Bayesian Analysis Using Power Priors with Application to Pediatric Quality of Care

Brian Neelon[1]* and A. James O'Malley[2]

[1]Nicholas School of the Environment, Duke University, Durham, North Carolina, USA
[2]Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts, USA

Abstract

We illustrate how power prior distributions can be used to incorporate historical data into a Bayesian analysis when there is uncertainty about the similarity between the current and historical studies. We compare common specifications of the power prior and explore whether it is preferable to condition on the power parameter, a_0, or to treat it as a random variable with a prior distribution of its own. We show that there are two natural ways of formulating the power prior for random a_0. The first approach excludes the historical data in all but extreme cases and may therefore be of limited practical use. The second approach, called the normalized power prior (NPP), provides a measure of congruence between the current and historical data, so that the historical data are downweighted more substantially as the studies diverge. While this is an intuitively appealing property, our experience suggests that in real world problems involving large datasets and models with several parameters, the NPP may lead to considerably more downweighting than desired. We therefore advise practitioners to consider whether such attenuation is desirable, or whether it is more appropriate to assign a_0 a fixed value based on expert opinion about the relevance of the historical data to the current analysis. We also extend the power prior to hierarchical regression models that allow covariate effects to differ across studies. We apply these methods to a pair of studies designed to improve delivery of care in pediatric clinics.

Keywords: Bayesian hierarchical model; Power prior; Normalized power prior; Historical data; Evidence synthesis

Introduction

Investigators conducting new research often have access to data from previous studies, and in such cases it is not only scientifically reasonable but also statistically advantageous to incorporate this information into the current analysis. Consider, for example, the common scenario in which a funding agency finances research incrementally, first requiring a small pilot or feasibility study before funding a more elaborate trial. In such cases, it can be beneficial to incorporate the pilot data into the subsequent analysis to increase the power to detect treatment effects. One strategy for synthesizing results across studies is through a Bayesian modeling approach. Because Bayesian methods can incorporate historical information through a prior distribution, they provide a natural framework for updating information across studies.

There are several advantages to a Bayesian analysis with informative priors elicited from historical data. First, informative priors can yield estimates consistent with accepted views of effects being studied [1]. For example, one might wish to restrict a model parameter to a reasonable range (e.g., a non-negative effect for smoking) to ensure a result consistent with established findings. Second, informative priors can improve the precision of estimates and increase the ability to detect treatment effects, even if the priors are only used to inform ancillary parameters. And finally, informative priors can be used in settings where diffuse priors may lead to computational difficulties or even improper posterior distributions [2].

While historical priors provide a useful analytic tool, there are times when investigators will want to limit the impact of historical data on the current analysis. For example, they may question the compatibility of the current and historical data, and may therefore wish to "downweight" the historical information to reduce its impact. Such downweighting may also be required by a regulatory agency, such as the US Food and Drug Administration (FDA), that wishes to limit the impact of historical information on the analysis of a new treatment.

Using an illustrative example, this paper explores common Bayesian approaches to incorporating historical information into regression models when there is uncertainty about the similarity between the current and historical studies. Our example involves a pair of studies designed to improve the delivery of care in pediatric clinics. New mothers were surveyed during baseline and follow-up periods and asked to rate their quality of care in a variety of areas; their responses were then aggregated into a binary measure representing high- or low-quality care. The aim was to evaluate the intervention effect comprehensively across both studies. However, while the two questionnaires overlapped substantially, they were not identical measures. To express our uncertainty about the compatibility of the two studies, we used the *power prior* methodology developed by Ibrahim and Chen [2] to mitigate the impact of the historical data in the event that the two studies differed in unobservable ways. The power prior introduces a parameter that explicitly controls the amount of weight assigned to the historical data. The approach has been applied to a variety of settings, including analyses involving cure rate models [3], environmental quality assessments [4], sample size determination [5], clinical trials [6], health-care evaluation [7] and cross-design meta-analysis [8,9].

Corresponding author: Brian Neelon, Children's Environmental Health Initiative, Nicholas School of the Envi-ronment, Box 90328 Duke University, Durham, NC 27708-0328, USA, E-mail: brian.neelon@duke.edu

In addition, we explore whether the data can determine how much weight to assign to the historical data, or whether this impact factor should be user-specified. We also propose an extension of the power prior to exchangeable hierarchical models. Here, we allow the covariate effects to differ across studies but assume that they share a common prior distribution. We then use the power prior to further attenuate the impact of the historical data in the event that the exchangeability assumption is itself too strong.

The remainder of this paper is organized as follows: Section 2 reviews common specifications of the power prior, explores strategies of assigning weights to the historical data, and describes the proposed extension to exchangeable hierarchical models; Section 3 provides background for the case study and presents analyses using the conventional and exchangeable power priors; and the final section summarizes the methods and offers guidelines for their use in practice.

The Power Prior

The conditional power prior

Let D and D_0 denote the data from current and historical studies, respectively. Ibrahim and Chen [2] define the *power prior* distribution for a set of parameters θ as

$$
\pi\left(\theta | D_0, a_0\right) \equiv \frac{L\left(D_0|\theta\right)^{a_0}\pi_0(\theta)}{\int_\Theta L\left(D_0|\theta\right)^{a_0}\pi_0(\theta)d\theta}
$$
$$
= \frac{L\left(D_0|\theta\right)^{a_0}\pi_0(\theta)}{g\left(a_0\right)}
$$
$$
\propto L\left(D_0|\theta\right)^{a_0}\pi_0(\theta),
$$
(2.1)

where $L(D_0|\theta)$ denotes the likelihood for the historical study, a_0 is (for now) a fixed, known constant ranging from 0 and 1, $g\left(a_0\right)=\int_\Theta L\left(D_0|\theta\right)^{a_0}\pi_0(\theta)d\theta$ is the normalizing constant, and $\pi_0(\theta)$ the *initial prior* assigned to θ before observing D_0. In many cases, $\pi_0(\theta)$ will be taken to be diffuse to re reflect lack of knowledge about θ prior to observing the historical data. We refer to prior (2.1) as the *conditional power prior*, since it is formed by conditioning on both D_0 and the fixed parameter a_0. The parameter a_0 governs the impact of the historical data on the current analysis, ranging from no influence when $a_0 = 0$ to parity when $a_0 = 1$, in which case (2.1) is just the usual posterior update of θ based on D_0. We note also that the power prior will be proper as long as the normalizing constant is finite.

Given (2.1), Bayes' theorem implies

$$
\pi\left(\theta | D, D_0, a_0\right) \propto L\left(D|\theta\right)\pi\left(\theta|D_0, a_0\right) \propto L\left(D|\theta\right)L\left(D_0|\theta\right)^{a_0}\pi_0(\theta),
$$
(2.2)

where $L(D|\theta)$ denotes the current-data likelihood. Note that, at this stage of the analysis, a_0 and D_0 are treated as hyperparameters that impact θ only through the power prior in (2.1). When $a_0 = 1$, the posterior for θ based on both D and D_0 can be expressed as $\pi\left(\theta|D,D_0,a_0=1\right) \propto L\left(D|\theta\right)L\left(D_0|\theta\right)\pi(\theta) = L\left(D,D_0|\theta\right)\pi_0(\theta)$ and hence assigning $a_0 = 1$ is tantamount to pooling the current and historical data and basing posterior inference on the aggregated data. Conversely, when $a_0 = 0$, the historical data are excluded altogether, and the prior for θ reduces to the initial prior, $\pi_0(\theta)$. And, when $0 < a_0 < 1$, the log-likelihood contributions of the historical subjects are downweighted by 100 x (1 - a_0)%. The parameter a_0 can also be viewed as a precision parameter that inversely affects the heaviness of the tails of the prior for θ (smaller a_0 values lead to heavier tails).

For generalized linear models, Ibrahim and Chen [2] recommend constructing the conditional power prior as

$$
\pi\left(\beta | D_0, a_0, \Sigma, \tau\right) \propto \prod_{i=1}^{n_0}\left[\int_{\Re^{q_m}}\prod_{j=1}^{n_{0i}}\left\{p\left(y_0 | X_0, z, \beta, b_0, \tau\right)\right\}^{a_0}\pi\left(b_0|\Sigma\right)db_0\right]\pi_0(\beta),
$$
(2.3)

where $P(y_{oij}|.)$ denotes the density function for y_{oij} ($j=1,\dots, n_0$; $i=1,\dots, n$); β is a $P{\times}1$ vector of fixed effect coefficients; b_{oij} is a $N_q(0,\Sigma)$ vector of random effects for the ith historical cluster; x_{0ij} and z_{0ij} are fixed and random effect covariate vectors; and τ is a dispersion parameter (perhaps known). Here, the power prior is placed on the historical likelihood given the random effects, rather than on the marginal likelihood formed by first integrating out the subject-specific effects. As Ibrahim and Chen [2] point out, conditioning on the random effects has several interpretive and computational advantages, including efficient implementation within a Markov chain Monte Carlo (MCMC) algorithm. We refer readers to section 4 of their paper for further details. Prior specification is completed by assuming an initial prior, $\pi_0(\beta)$, for β, as well as prior distributions for the dispersion parameter τ (if unknown) and the random-effect covariance matrix Σ.

Joint power priors

The ibrahim-chen joint power prior: Within a Bayesian framework, it is natural to express uncertainty about the value of a_0-and hence about the relevance of the historical data-by placing a prior distribution on a_0 and allowing the data to help determine its most likely value. Doing so defines a joint power prior, $\pi(\theta, a_0|D_0)$. Investigators have typically specified this joint power prior in one of two ways. Under the first approach, proposed in various articles by Ibrahim and Chen [e.g., 10,2,11], the joint prior is specified as

$$
\pi^{IC}\left(\theta, a_0 | D_0\right) \equiv \frac{L\left(D_0|\theta\right)^{a_0}\pi_0(\theta)\pi\left(a_0\right)}{\int_0^1\int_\Theta L\left(D_0|\theta\right)^{a_0}\pi_0(\theta)\pi\left(a_0\right)d\theta da_0}
$$
$$
\propto L\left(D_0|\theta\right)^{a_0}\pi_0(\theta)\pi\left(a_0\right),
$$
(2.4)

where $\pi(a_0)$ denotes a prior for a_0, such as a Be(λ, δ) prior, with support on [0,1]. We refer to prior (2.4) as the Ibrahim-Chen (IC) joint power prior. Chen et al. [10] derive sufficient conditions for its propriety, noting that for normal models, prior (2.4) is always proper, while for a binomial model with a logit link, λ must be greater than the rank of the design matrix to ensure propriety.

The corresponding joint posterior is given by

$$
\pi^{IC}\left(\theta, a_0 | D, D_0\right) = \frac{L\left(D|\theta\right)\pi\left(\theta, a_0|D_0\right)}{\int_0^1\int_\Theta L\left(D|\theta\right)\pi\left(\theta, a_0|D_0\right)d\theta da_0}
$$
$$
= \frac{L\left(D|\theta\right)L\left(D_0|\theta\right)^{a_0}\pi_0(\theta)\pi\left(a_0\right)}{\int_0^1\int_\Theta L\left(D|\theta\right)L\left(D_0|\theta\right)^{a_0}\pi_0(\theta)\pi\left(a_0\right)d\theta da_0}
$$
$$
\propto L\left(D|\theta\right)L\left(D_0|\theta\right)^{a_0}\pi_0(\theta)\pi\left(a_0\right).
$$
(2.5)

However, while this posterior is well defined, our experience suggests that the mass of the posterior distribution coalesces near zero as the prior variance of a_0 increases, effectively excluding D_0 under diffuse priors. In Appendix A, we show that the posterior mode occurs at either $a_0 = 0$ or at $a_0 = 1$ under a U[0,1] prior. (In most practical situations, the mode occurs at 0.) This result is evident in previous applications of the power prior, although, with the exception of Duan et al. [4], Duan and Ye [12] and more recently Neuenschwander et al. [13], it has not been explicitly

noted in the literature. In fact, this tendency for the posterior mass to concentrate near zero holds even when $D = D_0$, since a_0 is not affected by the commensurability between the current and historical data under the IC power prior.

To illustrate this last point, we conducted a small simulation study to examine how the posterior mean of $a_0, E\left(a_0 \mid \overline{y} \, \overline{y}_0\right)$, changes as $\pi(a_0)$ becomes more diffuse. We considered a simple normal model with unknown mean, μ and known variance σ^2; that is, $\overline{y} \sim N\left(\mu, \sigma^2/n\right)$ and $\overline{y}_0 \sim N\left(\mu, \sigma^2/n_0\right)$, where $\overline{y}$ and $\overline{y}_0$, the current and historical sample means, are sufficient statistics for the data. We assumed a beta prior for a_0 with mean $\mu a_0 = 0.50$ and a diffuse initial prior for $\mu, \pi_0(\mu) \propto 1$. Under these conditions, the IC joint power prior for μ and a_0 takes the form

$$\pi^{IC}\left(\mu, a_0 \mid \overline{y}_0\right) \alpha N\left(\overline{y}_0 \mid \mu, \sigma^2\right)^{a_0} Be\left(a_0 \mid \lambda, \delta\right), \qquad (2.6)$$

with λ equal to δ to ensure a prior mean for $a_0 = 0{:}50$. To explore the impact of sample size on the marginal posterior mean of $a_0, E\left(a_0 \mid \overline{y}, \overline{y}_0\right)$, we simulated 1, 2, 10 and 100 observations from a N(0,1) distribution and used these same observations for both D and D_0 (i.e., the two samples were identical). Below, we report the results based on a single simulated dataset from each of the four sample size settings. (As an informal check, we simulated datasets under different random-number generating seeds, and the results were similar in each case.) To examine the impact of increasing historical sample size, we inflated D_0 by factors of 1, 2 and 5 by replicating the current data an appropriate number of times.

Since the joint posterior $\pi\left(\mu, a_0 \mid \overline{y}, \overline{y}_0\right)$ does not have a closed form, we used the Bayesian software package WinBUGS [14] to conduct the simulations. We assigned a_0 a beta prior with fixed mean 0.50 and varying prior standard deviations (σ_{a_0}) that reflected increasing degrees of diffuseness. Specifically, we ran the model under 10 different choices for σ_{a_0}, ranging from 0 to 0.50, the upper bound for a beta random variable with mean 0.50. We ran three, initially-dispersed MCMC chains for 30,000 iterations

each, discarding the first 10,000 as a burn-in. MCMC diagnostics, such as trace plots and Gelman-Rubin statistics [15], indicated rapid convergence and efficient mixing of the chains.

Figure 1(a)-(d) plot $E\left(a_0 \mid \overline{y}, \overline{y}_0\right)$ as a function of the σ_{a_0} under various current and historical sample sizes. On the x-axis, we have highlighted the value 0.2887, the standard deviation corresponding to a U[0,1], or Be(1,1), prior. Figures (a)-(d) show the posterior means for current sample sizes $n = 1$, 2, 10 and 100, respectively. Within each figure, we plot $E\left(a_0 \mid \overline{y}, \overline{y}_0\right)$ when the historical data are inflated by factors of 1, 2 and 5.

The graphs suggest that $E\left(a_0 \mid \overline{y}, \overline{y}_0\right)$ is affected by both the prior standard deviation and the study sample sizes. When the current sample size $n = 1, E\left(a_0 \mid \overline{y}, \overline{y}_0\right)$ increases as σ_{a_0} and n_0 increase, so that when $n_0 = 5$ and $\sigma_{a_0} \geq 0.40, E\left(a_0 \mid \overline{y}, \overline{y}_0\right) \approx 1$. However, as the current study sample size n increases, $E\left(a_0 \mid \overline{y}, \overline{y}_0\right)$ decreases with increasing σ_{a_0}. In fact, even when $n = 10$ [Figure (c)], $E\left(a_0 \mid \overline{y}, \overline{y}_0\right)$ converges rapidly toward zero as σ_{a_0} increases. This effect is even more dramatic when $n = 100$ [Figure (d)]. Interestingly, when $n = 1$, inflating n_0 accelerates the converges to 1, while when n $\geq$ 10, increasing n_0 accelerates the convergence to 0. This reflects a somewhat counterintuitive feature of the IC prior that, for any realistically sized $n, E\left(a_0 \mid \overline{y}, \overline{y}_0\right)$ approaches zero as n_0 approaches infinity, excluding the historical data altogether. In order to hold $E\left(a_0 \mid \overline{y}, \overline{y}_0\right)$ fixed as n_0 increases, σ_{a_0} must decrease by a proportionately greater amount. This property contradicts what one would expect if a_0 accounted for the commensurability between the current and historical data. In our simulations, for example, $D = D_0$, and so we would expect $E\left(a_0 \mid \overline{y}, \overline{y}_0\right)$ to increase as σ_{a_0} increased, regardless of the value of n. Only a highly informative prior distribution bounds the posterior mean of a_0 from 0. For most applications, this entails choosing σ_{a_0} small enough to avoid overly discounting D_0. From a practical standpoint, this is essentially tantamount to assigning a_0 a fixed value.

The normalized power prior: To avoid potentially excessive attenuation of the historical data, Duan and Ye [12] defined a *normalized power prior* (NPP), which is formed by first normalizing the conditional power prior (2.1) and then multiplying this normalized density by an independent proper prior for a_0:

$$\pi^{NPP}\left(\theta, a_0 \mid D_0\right) \equiv \pi\left(\theta \mid D_0, a_0\right) \pi\left(a_0\right)$$
$$= \frac{L\left(D_0 \mid \theta\right)^{a_0} \pi_0(\theta) \pi\left(a_0\right)}{\int_\Theta L\left(D_0 \mid \theta\right)^{a_0} \pi_0(\theta) d\theta}$$
$$= \frac{L\left(D_0 \mid \theta\right)^{a_0} \pi_0(\theta) \pi\left(a_0\right)}{g\left(a_0\right)}, \qquad (2.7)$$

Where $\pi(a_0)$ is a proper distribution with support $A \subset [0, 1]$ such that $g(a_0) = \int_\Theta L\left(D_0 \mid \theta\right)^{a_0} \pi_0(\theta) d\theta$ is finite for all $a_0 \in A$. In general, any proper distribution truncated to A can be used for $\pi(a_0)$. If the denominator in (2.7) is finite for all $\pi(a_0) \in [0, 1]$, then a natural choice for $\pi(a_0)$ is Be (λ, δ). For fixed a_0, the NPP reduces to the conditional power prior given by expression (2.1).

While the NPP does not follow directly from Bayes' theorem, there is an intuitive rationale behind its construction: a) after observing D_0, update θ using Bayes' theorem (ignoring a_0 for the moment); b) to downweight D_0, introduce a_0 and form the conditional power prior (2.1); c) to express uncertainty about a_0, define a new joint power prior by first normalizing (2.1) with respect to θ and multiplying this expression by an independent initial prior for a_0;
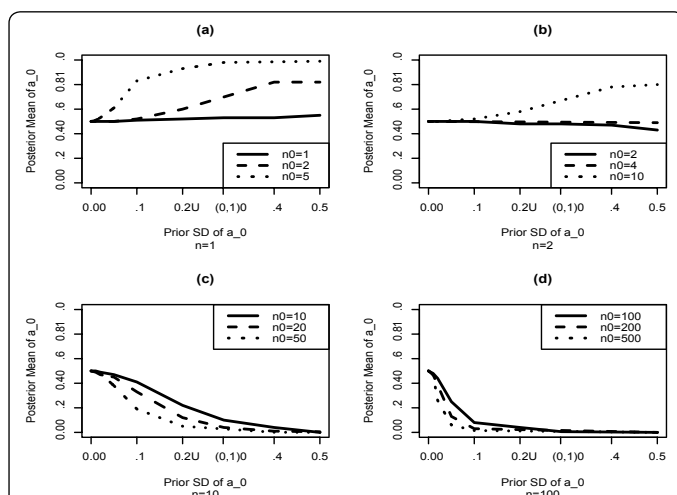

Figure 1: $E\left(a_0 \mid \overline{y}, \overline{y}_0\right)$ as a function of σ_{a_0} under the Ibrahim and Chen (IC) power prior for a normal model with unknown mean and known variance. Current data sample sizes are (a) $n = 1$, (b) $n = 2$, (c) $n = 10$ and (d) $n = 100$. Each plot presents trends when the historical sample size (n_0) is inflated by factors of 1 (solid), 2 (dashed) and 5 (dotted). In each plot, identical observations were used for both D and D_0 (replicates were used to inflate D_0). U [0,1] corresponds to the value 0.2887, the standard deviation for a uniform (0,1) random variable. As the figures indicate, for $n \geq 10$, $E\left(a_0 \mid \overline{y}, \overline{y}_0\right)$ decreases as σ_{a_0} increases, effectively excluding D_0.

and finally, d) treat (2.7) as a new prior distribution for analysis of the current data, and update (θ, a_0) jointly using Bayes' theorem. Under the NPP, the joint posterior for (θ, a_0) is expressed as

$$
\begin{aligned}
\pi^{NPP}\left(\theta, a_0 | D, D_0\right) &= L(D|\theta)\pi\left(\theta, a_0 | D_0\right) \\
&= \frac{L(D|\theta)L(D_0|\theta)^{a_0}\pi_0(\theta)\pi(a_0)/g(a_0)}{\int_A \int_\Theta L(D|\theta)L(D_0|\theta)^{a_0}\pi_0(\theta)\pi(a_0)/g(a_0)\,d\theta\,da_0} \quad (2.8) \\
&\propto L(D|\theta)L(D_0|\theta)^{a_0}\pi_0(\theta)\pi(a_0)/g(a_0).
\end{aligned}
$$

As Duan and Ye [12] note, unlike the IC prior, the NPP obeys the likelihood principle, since multiplying $L(D_0|\theta)$ by a factor K does not change the posterior. The NPP also has the virtue of "calibrating" the current and historical data so that more weight is given to D_0 when the two samples are similar. As the studies diverge, D_0 is increasingly attenuated. Likewise, as n_0 increases, D_0 is again downweighted, as evidence accumulates to suggest that the underlying populations differ. While this is an intuitively appealing feature-one would expect D_0 to be attenuated as it diverges from D-our experience indicates that unless the two studies are nearly identical, this attenuation tends to be quite excessive, obviating the need for the historical data for most realistically complex models. Although the rate of convergence to zero is less precipitous than under the IC power prior, it is substantial enough to question the general applicability of the NPP.

To illustrate this point, we return to the normal model above. Assuming, as above, a beta prior for a_0 and a diffuse initial prior for μ, the joint *NPP* is given by

$$
\begin{aligned}
\pi^{NPP}\left(\mu, a_0 / \overline{y}_0\right) &\equiv \frac{N\left(\overline{y}_0|\mu, \sigma^2/n_0\right)^{a_0} Be(a_0|\lambda, \delta)}{\int_{-\infty}^{\infty} N\left(\overline{y}_0|\mu, \sigma^2/n_0\right)^{a_0} d\mu}, \\
&= \left(\frac{\sigma^2}{n_0}\right)^{\frac{a_0-1}{2}} a_0^{\frac{1}{2}} N\left(\overline{y}_0|\mu, \sigma^2/n_0\right)^{a_0} Be(a_0|\lambda, \delta). \quad (2.9)
\end{aligned}
$$

As we outline in Appendix B, the marginal posterior for a_0 is given by

$$
\pi^{NPP}\left(a_0 / \overline{y}, \overline{y}_0\right) \propto N\left[\overline{y}-\overline{y}_0 \Big|0, \left(\frac{1}{n}+\frac{1}{a_0 n_0}\right)\sigma^2\right] Be(a_0|\lambda, \delta), \quad (2.10)
$$

Where $N\left(\overline{y}-\overline{y}_0|\mu, \sigma^2\right)$ denotes a normal density with mean μ and

variance σ^2 evaluated at $\overline{y}-\overline{y}_0$.[1] ([1]Duan et al. [4] derived the posterior for the more general case where σ^2 is unknown.) The variance term in equation (2.10) is intuitively appealing since it is the sum of the current and historical variances with the latter scaled by a_0.

Figures (2) (a)-(d) investigate the impact of sample size on $E\left(a_0|\overline{y}, \overline{y}_0\right)$ when $\overline{y}=\overline{y}_0$ (i.e., identical current and historical data for this special case). As the figures indicate, $E\left(a_0|\overline{y}, \overline{y}_0\right)$ is largely invariant to changes in sample size under the NPP. In all cases, $E\left(a_0|\overline{y}, \overline{y}_0\right)$ increases as a function of σ_{a_0}. When σ_{a_0} is near zero, corresponding to a highly informative prior for a_0, $E\left(a_0|\overline{y}, \overline{y}_0\right)$ is close to the prior mean of 0.50, while under a U[0,1] prior, $E\left(a_0|\overline{y}, \overline{y}_0\right) \approx 0.57$. Interestingly, $E\left(a_0|\overline{y}, \overline{y}_0\right)$ does not exceed 0.65 for any of the plots. (This upper bound is due in part to the fact that μa_0 was fixed at 0.50; larger values of μa_0 would likely yield larger posterior means.) As n_0 is inflated, $E\left(a_0|\overline{y}, \overline{y}_0\right)$ decreases slightly, but the effect is minimal.

Next, we examined the posterior behavior of a_0 under the NPP as the effect size $\overline{y}-\overline{y}_0$ increases. Figures 3(a)-(d) plot the marginal posterior distribution, $\pi\left(a_0|\overline{y}, \overline{y}_0\right)$, when a U[0,1] prior is assumed for a_0 (i.e., $\mu a_0 = 0.50$ and $\sigma_{a_0} = .2887$), $n = n_0 = 100$, $\sigma^2 = 1$, $\overline{y}_0 = 0$ and $\overline{y}$ ranges from 0 to 0.50, reflecting increasing disparity between the studies. When $\overline{y} = \overline{y}_0 = 0$ the posterior is diffuse with a posterior mode at 1, as expected. As the effect size increases, however, the distribution converges to zero. When $\overline{y} = 0.50$, the distribution is concentrated near 0, with a posterior mode of 0.04, a posterior mean of 0.20, and a posterior standard deviation of 0.18.

Figures 4(a)-(b) explore this phenomenon further by plotting $E\left(a_0|\overline{y}, \overline{y}_0\right)$ as function of the σ_{a_0} for increasing effect sizes. Figure 4(a) presents results for the case where $n = n_0 = 100$, $\overline{y}_0 = 0$ and $\overline{y}$ is allowed to range from 0 to 1.0. Figure 4(b) displays a similar graph for sample sizes $n = 2,531$ and $n_0 = 895$, which conform to the sample sizes for the case study presented later.

The results suggest that $E\left(a_0|\overline{y}, \overline{y}_0\right)$ is influenced by both the effect size and the study sample sizes n and n_0. As the discrepancy between $\overline{y}$ and $\overline{y}_0$ increases, the historical data are substantially

Figure 2: $E\left(a_0|\overline{y}, \overline{y}_0\right)$ as a function of σ_{a_0} under the normalized power prior (NPP) when D under the normalized power prior (NPP) when $D = D_0$. Current data sample sizes are (a) $n = 1$, (b) $n = 2$, (c) $n = 10$, and (d) $n = 100$. Each plot presents trends when historical sample sizes are inflated by factors of 1 (solid), 2 (dashed) and 5 (dotted). The figures suggest that, unlike the IC power prior, the NPP is relatively invariant to changes in n and n_0 when the data are identical.

Figure 3: The marginal posterior of a_0 under the NPP for increasing effect size $\overline{y}-\overline{y}_0$. For each figure, $\pi(a_0) = U(0; 1)$ (i.e., $\mu_{a0} = 0.50$ and $\sigma_{a_0} = .2887$), $n = n_0 = 100$, $\sigma^2 = 1$, $\overline{y}_0 = 0$, and $\overline{y}$ takes on four increasing values: (a) $\overline{y} = 0$, (b) $\overline{y} = 0.10$, (c) $\overline{y} = 0.25$, and (d) $\overline{y} = 0.50$. As y increases, reflecting increasing disparity between the studies, $\pi(a_0|\overline{y}, \overline{y}_0)$ converges to 0, thus downweighting the historical data.

downweighted, particularly as σ_{a_0} increases. In Figure 4(a), for example, when $\bar{y} = 0.75$ and U [0,1] is assumed for $\pi(a_0)$, $E(a_0|\bar{y}, \bar{y}_0) \approx 0.10$. The trends are even more evident as the sample sizes increase, as shown in Figure 4(b).

These results are not necessarily surprising, since simple *z*-tests suggest that the two studies comprise different population means when $\bar{y} \geq 0.50$ (or ≥ 0.25 for Figure 3(b)). We would therefore expect some attenuation of D_0. That said, given that the current and historical samples will almost always differ in some respect, it seems inevitable that the NPP will discount D_0 to a large extent for reasonably sized studies, even if the underlying data-generating process is the same for the current and historical studies. In a sense, this is the reverse of the classic p-value conundrum, in which clinically inconsequential differences become statistically significant as *n* increases. Under the NPP, as the sample sizes increase, potentially minor differences between the sufficient statistics can lead to substantial mitigation of the historical data. We note, however, that our simulations did not consider the case where σ_0^2 and σ^2 diverged; future work might examine the behavior of the NPP as the current and historical variances diverge under a univariate normal model.

Because the NPP accentuates between-study differences as *n* and n_0 increase, clinicians must decide whether discrepancies between D and D_0 are clinically meaningful enough to warrant extensively downweighting D_0. In many cases, it may be reasonable to override empirical differences between studies to avoid overly discounting D_0. Therefore, as a practical alternative, we recommend conditioning on a_0 by assigning it a range of fixed values elicited from expert opinion about the commensurability of D and D_0. We also recommend conducting "reference" analyses in which a_0 is set to 0 and 1. This conditional approach should prove especially useful for moderate to large studies, since both the IC and normalized power priors tend to assign little weight to the historical data in such situations. The conditional approach also extends naturally to complex regression models which differ due to departures in ancillary covariates (the case study presented in Section 3 is one such example). And finally, the conditional approach has the practical advantage of being straightforward to implement in standard Bayesian software, since
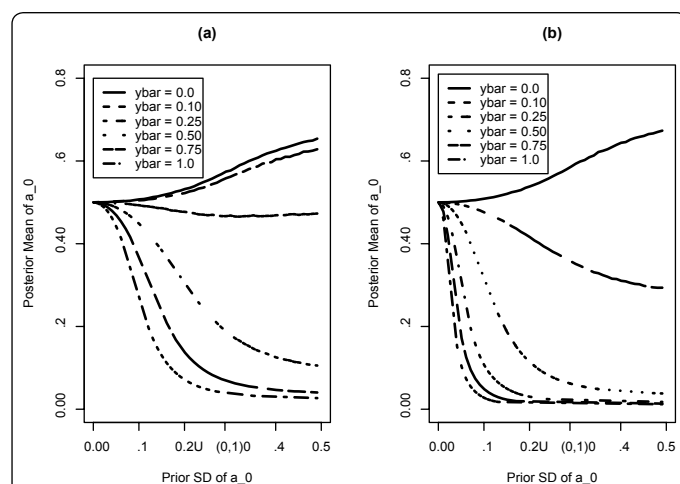
it avoids the burdensome integration inherent in the NPP. (To date, the NPP has been applied only to basic regression models in which the posteriors have closed forms.) Thus, the conditional approach is appealing from both an inferential and a computational standpoint. For these reasons, we focus on the conditional power prior in the application described in Section 3.

Power priors for exchangeable hierarchical models

In the regression setting, investigators may wish to relax the assumption of identical fixed effects between the current and historical studies, and instead link the two studies using a hierarchical modeling structure. Under this scenario, the studies are allowed to have distinct fixed-effect parameters, β and β_0, which are assumed in turn to share a common prior distribution according to an exchangeability condition. This hierarchical modeling approach is commonly used in Bayesian meta-analysis as a flexible alternative to assuming identical fixed effects across studies [16,7].
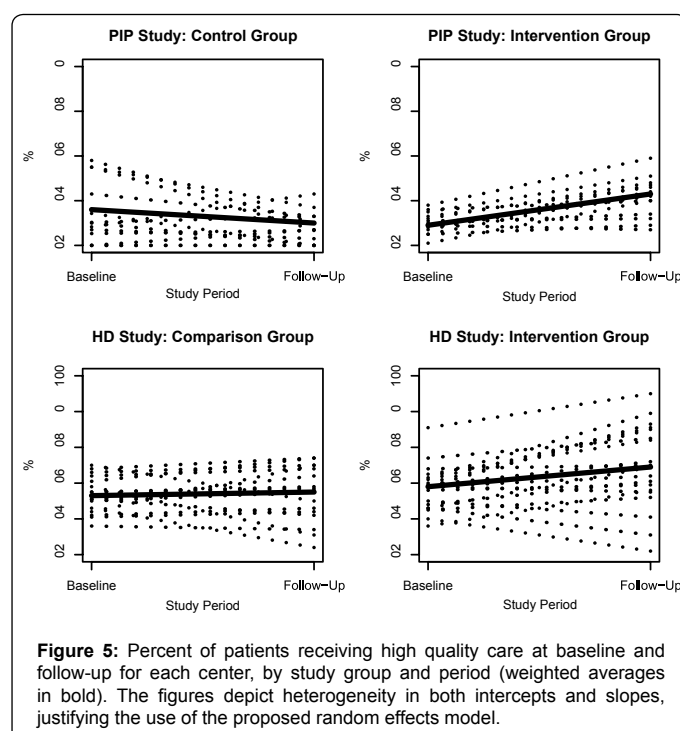
There are times, however, when the exchangeability assumption itself is questionable, and in such cases one can use the power prior to limit the impact of D_0. The conditional power prior can be adapted to accommodate the hierarchical nature of the model as follows

$$\pi(\beta, \beta_0, \gamma | D_0, a_0) \propto L(D_0 | \beta_0)^{a_0} \pi_0(\beta, \beta_0 | \gamma) \pi(\gamma), \qquad (2.11)$$

where $\pi_0(\beta, \beta_0 | \gamma)$ denotes the initial joint prior for β and β_0 and γ is a vector of hyperparameters. For regression models, a natural choice is to assign β and β_0 conditionally independent multivariate normal distributions; that is, $\pi_0(\beta, \beta_0 | \gamma) = N_p(\beta | \mu, S) \times N_p(\beta_0 | \mu, S)$, where $N_p(. | \mu, S)$ denotes a *p*-variate normal distribution with mean μ and covariance S. We refer to prior (2.11) as the *exchangeable power prior* (EPP) to distinguish it from the standard application of the power prior to models in which the treatment effect is fixed between studies. The conditional power prior given in expression (2.3) is a special case of the EPP in which (assuming a normal prior) $S = 0$ and $\beta = \beta_0 = \mu$. Strictly speaking, the EPP is a power prior applied to a hierarchical model with exchangeable cross-study effects; the power prior itself has no bearing on the validity of the exchangeability assumption. Typically, interest lies in estimating the current model parameters (β) while borrowing historical information in order to improve the precision of these inferences. However, depending on study aims, the focus could be on other parameters of interest, such as the grand mean μ.

Under the EPP, the models share only hyperparameters γ, and hence a_0 controls the impact of D_0 vis-a-vis these hyperparameters. When $a_0 = 1$, the D and D_0 contribute equal weight to the posterior for γ (and consequently β). When $a_0 < 1$, D_0 contributes $100 \times (1 - a_0)$ % less information to the log-posterior for γ, and when $a_0 = 0$, the historical data are excluded altogether. Because the historical data directly affect only γ, we can expect it to have less impact on the posterior of β than the power prior in a non-hierarchical model.

The connection between hierarchical models and the power prior has been explored by Chen and Ibrahim [11]. They develop expressions for a_0 to calibrate the power prior to a corresponding hierarchical model. Our focus here is slightly different: we use the power prior in conjunction with an exchangeable hierarchical model to guard against unreasonably optimistic assumptions about the relevance of D_0. We stress, however, that using the conditional power prior in this context is not tantamount to discounting the data twice, since the hierarchical structure and the power prior limit the influence of D_0 in different ways. The hierarchical structure evaluates

Figure 4: $E(a_0 | \bar{y}, \bar{y}_0)$ as a function of σ_{a_0} under the NPP as the current and historical data diverge. For figure (a), $n = n_0 = 100$; for figure (b), $n = 2,536$ and $n_0 = 895$, corresponding to the sample sizes for the case study presented later. For both figures, $\sigma^2 = 1$, $\bar{y}_0 = 0$, $\mu_{a0} = 0.50$, and y ranges from 0 to 1.0, reflecting increasing disparity between the current and historical data. As $\bar{y}$ and σ_{a_0} increase, the historical data are substantially downweighted.

Figure 5: Percent of patients receiving high quality care at baseline and follow-up for each center, by study group and period (weighted averages in bold). The figures depict heterogeneity in both intercepts and slopes, justifying the use of the proposed random effects model.

the similarities of the data in terms of the adjusted mean outcomes: similar population means lead to reduced between-study variability and more pooling. The conditional power prior, on the other hand, downweights the log-likelihood contributions of D_0 by a (fixed) factor of a_0 regardless of the congruence between the studies.

Application: The Healthy Development Study

Background

To illustrate the use of the conditional power prior, we analyzed a pair of studies designed to improve the delivery of preventive care in pediatric clinics. The "current" study, called the Healthy Development (HD) study, was an 18-month cohort study conducted from June 2004 to December 2005 [17]. Nineteen clinics from North Carolina and Vermont were given training sessions to improve their delivery of care in important clinical areas, such as child safety, maternal depression and substance-abuse. Seventeen non-randomized clinics served as a comparison group. Data were collected using the Promoting Healthy Development Survey (PHDS), a clinically validated measure of pediatric quality of care [18]. Parents were surveyed monthly and asked to rate their care in variety of areas. Due to low response during some months, the data were aggregated into a baseline period (June 2004-August 2004) and a follow-up period (June 2005-December 2005). To assess whether comprehensive, standardized care was provided across multiple dimensions of care, PHDS total scores were categorized into a binary outcome measure identifying patients who received "high quality" care. The primary aim was to determine whether intervention clinics showed greater improvement in delivering high quality care than comparison practices after adjusting for relevant patient and clinic covariates. A total of 2,536 surveys were used in the analysis.

As it happens, the investigators conducted a previous randomized quality-improvement trial, called the Partners in Practice (PIP) study [19]. The study consisted of 895 surveys in 26 centers (13 intervention and 13 control). As with the HD study, the primary outcome for the

Variable	PIP Study (n = 895) Posterior Mean (95% CrI)	HD Study (n = 2; 536) Posterior Mean (95% CrI)
Intercept	-2.27 (3.39, -1.28)	-0.54 (-0.90, -0.18)
Study Period	-0.47 (-1.69, 0.65)	0.02 (-0.41, 0.46)
Study Group	-0.85 (-1.63, -0.08)	0.20 (-0.22, 0.61)
Period × Group	**2.02 (0.77, 3.43)***	**0.57 (-0:03, 1.17)****
Education (0= < HS, 1= ≥ HS)	-0.03 (-0.77, 0.80)	-0.23 (0.44, -0.03)
Race (0=white, 1=other)	0.38 (-0.13, 0.87)	0.08 (-0.15, 0.31)
Random Intercept Variance	1.64 (0.50, 0.70)	0.25 (0.12, 0.47)
Random Slope Variance	2.20 (0.44, 6.25)	0.40 (0.14, 0.89)
Random Effect Covariance	-1.30 (-4.04, 0.09)	0.12 (-0.02, 0.31)

* $\Pr(\beta_4 > 0 | y_0) = 1.0$
** $\Pr(\beta_4 > 0 | y) = 0.97$

Table 1: Posterior means and 95% credible intervals (CrI) for separate analyses of PIP and HD studies.

PIP study was a binary measure of quality of care derived from parent survey items. However, the PIP study focused more on healthy child environments (e.g., exposure to second-hand smoke) and somewhat less on maternal psychosocial concerns as in the HD study. Thus, while the two study surveys measured overlapping areas of care, they were not identical measures.

To address our concerns about the compatibility of the two studies, we used a conditional power prior distribution to incorporate the PIP study data into the analysis of the HD study. We present results from this analysis in Section 3.3. As a comparison, we also applied the EPP approach described in Section 2.3; we present these results in Section 3.4.

Separate analysis of current and historical studies

We begin with separate analyses of the HD and PIP studies. Figure 5 displays the unadjusted percent of patients receiving high quality care in each center, by study group and period, for the PIP randomized trial and the HD cohort study. Highlighted in bold are the average percents across centers, weighted by center sample size. Since the plots indicated extensive between-clinic heterogeneity, both at baseline and over time, we proposed the following random effects model for the HD study:

$$\text{logit}\left\{\Pr\left(Y_{ij} = 1 | \beta, b_i\right)\right\} = x_{ij}'\beta + z_{ij}'b_i, \qquad j=1,\ldots,n_i, i=1,\ldots,n \quad (3.1)$$

where Y_{ij} takes the value 1 if the jth patient in the ith pediatric clinic reported that high-quality care was received, and 0 otherwise; x_{ij} is a 6×1 vector of fixed-effect covariates comprising an intercept term, an indicator for study group (1=intervention, 0=comparison), an indicator for study period (1 if follow-up, 0 if baseline), the period×group interaction, and two dichotomous patient-level covariates (mother's education, coded as 1=high school graduate, 0=non-graduate, and race, coded as 0=white, 1=other); β is a 6×1 vector of fixed-effect regression coefficients; $b_i \sim N_2(0,\Sigma)$ is a 2×1 vector comprising clinic-specific intercept and slope parameters; and $z_{ij}=(1, \text{study period}_{ij})'$. As a quick check of model fit, we ran a fixed-effects logistic regression and computed the Hosmer-Lemeshow goodness of fit statistic; the p-value was 0.79, suggesting no lack of fit.

We proposed a similar regression model for the PIP study:

$$\text{logit}\left\{\Pr\left(Y_{0ij} = 1 | \beta_0, b_{0i}\right)\right\} = x_{0ij}'\beta_0 + z_{0ij}'b_{0i}, \qquad j=1,\ldots,n_{0i}, i=1,\ldots,n_0, \quad (3.2)$$

where β_0 is a vector of fixed-effect regression coefficients, $b_{0i} \sim N(0, \Sigma_0)$ is a vector comprising random intercept and slope parameters specific to the ith clinic in the PIP study, and x_{0ij} and z_{0ij} are covariate vectors defined as in (3.1).

For both models, the intervention effect was measured by the period×group interaction term, denoted as β_4 for the HD model and β_{04} for the PIP model. For β and β_0, we assumed diffuse N_6 $(0,10^3 I_6)$ distributions, and for Σ and Σ_0 we assumed IW(I_2, 2) distributions. The analyses were run in WinBUGS [14]. We ran three, initially-dispersed MCMC chains for 110,000 iterations, discarding the first 10,000 as a burn-in. We retained every fiftieth draw to reduce autocorrelation. MCMC diagnostics, such as trace plots and Gelman-Rubin statistics [15], indicated rapid convergence and efficient mixing of the chains. A similar MCMC procedure was used for each of the analyses presented below.

While the directions of the estimates were similar across the studies, there were a few key differences (Table 1). In the PIP study, parents in the control group were somewhat less likely to receive high quality care at follow-up than at baseline (row 2). Meanwhile, parents in the control arm of the HD study remained relatively unchanged in their probability of receiving such care. Interestingly, parents in the PIP intervention arm began the study substantially less likely to receive high quality care (posterior mean=-0.85, 95% CrI=[-1.63,-0.08]), even though clinics in this study were randomized to study group. On the other hand, parents in the HD intervention group were somewhat more likely to report receiving high quality care at baseline (posterior mean=0.20, 95% CrI=[-0.22, 0.61]). The period×group interaction terms, highlighted in row 4, were used to assess the intervention effects. For both studies, the intervention groups demonstrated more improvement than the comparison groups, although this effect was most notable in the PIP study. The posterior probabilities of a positive intervention effect were close to 1 for both studies (see table footnotes). There was also more between- clinic variability in the PIP study than in the HD study, with a negative covariance between the random intercept and slope parameters (last row). Conversely, the HD study showed a positive random-effect covariance, although in both studies the 95% credible intervals overlapped 0.

Power prior analysis

Next, we used the conditional power prior approach described in Section 2.1 to borrow information from the PIP study while simultaneously expressing uncertainty about how much information should be shared. As above, we assumed random-effect logistic models for Y and Y_0, except here we set $\beta_0 = \beta$ and $\Sigma_0 = \Sigma$, thus allowing the models to share identical covariate effects and variance components. We then placed the conditional power prior distribution in (2.4) on the fixed effect coefficients, β. To complete the prior specification, we assumed a diffuse initial prior for β and an IW (I_2, 2) distribution for Σ.

Table 2 presents the results when $a_0 = 1$ which, together with the separate HD analysis presented in Table 1 (i.e., $a_0 = 1$), we considered to be our "reference" analysis. As expected, the estimates fall between the estimates for the two separate analyses (Section 3.2), reflecting the complete pooling of the two samples. There is substantial increase in the adjusted log-odds estimate for the interaction, from 0.57 in the separate HD analysis (Table 1, last column), to 0.89 for the pooled analysis. The random effect variance estimates also fall between those for the two separate analyses. In particular, the random effect covariance estimate is 0.06, indicating little association between the random intercept and slopes for the pooled model.

Next, we assumed a range of values for a_0 ($a_0 = 0.25, 0.50$ and 0.75). Table 3 presents the posterior summaries for the interaction term, β_4, under the different choices of a_0. The entries in the first row are identical to the results for the separate HD study presented in Table 1, while the last row repeats the result from Table 2. As more weight is given to D_0, there is a noticeable increase in the posterior means, reflecting the stronger intervention effect that was observed for the PIP study. When $a_0 \geq 0.25$, the credible intervals no longer overlap the null, strongly suggesting a benefit to the intervention.

EPP analysis

We next used the exchangeable power prior to jointly analyze the two studies. As in models (3.1) and (3.2), we allowed distinct fixed effect parameters, β and β_0, with joint prior distribution $\pi(\beta,\beta_0|\gamma)$. Specifically,

Variable	Parameter	Posterior Mean (95% CrI)
Intercept	β_1	-0.94 (-1.31, -0.55)
Study Period	β_2	-0.18 (-0.61, 0.24)
Study Group	β_3	-0.15 (-0.64, 0.28)
Period × Group	β_4	**0.89 (0.36, 1.47)***
Education (0 = < HS, 1= ≥ HS)	β_5	-0.26 (-0.44, -0.06)
Race (0=white, 1=other)	β_6	0.14 (-0.07, 0.34)
Random Intercept Variance	Σ_{11}	0.93 (0.52, 1.52)
Random Slope Variance	Σ_{22}	0.50 (0.19, 0.44)
Covariance	Σ_{12}	0.06 (-0.34, 0.41)

* Pr($\beta_4 > 0|$ y,y_0) = 1.0

Table 2: Posterior estimates for power prior model with $a_0 = 1$.

a_0	E($\beta_4|y$; y_0) (95% CrI)	Pr($\beta_4 > 0|$ y; y_0)
0	0.57 (-0.03, 1.17)	0.97
0.25	0.62 (0.11, 1.18)	0.98
0.50	0.69 (0.15, 1.22)	0.99
0.75	0.78 (0.25, 1.30)	1.0
1	0.89 (0.36, 1.47)	1.0

Table 3: Posterior results for the period×group interaction term (β_4) under various choices for a_0.

Variable	Parameter	Posterior Mean (95% CrI)
Intercept	β_1	-0.56 (-0.95, -0.15)
Study Period	β_2	-0.11 (-0.58, 0.34)
Study Group	β_3	0.16 (-0.33, 0.66)
Period × Group	β_4	**0.74 (0.11, 1.43)***
Education (0 = < HS, 1= ≥ HS)	β_5	-0.21 (-0.41, -0.02)
Race (0=white, 1=other)	β_6	0.11 (-0.12, 0.33)
Random Intercept Variance	Σ_{11}	0.35 (0.18, 0.64)
Random Slope Variance	Σ_{22}	0.51 (0.18, 1.10)
Covariance	Σ_{12}	0.06 (-0.20, 0.29)

* Pr($\beta_4 > 0|$y, y_0) = 0.99

Table 4: Posterior estimates for EPP model with $a_0 = 1$.

a_0	E($\beta_4|y$, y_0)(95% CrI)	Pr($\beta_4 > 0|$ y, y_0)
0	0.57 (-0.03, 1.17)	0.97
0.25	0.64 (0.07, 1.23)	0.98
0.50	0.68 (0.12, 1.30)	0.98
0.75	0.71 (0.11, 1.35)	0.99
1	0.74 (0.11, 1.43)	0.99

Table 5: Posterior results for the period×group interaction term (β_4) for EPP model under various choices for a_0.

$$\pi_0\left(\beta,\beta_0,\Sigma,\mu,S \mid D_0,a_0\right) \;\propto\; \prod_{i=1}^{n_0}\left[\int_{\mathfrak{R}^2}\prod_{j=1}^{n_{0i}}\left\{p\left(y_{0ij}\mid x_{0ij},z_{0ij},\beta_0,b_{0i}\right)\right\}^{a0}\pi\left(b_{0i}\mid\Sigma\right)db_{0i}\right]\times$$
$$\pi_0\left(\beta,\beta_0\mid\gamma\right)\pi(\gamma)\pi(\Sigma),$$

where β and β_0 are 6×1 vectors of fixed-effect coefficients with prior $\pi(\beta,\beta_0\mid\gamma)=N_6(\beta\mid\mu,S)\times N_6(\beta_0\mid\mu,S)$, and $S=\text{Diag}(\sigma_k^2)$ for $k=1,\ldots,6$. (Note that the EPP is expressed as a joint prior on all model parameters.) For the remaining parameters, we assumed the following prior distributions:

$$\pi(\mu)\;\propto\;1,$$
$$\pi(\sigma_k)\;=\;U(0,100)\ \text{or}\ N(0,1)1_{(\sigma_k>0)},$$
$$\pi(b_i\mid\sigma)\ \text{and}\ \pi(b_{0j}\mid\sigma)\;=\;N_2(0,\Sigma),\quad i=1,\ldots,n,\ j=1,\ldots,n_0,$$
$$\Sigma\;\sim\;IW(I_2,2).$$

Because our analysis consisted of only two trials, the posteriors of β and β_0 are likely to be affected by the choice of prior for σ_k. Therefore, as a sensitivity check, we considered two priors: a diffuse uniform prior, which assigns most mass to large between-trial variances; and a half-normal with standard deviation 1, which assigns about 5% prior probability to $\sigma_k>2$ (i.e., *a priori*, we assume similar effects for the two studies). Since the results were similar under both priors, we present only the results for the half-normal prior.

Table 4 presents the results for the EPP analysis when $a_0=1$, which is equivalent to a hierarchical model based on complete pooling of the two datasets. As the table indicates, the impact of D_0 is further attenuated relative to the non-exchangeable power prior. For example, the posterior mean for β_4 is 0.74 (95% CrI= [0.11, 1.43]), about halfway between the estimates for the separate HD analysis (Table 1, last column) and the non-exchangeable model (Table 3). Meanwhile, the posterior estimates for education and race closely resemble those for the separate HD analysis, suggesting that D_0 did not have much influence on them. Likewise, the random effect variance components are close to the values obtained in the HD analysis. This is not surprising given that only one historical study was used in this analysis.

Table 5 provides the posterior estimates for the interaction parameter, β_4, under various choices of a_0. The posterior estimates increase steadily as more weight is assigned to D_0: from 0.57 when $a_0=0$ to 0.74 when $a_0=1$. However, with the exception of $a_0=0.25$, the posterior estimates are smaller than under the standard power prior (Table 3), confirming that the EPP provides an additional attenuation of D_0 relative to the conventional approach. The credible intervals are also consistently wider under the EPP approach, reflecting the added heterogeneity imposed when the assumption of shared model parameters is relaxed. Despite this increased variability, when $a_0\geq0.25$, the credible intervals no longer overlap 0.

Discussion

Our aim has been to examine approaches to incorporating data from prior studies into a current analysis when there is uncertainty as to the similarities of the studies. We have emphasized the use of power prior distributions for this purpose. The power prior has gained increasing attention in recent years as a technique for synthesizing results across disparate studies. It avoids having to arbitrarily select observations in order to limit the impact of D_0 instead, all historical observations are included in the analysis, but each is downweighted equally. For normal models, this is equivalent to inflating the posterior variance of the model parameters by $1/a_0$ for $0<a_0\leq1$. However, the power prior will have a different impact on other densities, and therefore it cannot be viewed as a simple generalization of the variance-inflation approach to downweighting prior information.

In previous studies, investigators have recommended placing a prior distribution on the power parameter a_0. This approach is appealing from a Bayesian perspective, since it allows one to express prior uncertainty about a_0 and use the data to estimate its posterior. We considered two specifications of the power prior in which a_0 was treated as a stochastic parameter: the Ibrahim-Chen (IC) power prior, which normalizes jointly over θ and a_0 and the normalized power prior (NPP), which first normalizes with respect to θ (conditional on a_0) and then multiplies this normalized density by $\pi(a_0)$. Under both approaches, for any sizeable n, the posterior of a_0 concentrates near zero as the prior variance for a_0 increases. Our investigation revealed that the IC prior penalizes the historical data even when the current and historical data are identical, unless a very informative prior is used to bound a_0 away from zero. In contrast, the NPP provides a measure of commensurability between the studies, so that D_0 is substantially downweighted only as the studies diverge. As such, the NPP is a sensible methodological extension to the original power prior. Nevertheless, even under the NPP, D_0 will be heavily discounted when there are discrepancies between the sufficient statistics of D and D_0. Moreover, because the NPP has been formally developed only for basic normal and binomial models, it is not clear how it would perform in more complex settings. For example, in a multiple regression model, would the NPP downweight D_0 when auxiliary variables differed but the primary treatment effects were identical across studies? This is an open area of research. In addition, on a final practical note, the NPP is computationally challenging to implement due the integration inherent in its construction, requiring application of specific numerical integration methods or accurate approximations (e.g., the Laplace approximation) for evaluation. See Gajewski [20] and Neuenschwander et al. [21] for related comments on the computational challenges associated with the NPP.

An alternative strategy-the one we have adopted in this paper-is to assign a_0 a range of fixed values from 0 to 1 as part of a sensitivity analysis. This approach allows users more control over the impact of D_0 and is straightforward to implement in standard software packages such as WinBUGS. The trade-off is that the fixed approach requires external information, such as expert opinion, to determine a suitable range for a_0. If, for example, investigators question the commensurability of the studies, restricting a_0 to a "skeptical" range (e.g., $a_0\leq0.50$) may be reasonable. On the other hand, if there is strong belief in the compatibility of the studies, a more "optimistic" range, such as 0.50 to1, can be selected. In this case, $a_0=0.50$ could serve as the lower reference value. For our study, the credible intervals for β_4 did not overlap 0 when $a_0\geq0.25$. Thus, as long as investigators expressed some optimism about the studies' compatibility, we would infer a positive intervention effect. Conversely, if investigators expressed skepticism about the relevance of D_0, so that the value $a_0=0$ could not be ruled out, then we might exercise more caution when evaluating the efficacy of the intervention.

More generally, one could take a model averaging approach and assign a distribution, $\pi(a_0)$, to a_0, fit the model conditional on specific values for a_0 and then average inferences across $\pi(a_0)$ to obtain summary values. In the HD application, we effectively took $\pi(a_0)$ to be a discrete uniform distribution on {0; 0.25; 0.50; 0.75; 1}, although we did not perform model averaging, opting instead to tabulate the results separately for each value of a_0. In a sense, the IC and NPP priors are attempting to achieve this model averaging in one step by placing a joint prior on (a_0,θ) and jointly updating the

parameters. However, as we have seen, these priors tend to lead to rapid attenuation of D_0, which in many practical situations might be considered excessive.

Both the NPP and the conditional power prior are valid approaches depending on study aims. If the aim is to incorporate historical information only when $D \approx D_0$, then the NPP may be a reasonable choice. The fixed approach, on the other hand, is appealing if one wishes to incorporate historical information to aid inferences about the current model, or to provide a measure of the comprehensive treatment effect across studies, even when $D \not\approx D_0$, as long as external information suggests that the historical data are relevant the current study. The important point, from our perspective, is to understand the advantages and limitations to each approach, and to make an informed decision according to study aims.

We have also applied the power prior in the context of Bayesian hierarchical models that link the current and historical studies through an exchangeability condition. Here, we allocated unique parameters to the current and historical likelihoods and assumed a common prior distribution. We then used the conditional power prior to limit the contribution of D_0 to the posterior updates of the hyperparameters that link the two models. This exchangeable power prior, as we have called it, can prove useful if one has reservations about the underlying assumption of exchangeability between studies, or if there is some additional skepticism about the compatibility of the studies. The results from our case study suggest that this approach provides additional attenuation of D_0 over and above that obtained when the (conditional) power prior is applied to a non-hierarchical model. Since our example only involved two studies, inferences about between-trial parameters such as Σ are only weakly informed. However, because our focus was on β, the small number of historical studies was not a major concern.

In general, the methods described here can be extended to multiple historical studies, and may provide guidance for cross-design evidence synthesis using hierarchical models. They may also prove useful to regulatory agencies, such as the FDA, who wish to incorporate historical data when evaluating the efficacy of new medical devices and therapies. Meanwhile, future research might focus on applying the NPP to hierarchical regression models, such as the one described in our case study, to examine its behavior in complex settings.

Acknowledgements

References

1. Dunson DB (2001) Practical advantages of Bayesian analysis of epidemiologic data. Am J Epidemiol 153: 1222-1226.

2. Ibrahim JG, Chen MH (2000) Power prior distributions for regression models. Stat Sci 15: 46-60.

3. Chen M-H, Harrington D, Ibrahim J (2002) Bayesian cure rate models for malignant melanoma: a case study of Eastern Oncology Group trial E1690. Appl Stat 51: 135-150.

4. Duan Y, Ye K, Smith EP (2006) Evaluating water quality using power priors to incorporate historical information. Environmetrics 17: 95-106.

5. De Santis F (2007) Using historical data for Bayesian sample size determination. J R Stat Soc Ser A 170: 95-113.

6. Tan SB, Machin D, Tai BC, Foo KF, Tan EH (2002) Bayesian re-assessment of two phase II trials of Gemcitabine in metastatic nasopharyngeal cancer. Br J Cancer 86: 843-850.

7. Spiegelhalter DJ (2004) Incorporating Bayesian ideas into health-care evaluation. Stat Sci 19: 156-174.

8. Peters JL, Rushton L, Sutton AJ, Jones DR, Abrams KR, et al. (2005) Bayesian methods for the cross-design synthesis of epidemiological and toxicological evidence. J R Stat Soc Ser C Appl Stat 54: 159-172.

9. Congdon P (2008) Indirect area estimates of disease prevalence: Bayesian evidence synthesis with an application to coronary heart disease. J Data Sci 6: 15-32.

10. Chen MH, Ibrahim JG, Shao Qi-M (1999) Power prior distributions for generalized linear models. J Stat Plan Inference 84: 121-137.

11. Chen MH, Ibrahim JG (2006) The relationship between the power prior and hierarchical models. Bayesian Anal 1: 551-574.

12. Duan Y, Ye K (2008) Normalized power prior Bayesian analysis. The University of Texas at San Antonio College of Business Working Paper Series, WP# 0058MSS-432-2008.

13. Neuenschwander B, Branson M, Spiegelhalter DJ (2009) A note on the power prior. Stat Med 28: 3562-3566.

14. Spiegelhalter D, Thomas A, Best N, Lunn D (2003) WinBUGS Version 1.4: User Manual. Medical Research Council Biostatistics Unit, Cambridge.

15. Gelman A, Carlin J, Stern H, Rubin D (2004) Bayesian data analysis. (2nd edn), Chapman & Hall/CRC, Boca Raton, FL, USA.

16. Sutton AJ, Abrams KR (2001) Bayesian methods in meta-analysis and evidence synthesis. Stat Methods Med Res 10: 277-303.

17. Margolis PA, McLearn KT, Earls MF, Duncan P, Rexroad A, et al. (2008) Assisting primary care practices in using office systems to promote early childhood development. Ambul Pediatr 8: 383-387.

18. Bethell C, Peck C, Schor E (2001) Assessing health system provision of well-child care: The Promoting Healthy Development Survey. Pediatrics 107: 1084-1094.

19. Rosenthal MS, Lannon CM, Stuart JM, Brown L, Miller WC, et al. (2005) A randomized trial of practice-based education to improve delivery systems for anticipatory guidance. Arch Pediatr Adolesc Med 159: 456-463.

20. Gajewski BJ (2010) Power of BUGS. Letter to the Editor, Stat Med 29: 708-710.

21. Neuenschwander B, Branson M, Spiegelhalter DJ (2010) Reply to "Power of BUGS". Letter to the Editor, Stat Med 29: 708-710.

22. Hosmer DW, Lemeshow S (2000) Applied Logisitc Regression. (2nd edn), John Wiley & Sons, New York.