BAYESIAN DESIGN IN DRUG DEVELOPMENT-MULTIPLICITIES IN ASSESSING DRUG SAFETY

Maria Daya Roopa[1*], Nimitha John[2]

Article History: Received: 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract

The utilization of the Bayesian method in drug development offers several advantages. One of these is the ability to continuously update knowledge instead of restricting modifications to research design to major, isolated stages assessed in trials or phases. Additionally, the Bayesian approach is closely tied to decision-making within individual trials, drug development programs, and the broader context of developing a company's portfolio of medications. The future is expected to bring rapid advancement in clinical trial designs and analytics utilizing Bayesian methods. With political organizations and consumer advocacy groups calling for faster, safer, and more effective drug development, there is a risk of neglecting fundamental scientific concepts. However, adopting a Bayesian strategy can accelerate medication development and save money while maintaining sound research practices. The Bayesian approach is already gaining popularity in drug research and several therapeutic areas of medical device development, with variations influenced by the personalities involved. Notably, therapeutic areas where the clinical endpoint is detected early stand to benefit the most. Cancer and other diseases that have an increasing number of biomarkers available for modeling disease progression could benefit from the Bayesian approach. These biomarkers allow for more accurate tracking of a patient's progress and outcome determination. The use of Bayesian modeling is particularly useful in treatment areas where early signs of therapeutic effectiveness are evident.

Introduction: The trend in oncology treatment is shifting towards personalized medicine, where patients are matched with the most suitable treatments based on their prognostic factors [1,2]. This personalized approach has the potential to be highly beneficial for both patients and drug development. The initial step in evaluating the efficacy of a novel medication for a particular patient population in early Phase II trials is to determine whether the appropriate degree of efficacy has been achieved. In oncology, it is common to conduct a series of small screening trials in different patient subgroups, based on factors such as histology or a biomarker signature. However, these trials are often conducted independently of one another, without considering the possibility that some patient subpopulations may have similar therapeutic responses. The results of the trials in different subpopulations can provide insight into the treatment outcomes in other subpopulations. In Phase II cancer trials where a novel medication is being tested on various patient populations, Bayesian hierarchical models are used. Hierarchical modeling allows for the "borrowing" of information about the treatment effect in one group when predicting the treatment impact in another group [3]. Essentially, the estimated treatment effect for each group is reduced towards the average [4]. The degree of shrinkage is determined by the results, including the relative accuracy of the estimations in the various groups.

Keywords: Bayesian, Hierarchical, Empirical Data, Body system, Adaptive modeling, Optimal End points.

Corresponding Email: *maria.roopa@res.christuniversity.in
Email: nimitha.john@christuniversity.in

DOI: 10.31838/ecb/2023.12.s2.121

Novelty

When conducting research on adverse events (AEs), it is important to consider whether all three forms of AEs should be flagged. To answer this question, four factors need to be taken into account, including significance levels, overall number of AE kinds, rates for AEs not being considered, and biological links between different AEs. The frequentist approach to multiple comparisons only takes into account the first two factors.

A Bayesian approach considers all the aforementioned factors, and it matters if the three AEs with high drug-related rates are present in the same bodily system. The Bayesian approach is less concerned with type I error rates and is focused on determining the likelihood that the medicine may result in an AE using all available data. Bayesian methods may be more promising for safety evaluation. The simultaneous treatment of multiple AE kinds that are divided into body systems is addressed using a three-stage hierarchical mixture model.

This model offers a clear procedure for sharing data among various AE types, and regression is a result of the model's hierarchical structure. Various body systems exhibit a variety of AEs that are recorded in clinical studies, and a hierarchical model permits borrowing between AEs within the same body system.

A sensitivity analysis can be performed to show how crucial it might be to assign AEs to certain body systems. The assignment of AEs to body systems is based on biological or regulatory considerations, and if this assignment is ambiguous, the model can still be used for a variety of other plausible assignments.

2. Methodology

Bayesian Clinical Trials

The use of Bayesian statistical methods in clinical research has been on the rise, as they are well-suited to adapt to the information that accrues during a trial. This can potentially result in smaller, more informative trials and better treatment for patients. With Bayesian analyses, results can be assessed at any time, including continually, allowing for modifications to the trial design, such as slowing or stopping accrual, imbalancing randomization in favor of better-performing therapies, dropping or adding treatment arms, and changing the trial population to focus on patient subsets that respond better to experimental therapies. These analyses use available patient-outcome information, including biomarkers that may be related to clinical outcome, and can also incorporate historical data and trial data from relevant studies. In this article, we will explore the logic behind Bayesian clinical trials and how they could enhance the efficiency of drug development.

The multitude of medications, doses, and combinations can lead to various side effects. Data shows that around half of medications cause some degree of increase in the incidence of a specific adverse effect, but not all are impacted by drugs. Consequently, only a small proportion of medications will be statistically shown to be harmful in any one comparison. The question is, how can researchers distinguish the signal from the noise? How can they balance false positives (safe medications that are rejected) and false negatives?

Computational Techniques for Bayesian Analysis

The development of computational methods and the widespread usage of high-speed computers have contributed to the greater use of Bayesian methods in clinical research. Due to computing constraints, Bayesian procedures that always appeared right and proper could not be applied, but this is no longer the case. Almost any Bayesian design or analysis may be created and confirmed with the use of contemporary computer tools. Yet, frequentist software is far more advanced and accessible than Bayesian software. Writing custom Bayesian computer programmes is not difficult for statisticians, but it takes time. Moreover, the programmes will need validation. Online resources include a fantastic collection of applications known as WinBUGS (Windows version of Bayesian inference Using Gibbs Sampling) (see The BUGS Project in Further information). Moreover, SAS has some (mostly advanced) Bayesian macros and wants to include more Bayesian applications. Nevertheless, there is a lack of Bayesian software available. The previous distribution for the 2 parameter determines the borrowing amount. The model can range from assuming all treatments are the same to assuming no borrowing thanks to this parameter. A previous that we have specified places roughly a weight of 0.001 on an anticipated value of 0.1 for. The quantity of borrowing can be shaped by the data because this prior is weak. We demonstrate the probability of claiming efficacy and the predicted response rate across a number of prior distributions for the null and alternative situations in order to demonstrate the sensitivity of the results to the chosen prior. We display the findings for our same mean (0.1), with more weight (0.01), with less weight (0.0001), and with the same weight (0.001), but with larger and smaller means, respectively (0.01). Every prior still enables the data to influence the borrowing amount. As a result, there is little variation in the mean estimated likelihood of response, and the chance of claiming efficacy is constant across all priors.

As a result, even if the prior on 2 is crucial, we have chosen priors that are resistant to changes in the order of magnitude.

Bayesian Solution

It is a multiple testing scenario in the case study with 5 endpoints where we must test high versus low. Let y_{ij} be the vector of the ***i*** variables for ***j*** treatments and ***i*** = 1...5, ***j*** = 1…2

$$Y' = (y_{11}, y_{12}, y_{21}, y_{22} \ldots y_{52})$$

$$X = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \\ \vdots & \vdots \end{bmatrix}$$

$$\theta \sim \begin{bmatrix} \mu \\ \theta \end{bmatrix}$$

$$\text{Then } Y \sim N(X\theta, \Sigma)$$

Bayesian Approach:

Prior probabilities related to the multiplicities are used to account for any multiplicity adjustments that may be required. Generally, the more hypotheses that can be

considered, the lower the prior probabilities that each one is assigned.

Three stage hierarchical modelling

The three-stage hierarchical mixed model. There are B body systems. Within body system b there are k_b types of AE's labeled Ab_j, where b = 1,..., B and j= 1…k_b Of the N_c controls, Xbj experience Abj, and of the Nt patients in the treatment group, Y bj experience Abj. The probabilities of experiencing Abj are cbj and tbj, for control and treatment patients, respectively. We use logistic transformations:

$$\gamma_{bj} = log(\frac{c_{bj}}{1 - cbj})$$

and

$$\theta_{bj} = log(\frac{t_{bj}}{1 - t_{bj}}) - \gamma_{bj}$$

Three steps of the hierarchical prior are presented. The stage 1 priors are listed below. A normal previous distribution applies to the's:

$$\gamma_{bj} \sim N(\mu_{\gamma_b}, \sigma^2_{\gamma})$$

for b = 1;…B and *j*=1; … kb. Parameters bj are the log-odds ratios. If bj = 0 then the probability that a patient experiences Abj is the same for control and treatment; that is, $c_{bj} = t_{bj}$.

A positive probability is assigned to this possibility using the following mixture prior distribution:

$$\theta_{bj} \sim \pi_b I_{[0]} + (1 - \pi_b) N(\mu_{\theta_b}, \sigma^2_{\theta_b}),$$

for b = 1… B; j = 1 : : : kb.

The second stage of the prior structure is to assign a prior distribution to the hyper-parameters:

$$\mu_{\gamma_{bj}} \sim N(\mu_{\gamma 0}, \tau^2_{\gamma 0})$$

for b = 1 : : : B and

$$\sigma^2_{\gamma} \sim IG(\alpha_{\sigma\gamma}, \beta_{\sigma\gamma})$$

With early efficacy stopping

Simon's design does not have an early stopping point with a claim of efficacy. Stopping early may be preferable for efficacy, saving time and patient resources, and hastening the progression of the treatment to the following stage of development. In other situations, one might want more details about a treatment that seems to be working, but single-arm data with tumour response as the outcome is of limited use in deciding whether to proceed to Phase III. Reconsider the Bayesian hierarchical trial examples. With 10 patients in the first interim analysis of Example 1, group 2 could have ceased early for efficacy, and the other groups may have stopped early for efficacy as well at the second interim analysis. Despite the additional patients enrolled, the implication remained the same. This trial might have been stopped early, with the right conclusion, and most of the allocated patient resources would have been saved if there had been an early efficacy stopping criteria. Here, we contrast the three designs' operational traits when the Bayesian designs include early halting for both futility and efficacy. Early trial termination due to effectiveness claims often has no impact on the overall likelihood of trial success, but it does reduce the mean sample size. The mean sample size in each group tends to be smaller as a result of the Bayesian designs' addition of early efficacy stopping. There is a further reduction in sample size if the Bayesian hierarchical design offers a higher probability of trial success. The Bayesian hierarchical design has a larger mean sample size than the Independent Bayesian design in the "All in the Middle" scenario because it must reach the maximum sample size more frequently before declaring efficacy. Shrinkage in this scenario makes early efficacy stopping suitably more challenging because early stopping criteria are based on pmid for all groups and shrinkage in this scenario is towards pmid.

Predictive probabilities and trial design

The process of Bayesian updating has significant implications for the design of trials. One of its most valuable outcomes is the ability to predict what will happen in a trial from any given point, including the start of the trial, based on the currently available data. While future results cannot be predicted with certainty, the Bayesian approach provides a way to evaluate the future with the appropriate level of uncertainty. In the context of clinical trials, Bayesian hierarchical modeling has a wide range of applications. For instance, in the case of cancer research, several drugs that are effective in treating breast cancer may also be effective in other solid tumors. Traditionally, oncology drug development has focused on studying one cancer type at a time. However, a better approach would be to include patients with various cancer types in a single trial to assess the activity across different diseases. In a hierarchical model, one level of experimental unit could be cancer type, and another level could be patient within cancer type. If multiple trials are involved, a third level in the hierarchy can be added to include "trial". It is also important to model the potential role of biomarkers that may predict therapeutic benefits across diseases. Hierarchical modeling has several advantages, including providing a formal mechanism for adjusting the regression effect or "regression to the mean". For instance, some groups may have unusually large or small results, particularly for modest sample sizes. Additional data usually corrects these fluke observations, pulling them back toward the mean. Hierarchical modeling explicitly corrects for this by modeling the effect in all groups. This approach tends to produce more accurate estimates that are closer to the true values. The James-Stein estimator and other similar shrinkage estimators are more effective than no-borrowing estimates. It is even better to borrow measurements between entities that bear no relationship than to let them stand alone. Many authors have contributed to the literature on Bayesian hierarchical models and their relationships with empirical Bayesian methods. When borrowing hierarchically, groups that are extreme and those with greater uncertainty (i.e., those with smaller sample sizes) tend to experience greater shrinkage. The amount of borrowing is not determined in advance but rather by the data. If results across groups are very similar, there will be more borrowing. If results differ, there is less borrowing and greater uncertainty associated with the estimates. Three different design strategies are compared in this article. The first two approaches involve four separate trials, one for each patient group. The first approach employs Simon's Optimal Two-Stage design, while the second is Bayesian and adaptive, with results updated frequently for potentially stopping accrual early for futility. Each trial

involves many stages. Comparing the two approaches addresses the advantage of frequent monitoring versus having a single interim analysis. The third approach is a modification of the second, where the four groups are included in a single trial that employs Bayesian hierarchical modeling across the four groups in addition to frequent monitoring. Comparing the latter two approaches addresses the advantages of hierarchical modeling. This article focuses on a non-randomized single-arm trial with an endpoint of tumor response, but the same general approach can be applied to two-armed randomized trials and trials with other types of endpoints, including time-to-event endpoints.

Bayesian hierarchical adaptive design

The four patient groups are taken into account as a single, integrated trial in this design, and a Bayesian hierarchical model uses information from all four groups. In order to achieve this, we model the θ_i using a normal distribution whose mean μ and variance σ^2 is unknown.

$$\theta_i \sim \mathrm{N}(\mu, \sigma^2), i=1,2,3,4$$

Modeling the unidentified mean and variance involves using a second level of the distribution (hierarchy). A dynamic quantity of borrowing will be created dependent on the similarity between groups by the data across the groups shaping the posterior distribution for the mean and variance across groups. The following prior distributions of μ and σ^2 are

$$\mu \sim \mathrm{N}(-1.34, 10^2), \sigma^2 \sim \mathrm{Inverse-Gamma}(0.0005, 0.000$$

The parameter σ^2 is used to measure the level of diversity that exists among the patient groups. If σ^2 is 0, there will be complete pooling of the results across the patient groups, while still making adjustments for the targeted p1 rates in each group. On the other hand, if σ^2 is almost infinite, then no results can be shared across the groups. For values between these two extremes, there is some level of results sharing that is proportional to the variability among groups. The model can handle these different levels of diversity, so the selection of the prior for σ^2 significantly impacts the model results. Our choice of prior reflects a small degree of variability across the four groups. Assuming a prior estimate of σ=0.1, the prior for σ^2 is assigned with very little weight, only 0.1% of one observation. Considering the four patient groups to be observed in the trial, the posterior distribution provides only insignificant information to the overall posterior. The prior distribution of μ is weakly non-informative, with a prior mean that is close to the null hypothesis. While the model allows learning about μ and σ2 as the trial progresses, the information available for σ2 is limited when there are only four patient groups, even with large sample sizes. Hence, the prior distribution of σ2 plays a crucial role in determining the level of result sharing across groups, and its careful evaluation is essential. The model's sensitivity to the prior selection for σ2 is discussed further below.

The adaptive algorithm, including interim analyses, early stopping rules, and the final efficacy criteria, is the same as previously described for the independent-group Bayesian design. However, for the hierarchical model analysis, we apply the same early stopping rules and final evaluation criteria separately for each group. For instance, group 2 may stop early for futility, while groups 1, 3, and 4 continue to the maximum sample size. At the end of the trial, the efficacy criteria may be satisfied for group 1, but not for groups 3 and 4, enabling us to examine the effect of result sharing on error rates and mean sample size. For presentation convenience, we assume equal accrual to all groups, but in practice, accrual to the groups will differ, with some groups accruing more rapidly than others. Efforts may be made to balance accrual, such as opening additional sites for groups that are slower to accrue.

Estimation:

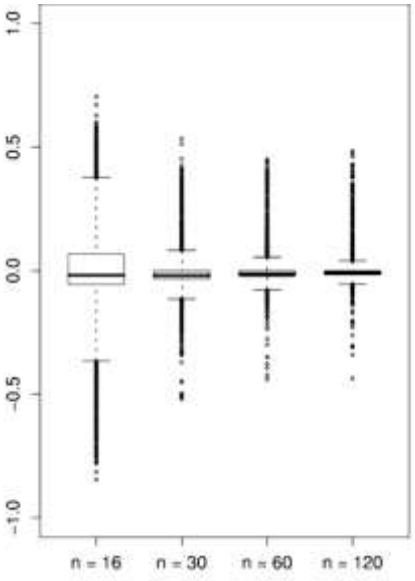
As previously mentioned, the degree of result sharing in the Bayesian hierarchical approach depends on the data: the more similar the groups, the greater the sharing. The "Null" and "Alternative" scenarios demonstrate that group similarity leads to extensive sharing and reduced uncertainty. In comparison to the other two designs, the standard deviations for the Bayesian hierarchical design are smaller in this situation. For example, the response rate for group 4 in the null scenario is 20%, but the Bayesian hierarchical model has a mean estimate of 14%. There is little or no bias for the other groups. As previously stated, there is a relatively greater amount of shrinkage for group 4 in this scenario, resulting in a particularly low Type I error rate.

Prior Information

- If the prior is reliable, the Bayesian approach is strong
- If the prior is appropriate, it is an improvement over non-Bayesian methods.
- If the specification is challenging, the approach may be less robust.
- The posterior may not be resistant to alternative priors.
- One could use a quasi-Bayesian prior that is not entirely specified to address this issue.

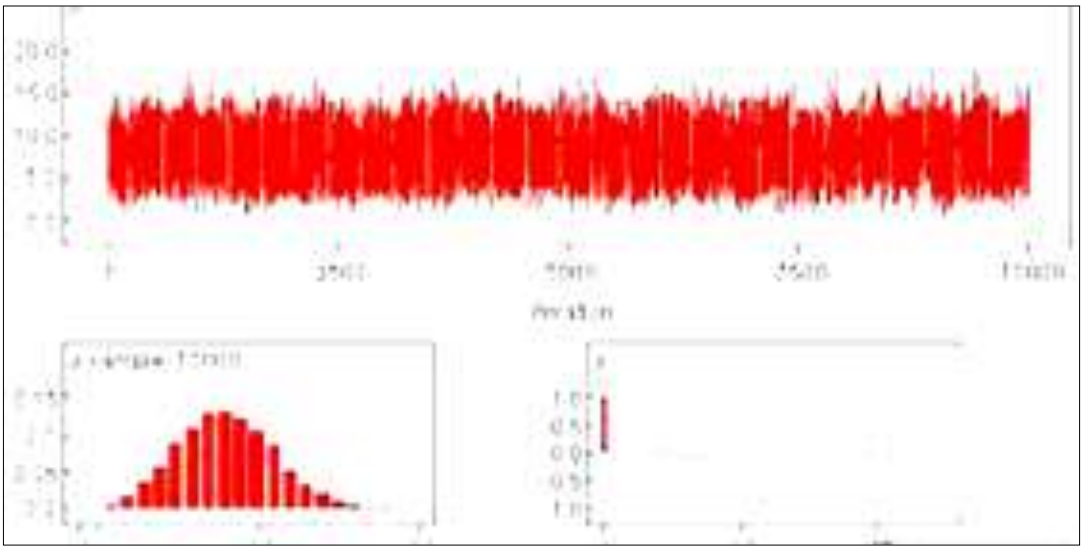
3. Results

	The ten 'signal' observations										#noise
n	-8.5	-5.4	-4.8	-2.6	-2.4	3.3	4.1	4.8	5.8	6.2	$p_i > .6$
10	1	1	1	.94	.89	.99	1	1	1	1	1
50	1	1	1	.71	.59	.94	1	1	1	1	0
500	1	1	1	.26	.17	.67	.96	1	1	1	2
5000	1	1.0	.98	.03	.02	.16	.67	.98	1	1	1



Node	Mean	SD	2.5%	97.5%
beta[0]	0.635	0.236	0.637	1.084
beta[1]	0.576	0.192	0.575	0.954
beta[2]	1.009	0.221	1.011	1.432
beta[3]	0.089	0.041	0.089	0.170

MLE approach		Bayesian approach	
Estimate	%RSE[a]	Estimate	%RSE
0.817	–	0.820	31.343
0.061	–	0.051	11.863
9.790	–	10.498	12.254
1.384	–	1.472	8.212
89.002	–	89.097	2.313
81.966	–	87.177	–
9.790	–	10.498	–



After 10,000 iterations, the convergence pattern for the data displays the MCMC outlier points.
The patient estimation data is modeled hierarchically with consideration of the parameters for patient observations, including missing data observations.

4. Conclusions

Bayesian methods offer the advantage of accounting for multiplicities, with hierarchical modeling being particularly efficient in this regard. Managing adverse events (AEs) in clinical trials poses a challenging statistical problem, especially when dealing with multiple comparisons. It is essential to model the available structure and information to conduct good science. My methodology shows that relocating an AE to a different body system can significantly affect conclusions, presenting both positive and negative aspects. The model benefits from exploiting information across related types of AEs, but it necessitates careful assignment of AEs to body systems, which should be done based on biological grounds and separate from the data. Assignments based on empirical correlations violate the modeling's spirit. For uncertain assignments, multiple runs of the model with different assignments should be performed to provide a sensitivity analysis of the conclusions. If an AE is assigned to the wrong body system, it can impact the model's conclusions, but we have not addressed how to make reparations. Any post-hoc corrections can be challenging to make without biasing the conclusions, representing another level of multiplicity similar to data dredging. Our model relies on marginal data, and while it offers precise conclusions about treatment effects, modeling dependent frequencies could yield even greater accuracy. Beyond drug AEs, our model has broad applications, including the analysis of cDNA microarray data, where the multiplicity problem is more pronounced, with tens of thousands of genes potentially involved. Categorizing genes into genetic pathways (similar to body systems for AEs) helps to identify differentially expressed genes implicated in diseases and their treatment. The developed Bayesian model accounts for categorical data and multiple comparisons to provide more accurate estimates.

5. Discussion

We conducted a comparison of two Bayesian adaptive approaches with Simon's Optimal Two-Stage design in the context of Phase II trials involving multiple patient groups. Bayesian adaptive designs can be configured to have similar operating characteristics to Simon's design in terms of Type I error and power. Such designs usually have smaller sample sizes because of more frequent interim analyses and the possibility of stopping early for efficacy. Utilizing a Bayesian hierarchical model to borrow across patient groups can further reduce Type I error, increase power, and decrease the mean sample size, making personalized medicine more feasible. Several pharmaceutical companies have successfully implemented this approach, which is sometimes referred to as a tumor-agnostic design, and frequently focuses on patients with a tumor positive for a specific biomarker regardless of the tumor site. Determining whether and how to borrow across groups depends on whether similar treatment effects are possible. The amount of borrowing in our model is based on an inverse gamma hyperprior on the variance term for the log-odds of the response rate, which can be adjusted according to clinical judgment during trial design. The choice of futility and efficacy thresholds can be made by examining simulation results and adjusting criteria to achieve optimal results. In our example, the Bayesian hierarchical design may be overpowered and could be refined to have lower power in each group. The definitions of p0, pmid, and p1 in the Bayesian designs are partially an artifact of the comparison with Simon's design, and it may be more natural to use a single target response rate, pgoal, instead. In conclusion, the Bayesian hierarchical design is an important alternative in Phase II trials, as it offers greater power and lower Type I error with a lower mean sample size, making personalized medicine more feasible.

6. References

Berry DA. Bayesian clinical trials. Nat Rev Drug Discov. 2006; 5(1):27–36. [PubMed: 16485344]

Berry DA. Bayesian approaches for comparative effectiveness research. Clin Trials. 2012; 9(1):37–47. [PubMed: 21878446]

Berry, SM.; Carlin, BP.; Lee, JJ.; Muller, P. Bayesian Adaptive Methods for Clinical Trials. CRC Press; Boca Raton, FL: 2011.

Berry SM, Ishak KJ, Luce BR, Berry DA. Bayesian meta-analysis for comparative effectiveness and informing coverage decisions. Med Care. 2010; 48(6):S137–44. [PubMed: 20473185]

Hamburg MA, Collins FS. The path to personalized medicine. N Engl J Med. 2010; 363(4):301–04. [PubMed: 20551152]

Jones H, Ohlssen D, Neuenschwander B, Racine A, Branson M. Bayesian models for subgroup analysis in clinical trials. Clin Trials. 2011; 8(2):129–43. [PubMed: 21282293]

Simon R. Clinical trials for predictive medicine: New challenges and paradigms. Clin Trials. 2010; 7(5):516–24. [PubMed: 20338899]