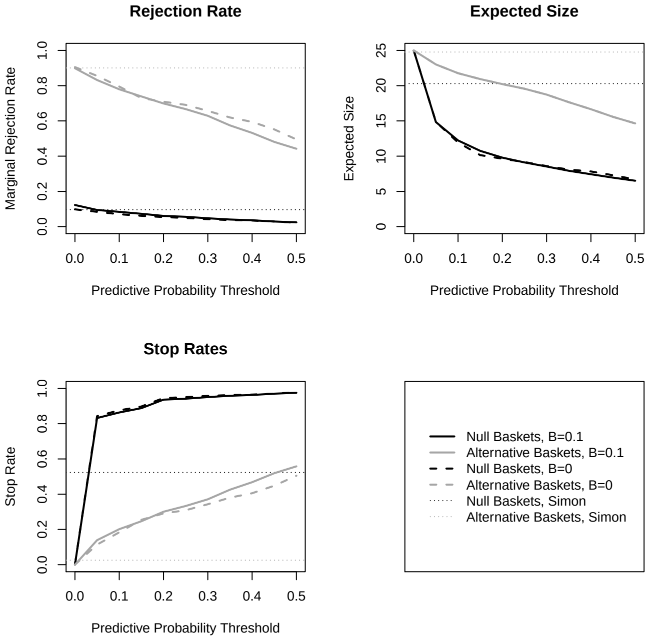
Expected Sample Size
Expected Sample Size
Predictive Probability Threshold
0
5
10
15
20
25
0.0
0.1
0.2
0.3
0.4
0.5
Null, Simon
Alt., Simon
Null, B=0
Alt., B=0
Null, B=0.1
Alt., B=0.1


Stop Rate
Stop Rate
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.1
0.2
0.3
0.4
0.5
Predictive Probability Threshold
Null, Simon
Alt., Simon
Null, B=0
Alt., B=0
Null, B=0.1
Alt., B=0.1


Stop Rate
Stop Rate
Predictive Probability Threshold
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4
0.5
Null, Simon
Alt., Simon
Null, B=0
Alt., B=0
Null, B=0.1
Alt., B=0.1


Stop Rate
Stop Rate
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.1
0.2
0.3
0.4
0.5
Predictive Probability Threshold
Null, Simon
Alt., Simon
Null, B=0
Alt., B=0
Null, B=0.1
Alt., B=0.1

Mixed Scenario Results



Discussion

- Ultimately, there is no free lunch
- The Simon two-stage design is inefficient with respect to many trial operating characteristics, but simple to implement
- Predictive probability monitoring can lead to a much lower expected sample size with only slightly lower power relative to the Simon design
- Information sharing with MEMs can increase power in the global scenario, but should be calibrated if other scenarios are expected (Kaizer et al., 2020)

Sources I

Hobbs, B., Kane, M., Hong, D., and Landin, R. (2018). Statistical challenges posed by uncontrolled master protocols: sensitivity analysis of the vemurafenib study. *Annals of Oncology*.

Kaizer, A. M., Koopmeiners, J. S., Chen, N., and Hobbs, B. P. (2020). Statistical design considerations for trials that study multiple indications. *Statistical Methods in Medical Research*.

Kaizer, A. M., Koopmeiners, J. S., and Hobbs, B. P. (2017). Bayesian hierarchical modeling based on multi-source exchangeability. *Biostatistics*.

Woodcock, J. and LaVange, L. M. (2017). Master protocols to study multiple therapies, multiple diseases, or both. *New England Journal of Medicine*, 377(1):62–70.

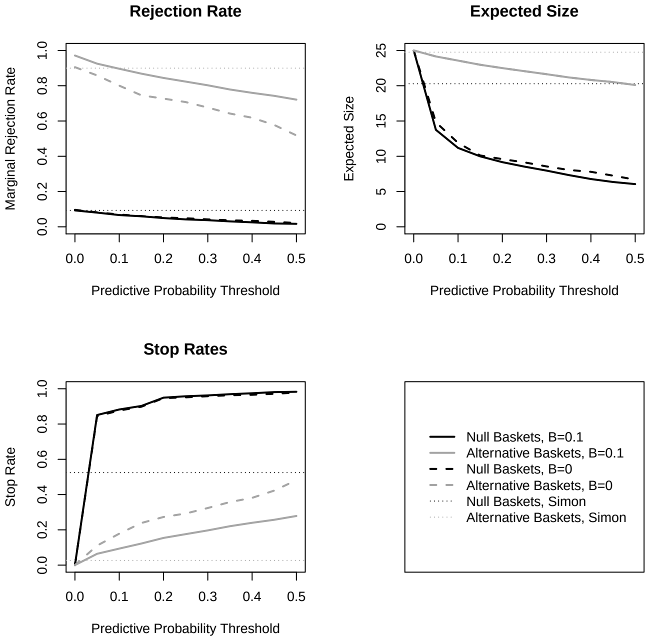
Tabular Results-Global Null

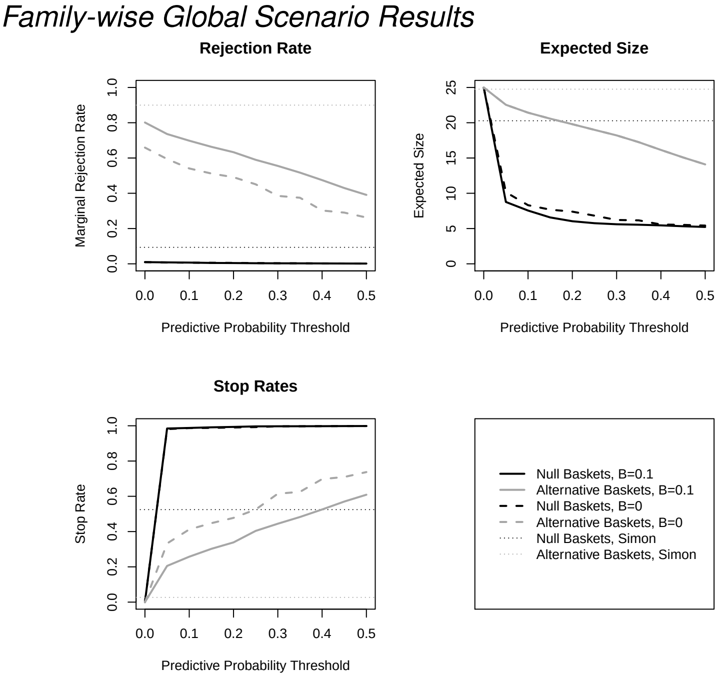
PP Thres hold	Rejection Rate		Family-wise Rate		Expected Size		Stop Rate		Prob. All Null Stop	
	B=0	B=0.1	B=0	B=0.1	B=0	B=0.1	B=0	B=0.1	B=0	B=0.1
0	0.096	0.093	0.635	0.559	25.0	25.0	0.000	0.000	0.000	0.000
0.05	0.082	0.081	0.563	0.543	14.8	13.8	0.844	0.851	0.211	0.242
0.1	0.070	0.067	0.508	0.478	11.9	11.2	0.877	0.882	0.282	0.320
0.15	0.061	0.060	0.453	0.442	10.1	10.0	0.898	0.902	0.355	0.393
0.2	0.053	0.050	0.418	0.386	9.6	9.2	0.947	0.950	0.582	0.614
0.25	0.049	0.042	0.391	0.340	9.1	8.5	0.951	0.957	0.609	0.660
0.3	0.042	0.038	0.348	0.295	8.6	8.0	0.958	0.962	0.652	0.705
0.35	0.037	0.031	0.312	0.247	8.1	7.3	0.963	0.969	0.688	0.753
0.4	0.034	0.025	0.297	0.210	7.8	6.8	0.966	0.975	0.703	0.790
0.45	0.028	0.019	0.253	0.166	7.3	6.4	0.972	0.981	0.747	0.834
0.5	0.021	0.017	0.198	0.145	6.7	6.1	0.979	0.983	0.802	0.855

Tabular Results-Global Alternative

PP Thres hold	Rejection Rate		Expected Size		Stop Rate		Prob. All Null Stop	
	B=0	B=0.1	B=0	B=0.1	B=0	B=0.1	B=0	B=0.1
0	0.905	0.971	25.0	25.0	0.000	0.000	0.000	0.000
0.05	0.859	0.925	23.7	24.2	0.111	0.064	0.000	0.000
0.1	0.801	0.896	22.3	23.6	0.178	0.094	0.000	0.000
0.15	0.746	0.869	21.0	23.0	0.239	0.122	0.000	0.000
0.2	0.727	0.844	20.7	22.5	0.273	0.154	0.000	0.000
0.25	0.707	0.824	20.3	22.1	0.293	0.176	0.000	0.000
0.3	0.676	0.802	19.7	21.6	0.324	0.197	0.000	0.000
0.35	0.642	0.779	19.1	21.2	0.358	0.221	0.000	0.000
0.4	0.619	0.760	18.6	20.8	0.381	0.240	0.000	0.000
0.45	0.577	0.743	17.7	20.5	0.423	0.257	0.000	0.000
0.5	0.518	0.721	16.4	20.1	0.482	0.279	0.000	0.000

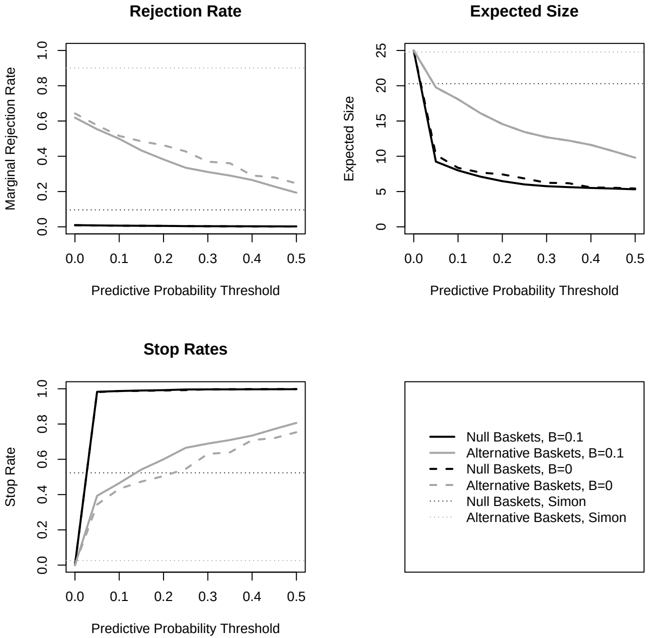
Global Scenario Results





Family-wise Global Scenario Results
Rejection Rate
Marginal Rejection Rate
Predictive Probability Threshold
Expected Size
Expected Size
Predictive Probability Threshold
Stop Rates
Stop Rate
Predictive Probability Threshold
Null Baskets, B=0.1
Alternative Baskets, B=0.1
Null Baskets, B=0
Alternative Baskets, B=0
Null Baskets, Simon
Alternative Baskets, Simon

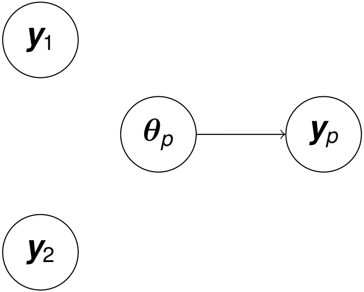
Family-Wise Mixed Scenario Results



Multi-Source Exchangeability Models (MEMs)

- A general Bayesian framework to enable incorporation of independent sources of supplemental information based on Bayesian model averaging across all possible combinations of exchangeability (Kaizer et al., 2017)
- Amount of borrowing determined by exchangeability of data (e.g., equivalent response rates)
- Exchangeability priors specified with respect to sources rather than models
- Using an empirical Bayes prior with single hyperparameter B, where $0 \leq B \leq 1$

Standard Analysis (No Borrowing)



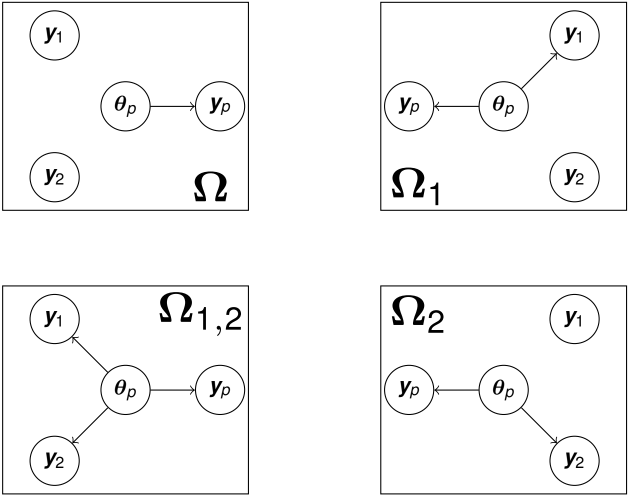
Notation

$\boldsymbol{y}_h$ Observable Data

θ_p Parameters of Interest

$\rightarrow$ Pooled data

MEM Framework



Notation

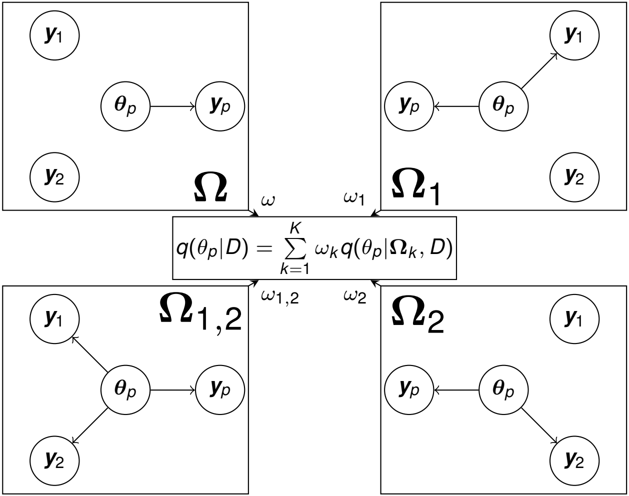
y_h Observable Data

θ_p Parameters of Interest

$\rightarrow$ Pooled data

Ω_k Model

MEM Framework



Notation

$\boldsymbol{y}_h$ Observable Data

$\boldsymbol{\theta}_{\mathrm{p}}$ Parameters of Interest

$\rightarrow$ Pooled data

$\boldsymbol{\Omega}_{\mathrm{k}}$ Model

$\boldsymbol{\omega}_{\mathrm{k}}$ Model weight

Building the MEM framework

- MEM framework leverages the concept of Bayesian model averaging
- Posterior model weights are

$$\omega_k = pr(\Omega_k|D) = \frac{p(D|\Omega_k)\pi(\Omega_k)}{\sum_{j=1}^{K} p(D|\Omega_j)\pi(\Omega_j)},$$

  where $p(D|\Omega_k)$ is the integrated marginal likelihood and $\pi(\Omega_k)$ is the prior belief that Ω_k is the true model
- MEM framework specifies priors with respect to the sources