ComPAS :A Bayesian drug combination platform trial design with adaptive shrinkage

Xinyi Lin, Jiawei Ye

Abstract

Cancer therapy usually involves concomitant use of drugs that potentially induce synergistic effects. However, the number of drug combinations grows exponentially given the vast amount of monotherapies and new drugs are constantly entering clinical trials. To address this challenge, this paper proposes a flexible Bayesian drug combination platform design with adaptive shrinkage (ComPAS). The adaptive shrinkage method aims at adaptively borrowing information across combinations based on Bayesian model selection and hierarchical models. Simulation studies show that ComPAS performs better than regular Bayesian hierarchical models, and is not sensitive to design parameters. ComPAS provides a novel design for screening drug combinations effectively in phase II studies.

Introduction

Drug combination therapy has become common practice for treating cancer and other diseases. Cancer therapies usually involves concomitant usage of drugs functioning through different molecular pathways and mechanisms, which potentially induces synergistic treatment effects. Since currently there are hundreds of available anti-cancer drugs, the number of combinations is huge. As the pharmaceutical industry develops new treatment regimens, more drugs and drug combinations may enter the trial, which requires a trial that enables new candidates to enter in a seamless and timely fashion.

The motivating trial of this article is a phase II trial in which three investigational compounds were initially planned to be combined with each of three backbone therapies. [1] The backbone drugs are defined as immunotherapies while the compound drugs are non-immunotherapy compounds. The drug combination in this article consists of a backbone therapy and a compound drug. The objective of this trial is to identify promising drug combinations for subsequent phase III trials.

Though there are a variety of clinical trial designs available, none of them are suitable for this problem. Traditional phase II designs such as Simon's two-stage design or two-arm randomized phase II design aim at evaluating candidate treatments one at a time. Multiarm trial designs such as screened selection design and "Pick a Winner" design does not incorporate the adaptive feature of platform designs and cannot allow the entering of new regimens during the trial. Platform trial design overcomes the above problems, but is not especially designed for testing drug combinations.

We introduce a flexible Bayesian drug combination platform design with adaptive shrinkage (ComPAS) for screening a set of drug combinations. [2] The shrinkage method aims at adaptively borrowing information from data acquired early in the trial and is based on Bayesian model selection and hierarchical models. [3] Our simulation results show that the proposed platform design has better performances in simulations and sensitivity analysis than some existing designs.

Methods

1. Model

1.1. Bayesian Hierarchical Model

We are interested in treatment effects of different combinations of compounds and backbones. Let(j,k)denotes the combination of compound j and backbone k. $n_{j,k}$denotes numbers of patients have been treated with (j,k), $y_{j,k}$ denotes numbers of patients have been responded with (j,k) and $p_{j,k}$denotes the true response rate associate with (j,k). Assume $y_{j,k}$ follows Binomial distribution, we can get the following model:

$$y_{j,k}|n_{j,k},p_{j,k} \sim Binom(n_{j,k},p_{j,k})$$

Based on our setting, the treatment effect may be rather heterogeneous across compounds and relatively similar across backbones. That means combinations with the same backbones and different compounds((j,k) and (j',k)) will have relatively different treatment effects and combinations with the same compounds and different backbones((j,k) and (j,k')) will have relatively similar treatment effects.

1.2. Adaptive Shrinkage

In Bayesian Hierarchical Model, we borrow information from other combinations to estimate treatment effects using the following model.

$$logit(p_{j,k}) = \mu_{j,k}$$
$$\mu_{j,k} \sim N(\mu_j,\sigma^2), k = 1,2,..,K_j$$
$$\mu_j \sim N(0,10^6), j = 1,2,..,J$$

where $\sigma^2$is the shrinkage parameter. A small value of σ^2 results in high probability of smaller difference between all $\mu_{j,k}$ and μ_j, which means $\mu_{j,k}$are similar with each other. This indicates strong information borrowing across the combinations, whereas a large value of σ^2 induces little information borrowing. A natural approach is estimate σ^2 is to assign σ^2 a noninformative or vague prior, such as the inverse gamma distribution, IG(0.001, 0.001), and let the data determine how much information we should borrow across $\mu_{j,k}$. This is often known as the fully BHM (FBHM).

The FBHM approach seems sensible but does not work well when the number of second-level units (ie, k's or the number of backbones in our example) is moderate, eg, $< 10.^{14-16}$. In this case, even when the sample size is large, we have limited information of shrinkage parameter $\sigma^2$and the estimated shrinkage parameters is not reliable. In particular, when the noninformative inverse gamma

distribution is used (ie, $\sigma 2 \sim$ IG(0.001, 0.001)), the BHM always induces strong shrinkage and information borrowing no matter whether the treatment effect is similar or not. In this case, the estimated shrinkage parameter is not reliable too. That's why in this paper, the authors introduce another way, adaptive shrinkage, to estimate the shrinkage parameter.

Instead of estimate σ^2 fully based on data, the adaptive shrinkage method first specifies L discrete values of σ^2, ${\sigma^2}_1 < {\sigma^2}_2 < \ldots < {\sigma^2}_L$. Let M_L denotes the lth BHM with $\sigma^2 = {\sigma^2}_l, l = 1,2,\ldots,L$. Let $Pr(M_l)$: the prior probability that M_l is true, the given observed interim data $D = \{(n_{jk,} Y_{jk}), j = 1,\ldots J, k = 1,\ldots,K_j\}$.

$$Pr(M_l|D) = \frac{Pr(M_l)L(D|M_i)}{\sum_{q=1}^{L} Pr(M_q)L(D|Mq)}$$

$$L(D|M_l) = \int \int f(y_j|n_{jk}, p_{jk}) f(logit(p_{jk})|\mu_j, {\sigma^2}_l) f(\mu_j) dp_{jk} d\mu_j$$

The model with the highest posterior probability be the optimal model M_{opt}

$$M_{opt} = argmax_{\{M_l\}}(Pr(M_l|D), l = 1,\ldots,L)$$

Then, we need to find out a way to specify value of σ^2. As in the BHM, the degree of shrinkage is determined by the ratio of first-level and second-level variances, to determine the appropriate value of σ^2, we introduce $\bar{y_{jk}}$ and apply normal approximation to y_{jk}, then

$$\bar{y_{jk}} = y_{jk}/n_{jk}$$

$$\bar{y_{jk}}|p_{jk} \sim N(p_{jk}, p_{jk}(1-p_{jk})/n_{jk})$$

So the first-level variance is approximately $p_{jk}(1-p_{jk})/n_{jk}$. We apply Delta method to equation of $\mu_{j,k}$, μ_j

$$Var(p_{jk}) = \{\frac{e^{\mu_{jk}}}{(1+e^{\mu_{jk}})^2}\}^2 Var(\mu_{jk}) = \{p_{jk}(1-p_{jk})\}^2 \sigma^2$$

Then, the degree of shrinkage is determined by the relative variance between the first level and second level of the model which is

$$w = \frac{\frac{p_{jk}(1-p_{jk})}{n_{jk}}}{\frac{p_{jk(1-p_{jk})}}{n_{jk}} + \{p_{jk}(1-p_{jk})\}^2\sigma^2} = \frac{1}{1 + n_{jk}p_{jk}(1-p_{jk})\sigma^2}$$

The value of w can be interpreted as the relative contribution (or weight) of the prior of p_{jk} to the posterior estimate of p_{jk} with respect to observed data y_{jk}. Therefore, by specifying several values of $w_l, l = 1,2,\ldots,L$. We can get

$${\sigma_l}^2 = \frac{1-w_l}{n_{jk}p_{jk}(1-p_{jk})w_l}$$

2. Trial design

In a ComPAS trial, $n_1 R$ patients are equally randomized to R experimental arms, each arm contains an experimental combination therapy. At sth interim, observed interim data D_s is fitted to L candidate BMHs $M_1,\ldots,M_L$, and the model with highest posterior probability is identified and denoted as M_{opt}.

Conditional on M_{opt}, the following futility stopping and superiority graduation rules are applied to each of the R arms.

a) Futility stopping rule: the combination therapy $(j,k)$is dropped if $P(p_{jk} > p_o|D_s, M_{opt}) < c_f$, where $p_o$is the null response rate specified by clinicians, and $c_f$is a prespecified probability cutoff. The probability cutoff $c_f$can be easily calibrated by simulation as follows: specify scenarios where some combinations are effective and some combinations are futile, then calibrate the value of $c_f$such that the probability of dropping the futile combinations is reasonably high and the probability of dropping effective combinations is reasonably low.
b) Superiority graduation rule: if $P(p_{jk} > p_o|D_s, M_{opt}) > c_g$, the combination $(j,k)$is graduated from ComPAS trial. The probability cutoff $c_g$is chosen by calibrating simulated probability of graduating futile combinations by controlling this probability under a prespecified type I error rate.

When a combination is dropped or graduated, a new combination can be added to the trial if available. The trial is stopped if all combinations are dropped or graduated, or the maximum sample size is reached. If there are still drug combinations not graduated or dropped, continue to randomize patients into the remaining arms and collect interim data, and repeat the above model selection and interim analysis.

For trials with a control group, the futility dropping rule and superiority graduation rule can use response rate of the control group,p_{ctr}, instead of p_o.

3. Defining additive effects, synergistic effects and antagonistic effects

Different models have been proposed to classify synergistic and antagonistic effects of drug combinations. In this article, additive effect is defined by the classical Bliss independence model. [4] Assume the effects of drug J and drug K are independent, and they have response rates of $p_j$and $p_k$respectively, then the Bliss independence model define the additive response rate as $p_{jk} = p_j + p_k - p_j \times p_k$. If the observed effect is greater than the additive response rate, then the combination is classified as synergistic, otherwise the combination is classified as antagonistic.

We further define a backbone-main scenario and a compound-main scenario. Backbone-main scenario refers to the scenario where the response rates of the combinations varies across backbones but are relatively stable for different compounds, compound-main scenario refers to the scenario where the response rates of the combinations varies across compounds but are relatively stable for different backbones. Examples of these two scenarios are illustrated in the below results section.

Results

1. Simulations

We choose three scenarios, backbone-main, compound-main and additive scenarios, to test the performance of ComPAS. Among those, backbone-main and compound-main scenarios reflects the characteristics of drug combination based on backbone type drug and compound type drug.

Table 1 True Treatment Effects of Combinations in Backbone-main Scenario

	Backbone 1	Backbone 2	Backbone 3
Compound 1	0.40	0.50	0.60
Compound 2	0.42	0.52	**0.62**
Compound 3	0.45	0.55	**0.65**

Table 2 True Treatment Effects of Combinations in Compound-main Scenario

	Backbone 1	Backbone 2	Backbone 3
Compound 1	0.40	0.42	0.45
Compound 2	0.50	0.52	0.55
Compound 3	0.60	**0.62**	**0.65**

Table 3 True Treatment Effects of Combinations in Additive Scenario

	Backbone 1	Backbone 2	Backbone 3
Compound 1	0.73	0.64	0.19
Compound 2	0.85	0.80	0.55
Compound 3	0.94	**0.92**	0.82

We simulated these three scenarios using ComPAS and regular BHM models and calculated percentages of correct selection. Results are shown as following.

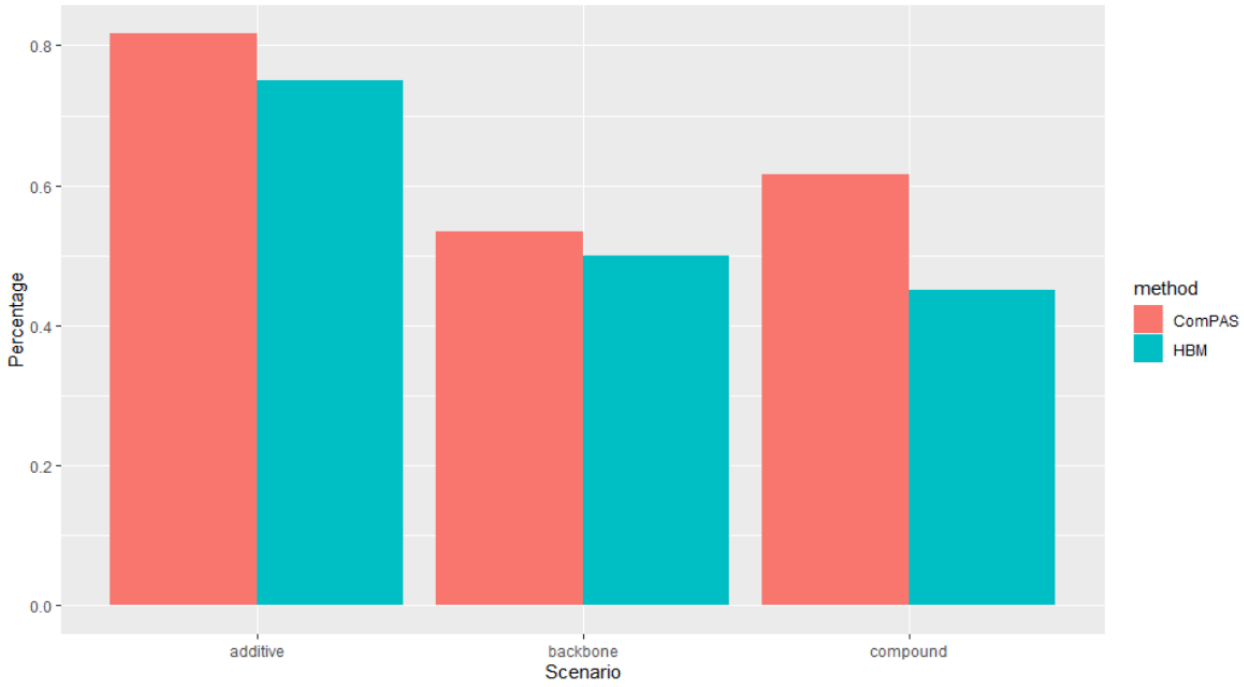

Figure 1 Percentages of correct selection based on ComPAS and BHM models

According to the above plot, we can find that in all three scenarios, ComPAS performs better than regular BHM model. In compound-main scenario, the correct selection percentage of ComPAS is significantly higher than BHM. This result is reasonable as the model ComPAS assume compound main treatment effects. Even this assumption is violated, ComPAS also has better performance than regular BHM. But overall, ComPAS performs better in the additive scenario. As ComPAS is designed for drug combination selection, we expected it performs best on scenarios which reflect drug combination characteristics like backbone-main scenario or compound-main scenario. However, both ComPAS and regular BHM model show better performance in the additive scenario. We think it might cause by discrete distribution of treatment effects in additive scenario. In other words, performance of ComPAS is mainly determined by the distribution or treatment effect instead of how drugs cooperate with each other.

2. Sensitivity analysis

Sensitivity analysis explores how changes of parameters influence the performance of model. Using simulation, we do sensitive analysis to find out how changes of L influence performance of ComPAS. We set L equal to 3, 4, 5 respectively and calculated percentages of correct selection in three scenarios, backbone-main, compound-main and additive scenarios. Percentages of correct selection are shown below.

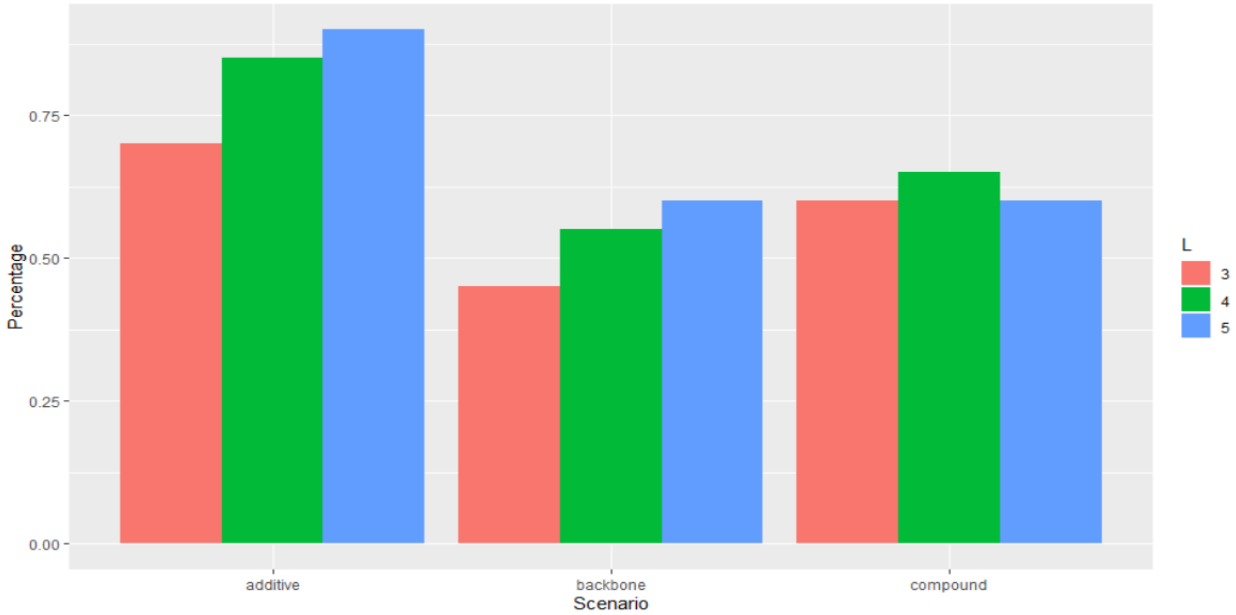

Figure 2 Percentages of correct selection with different L

According to the above plot, we can know that except for compound-main scenario, with value of L increases, performance of ComPAS increase. ComPAS still have the best performance in additive scenario. In future, if we would better choose larger L in additive and backbone-main scenarios to increase accuracy and choose L equals to 3 in compound-main scenario to increase efficiency.

Conclusion

According to simulation results, we can find that ComPAS performs better than regular BHM model in three scenarios, backbone-main, compound-main and additive scenarios. In compound-main scenario, the correct selection percentage of ComPAS is significantly higher than BHM. Overall, performance of ComPAS is mainly determined by the distribution or treatment effect instead of how drugs cooperate with each other.
According to sensitivity analysis, larger L in additive and backbone-main scenarios increase accuracy while choosing L equals to 3 in compound-main scenario increases efficiency.
Since we only do simulation for number of compounds equals to 3 and number of backbone equals to 3, we believe that when the number of second-level units (ie, k's or the number of backbones in our example) is moderate, ComPAS have better performance than regular BHM model.
Overall, ComPAS test drug combinations under platform trial setting, which is novel and meaningful in clinical trial design area. This paper is a good example in developing a new biostatistical method as it 1) states valid motivation in clinical trial setting; 2) shows clear model and trial design; 3) shows simulation process; 4) provides sensitivity analysis. However, as we mentioned above, performance of ComPAS mainly determined by the distribution or treatment effect instead of how drugs cooperate with each other, so we believe the characteristics of drug combinations are not as obvious as the paper claims

References

1. Larkin J, Chiarion-Sileni V, Gonzalez R, et al. Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. N Engl J Med. 2015;373(1):23-34.

2. Tang R, Shen J, Yuan Y. ComPAS: A Bayesian drug combination platform trial design with adaptive shrinkage. Statistics in Medicine. 2019;38(7):1120-1134

3. Brock K, Hierarchical Bayesian Model for Binary Responses (link)

4. Bliss CI. The toxicity of poisons applied jointly. Ann Appl Biol. 1939;26:585–615.