Journal of Data Science 23 (1), 52–55
January 2025
DOI: 10.6339/25-JDS1161G
Discussion

Discussion of "Power Priors for Leveraging Historical Data: Looking Back and Looking Forward"☆

Chenguang Wang[1]

[1]*Biostatistics and Data Management, Regeneron, Inc., USA*

The power prior, a sophisticated yet elegantly simple statistical method, has been a focal point of research for statisticians over the past twenty-five years since its introduction. As we reach this milestone, it is fitting to reflect on the past achievements and future potential of this method. Chen et al. (2025) offers an outstanding review of the method's historical development and explores future directions, serving as an insightful addition to the existing literature. It is my honor to be part of this meaningful effort as a discussant.

In this discussion, I will share a few thoughts on utilizing the power prior in Bayesian clinical trial design and analysis, particularly in the context of leveraging real world data (RWD) and real world evidence (RWE). Additionally, I will discuss a possible expansion of the power prior for addressing temporal changes.

1 Power Prior for Designing Clinical Trials

As highlighted in Chen et al. (2025), leveraging RWD and RWE in clinical trial design and analysis has become increasingly popular, and the US FDA has been a major driving force behind various related initiatives. Typical statistical thinking, rooted in the design and analysis of observational studies, focuses on addressing the potential biases introduced by RWD/RWE. In the Bayesian realm, methods have been developed to adaptively or dynamically 'borrow' from RWD for the analysis of the current clinical trial, i.e., the clinical trial of primary interest. Chen et al. (2025) provides a comprehensive summary of the power prior-based dynamic borrowing methods that have been published in recent years.

Through dynamic borrowing, the extent of borrowing is determined by the similarity or heterogeneity between the current trial and the RWD, for instance, through a commensurate parameter (Hobbs et al., 2012). However, the appropriate amount of borrowing often depends on factors beyond statistical considerations and the observed data, particularly in regulatory decision-making. For example, the relevance of the RWD to the current trial, the quality of the RWD, and the unmet medical need are critical factors that are not fully captured by the observed data but are essential for evaluating the appropriateness of the amount of borrowing. These considerations ensure that the borrowed information is not only statistically sound but also clinically meaningful and aligned with regulatory standards.

For confirmatory trials, joint input from clinicians, regulators, statisticians, and other experts and stakeholders should be sought at the design stage to determine the appropriate amount of borrowing, or at least to establish a benchmark value for the borrowing amount. Moreover, the amount of borrowing should be quantified using a metric that is easily interpretable and communicable among all parties, regardless of their statistical background. In other words, the discussion should remain straightforward and not be complicated by the potential statistical

☆Main article: https://doi.org/10.6339/24-JDS1161.

methods that will be applied to execute the borrowing. For instance, it is less ideal to convey the amount of borrowing as 'the weight on a mixture component of a mixture prior,' compared to 'the number of patients that will be added through borrowing.' After all, statistical thinking should address the questions of interest rather than determine the questions themselves, a general principle recently underscored by the estimand concept introduced in the FDA guidance, "E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials."

'Determining' the amount of borrowing at the design stage is a practical and critical aspect of designing confirmatory clinical trials that plan to leverage RWD/RWE. This arguably requires more attention from the field. Notably, the elegant simplicity of the power prior makes it a uniquely powerful tool for this purpose.

To illustrate this, consider a simplified Kociba setting without dose levels. Let y denote the number of responders out of m animals, and let y_0 denote the number of responders out of m_0 animals in historical data. For the power prior for the response rate p with power parameter a and an initial Beta(1, 1) prior, the posterior distribution of p is given by:

$$\pi(p|y, m, y_0, m_0) = \text{Beta}(y + ay_0 + 1, m + am_0 + 1).$$

Evidently, the power prior introduces an additional am_0 animals to the final analysis. Instead of delving into the specifics of the power parameter, the cross-disciplinary discussion can focus on the number of animals, am_0, that can be borrowed from historical data. This might seem trivial from a statistical perspective, but the practical benefits are substantial. Indeed, it is generally recommended that all methods developed for borrowing RWD/RWE include a built-in feature that facilitates straightforward discussions on the desirable amount of borrowing during the design stage.

Once the appropriate amount of borrowing, or a benchmark value, is determined, statistical methods can be applied to address potential biases introduced by leveraging RWD/RWE and execute the borrowing based on bias-adjusted data. It is important to note that, in addition to the selection bias (i.e., differences in patient populations) comprehensively covered in Chen et al. (2025), the operational differences between the current trial and the environment where the RWD was collected should also be carefully addressed. This aspect appears to be relatively under-discussed in the literature, especially from a design perspective.

2 Power Prior for Analyzing Clinical Trials

The intrinsic simplicity of the power prior can not only aid in the design but also enhance the interpretation of the analysis results.

In the previous section, we emphasized the importance of determining the benchmark amount of borrowing based on the collective input from all stakeholders. However, the actual amount of borrowing in the final analysis will need to be data-dependent for nearly all existing methods, whether the borrowing is dynamic or static. Therefore, it is essential to evaluate the robustness of the final analysis results by exploring the neighborhood of the benchmark amount of borrowing.

Consider a clinical trial where the endpoint is the number of certain recurrent event. Let p denote the log-hazard ratio (HR) between treatment and control. Suppose there exists a historical trial, from which p is estimated to be $\widehat{p}_0$ with a standard error of $\widehat{\sigma}_0^2$.

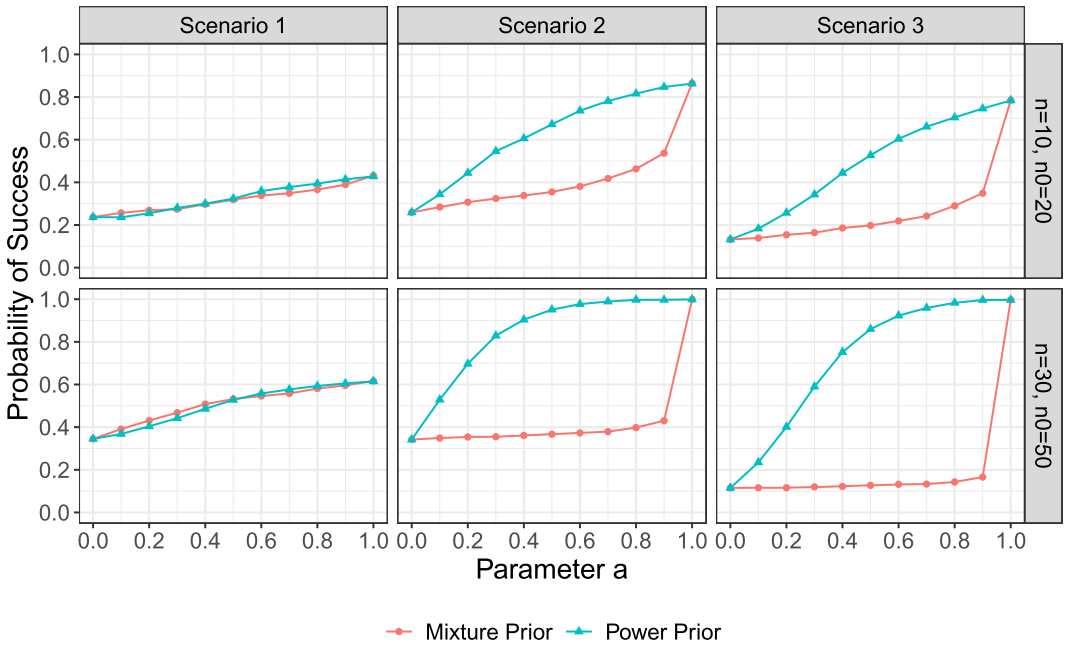

Figure 1: Simulation results for power prior vs. mixture prior.

For borrowing information from the historical trial, one may consider constructing a power prior as

$$\pi\left(p|\widehat{p}_0,\widehat{\sigma}_0^2\right) \propto \left[\phi_{\widehat{p}_0,\widehat{\sigma}_0^2}(p)\right]^a \phi_{0,B}(p)$$

where $\phi_{\mu,\sigma^2}(\cdot)$ is the density function for the normal distribution with mean μ and variance σ^2, and B is a large number that makes $\phi_{0,B}(p)$ non-informative. Alternatively, one may construct a mixture prior as

$$\pi\left(p|\widehat{p}_0,\widehat{\sigma}_0^2\right) = a\phi_{\widehat{p}_0,\widehat{\sigma}_0^2}(p) + (1-a)\phi_{0,B}(p).$$

Under these settings, we construct three data generation scenarios: (1) the HR for both the current and historical trials is 0.7; (2) the HR for the current trial is 0.7 and for the historical trial is 0.1; and (3) the HR for the current trial is 1 and for the historical trial is 0.1. Furthermore, we consider the current trial to have $n = 10$ or 30 patients per arm, and the historical trial to have $n_0 = 20$ or 50 patients per arm. We set the study success criterion to be $P(p > 0|\text{Data}) > 0.9$. We then evaluate the study success probability for different values of a based on 1000 replications and report the results in Figure 1.

In Scenario 1, when the current trial and the historical trial have the same HR, both priors perform almost identically. However, when the HRs differ between the two studies in Scenarios 2 and 3, the change in the probability of success with respect to the choice of a is much smoother for the power prior. Practically, this smoothness allows investigators and regulators to more easily interpret the impact of the amount of borrowing on the study results. Consequently, this further exemplifies the advantages the power prior can bring.

3 Power Prior for Addressing Temporal Changes

In addition to the review of the evolution of the power prior, Chen et al. (2025) also provides insightful thoughts on the future directions of research on the power prior. It appears that handling temporal changes in RWD could be added to the list of potential expansions.

In RWD, shifts in patients' characteristics, standard of care, and other clinical and operation features may occur over time, particularly when the RWD collection spans a long period. Addressing these temporal changes presents a challenge statistically. One solution is to parametrically model the temporal effect. However, this approach is prone to model misspecification and is difficult to be incorporated into the trial design.

Wang et al. (2023) proposed a non-parametric Bayesian approach for handling temporal changes. Their proposal constructs the prior by periodically mixing the posterior with a weak distribution. The power prior can be adopted to accomplish the same objectives. Specifically, let D_0 denote the RWD, and let $\{D_{0,1}, \ldots, D_{0,K}\}$ be a mutually exclusive partition of D_0 that follows the temporal order such that $D_{0,k}$ was observed prior to $D_{0,k'}$ for all $1 \leqslant k < k' \leqslant K$. Let p denote the parameter of interest. Then, the prior $\pi(p|D_0)$, can be constructed iteratively as

$$\begin{aligned}
\pi(p|D_{0,1}) &\propto L(p|D_{0,1})^{a_1}\pi_0(p) \\
\pi(p|D_{0,2}, D_{0,1}) &\propto \left[L(p|D_{0,2})\pi(p|D_{0,1})\right]^{a_2} \\
&\vdots \\
\pi(p|D_0) = \pi(p|D_{0,K}, \ldots, D_{0,1}) &\propto \left[L(p|D_{0,K})\pi(p|D_{0,K-1}, \ldots, D_{0,1})\right]^{a_K}.
\end{aligned}$$

Such construction enables the older data to be discounted more heavily compared to data that is closer to the current time, thereby addressing the temporal changes in a non-parametric manner. Moreover, as a special case, when $\{D_{0,1}, \ldots, D_{0,K}\}$ evenly split D_0 and $a_k = a$ for all k, the prior $\pi(p|D_0)$ will exhibit a 'memoryless' property. This means that the choice of a only needs to account for the temporal changes between two consecutive time intervals, regardless of the overall span of the RWD. This can be a beneficial feature for determining a, especially at the design stage.

Of course, the technical details about $\pi(p|D_0)$, such as posterior sampling, need to be explored further. However, I hope the idea can once again exemplify the potential power possessed by the power prior for addressing various issues.

References

Chen MH, Guan Z, Lin M, Sun M (2025). Power priors for leveraging historical data: looking back and looking forward. *Journal of Data Science*, 23(1): 1–30. https://doi.org/10.6339/24-JDS1161

Hobbs B, Sargent D, Carlin B (2012). Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Analysis*, 7(3): 639–674. https://doi.org/10.1214/12-BA722

Wang C, Lin M, Rosner GL, Soon G (2023). A bayesian model with application for adaptive platform trials having temporal changes. *Biometrics*, 79(2): 1446–1458. https://doi.org/10.1111/biom.13680