

Comparative Study of Bayesian Information Borrowing Methods in Oncology Clinical Trials

Liwen Su, PhD¹; Xin Chen, MS¹; Jingyi Zhang, MS¹; and Fangrong Yan, PhD¹

PURPOSE With deeper insight into precision medicine, more innovative oncology trial designs have been proposed to contribute to the characteristics of novel antitumor drugs. Bayesian information borrowing is an indispensable part of these designs, which shows great advantages in improving the efficiency of clinical trials. Bayesian methods provide an effective framework when incorporating information. However, the key point lies in how to choose an appropriate method for complex oncology clinical trials.

METHODS We divided the borrowing information scenarios into concurrent and nonconcurrent scenarios according to whether the data to be borrowed are observed at the same time as in the current trial or not. Then, we provided an overview of the methods in each scenario. Performance comparison of different methods is carried out with regard to the type I error and power.

RESULTS As demonstrated by the simulation results in each borrowing scenario, the Bayesian hierarchical model and its extensions are more appropriate for concurrent borrowing. The simulation results demonstrate that the Bayesian hierarchical model shows great advantages when the arms are homogeneous. However, such a method should be adopted with caution when heterogeneity exists. We recommend the other methods, considering heterogeneity. Borrow information from informative priors is more suggested for nonconcurrent borrowing scenarios. Multisource exchangeability models are more suitable for multiple historical trials, while meta-analytic-predictive prior should be carefully applied.

CONCLUSION Bayesian information borrowing is useful and can improve the efficiency of clinical trial designs. However, we should carefully choose an appropriate information borrowing method when facing a practical innovative oncology trial, as an appropriate method is essential to provide ideal design performance.

JCO Precis Oncol 6:e2100394. © 2022 by American Society of Clinical Oncology

Creative Commons Attribution Non-Commercial No Derivatives 4.0 License 

INTRODUCTION

Recently, with the development of cancer molecular biology, antitumor therapy has entered the era of precision medicine. A number of innovative therapies have emerged, including immunotherapies, targeted therapies, and cancer vaccines.¹⁻⁵ The emergence of these innovative drugs has introduced new challenges to the design of oncology clinical trials.⁶ More innovative methods have been proposed, and some of them are also encouraged by the US Food and Drug Administration (FDA), such as master protocol trials⁷ and Complex Innovative Trial Designs,⁸ which raised several discussions.

In contrast to traditional trial designs, these innovative methods encourage the use of Bayesian methods.⁹ One of the most common scenarios is borrowing information when multiple parallel arms exist. For example, we designed a master protocol clinical trial that encourages borrowing information within multiple substudies. The

I-SPY2 trial is a success story for a platform trial that uses a Bayesian hierarchical model (BHM) to adaptively borrow information between running arms.¹⁰ Another innovative paradigm is that historical information from external data forms the design of the current trial. Using large-scale real-world clinical data sets and high-quality completed medical data has become a new trend in clinical trials. By incorporating historical information, these methods offer the possibility of a substantially reduced sample size because of the efficient usage of external data.

Information borrowing is an innovative technique of great importance in current clinical trials. It can accelerate the trial process and reduce the cost, thus ultimately improving the efficiency of clinical trials.¹¹ According to the needs of current cancer clinical trials, information borrowing strategies can be divided into two strategies according to when the information to be borrowed is observed, similar to the division of the

ASSOCIATED CONTENT

Data Supplement

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on January 28, 2022 and published at ascopubs.org/journal/po on March 9, 2022; DOI <https://doi.org/10.1200/P0.21.00394>

CONTEXT

Key Objective

How do we choose an appropriate information borrowing method with so many methodologies when conducting a complex innovative oncology trial?

Knowledge Generated

Borrowing information can increase power, reduce type I error, and improve the efficiency when trials are homogeneous. However, borrowing methods should be applied with caution when heterogeneity exists.

Relevance

When designing new clinical trials which can borrow external information to accelerate the drug development, clinical investigators can refer to this study to select more efficient and appropriate methods.

control group in ICH-E10.¹² The first strategy is concurrent borrowing that is applied in the master protocol. Concurrent borrowing occurs when multiple parallel arms exist, such as basket and platform trials. Arms under these circumstances are of equal importance and are analyzed simultaneously without a chronologic order. We focus on the estimation of all these parameters by using all arm information. Another strategy for incorporating external data is nonconcurrent borrowing. For nonconcurrent borrowing, there is only one primary trial and others are recognized as supplementary trials, which are not of main interest. Analyses were conducted in proper sequence, and we mainly focused on estimation of parameters in the primary trial. In this case, these supplementary trials can be regarded as external or historical data. In some seamless clinical trial designs, borrowing information between different stages also belongs to this situation.

Several Bayesian information borrowing methods have been developed, including concurrent and nonconcurrent borrowing. However, the proper method should be adopted to improve the efficiency of our trial. On the basis of the existing research, we conduct a more in-depth comparison and recommendation for application scenarios on methods for borrowing information.

METHODS

As discussed above, there are two scenarios for information borrowing: concurrent borrowing and nonconcurrent borrowing. The Bayesian method shows nonsubstitutable advantages because it naturally incorporates historical information through the setting of prior distribution. Some FDA guidelines also encourage the use of the Bayesian method in this situation.⁹ Ibrahim first proposed the power prior (PP) method, which is a pioneering work in this field.¹³ Since then, a large number of methods have been proposed. The PP raises the likelihood of historical data to the prespecified power α . Many similar methods have been developed under this framework, such as the modified power prior (MPP),¹⁴ calibrated power prior (CPP),¹⁵ P value–based power prior (PvPP),¹⁶ commensurate power

prior (CP),¹⁷ meta-analytic-predictive prior (MAP),¹⁸ robust meta-analytic-predictive prior (RMAP),¹⁹ and multisource exchangeability models (MEMs).²⁰ These methods mainly borrow information from informative priors, which are used for nonconcurrent borrowing to a great extent. Another important method of borrowing information is BHM,²¹ proposed by Berry, using variance in a hierarchical model to control the extent of borrowing. BHM has been developed into different variants, such as the calibrated Bayesian hierarchical model (CBHM),²² Bayesian hierarchical classification and information sharing (BaCIS),²³ and Bayesian cluster hierarchical model (BCHM).²⁴ Although MEMs are not variants of BHM, MEMs have been extended to basket trials.²⁵⁻²⁶ Application scenarios and the features of each method are listed in Table 1. In addition, the details of these methods are described in the Data Supplement.

Concerning a novel clinical trial to evaluate the efficacy of an anticancer drug candidate, the primary end point was assumed to be the objective response rate (ORR) as determined by RECIST version 1.1. First, we investigated the scenarios in which concurrent borrowing occurs, along with other drug candidates to compose a parallel multiarm trial. A basket trial was hypothesized in this situation. By contrast, assuming that external data or supplementary trials exist, the efficacy was studied through nonconcurrent borrowing, including single historical trials and multiple historical trials.

Detailed configurations of competing scenarios are provided in the Data Supplement. We examined eight scenarios of concurrent borrowing, as shown in the Data Supplement. Scenarios 1, 2, and 7 denote homogeneous efficacy between the four arms, and scenarios 3, 4, 5, 6, and 8 denote heterogeneous efficacy to different extents. For a single historical trial in nonconcurrent borrowing, the continuous change in type I error and power was investigated with regard to different response rates. The other six scenarios with manifold heterogeneity modes are considered in multiple historical trials, as shown in the Data Supplement. The number of simulated trials is 10,000.

TABLE 1. Comparison of Typical Borrowing Information Methods

Application Scenarios	Methods	Features
Concurrent borrowing	BHM	Variance in hierarchical model reflects heterogeneity and estimate it using fully Bayesian method
	CBHM	Variance is a function of heterogeneity measurement
	BaCIS	Dichotomous cluster between arms and borrow information within each cluster
	BCHM	Nonparametric clustering method is used to dynamically determine the number of clusters
	MEMs	Specify all possible pairwise exchangeability models among arms by a symmetric matrix and weight above models
Nonconcurrent borrowing	PP	Downweigh the historical trial by a prespecified power α
	MPP	α follows a vague prior and is estimated by combined data
	CPP	α is a function of heterogeneity measurement
	CP	Parameter in the current trial centers around the historical trial
	MAP	Parameters in historical trials and the current trial come from the same normal distribution with variance reflecting heterogeneity
	RMAP	A hybrid prior is constructed by weighted noninformative prior and MAP
	MEMs	Specify all possible pairwise exchangeability models among trials by a vector and weight above models

Abbreviations: BaCIS, Bayesian hierarchical classification and information sharing; BCHM, Bayesian cluster hierarchical model; BHM, Bayesian hierarchical model; CBHM, calibrated Bayesian hierarchical model; CP, commensurate power prior; CPP, calibrated power prior; MAP, meta-analytic-predictive prior; MEM, multisource exchangeability model; MPP, modified power prior; PP, power prior; RMAP, robust meta-analytic-predictive prior.

RESULTS

Concurrent Scenarios

The methods to be compared include independent analysis, BHM, CBHM, BaCIS, BCHM, and MEMs. The probabilities of rejecting the null hypothesis for each indication are shown in [Table 2](#) and the Data Supplement. Furthermore, additional sensitivity analyses can be found in the Data Supplement.

Compared with independent analysis, all five hierarchical models can control type I error and increase power when the arms are homogeneous (scenario 1, 2, and 7). BHM gives better results, followed by MEMs. However, in heterogeneous scenarios (scenarios 3-8), BHM is inferior to other models. For example, in scenarios 4 and 6, where most arms are effective and there is only one ineffective arm, BHM reveals unacceptable type I error inflation, up to 0.2825 and 0.2497, respectively. When the number of effective arms is dominant (scenarios 4, 6, and 8), MEMs show great advantages as MEMs reach a high performance in power similar to BHM, while type I error is much smaller than BHM. However, in scenario 5 where most arms are ineffective, information borrowing from ineffective arms does not bring power advantages to MEMs. BaCIS and CBHM show higher efficiency, exhibiting high power while maintaining type I error in a relatively acceptable range, which is followed by BCHM. CBHM tends to maintain the lowest type I errors. This is caused by the parameter calibrating for a and b in $\exp\{a + b \times \log(\chi^2)\}$. The control can also be relaxed by a different parameter setting according to risk preference. In scenario 3, where the number of effective and ineffective arms is balanced, BaCIS is the most effective, reflecting the advantage of dichotomous clustering. It is worth mentioning that, with regard to indications 2 and 4 in scenario 6, borrowing information simultaneously from arms with response

rates of 0.15 and 0.4, the biases in both arms offset each other, leading to an unbiased result. This is unreasonable and requires special attention in practice. The results for BCHM lie in between BaCIS and CBHM. Theoretically, BCHM allows adaptive determination of the number of clusters. However, dividing the arms into effective and ineffective arms by a hypothesis test is actually a dichotomous result. Under such a hypothesis test, more clustering shows no advantages. Detailed results for the root mean square error of estimation for ORR for each indication can be found in the Data Supplement.

Single Historical Trial in Nonconcurrent Scenarios

The operating characteristics of each method are shown in [Figure 1](#). Additional sensitivity analyses can be found in the Data Supplement. When the ORR in the current control arm is the same as the ORR in the historical trial (30%), borrowing information can result in a reduced type I error ([Fig 1A](#)), increased power ([Fig 1C](#)), and an unbiased estimation of the effect size ([Figs 1B](#) and [1D](#)). This means that in case of homogeneity between the current control arm and the historical trial, all information borrowing methods are favorable.

However, when trials were heterogeneous, the operating characteristics differed between the methods. Taking PP ($\alpha = 0$), that is, independent analysis, as the benchmark, PP with $\alpha \geq .5$ can lead to a sharp change in type I error, power, and bias. The range of change was positively correlated with α . In the most extreme cases, PP ($\alpha = 1$), that is, pooled analysis, can produce a type I error higher than 0.3 ([Fig 1A](#)) and a power lower than 0.6 ([Fig 1C](#)). The PP method is too sensitive to heterogeneity, which is unacceptable in practice. In the other five methods that consider the assessment of heterogeneity, the curves are typically S-shaped and control type I/II errors within an acceptable

TABLE 2. Rejection Rate of the Null Hypothesis for Concurrent Scenarios

Scenario	Method	ORR of Each Indication				Scenario	Method	ORR of Each Indication			
		1	2	3	4			1	2	3	4
1		0.15	0.15	0.15	0.15	5		0.15	0.15	0.15	0.3
	Independent	0.0499	0.0477	0.0483	0.0480		Independent	0.0499	0.0477	0.0483	0.7887
	BHM	0.0215	0.0212	0.0214	0.0229		BHM	0.0789	0.0755	0.0761	0.5584
	CBHM	0.0603	0.0591	0.0594	0.0594		CBHM	0.0625	0.0605	0.0595	0.7866
	BaCIS	0.0487	0.0461	0.0465	0.0470		BaCIS	0.0680	0.0652	0.0664	0.7819
	BCHM	0.0346	0.0346	0.0355	0.0354		BCHM	0.0605	0.0594	0.0614	0.7304
	MEMs	0.0544	0.0536	0.0544	0.0538		MEMs	0.0850	0.0884	0.0803	0.7691
2		0.3	0.3	0.3	0.3	6		0.4	0.3	0.15	0.3
	Independent	0.7868	0.7880	0.7922	0.7887		Independent	0.9851	0.7880	0.0483	0.7887
	BHM	0.9562	0.9574	0.9605	0.9587		BHM	0.9931	0.8814	0.2497	0.8842
	CBHM	0.8353	0.8351	0.8419	0.8360		CBHM	0.9866	0.7910	0.0624	0.7924
	BaCIS	0.8623	0.8623	0.8650	0.8631		BaCIS	0.9909	0.8709	0.1053	0.8712
	BCHM	0.8494	0.8523	0.8554	0.8529		BCHM	0.9887	0.8170	0.0898	0.8141
	MEMs	0.9187	0.9170	0.9161	0.9185		MEMs	0.9935	0.8553	0.1421	0.8612
3		0.3	0.15	0.15	0.3	7		0.1	0.1	0.15	0.15
	Independent	0.7868	0.0477	0.0483	0.7887		Independent	0.0025	0.0023	0.0483	0.0480
	BHM	0.7784	0.1490	0.1520	0.7825		BHM	0.0002	0.0002	0.0055	0.0080
	CBHM	0.7941	0.0655	0.0657	0.7960		CBHM	0.0044	0.0041	0.0494	0.0495
	BaCIS	0.8152	0.0815	0.0855	0.8184		BaCIS	0.0023	0.0023	0.0404	0.0410
	BCHM	0.7983	0.0695	0.0731	0.8004		BCHM	0.0016	0.0015	0.0263	0.0282
	MEMs	0.8322	0.1245	0.1221	0.8320		MEMs	0.0035	0.0035	0.0442	0.0479
4		0.15	0.3	0.3	0.3	8		0.3	0.3	0.25	0.35
	Independent	0.0499	0.7880	0.7922	0.7887		Independent	0.7868	0.7880	0.5293	0.9324
	BHM	0.2825	0.8855	0.8931	0.8892		BHM	0.9466	0.9447	0.8488	0.9881
	CBHM	0.0938	0.8038	0.8069	0.8044		CBHM	0.8217	0.8234	0.5875	0.9512
	BaCIS	0.0937	0.8466	0.8507	0.8483		BaCIS	0.8634	0.8651	0.6617	0.9595
	BCHM	0.0925	0.8240	0.8289	0.8225		BCHM	0.8393	0.8458	0.6325	0.9535
	MEMs	0.1718	0.8807	0.8844	0.8889		MEMs	0.9070	0.9066	0.7376	0.9804

NOTE. Bold values represent power in effective arms, whereas others represent type I error. The results are computed on the basis of 10,000 simulated trials under each of the eight scenarios shown in the Data Supplement. The maximum sample size is 43 for each arm.

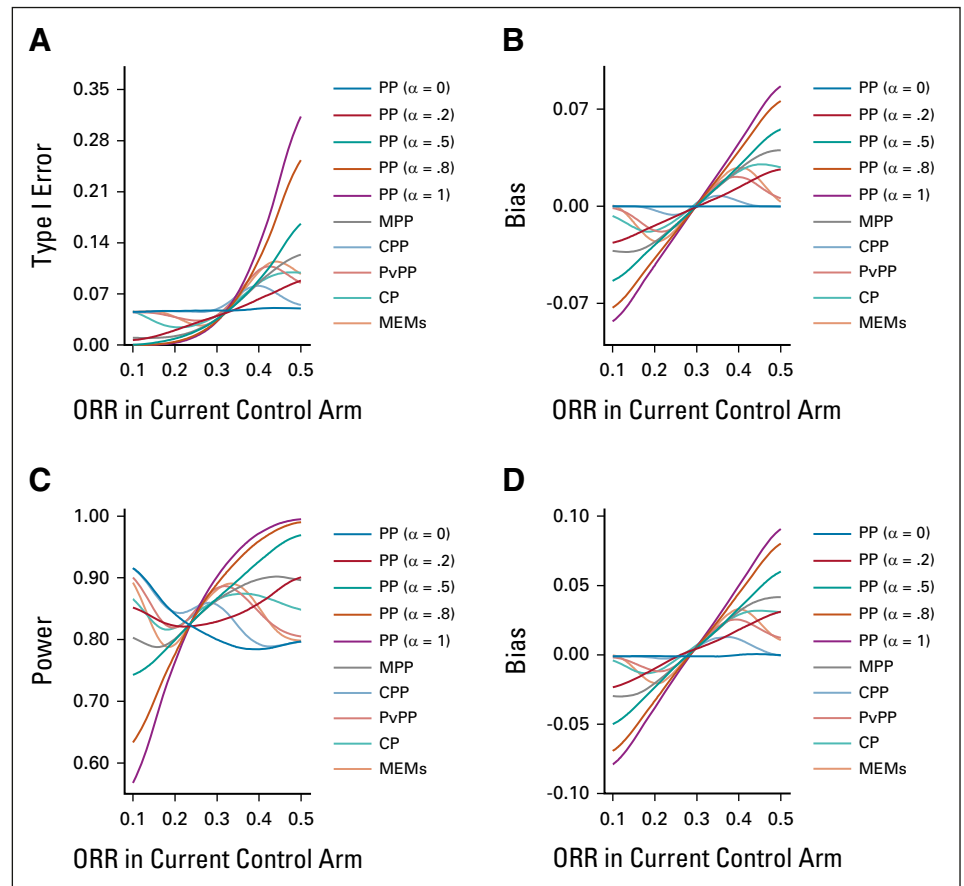
Abbreviations: BaCIS, Bayesian hierarchical classification and information sharing; BCHM, Bayesian cluster hierarchical model; BHM, Bayesian hierarchical model; CBHM, calibrated Bayesian hierarchical model; MEM, multisource exchangeability model; ORR, objective response rate.

range. This is because these methods can reduce the amount of borrowed information when the current trial data show great heterogeneity from that of historical trials. Concerning bias, the changes in these curves are similar to those of the type I/II error. The curve of bias in PP, except PP with $\alpha = 0$, has a linear relationship with p_c , whereas the curves of bias in the five methods that consider heterogeneity are S-shaped and the estimation bias will not be outrageous, reflecting a tradeoff between precision gains and bias loss.

As for the five models that consider heterogeneity, differences were detected between them in this simulation. First, in MPP, α is estimated by only one historical trial and one

current control arm and will not be accurate in most cases. As a result, MPP does not have a strong control on type I/II error when heterogeneity exists; so, type I error continuously increases, as shown in Figure 1. By contrast, MPP is a data-driven method that restricts researchers' opinions on information borrowing. CP has a smaller Type I error than MPP and finally reaches a plateau in that it introduces commensurability parameters to measure heterogeneity. In CPP, P_vPP, and MEMs, ideal statistical performance can be achieved as long as we carefully set those design parameters. Theoretically, these three methods can achieve very similar tradeoff profiles after sufficient calibrations of the parameters. When the ORR in the current control arm is

FIG 1. Simulation results for scenarios of a single historical trial. (A) Type I error under the null hypothesis against different ORR in current control arm. (B) Bias under the null hypothesis against different ORR in current control arm. (C) Power under the alternative hypothesis against different ORR in current control arm. (D) Bias under the alternative hypothesis against different ORR in current control arm. CP, commensurate prior; CPP, calibrated power prior; MAP, meta-analytic-predictive prior; MEMs, multisource exchangeability models; MPP, modified power prior; ORR, objective response rate; PP, power prior; PvPP, P value-based power prior.



substantially larger than that in the historical arm, they can identify the difference and reduce the borrowing, resulting in decreased type I error. Among these methods, CPP poses more control on type I error, whereas PvPP and MEMs achieve higher power. It should be pointed out that, when using CPP, it is sometimes difficult to specify the boundaries between fully homogeneous and fully heterogeneous situations. Therefore, CPP is more suitable for bioequivalence trials, in which we have an equivalence boundary usually defined by regulators [0.8, 1.25]. In a single historical trial, the performance of MEMs is similar to that of PvPP. However, the difference between them is more obvious in multiple historical trials, owing to the different borrowing mechanism.

Multiple Historical Trials in Nonconcurrent Scenarios

For multiple historical trials, the performance metrics were the same as those in the simulation of a single historical trial. The simulation results are presented in Table 3. Additional sensitivity analyses can also be found in the Data Supplement.

PP, MPP, PvPP, and CP produced compromise results between independent and pooled analysis, similar to those in single historical trials. In most scenarios (Data Supplement), the type I error and power curves for MEMs, CPP, MAP, and RMAP are bounded by independent analysis and pooled analysis, reflecting the nature of borrowing

information as a discounting of historical data, except for scenario 6 (Data Supplement), as well as scenarios 2 and 5 (Data Supplement). Considering scenario 6, MAP and RMAP have a larger type I error than the independent analysis. One possible interpretation is that the prior mean for scenario 4 is closer to 0.3 than that in scenario 5, which is an asymmetric mapping (Table 4). Therefore, for scenario 6, MAP and RMAP borrow more from historical trials that are lower than 0.3, which results in type I error inflation. For scenario 2, CPP outperforms independent analysis because the parameter settings in CPP allow it to regard a difference of 0.1 as almost totally heterogeneous (Data Supplement). Therefore, in scenario 2, CPP can exclude historical trials 1 and 2, and only borrow information from historical trial 3, making higher power than the independent method. MEMs fail to identify a difference of 0.1, so they tend to borrow information and have a relatively low power in scenario 2. Regarding scenario 5, where the difference can reach up to 0.2, MEMs successfully identify homogeneous trials and exclude the heterogeneous one (Data Supplement). This scenario is more meaningful for practical applications, as it is difficult to statistically explain whether a smaller difference (such as 0.1) should be borrowed. As a result, MEMs have an excellent performance in scenario 5, which is significantly better than the independent analysis, along with CPP. In the other scenarios in the Data Supplement, the methods mainly lie

TABLE 3. Results for Scenarios of Multiple Historical Trials

Scenario	Method	$P_T = .3$		$P_T = .5$		Scenario	Method	$P_T = .3$		$P_T = .5$	
		Type I Error	Bias	Power	Bias			Type I Error	Bias	Power	Bias
1	PP ($\alpha = 0$)	0.046	0.000	0.803	0.001	4	PP ($\alpha = 0$)	0.046	0.000	0.803	0.001
	PP ($\alpha = .2$)	0.039	0.000	0.832	-0.003		PP ($\alpha = .2$)	0.059	0.009	0.859	0.006
	PP ($\alpha = .5$)	0.035	0.000	0.864	-0.002		PP ($\alpha = .5$)	0.070	0.020	0.921	0.017
	PP ($\alpha = .8$)	0.033	0.001	0.887	-0.002		PP ($\alpha = .8$)	0.086	0.027	0.954	0.025
	PP ($\alpha = 1$)	0.032	0.001	0.899	-0.001		PP ($\alpha = 1$)	0.098	0.031	0.966	0.029
	MPP	0.035	0.000	0.863	-0.002		MPP	0.067	0.016	0.903	0.014
	CPP	0.048	0.000	0.860	-0.002		CPP	0.047	0.001	0.847	-0.002
	PvPP	0.043	0.001	0.881	-0.001		PvPP	0.087	0.005	0.874	0.003
	CP	0.040	0.002	0.862	0.000		CP	0.074	0.019	0.888	0.017
	MEMs	0.033	0.001	0.899	-0.001		MEMs	0.052	0.006	0.886	0.004
	MAP	0.042	0.002	0.835	0.000		MAP	0.089	0.018	0.915	0.015
RMAP	0.050	0.002	0.844	-0.001	RMAP	0.091	0.017	0.916	0.014		
2	PP ($\alpha = 0$)	0.046	0.000	0.803	0.001	5	PP ($\alpha = 0$)	0.046	0.000	0.803	0.001
	PP ($\alpha = .2$)	0.026	-0.012	0.790	-0.014		PP ($\alpha = .2$)	0.028	0.010	0.800	-0.012
	PP ($\alpha = .5$)	0.013	-0.023	0.779	-0.025		PP ($\alpha = .5$)	0.016	0.019	0.798	-0.021
	PP ($\alpha = .8$)	0.009	-0.031	0.771	-0.033		PP ($\alpha = .8$)	0.011	0.025	0.795	-0.028
	PP ($\alpha = 1$)	0.007	-0.035	0.767	-0.037		PP ($\alpha = 1$)	0.009	0.029	0.795	-0.031
	MPP	0.014	-0.022	0.780	-0.025		MPP	0.019	0.017	0.800	-0.019
	CPP	0.046	-0.002	0.825	-0.005		CPP	0.047	0.001	0.846	-0.002
	PvPP	0.037	-0.023	0.795	-0.024		PvPP	0.035	0.005	0.810	-0.008
	CP	0.024	-0.018	0.790	-0.019		CP	0.025	0.015	0.804	-0.017
	MEMs	0.021	-0.028	0.776	-0.030		MEMs	0.032	-0.008	0.843	-0.009
	MAP	0.035	-0.007	0.803	-0.010		MAP	0.038	-0.004	0.807	-0.007
RMAP	0.038	-0.007	0.808	-0.010	RMAP	0.040	-0.004	0.809	-0.006		
3	PP ($\alpha = 0$)	0.046	0.000	0.803	0.001	6	PP ($\alpha = 0$)	0.046	0.000	0.803	0.001
	PP ($\alpha = .2$)	0.059	0.009	0.859	0.006		PP ($\alpha = .2$)	0.039	0.000	0.832	-0.003
	PP ($\alpha = .5$)	0.070	0.020	0.921	0.017		PP ($\alpha = .5$)	0.035	0.000	0.864	-0.002
	PP ($\alpha = .8$)	0.086	0.027	0.954	0.025		PP ($\alpha = .8$)	0.033	0.001	0.887	-0.002
	PP ($\alpha = 1$)	0.098	0.031	0.966	0.029		PP ($\alpha = 1$)	0.032	0.001	0.899	-0.001
	MPP	0.070	0.019	0.912	0.017		MPP	0.038	0.001	0.842	-0.003
	CPP	0.073	0.002	0.836	0.001		CPP	0.047	0.001	0.831	-0.003
	PvPP	0.087	0.020	0.874	0.018		PvPP	0.043	0.000	0.881	-0.002
	CP	0.074	0.019	0.888	0.017		CP	0.040	0.002	0.862	0.000
	MEMs	0.095	0.026	0.925	0.025		MEMs	0.056	-0.001	0.812	-0.002
	MAP	0.090	0.020	0.928	0.017		MAP	0.081	0.006	0.883	0.004
RMAP	0.093	0.019	0.928	0.016	RMAP	0.083	0.006	0.886	0.004		

NOTE. The results are computed on the basis of 10,000 simulated trials under the six scenarios whose specific details are shown in the Data Supplement. Assume there are three historical trials with the sample size of each historical trial being 20. Again, the maximum sample size for the current trial is 72.

Abbreviations: CP, commensurate power prior; CPP, calibrated power prior; MAP, meta-analytic-predictive prior; MEM, multisource exchangeability model; MPP, modified power prior; PP, power prior; PvPP, P value-based power prior; RMAP, robust meta-analytic-predictive prior.

between independent and pooled analysis. In conclusion, pairwise comparison strategies (such as MEMs) are more effective in multiple historical trials than combined comparisons (such as PP and PvPP).

No significant advantages were observed for MAP or RMAP compared with other methods in these scenarios. One possible reason is the bias of the prior mean. The other reason is that MAP is not sensitive to the standard

TABLE 4. MAP Prior Generated in Each Scenario

Scenarios	P_{H1}	P_{H2}	P_{H3}	MAP Prior	Prior Mean	Prior Standard Deviation
1	.3	.3	.3	0.010Beta(1.66, 45.1) + 0.990Beta(3.67, 7.93)	0.305	0.114
2	.42	.42	.3	0.010Beta(2.04, 45.2) + 0.990Beta(3.88, 5.94)	0.384	0.126
3	.2	.2	.3	0.761Beta(6.65, 21.9) + 0.239Beta(1.07, 2.45)	0.242	0.112
4	.1	.3	.3	0.685Beta(4.88, 16.7) + 0.315Beta(0.90, 2.03)	0.244	0.129
5	.5	.3	.3	0.013Beta(1.67, 45.26) + 0.987Beta(3.17, 5.05)	0.372	0.138
6	.1	.3	.5	0.547Beta(2.90, 7.52) + 0.453Beta(0.86, 1.48)	0.311	0.176

Abbreviation: MAP, meta-analytic-predictive prior.

deviation in case of different heterogeneous scenarios owing to the smaller number of historical trials, and mutual transformation of binary variables and continuous variables (Table 4). Taking scenario 1 and 3 as an example, we expected that MAP would have a smaller standard deviation than RMAP, but the truth is the opposite. We could not obtain a precise variance because there were only three historical trials. By contrast, when the binary variable ORR is converted into continuous variable and hierarchical models for meta-analysis are constructed, the logit function has a natural property that $\text{logit}(p)$ varies in a relatively small range, although p has significantly changed. The possibility of detecting heterogeneity was further compressed after the transformation. Compared with MAP, RMAP can solve this problem to some extent by introducing a noninformative part of the prior. As a result, we can see a smaller bias in RMAP than in MAP as shown in Table 3, when the historical trials are not completely homogeneous.

DISCUSSION

In conclusion, Bayesian information borrowing is very useful in improving the performance of oncology clinical trial designs. Given the above results, we arrive at the decision-making conclusion summarized in Figure 2. For multiple parallel arm trials, such as master protocol trials, because subtrials are conducted simultaneously and the efficacy in all running arms is considered, it is recommended to choose a BHM and its extensions. When all arms are homogeneous (either effective or ineffective), the BHM method gains greater power than other methods. However, in practice, it is more common that we are not sure whether all arms are homogeneous, and methods with dynamic clustering are recommended, such as BaCIS and MEMs, to avoid type I error inflation. Furthermore, when the majority of the arms are expected to be effective, MEMs are recommended because of their excellent power performance. When the end point is ordinal or the division of the effective population is more refined, BCHM is recommended. For trials in which

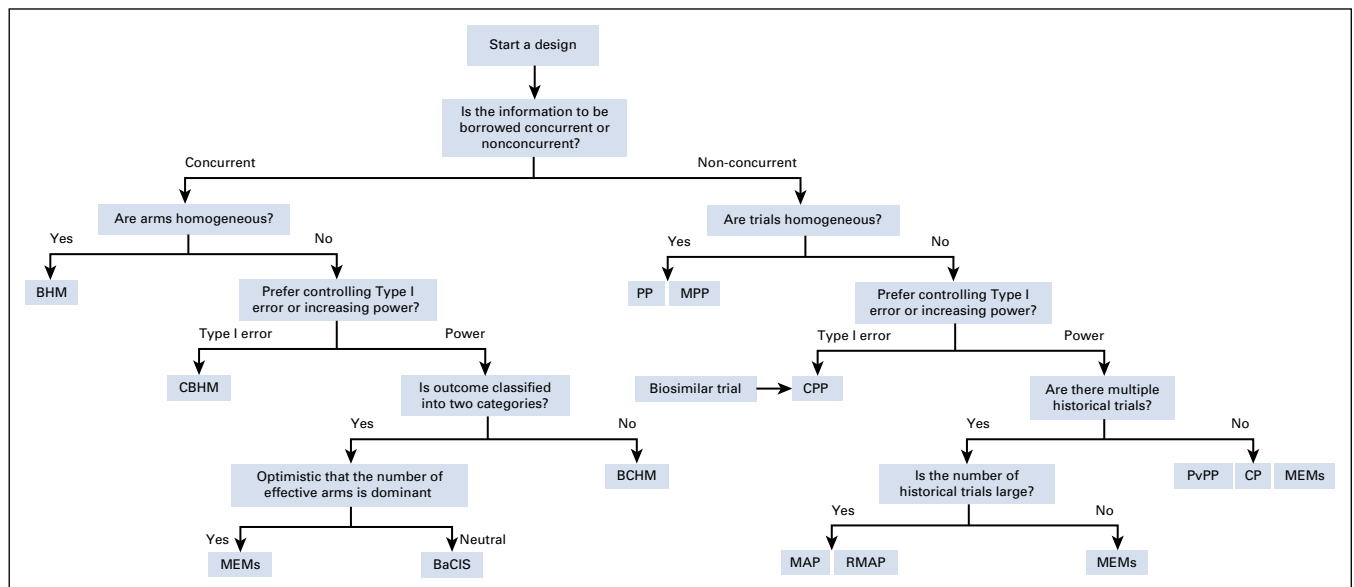


FIG 2. Decision-making diagram for how to choose an appropriate borrowing information method. BaCIS, Bayesian hierarchical classification and information sharing; BCHM, Bayesian cluster hierarchical model; BHM, Bayesian hierarchical model; CBHM, calibrated Bayesian hierarchical model; CP, commensurate prior; CPP, calibrated power prior; MAP, meta-analytic-predictive prior; MEM, multisource exchangeability model; MPP, modified power prior; PP, power prior; PvPP, P value–based power prior; RMAP, robust meta-analytic-predictive prior.

external data or supplementary trials exist, nonconcurrent borrowing methods are more appropriate, because historical data are available and we are mainly concerned with the current trial. PP and MPP failed to control type I errors when heterogeneity existed; thus, they were not appropriate for most scenarios. Methods other than PP take heterogeneity into consideration and gain an advantage in power. Specifically, the choice of method depends on the tradeoff between bias and variance. CPP provides a stricter criterion for controlling type I error, whereas MEMs place more weight on power gains. The performance of other methods lies between that of CPP and MEMs. In particular, CPP is naturally suitable for bioequivalence trials, whereas MEMs show an advantage in multiple historical trials. MAP and RMAP should be used with caution when the number of multiple historical trials is small.

In addition, sometimes, patient covariates are different, leading to heterogeneity in efficacy. This is more common in nonconcurrent scenarios, which usually have a large time span. Some methods, which have characteristics similar to those of nonconcurrent borrowing, have been proposed on the basis of covariate adjustment.²⁷⁻³¹

Regarding the borrowing strength parameters of each method, we used the recommended values in the literature. In fact, there are two common ways to determine these parameters: empirically specifying and objectively estimating. For the empirically specifying method, we can communicate with

clinical investigators to know how much power they intend to borrow. In addition, we can refer to sensitivity analysis in published articles to obtain the recommended values or select particular settings according to the required performance of the new trial. For the objectively estimating method, we can either estimate these parameters²⁵ or use model averaging/model selection to fit the most likely value.³² The above are general suggestions. For more details on the setting of key parameters in each method, we can refer to Viele's work.³³

It is worth noting that the difference between concurrent and nonconcurrent borrowing in our study is the information source. Methods commonly applied in two scenarios are listed in Table 1. However, in terms of statistical models, their use is not so strictly differentiated. Some methods can be applied to both concurrent and nonconcurrent scenarios, such as the MEMs in Table 1. In fact, other methods such as MAP and BHM are not only applicable to the recommended scenario, but can also be applied to the other scenario in terms of modeling construction when the source of information is properly explained. The specification of statistical models is actually flexible as long as it satisfies the practical clinical requirements.

In conclusion, whatever method one finally chooses when conducting an oncology trial, sufficient simulations should be performed to explore the statistical performance under various scenarios, and adequate communication should be performed to obtain the regulator's approval.

AFFILIATION

¹Research Center of Biostatistics and Computational Pharmacy, China Pharmaceutical University, Nanjing, China

CORRESPONDING AUTHOR

Fangrong Yan, PhD, Research Center of Biostatistics and Computational Pharmacy, China Pharmaceutical University, #639 Longmian Avenue, Nanjing, Jiangsu, P.R.China, 211198; e-mail: f.r.yan@163.com.

SUPPORT

Supported by the National Natural Science Foundation of China (81973145).

AUTHOR CONTRIBUTIONS

Conception and design: Liwen Su, Fangrong Yan

Financial support: Fangrong Yan

Collection and assembly of data: Liwen Su

Data analysis and interpretation: Liwen Su, Xin Chen, Jingyi Zhang

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/po/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians ([Open Payments](http://OpenPayments)).

No potential conflicts of interest were reported.

REFERENCES

1. Tang J, Pearce L, O'Donnell-Tormey J, et al: Trends in the global immuno-oncology landscape. *Nat Rev Drug Discov* 17:783-784, 2018
2. Shuford RA, Cairns AL, Moaven O: Precision approaches in the management of colorectal cancer: Current evidence and latest advancements towards individualizing the treatment. *Cancers* 12:3481, 2020
3. Bristow RG, Alexander B, Baumann M, et al: Combining precision radiotherapy with molecular targeting and immunomodulatory agents: A guideline by the American Society for Radiation Oncology. *Lancet Oncol* 19:e240-e251, 2018
4. Alexander NS, Allyson JO, Manish AS, et al: First-in-human trial of a novel anti-Trop-2 antibody-SN-38 conjugate, sacituzumab govitecan, for the treatment of diverse metastatic solid tumors. *Clin Cancer Res* 21:3870-3878, 2015

5. Charles J, Chaperot L, Hannani D, et al: An innovative plasmacytoid dendritic cell line-based cancer vaccine primes and expands antitumor T-cells in melanoma patients in a first-in-human trial. *Oncoimmunology* 9:1738812, 2020
6. Goldberg RM, Wei L, Fernandez S: The evolution of clinical trials in oncology: Defining who benefits from new drugs using innovative study designs. *Oncologist* 22:1015-1019, 2017
7. FDA: Master Protocols: Efficient Clinical Trial Design Strategies to Expedite Development of Oncology Drugs and Biologics. 2018. <https://www.fda.gov/media/120721/download>
8. FDA: Interacting With the FDA on Complex Innovative Trial Designs for Drugs and Biological Products. 2020. <https://www.fda.gov/media/130897/download>
9. FDA: A Bayesian Approach to Detection of Treatment-Related Effects in Toxicology Studies by Borrowing Information From Historical Control Animals. 2021. <https://www.fda.gov/science-research/fda-science-forum/bayesian-approach-detection-treatment-related-effects-toxicology-studies-borrowing-information>
10. Nanda R, Liu MC, Yau C, et al: Effect of pembrolizumab plus neoadjuvant chemotherapy on pathologic complete response in women with early-stage breast cancer: An analysis of the ongoing phase 2 adaptively randomized I-SPY2 trial. *JAMA Oncol* 6:676-684, 2020
11. van Rosmalen J, Dejaridin D, van Norden Y, et al: Including historical data in the analysis of clinical trials: Is it worth the effort? *Stat Methods Med Res* 27:3167-3182, 2018
12. ICH: Choice of control group and related issues in clinical trials E10, 2000. https://database.ich.org/sites/default/files/E10_Guideline.pdf
13. Chen MH, Ibrahim JG: Power prior distributions for regression models. *Stat Sci* 15:46-60, 2000
14. Duan Y, Ye K, Smith EP: Evaluating water quality using power priors to incorporate historical information. *Environmetrics* 17:95-106, 2006
15. Pan H, Yuan Y, Xia J: A calibrated power prior approach to borrow information from historical data with application to biosimilar clinical trials. *J R Stat Soc Ser C Appl Stat* 66:979-996, 2017
16. Liu GF: A dynamic power prior for borrowing historical data in noninferiority trials with binary endpoint. *Pharm Stat* 17:61-73, 2018
17. Hobbs BP, Carlin BP, Mandrekar SJ, et al: Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics* 67:1047-1056, 2011
18. Neuenschwander B, Capkun-Niggli G, Branson M, et al: Summarizing historical information on controls in clinical trials. *Clin Trials* 7:5-18, 2010
19. Schmidli H, Gsteiger S, Roychoudhury S, et al: Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics* 70:1023-1032, 2014
20. Kaizer AM, Koopmeiners JS, Hobbs BP: Bayesian hierarchical modeling based on multisource exchangeability. *Biostatistics* 19:169-184, 2018
21. Thall PF, Wathen JK, Bekele BN, et al: Hierarchical Bayesian approaches to phase II trials in diseases with multiple subtypes. *Stat Med* 22:763-780, 2003
22. Chu Y, Yuan Y: A Bayesian basket trial design using a calibrated Bayesian hierarchical model. *Clin Trials* 15:149-158, 2018
23. Chen N, Lee JJ: Bayesian hierarchical classification and information sharing for clinical trials with subgroups and binary outcomes. *Biom J* 61:1219-1231, 2019
24. Chen N, Lee JJ: Bayesian cluster hierarchical model for subgroup borrowing in the design and analysis of basket trials with binary endpoints. *Stat Methods Med Res* 29:2717-2732, 2020
25. Hobbs BP, Landin R: Bayesian basket trial design with exchangeability monitoring. *Stat Med* 37:3557-3572, 2018
26. Kaizer AM, Koopmeiners JS, Kane MJ: Basket designs: Statistical considerations for oncology trials. *JCO Precis Oncol* 3:1-9, 2019
27. Wang C, Li H, Chen WC, et al: Propensity score-integrated power prior approach for incorporating real-world evidence in single-arm clinical studies. *J Biopharm Stat* 29:731-748, 2019
28. Chen WC, Wang C, Li H, et al: Propensity score-integrated composite likelihood approach for augmenting the control arm of a randomized controlled trial by incorporating real-world data. *J Biopharm Stat* 30:508-520, 2020
29. Wang C, Lu N, Chen WC, et al: Propensity score-integrated composite likelihood approach for incorporating real-world evidence in single-arm clinical studies. *J Biopharm Stat* 30:495-507, 2020
30. Liu ML, Bunn V, Hupf B, et al: Propensity-score-based meta-analytic predictive prior for incorporating real-world and historical data. *Stat Med* 40:4794-4808, 2021
31. Kotalik A, Vock DM, Donny EC, et al: Dynamic borrowing in the presence of treatment effect heterogeneity. *Biostatistics* 22:789-804, 2020
32. Tang R, Shen J, Yuan Y: ComPAS: A Bayesian drug combination platform trial design with adaptive shrinkage. *Stat Med* 38:1120-1134, 2019
33. Viele K, Berry S, Neuenschwander B, et al: Use of historical control data for assessing treatment effects in clinical trials. *Pharm Stat* 13:41-54, 2014

