Journal of Data Science 23 (1), 64–69
January 2025

DOI: 10.6339/25-JDS1172
Statistical Data Science

# Rejoinder: "Power Priors for Leveraging Historical Data: Looking Back and Looking Forward"☆,☆☆

Ming-Hui Chen*, Zhe Guan, Min Lin, and Max Sun, for the Alzheimer's Disease Neuroimaging Initiative

*Department of Statistics, University of Connecticut, Storrs, CT 06269-4120, United States*

We would like to thank Dr. Fang K. Chen; Drs. Margaret Gamalo, Heliang Shi, Yuxi Zhao, and Maria Kudela; Dr. Lei Nie; Dr. Chenguang Wang; Dr. Guohui Wu; Professor Minge Xie; and Professor Panpan Zhang for their extremely helpful comments and discussions for this paper. Additional empirical investigations carried out in their discussion shed further light on how to more effectively borrow or more appropriately extract the information from external data. Their inputs add several theoretical and practical perspectives on power priors and their variations, with the hope of stimulating broader interest in future research. We address a few of the discussants' comments as follows.

**Simpson's Paradox** Professor Xie points out an interesting phenomenon of outlying (or discrepant) posterior distributions that arise in certain Bayesian analyses of clinical trials. In the same spirit as Xie et al. (2013), we consider two data sets from randomized controlled trials: a hypothetical historical migraine headache dataset and the real current migraine headache dataset, which is summarized in Table R.1. In Table R.1, $m_{00}$ and $m_{01}$ denote the numbers of subjects and $y_{00}$ and $y_{01}$ denote the numbers of responders in the control group ($t = 0$) and the treatment group ($t = 1$), respectively, in the historical trial; and $m_{10}$ and $m_{11}$ denote the numbers of subjects and $y_{10}$ and $y_{11}$ denote the numbers of responders in the control group ($t = 0$) and the treatment group ($t = 1$), respectively, in the current trial. Let $p_{00}$ and $p_{01}$ denote the responder rates of the control group and the treatment group, respectively, for the historical data; and also let $p_{10}$ and $p_{11}$ denote the responder rates of the control group and the treatment group, respectively, for the current data. For the current data, the maximum likelihood estimate (MLE) of the treatment effect $p_{11} - p_{10}$ is $33/59 - 31/68 = 0.103$, and for the historical data, the MLE of the treatment effect $p_{01} - p_{00}$ is $46/50 - 14/18 = 0.142$. However, in the pooled data, which are obtained by collapsing the historical and current data, shown in Table R.1, the MLE of the difference in the response rates of the two treatment groups is $79/109 - 45/86 = 0.202$. This pooled estimate is much greater than both individual estimates. Hence, the data in Table R.1 exhibit a sort of Simpson's paradox (Simpson, 1951), where the individual trends are magnified. This occurs because the MLEs of $p_{00}$ and $p_{10}$ are 0.778 and 0.456, respectively, showing substantial heterogeneity in the responder rates in the control groups between the historical and current data.

☆Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

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

*Corresponding author. Email: ming-hui.chen@uconn.edu.



Received January 25, 2025; Accepted January 28, 2025

Table R.1: Migraine headache data.

| | Historical | | Current | | Pooled | |
|---|---|---|---|---|---|---|
| $t$ | $y_{0t}$ | $m_{0t}$ | $y_{1t}$ | $m_{1t}$ | $y_t^*$ | $m_t^*$ |
| 0 | 14 | 18 | 31 | 68 | 45 | 86 |
| 1 | 46 | 50 | 33 | 59 | 79 | 109 |

We can reproduce and subsequently alleviate this phenomenon by conducting Bayesian analysis through the power prior (PP) and partial borrowing power prior (pPP) frameworks. The power prior can be constructed based on the historical data in Table R.1 as follows

$$\pi(p_{10}, \delta | D_0, a_0) \propto \left[ \prod_{t=0}^{1} (p_{10} + t\delta)^{y_{0t}} (1 - p_{10} - t\delta)^{m_{0t} - y_{0t}} \right]^{a_0} \pi_0(p_{10}, \delta), \quad \text{(R.1)}$$

where $D_0 = \{(y_{0t}, m_{0t}), t = 0, 1\}$ denotes the historical data and $\pi_0(p_{10}, \delta)$ is an initial prior on a constrained parameter space defined by $\Theta_{\mathrm{PP}} = \{(p_{10}, \delta) : 0 < p_{10} < 1, -p_{10} < \delta < 1 - p_{10}\}$. We note that in Equation (R.1), $p_{10} + \delta$ is a reparametrized responder rate for the treatment group and the same responder rate $p_{10} = p_{00}$ for the control group and we assume the same treatment effect $\delta$ for both the historical and current data. The parameter of interest is the treatment effect $\delta$. Similarly, the partial borrowing power prior can be expressed as

$$\pi(p_{10}, \delta | D_0, a_0) \propto \int \left[ \prod_{t=0}^{1} (p_{00} + t\delta)^{y_{0t}} (1 - p_{00} - t\delta)^{m_{0t} - y_{0t}} \right]^{a_0} \pi_0(p_{00}, p_{10}, \delta) dp_{00}, \quad \text{(R.2)}$$

where $\pi_0(p_{00}, p_{10}, \delta)$ is an initial prior on a constrained parameter space defined by $\Theta_{\mathrm{pPP}} = \{(p_{00}, p_{10}, \delta) : 0 < p_{00}, p_{10} < 1, -\min(p_{00}, p_{10}) < \delta < \min(1 - p_{00}, 1 - p_{10})\}$. Similarly, in Equation (R.2), $p_{01}$ is assumed to be $p_{00} + \delta$. In Equation (R.2), the responder rates $p_{00}$ and $p_{10}$ for the control group are assumed to be different between the historical and current data while the treatment effect $\delta$ is shared by both the historical and current data.

Table R.2 displays the posterior means, standard deviations (SD), and 95% highest posterior density (HPD) intervals for $\delta$ and $p_{10}$ using $\pi_0(p_{10}, \delta) \propto 1\{\Theta_{\mathrm{PP}}\}$, where $1\{A\}$ denotes the indicator function, which takes a value of 1 if $A$ is true and 0 if otherwise, and $\pi_0(p_{00}, p_{10}, \delta) \propto 1\{\Theta_{\mathrm{pPP}}\}$.

Table R.2: Posterior estimates of $\delta$ and $p_{10}$.

| | | $\delta$ | | | $p_{10}$ | | |
|---|---|---|---|---|---|---|---|
| Prior | $a_0$ | Estimate | SD | 95% HPD | Estimate | SD | 95% HPD |
| No Borrow | 0 | 0.100 | 0.086 | (−0.070, 0.268) | 0.457 | 0.059 | (0.343, 0.572) |
| PP | 0.1 | 0.120 | 0.084 | (−0.044, 0.282) | 0.465 | 0.058 | (0.352, 0.579) |
| | 0.5 | 0.169 | 0.076 | (0.020, 0.317) | 0.494 | 0.056 | (0.385, 0.604) |
| | 1.0 | 0.198 | 0.067 | (0.069, 0.332) | 0.522 | 0.053 | (0.420, 0.626) |
| pPP | 0.1 | 0.073 | 0.076 | (−0.064, 0.235) | 0.467 | 0.056 | (0.358, 0.578) |
| | 0.5 | 0.085 | 0.066 | (−0.041, 0.219) | 0.462 | 0.054 | (0.357, 0.565) |
| | 1.0 | 0.097 | 0.059 | (−0.016, 0.215) | 0.456 | 0.051 | (0.356, 0.557) |

Under the power prior, we are effectively "pooling" the current and historical data. In particular, for $a_0 = 1$, the posterior estimate for $\delta$ is 0.198, which nearly replicates the MLE obtained by collapsing the historical and current data. In short, the power prior is not an effective tool to resolve the Simpson's paradox in this case. The partial borrowing power prior accounts for the "stratification effect" of the historical and current data by assuming a different historical control event rate $p_{00}$ from the current $p_{10}$. As a result, the posterior estimates of $\delta$ do not exhibit any inflation — in fact, we observe a reduction in the estimates relative to No Borrow. Thus, the partial borrowing power prior is an effective remedy for the Simpson paradox.

The posterior estimates of $p_{10}$ in Table R.2 also facilitate a better understanding of the impact of the partial borrowing power prior on the non-borrowing parameters raised by Dr. F. Chen in his discussion. We note that in Equation (R.2), $\delta$ is the common parameter shared by both the historical and current data, while $p_{10}$ is a non-borrowing parameter for the current data. The posterior estimates for $p_{10}$ are quite similar, ranging from 0.467 to 0.456, under pPP with $a_0 = 0.1$, 0.5, and 1.0. These estimates are also similar to the one under No Borrow. Meanwhile, the posterior estimates for $\delta$ under partial borrowing vary from 0.073 to 0.097. These empirical results demonstrate that the impact of the partial borrowing power prior on the non-borrowing parameter is moderate, since the treatment effects for historical and current data are similar.

**Tipping Point Analysis** We are appreciative of Dr. Nie for broaching the subject of tipping point analysis and also Drs. Gamalo et al. for mentioning the topic as well. To illustrate the tipping point analysis of the ADNI data, we exchange the role of ADAS_bl and MCI in Equation (3) in Chen et al. (2025) so that $\gamma$ corresponds to the regression coefficient for the covariate ADAS_bl. In Table 4 in Chen et al. (2025), for the response variable, the changes in ADAS, the OLS estimates for the coefficient of ADAS_bl are −0.109 and −0.090 with $p$-values of 0.075 and 0.080 respectively, within the historical and current studies, and they are not significantly different from zero at a significance level of 0.05. We carry out the tipping point analysis with the power prior, the partial borrowing power prior, the borrowing-by-parts power prior ($\overline{\text{p}}$PP), and the partial borrowing-by-parts power prior (p$\overline{\text{p}}$PP) defined in Section 6.2 of Chen et al. (2025). For consistency, we set the discounting coefficients for the variance parameter as zero for $\overline{\text{p}}$PP and p$\overline{\text{p}}$PP. Following the work of Best et al. (2021), we define the tipping point in our context as the minimal value of $a_0$ (or $a_{01}$) so that the 95% HPD interval of $\gamma$ does not contain zero. We carry out the tipping point analysis via a grid search from 0 to 1 with difference 0.01. The results are displayed in Table R.3. We observe that even $a_0 = 1$, PP and pPP cannot reach the tipping point, while $\overline{\text{p}}$PP and p$\overline{\text{p}}$PP have tipping points at $a_{01} = 0.76$ and 0.18, respectively. We further note that p$\overline{\text{p}}$PP yields a posterior mean of $\gamma$ that is very close to the one produced under No Borrow, with a smaller posterior standard deviation.

**Covariate Mismatch** As pointed out by Drs. Gamalo et al. and Dr. Wang, covariate mismatch between the current and historical control arms may undermine the reliability of the power prior. We consider two possible directions for improvement. If individual patient covariates are available, the strengths of the propensity score (PS) techniques can be combined with variations of the power prior, as briefly described in Section 4.4 of Chen et al. (2025). If individual patient covariates are not available, but a temporal order $D_0 = \{D_{0,1}, \ldots, D_{0,K}\}$ of the external data set $D_0$ is available, where $D_{0,k}$ was observed before $D_{0,k'}$ for all $1 \leqslant k \leqslant k' \leqslant K$, Dr. Wang proposed an interesting sequentially down-weighted power prior to account for temporal effects. However, the initial prior $\pi_0(p)$ is discounted multiple times, and a slightly modified version might appear

Table R.3: Tipping point analysis.

| Prior | $a_0$ | Estimate | SD | 95% HPD |
|---|---|---|---|---|
| No Borrow | – | −0.090 | 0.051 | (−0.190, 0.011) |
| PP | 1 | −0.076 | 0.040 | (−0.154, 0.002) |
| pPP | 1 | −0.076 | 0.039 | (−0.153, 0.001) |
| | $(a_{01}, 0)$ | | | |
| $\overline{\text{p}}$PP | (0.76, 0) | −0.076 | 0.039 | (−0.152, −0.000) |
| p$\overline{\text{p}}$PP | (0.18, 0) | −0.093 | 0.047 | (−0.185, −0.000) |

more appealing:

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

In a similar spirit, Gamalo et al. (2014) consider a case when an ordering of priority or relevance of the historical data sets is present.

**Propensity-Score-Based Power Priors** Dr. F. Chen raises several interesting and important issues regarding the propensity-score (PS) based power priors. As in all PS-based methods, the validity of PS-based variants of power priors are also subject to the validity of the propensity model. How to properly specify a propensity model is out of scope of the current rejoinder, and it is an ongoing endeavor to understand the impact of model misspecification of the propensity scores on the PS-based power priors. It is worth noting that, under the additive outcome model assumption and a certain permutation invariance property, Baron et al. (2024) show that the biases introduced by the PS stratification procedure can be canceled out within each stratum, even if unmeasured confounders are present.

**Smoothness and Mixing Prior** Dr. Wang points out the local smoothness property of the power prior and we appreciate his simulation study illustrating this property. The mixture prior, in contrast, may produce a surprising jump in the estimated probability of success when the tuning parameter increases a little.

**Interpretation of the Priors and the Amount of Information Borrowing** Dr. Wu asks the two interesting questions regarding the interpretation of $a_0$ and the amount of information borrowing in the normalized power prior. Letting $L(\boldsymbol{\beta}|D)$ denote the likelihood function of the current data $D$, the posterior distribution of $\boldsymbol{\beta}$ under the normalized power prior given in Equation (11) of Chen et al. (2025) can be written as

$$\pi(\boldsymbol{\beta}|D, D_0) \propto L(\boldsymbol{\beta}|D)\int_0^1 \pi(\boldsymbol{\beta}, a_0|D_0)da_0.$$

Thus, $a_0$ is similar to a non-borrowing parameter $p_{00}$ assumed for the historical data in the partial borrowing power prior in Equation (R.2). Therefore, $a_0$ can be neither directly interpreted nor used to quantify the amount of information borrowing. As discussed in Section 6.1 of Chen et al. (2025), the posterior densities under power prior with $a_0 = 0.1$ and the normalized power prior are very close to each other, which is expected since the posterior mean $a_0$ under the normalized power prior is 0.121, which is close to 0.1. In this case, the normalized power prior as a whole may be well approximated by the power prior with $a_0 = 0.121$.

Dr. F. Chen raises an insightful point regarding the interpretation of $a_0$ in the context of partial borrowing power priors. Unlike regular power priors, where $a_0$ can be intuitively understood as the proportion of an equivalent historical subsample (as illustrated in Figure 3 of his discussion), such an analogy becomes less straightforward under partial borrowing. This is because the partial borrowing power prior operates on the marginal prior for a subset of parameters, effectively projecting onto a subspace of the parameter space. The integration over nuisance parameters alters the coordinate system, complicating direct interpretation of $a_0$. To address this complexity, the concept of effective sample size could provide a more meaningful measure for quantifying the influence of $a_0$ in partial borrowing settings. Specifically, Dr. Wang, in the first section of his discussion, notes that the effective sample size of the regular power prior for the probability parameter under a binary outcome would be $a_0$ multiplied by the sample size of the historical data set. The general definition of the effective sample size for variants of the power prior is still under investigation, and further exploration is needed to adapt this concept for partial borrowing scenarios.

**Best Practices of Using Power Priors** Professor Zhang wants us to discuss an optimal way of choosing the discounting parameter $a_0$. However, as pointed out by Dr. Wang, the appropriate amount of borrowing often depends on factors beyond statistical considerations and the observed data. It is essential to include all stakeholders into discussion for determining a benchmark $a_0$ or the maximum $a_0$. The extension of normalized power priors for multiple historical data sets is still under-developed. An extended, in-depth discussion, and empirical investigation on this topic deserves to be a future research project.

**Variational Bayes** Dr. Wu mentioned accelerating model fitting with the normalized power prior via variational Bayes (VB). VB is indeed a significantly faster alternative to MCMC. However, the use of historical borrowing typically implies moderate to small sample sizes, a scenario where the reliability of VB is unclear and necessitates further research. In particular, it is well known that VB generally underestimates posterior variance (Blei et al., 2017) due to the independence simplification of the mean-field family.

## References

Baron E, Lin M, Zhu J, Tang RS, Chen MH (2024). Enhancing randomized controlled trials: A Bayesian divide-and-conquer approach for borrowing external control data. *Statistics in Biosciences*, https://doi.org/10.1007/s12561-024-09465-2.

Best N, Price RG, Pouliquen IJ, Keene ON (2021). Assessing efficacy in important subgroups in confirmatory trials: An example using Bayesian dynamic borrowing. *Pharmaceutical Statistics*, 20(3): 551–562. https://doi.org/10.1002/pst.2093

Blei DM, Kucukelbir A, McAuliffe JD (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518): 859–877. https://doi.org/10.1080/01621459.2017.1285773

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

Gamalo MA, Tiwari RC, LaVange LM (2014). Bayesian approach to the design and analysis of non-inferiority trials for anti-infective products. *Pharmaceutical Statistics*, 13(1): 25–40. https://doi.org/10.1002/pst.1588

Simpson EH (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, Series B (Methodological)*, 13(2): 238–241. https://doi.org/10.1111/j.2517-6161.1951.tb00088.x

Xie M, Liu RY, Damaraju CV, Olson WH (2013). Incorporating external information in analyses of clinical trials with binary outcomes. *Annals of Applied Statistics*, 7(1): 342–368. https://doi.org/10.1214/12-AOAS585