Comparison of Several Bayesian Methods for Basket Trials When a Control of Subgroup-Wise Error Rate Is Required

301131

Gakuto Ogawa[1], Shogo Nomura[2]

1: Biostatistics Division, Center for Research Administration and Support, National Cancer Center Japan. 2: Department of Biostatistics and Bioinformatics, Graduate School of Medicine, The University of Tokyo

Introduction

Notation: Assuming RR is $\pi_j^{(0)}$ when the drug is inefficacious, $\pi_j^{(1)}$ when the drug is efficacious.

- Subgroup-specific analysis (SS) is the conventional approach in basket trials that assess the efficacy of a new agent across multiple histological subtypes in one trial.
- The notable power gain is expected if one assumes homogeneity of response rates (RRs) in each subgroup and borrows information across subgroups by using a hierarchical Bayesian model (HBM)[1, 5].
- **However, the power gain is seriously lost when "subgroup-wise" (type-I) error rate (SWER) needs to be controlled in a strong sense[2].**
  - For example, in BRAF V600E trial with 6 cancer diseases (described at the later section: "Examples"), the null scenario for CRC.v has total of 32 ($= 2^{6-1}$) variations. The strong control of SWER indicates that the type-I error rate in CRC.v is controlled at the nominal level under all possible variations that investigators concern.
  - Note that, SWER rises in proportion to η_j, which is the threshold value for the decision making (defined at "Methods" section). Also, in a simple situation where RRs in each subset are either $\pi_j^{(0)}$ or $\pi_j^{(1)}$, the SWER is maximized when RR of CRC equals $\pi_j^{(0)}$ and those of the rest equal $\pi_j^{(1)}$ ($\pi_j^{(0)}$ and $\pi_j^{(1)}$ is defined at "Methods" section) .
  - Therefore, to control SWER in a strong sense, more stringent values of η_j must be used.
  - Unfortunately, according to the numerical studies by Freidlin and Korn[2], the power of HBM was equivalent or worse than that of SS in various scenarios when a strong control of SWER is required.
- From a regulatory and patients' perspective, an unnecessary recommendation of inefficacious treatment for further evaluation in future clinical trials or drug labeling should be avoided, and thus, a strong control of SWER is one of vital aspects for actual applications.
- Several newer methods have been proposed; e.g. exchangeability-nonexchangeability model (EXNEX)[4] and multisource exchangeability model (MEM)[3], but their performance was not fully investigated in the situation where a strong control of SWER is required.

Objective

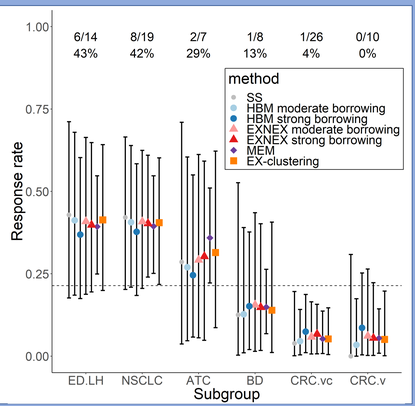
- **To compare the performance of EXNEX and MEM under the control of SWER, setting HBM and SS as a benchmark.**

Examples

- BRAF V600E trial[6] (for case study) (NCT01524978)
  - The primary endpoint was RR based on an imaging at 8 weeks.
  - For illustrative purposes, we here focus on the results of NSCLC, cholangiocarcinoma (bile duct or BD), Erdheim-Chester disease or Langerhans' cell histiocytosis (ED.LH), anaplastic thyroid cancer (ATC), colorectal cancer with monotherapy (CRC.v) and CRC with a combined therapy (CRC.vc).
- TMB-H basket trial[7] (for simulation study) (UMIN000033182)
  - The primary endpoint is RR.
  - Patients are classified to four subgroups.
  - Approximately 70 patients are planned to be enrolled.

Case study

- We applied the methods described in "Methods" to BRAF V600E study.
- Common values of $\pi_{0j} = 15\%$ and $\pi_{1j} = 45\%$ were considered.
- We compared following methods (mean priors were modified)
  - SS: RR is S_j/n_j and Clopper-Pearson 95% CI is estimated
  - HBM moderate borrowing
    - $\mu \sim N(-0.847, 10)$ (centered at logit(0.3))
  - HBM strong borrowing
    - Same mean prior as HBM moderate borrowing
  - EXNEX moderate borrowing
    - $\mu_{EX0} \sim N(-1.735, 6.843)$ (centered at logit(0.15))
    - $\mu_{EX1} \sim N(-0.201, 3.040)$ (centered at logit(0.45))
    - NEX model: $\theta_j \sim N(-0.847, 4.762)$ (centered at logit(0.3))
  - EXNEX strong borrowing
    - Same mean priors as EXNEX moderate borrowing
  - MEM
  - EX-clustering
    - Clustering model:
      - $\mu_{EX0} \sim N(-1.735, 3.76)$ and $\mu_{EX1} \sim N(-0.201, 3.76)$
    - Analytical model:
      - HBM strong borrowing for the efficacious cluster
      - EXNEX moderate borrowing for the inefficacious cluster



Methods

Notation

- Let π_j be the RR in the jth subgroup ($j = 1, \ldots, J$) and π be the RR in the pooled study population.
- $\pi_j^{(M)} = \frac{\pi_j^{(0)} + \pi_j^{(1)}}{2}$, which is the mean of $\pi_j^{(0)}$ and $\pi_j^{(1)}$
- S_j and n_j denote the number of responders and patients in the jth subgroup, respectively.

SS (Subgroup-Specific analysis)

- Defining subgroup-specific threshold and expected values according to disease subgroups and tests the subgroup-specific null hypotheses, $H_{0j}: \pi_j = \pi_{0j}$ versus $H_{1j}: \pi_j = \pi_{1j}$.

HBM (Hierarchical Bayesian Model)[5]

$\theta_1, \ldots, \theta_J \sim i.i.d.\ N(\mu, \tau^{-1})$, where $\theta_j = logit(\pi_j)$.

- Hyperprior distribution
  - $\mu \sim N(logit\left(\pi_j^{(M)}\right), 10)$
  - Moderate borrowing: $\tau^{-1} \sim Gamma(2, 20)$
  - Strong borrowing: $\tau^{-1} \sim Gamma(2, 2)$
- As a criterion to declare a superiority, the following decision rule on the posterior probability given the observed data D is basically used:

$$\Pr(\pi_j > \pi_{0j} | D) > \eta_j$$

MEM (Multisource Exchangeability Model)[3]

- Considering whether all possible subgroup combinations are regarded as exchangeable or non-exchangeable
- The pairwise indicator of exchangeability is represented by a symmetric $J \times J$ matrix $\boldsymbol{\Omega}$; $\Omega_{j,i} = \Omega_{i,j} \in (0,1)$
- Given Ω_j (jth row of $\boldsymbol{\Omega}$), conjugate conditional posterior distribution with the initial prior of $\pi_j \sim beta(0.5, 0.5)$ is yielded as: $q(\pi_j | \boldsymbol{S}, \boldsymbol{\Omega_j}) \propto Beta(0.5 + \sum_{h=1}^{J} \Omega_{j,h} S_h, 0.5 + \sum_{k=1}^{J} \Omega_{j,k}(n_k - S_k))$
- By Bayesian model averaging around $\Pr(\Omega_j = \omega_g | S)$, the marginal posterior distribution can be represented as

$$q(\pi_j | S) \propto \sum_{g=1}^{G} q(\pi_j \mid S, \Omega_j = \omega_g) \, Pr(\Omega_j = \omega_g \mid S)$$

- Initial non-informative probability $\Pr(\Omega_{j,i} = 1) = 0.5$

EXNEX (EXchangeability-NonEXchangeability model)[4]

- Adaptively using multiple exchangeable and non-exchangeable models (EX or NEX models) according to a non-fixed mixture weight p_j for posterior sampling.
- We used two EX models and one NEX model:
  - Null EX model: $\theta_j \sim N(\mu_{EX0}, \tau_{EX}^2)$
    - $\mu_{EX0} \sim N(logit\left(\pi_j^{(0)}\right), \frac{1}{\pi_j^{(0)}} + \frac{1}{1-\pi_j^{(0)}} - 1)$
  - Alternative EX model: $\theta_j \sim N(\mu_{EX1}, \tau_{EX}^2)$
    - $\mu_{EX1} \sim N\left(\pi_j^{(1)}\right), \frac{1}{\pi_j^{(1)}} + \frac{1}{1-\pi_j^{(1)}} - 1)$
  - NEX model: $\theta_j \sim N\left(logit\left(\pi_j^{(M)}\right), \frac{1}{\pi_j^{(M)}} + \frac{1}{1-\pi_j^{(M)}}\right)$
- Two patterns of hyperprior τ_{EX}:
  - Moderate borrowing: $\tau_{EX} \sim Half - Normal(1)$
  - Strong borrowing: $\tau_{EX} \sim Half - Normal(0.5)$
- Initial values for mixture weight $p_j = (0.25, 0.25, 0.50)$

EX-clustering

- First, dividing all subgroups into an efficacious cluster and an inefficacious cluster based on EXNEX with null and alternative EX model without the NEX model.
  - EX-clustering selects the null EX model or the alternative EX model with p_j or $1 - p_j$.
  - When a posterior weight p_j exceeds ξ, subgroup j is labeled as the efficacious cluster, and vice versa.
  - All subgroups are labeled by the above procedure and divided into two clusters as a result.
- Second, in efficacious cluster, HBM with strong borrowing are performed, and EXNEX with moderate borrowing are done in inefficacious cluster at the same time.

Settings

- We performed the simulation under the situation of TMB-H basket trial:
  - Total sample size: 70
  - Number of subgroups: 4
- We considered total of 8 scenarios summarized in the right table.
  - In addition to the threshold and expected values (10% and 30%), mixed situations where RR of 50% were also considered.
- SWER is strongly controlled at 10% in scenario 1.
  - η_j for each method was calibrated.
- Under the same enrollment rate across all subgroups, patient's binary response was generated from Bernoulli distribution in each scenario. We repeated generating datasets for 1,000 times.
- Performance measures: Proportion of claiming efficacy and $\Pr(\pi_j > \pi_{0j} | D)$
  - (data not shown: bias and CI)

	Response rate by each subgroup			
	1	2	3	4
Scenario 1	0.1	**0.3**	**0.3**	**0.3**
Scenario 2	0.1	0.1	**0.3**	**0.3**
Scenario 3	0.1	0.1	0.1	**0.3**
Scenario 4	0.1	0.1	0.1	0.1
Scenario 5	**0.3**	**0.3**	**0.3**	**0.3**
Scenario 6*	0.1	**0.5**	**0.5**	**0.5**
Scenario 7*	0.1	**0.3**	**0.5**	**0.5**
Scenario 8*	0.1	0.1	**0.3**	**0.5**

Results

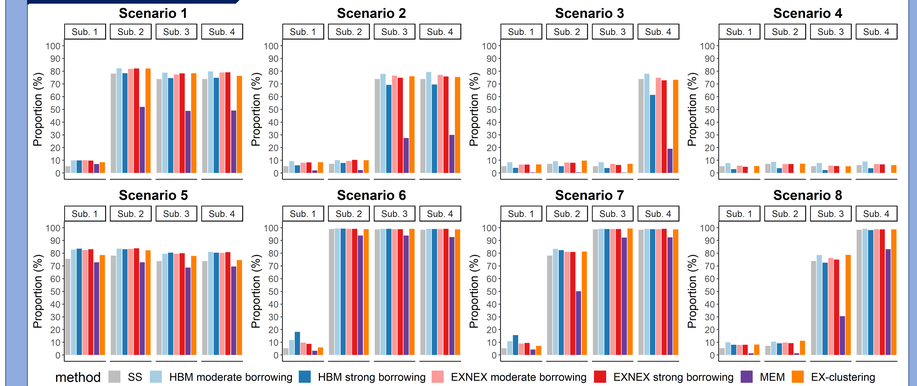


Fig. 1: Proportion of claiming efficacy

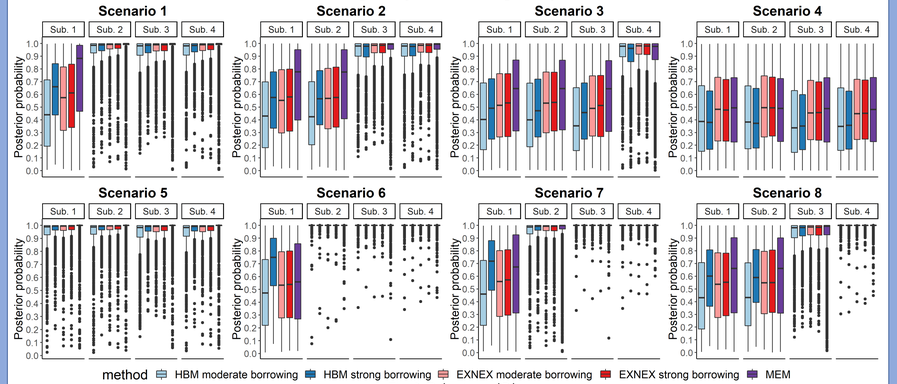


Fig. 2: $\Pr(\pi_j > \pi_{0j} | D)$

Conclusion

- **HBM and EXNEX with moderate borrowing can achieve relatively high power when a control of SWER is required.**
- **Considering a control of type-I error rate in the situations where RR is higher than expected, EXNEX might be a good choice.**

Reference

1. Berry SM, Broglio KR, et al. Bayesian hierarchical modeling of patient subpopulations: efficient designs of Phase II oncology clinical trials. Clin Trials. 2013;10(5):720-34.
2. Freidlin B, Korn EL. Borrowing information across subgroups in phase II trials: is it useful? Clin Cancer Res. 2013;19(6):1326-34.
3. Hobbs BP, Landin R. Bayesian basket trial design with exchangeability monitoring. Stat Med. 2018;37(25):3557-72.
4. Neuenschwander B, Wandel S, et al. Robust exchangeability designs for early phase clinical trials with multiple strata. Pharm Stat. 2016;15(2):123-34.
5. Thall PF, Wathen JK, et al. Hierarchical Bayesian approaches to phase II trials in diseases with multiple subtypes. Stat Med. 2003;22(5):763-80.
6. Hyman DM, Puzanov I, et al. Vemurafenib in Multiple Nonmelanoma Cancers with BRAF V600 Mutations. N Engl J Med. 2015;373(8):726-36.
7. Nakamura Y, Komatsu Y, et al. bTMB-High Basket trial: A multicenter phase II trial of nivolumab monotherapy in patients with advanced gastrointestinal cancers with high blood tumor mutational burden (bTMB). ASCO Annual Meeting; 2019.