Bayesian Frameworks And Their Relationships – An Application In Pediatric Drug Development

Amarjot Kaur[1], Qing Li[1], and Ruoyang Zhang[2],
[1]Biostatistics and Research Decision Sciences, MRL, Merck & Co., Inc.
[2]Department of Statistics, University of Florida,

Abstract

Pediatric trials are an important component of drug development and are typically conducted when the efficacy and safety has been established in adult population. Many disease areas are presented with inherent difficulties in conducting large pediatrics trials making such development highly challenging. When pediatric trials are conducted after adult trials, it allows for utilization of prior adult information to help enhance efficiencies for scenarios that justify borrowing prior information. Innovative clinical trials using Bayesian framework to leverage historical information have been discussed extensively in recent years. This framework helps alleviate logistical and ethical challenges in conducting large clinical trials in children by providing avenues for designing pediatric trials with feasible sample size which in turn facilitate timely decision making and treatment access to children. We consider three frequently used Bayesian frameworks including hierarchical model, power prior, mixture prior, and explore their relationship with one another. Chen & Ibrahim (2006) established an 1-1 correspondence between power prior and hierarchical model under certain settings. In this paper, we further established relationship between power prior and mixture prior using their operating characteristics. This newly established relationship allows unifying the three commonly used Bayesian frameworks: hierarchical model, power prior, and mixture priors. We demonstrate the correspondence across three methods using the key parameters that controls the amount of borrowing of historical information via simulations. Although discussed in the context of pediatrics development, the application is this work is applicable for any other scenario of borrowing prior information.

Key Words: Innovative clinical trials, Bayesian framework, Mixture prior; Power prior; Pediatric clinical trials, Historical information

1. Introduction

Pediatric trials are an important component of drug development. But it often encounters the challenge of not being able to enroll sufficient number of subjects to meet the stringent efficacy criterion. The enrollment challenges for the pediatric trials are many, including but not limited to rarity of disease, reluctance of exposing children to new drug, or other ethical reasons. Innovative ideas become even more necessary when face with practical challenges that could delay access of important treatments to children. Augmenting information from historical data, whenever appropriate, becomes an important avenue to demonstrate efficacy and move forward the pediatrics

development. The discussion here focuses on the efficacy evaluation only, as the safety evaluation is based on regulatory requirement and is out of scope of this manuscript.

Bayesian framework allows augmentation of available information from historical trial(s) and that can help reduce sample size. Borrowing information from historical data hinges on the fact that there is sufficient evidence that the treatment responses are likely to be similar across pediatric and the prior adult population, based on clinical rationales, pharmacokinetics, and pharmacodynamics of the medical product [1, 2]. Spiegelhalter (2004) provided a systemic review for Bayesian methods in health care evaluation [3,4]. Bayesian paradigm provides a natural framework for extrapolation and there are many frameworks to incorporate results from historical trials to different populations. To combine information from different sources, Viele (2014) discussed various scenarios of borrowing including static and dynamic borrowing, and Schmidt (2014) discussed borrowing from multiple sources utilizing meta analytical approach [5,6]. Schoenfeld et al. (2009) considered hierarchical models to augment information from adult trial to pediatric trial [7]. Kaur et al. (2018) utilized similar approach in design of pediatric study and sample size determination by incorporating information from previously completed adult treatment effect, and Jin et al. (2020) further extended the approach for various scenarios including that for binary data as well as incorporating directly information from multiple historical trials[8,9]. Gamalo-Siebers et al. (2017) discussed the sample size problem in terms of extrapolation to extend information from subgroups of the patient from source population (e.g. adult population), to make inferences for another subgroup of the target population (e.g. pediatric population), allowing for dynamic borrowing [10]. Further review of borrowing external data has been discussed for medical devices, rare diseases, and various other scenarios and many recent regulatory guidance have discussions on utilizing Bayesian framework under various settings [11-14]. A recent impact paper mentions that *"…CDER statisticians are applying Bayesian hierarchical models to other critical areas in drug evaluation as well, such as in the evaluation of treatments for children. Considering the adult and pediatric data together improves the quality of decision-making in the pediatric setting by borrowing from the adult results. The amount of borrowing from adults (the weight that can be given to the adult data) is based on an evaluation of all available data and depends on ratio of variability in adult and children's data to variability between subgroups. Such an approach can be especially helpful for pediatric indications where recruiting pediatric patients for clinical trials can be difficult…"* [15].

The most common approach for implementing Bayesian methods is to build a prior distribution on the treatment effect by utilizing available information to design and analyze the trial. Three most commonly used Bayesian methods that down-weight data from the source population include power prior, hierarchical models, and mixture priors. Power priors discussed by Ibrahim & Chen (2000) are formed by raising the likelihood of the historical data to a power parameter $a_p \in [0, 1]$ [16]. The parameter a_p controls the extent of borrowing, and its larger value means more extrapolation from prior data. Hierarchical models have been around for a while, and Schoenfeld et al. (2009) discussed this model assuming the parameter of interest, say treatment effect, from the adult and the pediatric populations with a common normal distribution $N(\mu, v^2)$, where the treatment effect μ has a non-informative prior and the variance of the prior distribution v^2, representing the heterogeneity across the adult and pediatric population, is estimated from available information. A smaller value of v^2 implies similarity between adult and pediatric population thus allowing for borrowing more extrapolation. Mixture prior (Rover et al. 2019, Ye & Travis 2017, Islas et al. 2017) is formed by

mixing two distributions with weight a_m, i.e., $\pi(\theta) = (1 - a_m)\pi_1(\theta) + a_m\pi_2(\theta)$ [17-19]. The mixing weight a_m can be viewed as the applicability probability of historical trials and can be solicited through expert opinion. Compared with power priors and hierarchical models, mixture prior is less studied in adult-to-pediatric extrapolation, though it has good interpretive properties.

Chen & Ibrahim (2006) showed 1-1 correspondence, under certain settings, between the power parameter a_p, when using power prior framework, and variance parameter v^2, when using hierarchical modeling, [20] and is discussed in detail in subsequent sections. We further establish a 1-1 correspondence between the mixing weight a_m and power parameter a_p for mixture prior and power prior in the context of the operation characteristics. Together, with the already established relationship between power prior and hierarchical model and newly established relationship between power prior and mixture prior, the three commonly used Bayesian methods can be viewed under the same framework. We demonstrate the correspondence across these three Bayesian methods in terms of the key parameters that affect the amount of borrowing.

The remainder of the manuscript is organized as follows. In Section 2, we provide background of the three commonly used methods, hierarchical model, power prior, mixture prior, and discuss the key parameters that control borrowing of historical information. In Section 3, we review the established relationship across power priors and hierarchical models, and establish new relationship across power prior and mixture priors, along with evaluation of these relationships via simulations. The concluding remarks are provided in Section 4.

2. Methods

Let's consider a single historical dataset with two arms and equal randomization ratio similar to that in the planned new trial. If there is more than one historical dataset, one may obtain the parameter estimate of historical treatment effect from a meta-analytical approach. Let θ and θ_0 denote the treatment effect (difference between treatment and control arm) for current and historical trials, respectively. Denote the estimator $\bar{Y}$ for θ, $\bar{Y}_0$ for θ_0. A commonly used data model is

$$\bar{Y}|\theta \sim N(\theta, \sigma^2/n), \quad \bar{Y}_0|\theta_0 \sim N(\theta_0, \sigma_0^2/n_0),$$

Where n and n_0 are sample sizes per arm for the new (e.g. pediatric) and historical (e.g. adult) trial, $\sigma^2/2$ and $\sigma_0^2/2$ are the variances for current and historical data, respectively. Further assuming the threshold for testing is C, the hypothesis test can be expressed as

$$H_0 : \theta \leq C \quad \text{vs} \quad H_1 : \theta > C$$

If H_0 is rejected when $\bar{Y} > y_r$, then the power function is

$$g(\theta) = P(\bar{Y} > y_r|\theta) = P\left(\frac{\bar{Y} - \theta}{\sigma/\sqrt{n}} > \frac{y_r - \theta}{\sigma/\sqrt{n}}\right) = \Phi\left(\frac{\theta - y_r}{\sigma/\sqrt{n}}\right)$$

Hierarchical Model:
Let's consider the normal prior distribution of i.i.d. parameters as follows:

$$\theta\ ,\ \theta_0 \sim N(\mu, v^2),$$

where v^2 measures the variation between the treatment effect in the pediatric and adult trials and; larger v^2 indicates larger variation between the populations, thus allows for less borrowing. Schoenfeld et al, (2009) indicated that v could be elicited from the prior evidence using a data driven approach along with clinical input, e.g. $\hat{v} = \left|\hat{\theta}_0 - \hat{\theta}\right|/\sqrt{2}$.

The parameter μ usually has a non-informative prior, e.g., $N(0, \tau^2)$, $\tau^2 \to \infty$.

Power Priors:
The prior is expressed by down-weighting the likelihood of historical (say adult trial) information

$$\pi\left(\theta|\bar{Y}_0, a_p\right) \propto L(\theta|\bar{Y}_0)^{a_p}$$

$0 \le a_p \le 1$. The discounting parameter a_p controls amount of information borrowed, where $a_P = 1$ implies full borrowing, i.e., equivalent to pooling data across adult and pediatric trials, and $a_P = 0$ implies no borrowing , i.e., equivalent to using only pediatric data. However, the challenge of this approach is to how to elicit a_p as that would determine the extent of borrowing from the historical data [16]

Mixture Prior Model:
The prior is expressed by down-weighting by mixing a skeptical part as

$$\pi(\theta|\bar{Y}_0, a_m) = (1 - a_m)\pi_{\text{skep}}(\theta) + a_m L(\theta|\bar{Y}_0)$$

Where $\pi_{skep} \sim N(0, k^2)$ measures skepticism on the pediatric treatment effect ($\delta = 0$), where larger k^2 implies less skepticism, and smaller k^2 implies more skepticism towards prior (or adult) data. Further, the mixing weight a_m controls the amount of information borrowed, where $a_m = 1$ implies full borrowing, i.e., equivalent to pooling, and $a_m = 0$ implies no borrowing, and that turns out to be more conservative than a standalone trial. The mixing parameter a_m can be interpreted as the evidence of the applicability of adult results [18] and could be estimated by elicitation through expert opinion through properly designed survey questions. A schematic representation of mixture prior is presented in Figure 1, when $\bar{Y}_0 = 1$, $\sigma_0^2 = 9$, $\mathrm{n}_0 = 100$, $k^2 = 1$, $a_m = 0.3$, and $k^2 > \frac{\sigma_0^2}{n_0}$.

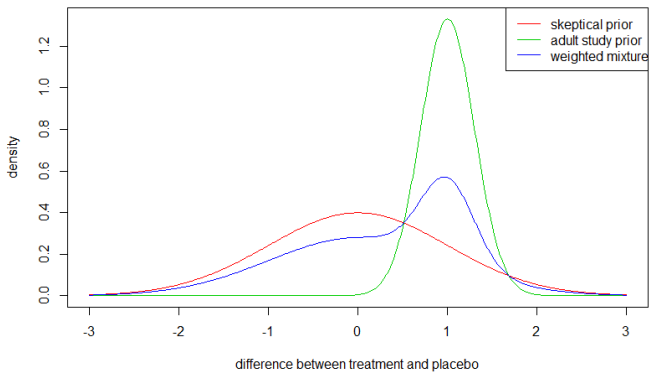

Figure 1: A schematic representation of mixture prior

3. Relations Across Bayesian Frameworks

Chen & Ibrahim (2006) established 1-1 correspondence between a_P in power prior and variance v^2 in hierarchical model (Chen & Ibrahim, 2006)

$$a_P = 1/\left(\frac{2v^2 n_0}{\sigma_0^2} + 1\right)$$

This correspondence between a_P and v^2 is established by matching the posteriors of power prior and hierarchical models.

We further explored the relationship between the a_P in power prior and the a_m in the mixture prior. When the posteriors of power prior and hierarchical model are both normal then one could match them exactly under some settings. In contrast, the posterior distribution in the mixture prior is always a mixture and therefore it is not feasible to match it exactly with the normally distributed posterior distribution of power prior. Hence the equivalence between the mixture prior and the power prior was established in terms of power and Type I error. We derived (calculations details not shown) the relationship between the mixture prior mixing weight a_m and the power prior discounting factor a_P, and this correspondence is presented in Figure 2. The relationship changes as the skepticism parameter k changes.

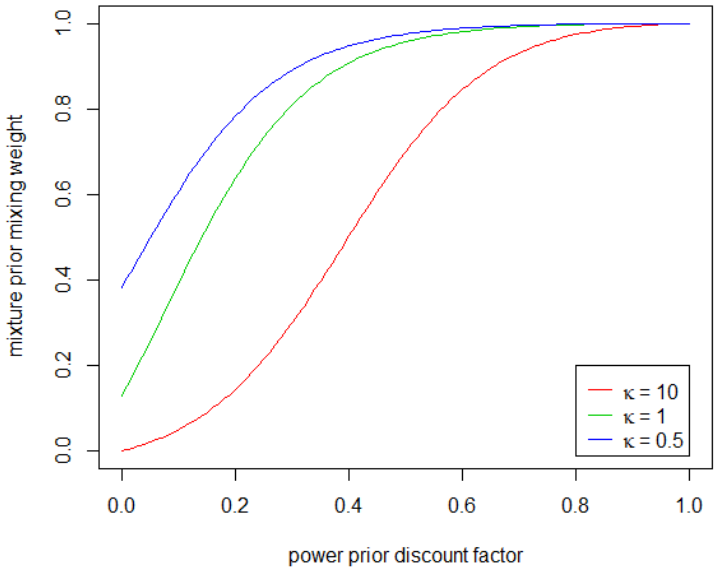

Figure 2: A relationship between the mixture prior mixing weight $a_m$and the power prior discounting factor a_P, when $\bar{Y}_0 = 1, \sigma_0^2 = 9, n_0 = 100, \sigma^2 = 8, n = 50, \delta = 0, \alpha = 0.05$

A simulation study was conducted to explored power and Type I error both for binary data as well as its normal approximation for testing superiority hypothesis; the size of the adults study (historical information) $n_{\mathrm{H}} = 400$ per arm, and size of the pediatric study was $n = 200$ per arm. Table 1 provides power and Type I error for various scenarios of v from hierarchical model and its equivalent values a_{p} from power prior, and a_{m} from mixture prior, respectively. Frequentist column represents the case of no borrowing and the Binomial column provides results without normal approximation of binary data. The simulation results indicate the power increase when relatively more information is borrowed from the historical data and remains higher than the frequentist approach. The Type I error is maintained under frequentist approach though it gets inflated under Bayesian methods. The inflation of Type I error is unavoidable when borrowing "successful" information from historical data in Bayesian framework. The evaluation of Bayesian methods therefore needs to be made not only based on Type I error but also by all considering of overall benefit of this approach along with other related perspective.

Table 1: Power and Type I Error Under Various Scenarios of Borrowing for Hierarchical Model, Power Prior, Mixture Priors

p_1	p_2	p_{1H}	p_{2H}	$\boldsymbol{\nu}$	$\mathbf{a_p}$	$\mathbf{a_m}$	Power (%)			Type 1 Error (%)		
							Hierarchical*	Binomial	Frequentist	Hierarchical*	Binomial	Frequentist
0.5	0.4	0.5	0.4	0.05	0.20	0.27	68.8	68.1	52.0	6.4	7.1	2.5
0.5	0.4	0.5	0.4	0.1	0.06	0.07	56.9	56.6		3.3	3.3	
0.2	0.1	0.2	0.1	0.05	0.11	0.27	89.7	88.9	80.0	5.7	5.7	2.5
0.2	0.1	0.2	0.1	0.1	0.03	0.06	83.0	82.6		3.2	2.6	

*Equivalent power for hierarchical model, power prior, mixture prior for parameters v, a_{p}, and a_{m}, respectively.
Frequentist column is the case of no borrowing - i.e., only pediatric data is used; Binomial column is the counterpart without normal approximation [9].
In mixture prior computations, $\mathrm{k}^2 = 0.09$ (about 80-150 times $\frac{\sigma_0^2}{\mathrm{n}_0}$).

4. Summary

Bayesian framework provides a reasonable solution for pediatrics development when the large trials are not possible. There are multiple ways to borrow information with different models (hierarchical model, power priors, mixture priors, etc.). There is well established correspondence between hierarchical model and power priors [20]. We established correspondence between power prior and mixture priors. This helps unify three Bayesian frameworks under normal data assumption; hierarchical model, power priors, and mixture priors. Borrowing information from historical data requires upfront justification and transparency of underlying assumption and discussion with regulatory agencies for agreement on the extent of information that can borrowed.

References

1. Food US, Administration Drug. Exposure-response relationships—study design, data analysis, and regulatory applications (2018). https://www.fda.gov/regulatory-information/search-fda-guidance-documents/exposure-response-relationships-study-design-data-analysis-and-regulatory-applications.
2. Mulugeta Y, Barrett JS, Nelson R, Eshete AT, Mushtaq A, Yao L, Glasgow N, Mulberg AE, Gonzalez D, Green D, Florian J. Exposure matching for extrapolation of efficacy in pediatric drug development. J Clin Pharmacol. (2016);56(11):1326–34.
3. Spiegelhalter JD, Abrams RK, Myles JP. Bayesian Approaches to Clinical Trials and Health-Care Evaluation, vol. 13. Hobo-ken: Wiley; 2004.
4. Spiegelhalter David J, et al. Incorporating Bayesian ideas into health-care evaluation. Stat Sci. 2004;19(1):156–74.Spiegelhalter, Abrams, and Myle (2004). Bayesian Approaches to Clinical Trials and Health- care Evaluation. *Wiley, Chichester*.
5. K. Viele *et al*. Use of historical control data for assessing treatment effects in clinical trials. Pharmaceut. Statist. (2014), 13, 41-54.
6. Schmidli, H., Gsteiger, S., Roychoudhury, S., O'Hagan, A., Spiegelhalter, D. & Neuenschwander, B. (2014), `Robust meta-analytic-predictive priors in clinical trials with historical control information', Biometrics 70(4), 1023{1032}.
7. Schoenfeld, D., Zheng, H., and Finkelstein, D. (2009). Bayesian design using adult data to augment pediatric trials. *Clin Trials*, 6, 297{304}.
8. Kaur, Skoner, Ibrahim, Li, Lockey, Blaiss, Bufe, Andersen, Canonica, Nolte (2018). Effect of grass sublingual tablet immunotherapy is similar in children and adults: A Bayesian approach to design pediatric sublingual immunotherapy trials. Journal of Allergy and Clin. Immunol, 141:1744-1749.
9. Jin, M., Li, Q., Kaur, A. Bayesian design for pediatric clinical trials with binary endpoints when borrowing historical information of treatment effect (2020). Therapeutic Innovation & Regulatory Science. https://doi.org/10.1007/s43441-020-00220-5.
10. Gamalo-Siebers, M., Savic, J., Basu, C., Zhao, X., Gopalakrishnan, M., Gao, A., Song, G., Baygani, S., Thompson, L., Xia, H.A. and Price, K. (2017). Statistical modeling for Bayesian extrapolation of adult clinical trial information in pediatric drug evaluation. *Pharmaceutical statistics*, *16*(4), pp.232-249.
11. FDA Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials - Guidance for Industry and FDA Staff (2010). http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071121.pdf.
12. EMA Concept paper on extrapolation of efficacy and safety in medicine development (2012). http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/06/WC500129285.pdf.
13. FDA Guidance for Industry (Draft). Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products (2020). https://www.fda.gov/media/130897/download.
14. FDA Duke Margolis Convening on Utilizing Innovative Statistical Methods and Trial Designs in Rare Disease Drug Development (2018) https://healthpolicy.duke.edu/events/utilizing-innovative-statistical-methods-and-trial-designs-rare-disease-drug-development.
15. Regulatory Science Impact Story: Using innovative statistical approaches to provide the most reliable treatment outcomes information to patients and clinicians. https://www.fda.gov/drugs/regulatory-science-action/impact-story-using-innovative-statistical-approaches-provide-most-reliable-treatment-outcomes.
16. Chen, Ming-Hui; Ibrahim, Joseph G. The relationship between the power prior and hierarchical models. Bayesian Anal. 1 (2006), no. 3, 551--574. doi:10.1214/06-BA118.
17. Rover, C., Wandel, S. & Friede, T. (2019), `Model averaging for robust extrapolation in evidence synthesis', Statistics in Medicine 38(4), 674{694}.
18. Ye and Travis, A Bayesian approach to incorporating adult clinical data into pediatric clinical trials.
19. Islas, C. D., Mander, A., Best, N. and Turner, R. (2017). A Bayesian framework for extrapolation using mixture priors: Challenges and some recommendations. https://www.psiweb.org/events/past-psi-events/2017/11/22/default-calendar/psi-one-day-meeting-extrapolation..
20. Chen, Ming-Hui; Ibrahim, Joseph G. The relationship between the power prior and hierarchical models. Bayesian Anal. 1 (2006), no. 3, 551--574. doi:10.1214/06-BA118.