Received: 31 March 2018 | Revised: 8 September 2018 | Accepted: 4 October 2018

DOI: 10.1002/sim.8019

RESEARCH ARTICLE

WILEY Statistics in Medicine

Modified power prior with multiple historical trials for binary endpoints

Akalu Banbeta[1,2] | Joost van Rosmalen[3] | David Dejardin[4] | Emmanuel Lesaffre[5]

[1]I-Biostat, UHasselt, Hasselt, Belgium

[2]Department of Statistics, Jimma University, Jimma, Ethiopia

[3]Department of Biostatistics, Erasmus University Medical Center, Rotterdam, The Netherlands

[4]F. Hoffmann-La Roche AG, Basel, Switzerland

[5]I-Biostat, KU Leuven, Leuven, Belgium

Correspondence
Akalu Banbeta, I-Biostat, UHasselt, Hasselt, Belgium.
Email: akalubanbetastat@gmail.com

Funding information
Hercules Foundation; EWI Department of the Flemish Government; Jimma University Inter-University Cooperation (IUC-JU); BOF-BILA of UHasselt

Including historical data may increase the power of the analysis of a current clinical trial and reduce the sample size of the study. Recently, several Bayesian methods for incorporating historical data have been proposed. One of the methods consists of specifying a so-called power prior whereby the historical likelihood is downweighted with a weight parameter. When the weight parameter is also estimated from the data, the modified power prior (MPP) is needed. This method has been used primarily when a single historical trial is available. We have adapted the MPP for incorporating multiple historical control arms into a current clinical trial, each with a separate weight parameter. Three priors for the weights are considered: (1) independent, (2) dependent, and (3) robustified dependent. The latter is developed to account for the possibility of a conflict between the historical data and the current data. We analyze two real-life data sets and perform simulation studies to compare the performance of competing Bayesian methods that allow to incorporate historical control patients in the analysis of a current trial. The dependent power prior borrows more information from comparable historical studies and thereby can improve the statistical power. Robustifying the dependent power prior seems to protect against prior-data conflict.

KEYWORDS

Bayesian inference, dependent weights, modified power prior, multiple historical trials

1 | INTRODUCTION

The randomized controlled trial (RCT) is considered the most appropriate way to establish a cause-effect relationship between a treatment and an outcome.[1] In the majority of RCTs, historical data are only used as guidance to set up a new study.[2] Explicitly including historical data into the analysis of the current trial data may have ethical and economic advantages. This is true when the characteristics of the control arm of subsequent studies remain basically the same. In that case, including historical controls into the current study allows to reduce the number of control patients, and conclusions may be reached earlier.[3-5] However, gains will only be obtained when the historical controls are comparable with the current control treatment.[6]

Due to possible differences in patient populations across different trials, it is inappropriate to simply lump all historical controls into the current trial data. Pocock[6] proposed formal methods for incorporating historical controls in both

the design and the analysis of RCTs. Since then, several Bayesian methods have been proposed for the inclusion of historical data into the analysis of current data, especially in clinical trials.[7-12] The main approaches are based on the meta-analytic-predictive (MAP) prior,[7] the power prior,[10] and the commensurate prior.[11] These methods share the same feature that they discount historical data to account for between-trial heterogeneity in the context of a single historical study or multiple historical studies.

The MAP prior proposed by Neuenschwander et al[7] is a popular meta-analytic approach when several historical trials are available. To make use of information contained in historical controls, the MAP prior assumes that the control parameters of all trials are exchangeable and are drawn from the same distribution. If the meta-analysis is performed at the design stage of a new trial, the predictive distribution for the parameter(s) of interest of the new control can be derived from the historical controls. This distribution summarizes the available knowledge about the control arm in the new trial and provides an informative prior (MAP prior) to be used in the analysis of that trial. The meta-analytic approach can also be used at the analysis stage of the new trial and is then referred to as the meta-analytic combined (MAC) approach.

Including historical information into the analysis of a current trial needs to be done with care, especially when there is the risk of a prior-data conflict. That is, when the historical data support vastly different parameter values than the current data. Incorporating the historical data could then mislead inference,[13] ie, the inference may not be robust. Prior-data conflicts may be due to (unanticipated) differences of the historical and current trials in study design, conduct, or patient population. In that case, it is probably best to drastically discount or even discard the prior information. To acknowledge the possibility of prior-data conflict, Schmidli et al[3] proposed a robust version of the MAP prior by adding a weakly informative component. This robust prior is a mixture of the original MAP prior and a vague prior with weights fixed in advance. When the historical and current control data are in clear conflict, the robust MAP prior will essentially discard the historical information.

The power prior, introduced by Ibrahim and Chen,[10] provides another way to incorporate and downweight historical data by raising the historical likelihood to a power smaller than 1. The power parameter may be fixed in advance or estimated from the data. In the latter case, the power prior had to be modified to satisfy the likelihood principle leading to the modified power prior (MPP); see Duan et al.[14] The (modified) power prior has been suggested for a single historical study, but Duan et al[14] also formulated some initial ideas when there are multiple historical studies. A straightforward generalization of the MPP to multiple historical studies is to assume different and independent weight (power) parameters. However, as with the MAP approach, it is reasonable to assume in first instance that the historical studies are not too different from each other. This leads to the dependent prior of the weights referred to here as the dependent MPP (DMPP). The performance of the DMPP with a binary outcome is evaluated in this paper, both analytically as well as via real data sets and simulation studies. We also suggest a robustified version of the DMPP to be used when there is possible conflict between the historical and current data. Again, the robustified version consists of adding a component that aims to ignore the historical information.

In this paper, we consider the MAP methods as competitor to the different versions of the power prior approach. We did not consider the commensurate prior, because this method was primarily designed for the context of a single historical trial and a single current trial, which is not in line with the purpose of our manuscript.

In Section 2, we describe two real-life data sets to be analyzed in the study: the HOVON data set and a data set from a proof-of-concept study. The Dutch-Belgian Hemato-Oncology Cooperative Group (HOVON) has organized over the last two decades a series of RCTs to evaluate investigational treatments for acute myeloid leukemia (AML) in comparison to a control arm. The second application is a phase II proof-of-concept trial for the treatment of ulcerative colitis (UC). Section 3 introduces the MPP for a single historical trial. Section 4 reviews methods that incorporate multiple historical trials into the design and analysis of a current trial. Section 5 focuses on the DMPP. In Section 6, several methods for the inclusion of multiple historical data are applied to the real-life data sets. In Section 7, simulation studies based on binary outcomes evaluate the performance of these methods across various scenarios and settings. This paper concludes with a general discussion in Section 8. Derivations and additional results are given in the Appendix.

2 | CLINICAL TRIAL EXAMPLES

This section introduces two clinical trial data sets to be used in the study. In Section 2.1, we describe the HOVON data set. Since the control treatment remained essentially the same (apart from the standard treatment of care), the question arises whether the control data can be used for future trials in this context. In this section, we also check which of the

TABLE 1 HOVON acute myeloid leukemia trial: descriptive statistics of acute myeloid leukemia patients with respect to the trials. N = number of patients, CR = complete response

Trial	Group	Year	N	CR (%)
HOVON 4	Control	1988-1992	359	279 (77.7)
HOVON 4A	Control	1992-1993	252	208 (82.5)
HOVON 29	Control	1997-2000	693	598 (86.3)
HOVON 42	Control	2002-2004	437	358 (81.9)
HOVON 42A	Control	2004-2006	259	214 (82.6)
HOVON 42A	Treatment	2004-2006	252	211 (83.7)

control data can be used thereby making use of Pocock's criteria.[6] Section 2.2 introduces the proof-of-concept trials that were used by Neuenschwander et al[7] to illustrate the MAP prior.

2.1 | The HOVON AML trial

Patients suffer from AML when their bone marrow produces immature white blood cells (blasts). For the evaluation of induction treatment and treatment strategies, complete remission or complete response (CR) is an important dichotomous outcome. Complete AML remission implies no evidence of leukemia after 4 weeks defined using the following criteria[15,16]:

- Normal values for absolute neutrophil count ($> 1000/\mu L$) and platelet count ($> 100,000/\mu L$), and independence from red cell transfusion.
- A bone marrow biopsy that reveals no clusters or collections of blast cells. Extramedullary leukemia (eg, central nervous system or soft tissue involvement) must be absent.

The HOVON data set is obtained from a number of RCTs conducted by HOVON since 1988, ie, the trials called HOVON 4, HOVON 4A, HOVON 29, HOVON 42, and HOVON 42A. All of these trials had essentially the same control treatment consisting of one cycle of induction with an anthracycline (daunorubicin or idarubicin) in combination with cytarabine (200 mg/m^2 for 7 days) and a second cycle of amsacrine with intermediate-dose cytarabine (1000 mg/m^2 every 12h for 6 days). In Table 1, some basic information of the HOVON control data is given. The question is whether and how these historical control data can be used for the evaluation of the investigational treatment in the HOVON 42A trial. This is the most recently conducted trial, where the effect of priming has been investigated using a granulocyte colony-stimulating factor (G-CSF) in the remission induction chemotherapy course for treatment of AML. Whether the historical control data from the HOVON 4, HOVON 4A, HOVON 29, and HOVON 42 trials can be used for the analysis of the HOVON 42A trial was evaluated in the work of van Rosmalen et al[17] using Pocock's criteria.[6] These are six criteria to evaluate whether the circumstances in which the historical studies have been performed are similar to those of the current study. The HOVON 29 and HOVON 42 trials have been selected using these criteria. Here, we consider these historical trials for the analysis of the data of the most recent study, HOVON 42A, which was set up to test whether G-CSF priming (investigational treatment) improves the CR rate of AML patients. The CR rate of the patients in the HOVON trials ranges between 77.7% and 86.3%, see Table 1. The response rate of patients with G-CSF priming of HOVON 42A is 83.7% whereas the rate is 82.6% for the control patients.

2.2 | The proof-of-concept trial

This application is a phase II proof-of-concept trial for the treatment of UC. The data are from four external RCTs with similar placebo arms (obtained from the works of Van Assche et al,[18] Feagan et al,[19] and two trials from Rutgeerts et al[20]). We refer to the four trials as Van Assche, Feagan, Rutgeerts-1, and Rutgeerts-2, respectively. Neuenschwander et al[7] made use of the placebo data of the four trials to illustrate the performance of the MAP prior. In this study, we consider Rutgeerts-2 as a current trial to compare the effect of 5 mg of Infliximab with placebo on the remission rate at week 8 after randomization. The placebo data from the other three trials are used as historical control data in the study (see Table 2). The remission rate at week 8 of the placebo patients ranges between 5.7% and 14.9%. In the current trial (Rutgeerts-2), the remission rate is 5.7% for the control group and 33.9% for the treatment group.

TABLE 2 Proof-of-concept trial: descriptive statistics of ulcerative colitis patients for each trial arm

Trial	Group	Year	N	Remission at Week 8 (%)
Van Assche	Control	2003-2004	56	6 (10.7)
Feagan	Control	2000-2003	63	9 (14.3)
Rutgeerts-1	Control	2002-2005	121	18 (14.9)
Rutgeerts-2	Control	2002-2005	123	7 (5.7)
Rutgeerts-2	Treatment	2002-2005	121	41 (33.9)

3 | THE POWER PRIOR

In this section, we review the power prior and its modified version for the inclusion of single historical control data into the analysis of a current trial. Let D_0 denote the historical data, D the current data, and $L(.)$ the likelihood function. In the power prior approach, we assume the same model parameter θ_C in both historical and current control data, but the historical data are downweighted with a power parameter δ.

3.1 | The power prior

Ibrahim and Chen[10] defined the power prior for θ_C of the current study as

$$\pi(\theta_C \mid D_0, \delta) \propto L(\theta_C \mid D_0)^{\delta} p(\theta_C). \tag{1}$$

The power parameter δ in (1) controls the degree of borrowing from historical data. Initially, δ was a fixed value between 0 and 1, with $\delta = 0$ meaning that historical data should be neglected, whereas with $\delta = 1$, the historical data are fully incorporated into the analysis. Since it is difficult to choose a particular value for δ, it was suggested to also give δ a prior yielding then the joint prior for (θ_C, δ) defined as[10,21]

$$\pi(\theta_C, \delta \mid D_0) \propto L(\theta_C \mid D_0)^{\delta} p(\theta_C) p(\delta). \tag{2}$$

For a dichotomous outcome, the historical data D_0 consist of y_0 "successes" out of n_0 subjects. Assuming a binomial distribution for y_0 with a Beta$(\alpha_\theta, \beta_\theta)$ prior for the success rate θ_C, the joint power prior becomes

$$\begin{aligned}
\pi(\theta_C, \delta \mid D_0) &\propto L(\theta_C \mid y_0, n_0)^{\delta} p(\theta_C) p(\delta) \\
&\propto \binom{n_0}{y_0}^{\delta} \theta_C^{\delta y_0} (1-\theta_C)^{\delta(n_0-y_0)} \frac{\theta_C^{\alpha_\theta - 1}(1-\theta_C)^{\beta_\theta - 1}}{B(\alpha_\theta, \beta_\theta)} p(\delta) \\
&\propto \frac{\binom{n_0}{y_0}^{\delta}}{B(\alpha_\theta, \beta_\theta)} \theta_C^{\delta y_0 + \alpha_\theta - 1} (1-\theta_C)^{\delta(n_0-y_0)+\beta_\theta - 1} p(\delta),
\end{aligned} \tag{3}$$

where $B(\alpha_\theta, \beta_\theta)$ is the beta function evaluated in α_θ and β_θ.

3.2 | The MPP

The power prior in (2), however, violates the likelihood principle.[22] In addition, the posterior distribution of δ tends to zero regardless of the compatibility between the historical data and the current data.[14,23] To fix this problem, the modified or normalized power prior has been proposed given by

$$\pi(\theta_C, \delta \mid D_0) \propto \frac{L(\theta_C \mid D_0)^{\delta} p(\delta) p(\theta_C)}{C(\delta)}. \tag{4}$$

The left-hand side of (4) can be written as $\pi(\theta_C, \delta \mid D_0) = \pi(\theta_C \mid D_0, \delta) p(\delta)$, where the conditional prior $\pi(\theta_C \mid D_0, \delta)$ is equivalent to the power prior in (1). However, a normalizing constant should be introduced that depends on δ to satisfy the likelihood principle. To this end, the scaling constant $C(\delta) = \int_{\theta_C} L(\theta_C \mid D_0)^{\delta} p(\theta_C) d\theta_C$ is taken as denominator.[14] The computation of $C(\delta)$ with many model parameters can be challenging, but it can be implemented using an algorithm based on the principle of path sampling[17,24,25] or by making use of a Laplace approximation.[26]

For the above binomial example, $C(\delta)$ can be computed analytically as

$$\int_{\theta_C} L(\theta_C \,|\, y_0, n_0)^{\delta} p(\theta_C) d\theta_C = \frac{\binom{n_0}{y_0}^{\delta}}{B(\alpha_\theta, \beta_\theta)} B(\delta y_0 + \alpha_\theta, \delta(n_0 - y_0) + \beta_\theta). \tag{5}$$

The numerator in formula (4) can be given as in (3) and substituting (5) for the denominator in (4), the MPP for binary data, $\pi(\theta_C, \delta \,|\, y_0, n_0)$, can be computed as

$$\frac{\theta_C^{\delta y_0 + \alpha_\theta - 1} (1 - \theta_C)^{\delta(n_0 - y_0) + \beta_\theta - 1} p(\delta)}{B(\delta y_0 + \alpha_\theta, \delta(n_0 - y_0) + \beta_\theta)}. \tag{6}$$

4 | THE META-ANALYTIC APPROACH

The meta-analytic prior is the most commonly used approach to incorporate multiple historical controls into the analysis of a current trial. In this section, we review the MAP prior and its robustified version.

Let $\underline{\boldsymbol{D_0}} = \{D_{01}, \ldots, D_{0K}\}$ represent the control data from K historical studies. Further, the parameters $\theta_{C_1}, \ldots, \theta_{C_K}$ express the "success" probabilities in each of the K historical control arms. Let D denote the current data (investigational + control) with parameter θ_T for the investigational arm and θ_{CC} for the control arm. The ultimate aim is to compare the efficacy of the investigational arm with that of the control arm in the current trial expressed by Δ, which is the difference of θ_T and θ_{CC} ($\theta_T - \theta_{CC}$) on the original probability scale or a transformed scale thereof.

4.1 | The meta-analytic prior

The meta-analytic approach can be applied at the design and the analysis stage of a new trial. All trials, here all control arms, are assumed to be exchangeable and to have been drawn from the same population.[7] In the first case, a Bayesian meta-analysis produces the predictive distribution for θ_{CC}, which is the MAP prior. In case the current study has already been completed, the meta-analytic approach consists in incorporating the current data (investigational + control) into a Bayesian meta-analysis of all control data. In that case, we speak of the MAC approach.

The original MAP approach assumes a Gaussian distribution of the control parameters. With a dichotomous outcome, it is advised to transform the parameters θ_{C_j}, ($j = 1, \ldots, K$) to, say, $\psi_{C_j} = \text{logit}(\theta_{C_j})$ and $\psi_{CC} = \text{logit}(\theta_{CC})$. Then, assume that the control data have a sampling distribution F, and the transformed control parameters have a Gaussian distribution. That is,

$$D_{0j} \sim F(\psi_{C_j}), (j = 1, \ldots, K), \tag{7}$$

$$\psi_{C_1}, \ldots, \psi_{C_K}, \psi_{CC} \,|\, \mu, \tau^2 \sim N(\mu, \tau^2), \tag{8}$$

where μ is the population mean and τ the standard deviation (SD) of the control parameters. The posterior distribution with the MAP prior can be formulated as

$$p_{\text{MAP}}(\psi_{C_1}, \ldots, \psi_{C_K}, \psi_{CC}, \Delta, \mu, \tau^2 \,|\, \underline{\boldsymbol{D_0}}, D) \propto L(\psi_{CC}, \Delta \,|\, D) \times$$
$$p(\psi_{CC} \,|\, \mu, \tau^2) \left(\prod_{j=1}^{K} L(\psi_{C_j} \,|\, \underline{\boldsymbol{D_0}}) p\left(\psi_{C_j} \,|\, \mu, \tau^2\right) \right) p(\Delta) p(\mu) p(\tau^2). \tag{9}$$

We note that, if the control arms are heterogeneous, the variance τ^2 will be increased, which reduces the role of the historical data for the analysis of the current data.

4.2 | The robust meta-analytic prior

The MAP is relatively robust to discrepant controls, as indicated in the previous section. However, the MAP for ψ_{CC} can be inappropriate when the historical data tell a different story than the current data, ie, when there is a prior-data conflict. At the design stage of a new trial, one is never sure of a possible prior-data conflict. So to be on the safe side,

Schmidli et al[3] suggested a robustified version of the MAP. They suggested to use a mixture prior with one component the MAP and the other component a weakly informative component. More specifically, it is assumed that

$$\psi_{CC} \mid \mu, \tau^2 \sim (1 - w_R) \times N(\mu, \tau^2) + w_R \times \pi_R, \tag{10}$$

where π_R is the robust (actually vague prior) component and w_R is the proportion of this component. Prior (10) is called the robustified MAP prior, because in case of prior-data conflict, the weakly informative component will dominate the posterior. Hence, the robust version of the MAP prior largely ignores the historical information when there is a prior-data conflict. To realize a large variance for π_R, a $N(\mu, \kappa_M \tau^2)$ distribution is chosen, with κ_M large. Classically, a fixed w_R but relatively small value (≈ 0.1) is taken depending on the relevance of the historical data.

In conclusion, the robustified prior will ignore all historical controls if they are in conflict with the current control.

5 | THE POWER PRIOR FOR MULTIPLE HISTORICAL CONTROLS

We now discuss the MPP introduced in Section 3 when multiple historical studies are available. The original MPP can be generalized by assuming different weight parameters δ_j for each D_{0j}. We use the term "weight parameter" throughout this paper to refer to the power parameter of the power prior. We consider three versions of MPP adapted to multiple historical controls. In Section 5.1, we assume independently distributed δ_j. In Section 5.2, we assume dependently distributed weights leading to the DMPP. Finally, in Section 5.3, the robustified version of the DMPP is investigated. Important to note is that, in the context of the power prior, the model parameter is assumed to be the same for all historical controls and the current control, namely, $\theta_{C_1} = \cdots = \theta_{C_K} = \theta_{CC} = \theta_C$. This is a generalization of the original assumption made by Chen et al[10,21] and was also assumed by Duan et al[14] in their suggestions of the MPP for multiple controls. Differential weighting of the different historical controls is achieved by having a different weighting parameter for each of the historical controls, $\delta_j (j = 1, \ldots, K)$. We denote the total vector of weights by $\underline{\boldsymbol{\delta}} = \{\delta_1, \ldots, \delta_K\}$.

Note that the assumption of equal θ's in the MPP may be considered as not very realistic. However, the assumption could be considered as a consequence of Pocock's conditions. The MAP prior uses a modeling approach where the θ's are allowed to differ according to a hierarchical prior. In the power prior approach, rather than a modeling approach, an algorithmic approach is adopted regulated via the prior variance, which essentially comes down to varying the value of the power.

5.1 | Independent power parameters

Chen et al[21] suggested the joint power prior for multiple historical data, and later, Duan et al[14] extended the MPP for i.i.d. distributed δ_j. A natural prior for δ_j is a beta prior with hyperparameters α and β set to fixed positive values. The MPP for multiple historical data with i.i.d. distributed δ_j is given by

$$\pi_{\text{MPP}}(\theta_C, \underline{\boldsymbol{\delta}} \mid \underline{\boldsymbol{D}}_{\mathbf{0}}) \propto \frac{\left[\prod_{j=1}^{K} L(\theta_C \mid D_{0j})^{\delta_j} p(\delta_j)\right] p(\theta_C)}{C(\underline{\boldsymbol{\delta}})}, \tag{11}$$

with the scaling constant $C(\underline{\boldsymbol{\delta}})$ defined as

$$C(\underline{\boldsymbol{\delta}}) = \int_{\theta_C} \left[\prod_{j=1}^{K} L(\theta_C \mid D_{0j})^{\delta_j)}\right] p(\theta_C) d\theta_C \tag{12}$$

and introduced such that the likelihood principle is satisfied; see Appendix A.1.

The posterior of θ_C and Δ after having collected the current data D, which is the MPP equivalent to the MAC approach, is given by

$$p_{\text{MPP}}(\theta_C, \Delta, \underline{\boldsymbol{\delta}} \mid \underline{\boldsymbol{D}}_{\mathbf{0}}, D) \propto L(\theta_C, \Delta \mid D) \pi_{\text{MPP}}(\theta_C, \underline{\boldsymbol{\delta}} \mid \underline{\boldsymbol{D}}_{\mathbf{0}}). \tag{13}$$

For a dichotomous response with $\underline{\boldsymbol{y}}_{\mathbf{0}} = \{y_{01}, \ldots, y_{0K}\}$ the number of events out of $\underline{\boldsymbol{n}}_{\mathbf{0}} = \{n_{01}, \ldots, n_{0K}\}$ subjects, the numerator and denominator in (11) can be analytically derived (see Appendix A.2.1). The MPP for multiple historical

data sets with a dichotomous outcome and with independent weight parameters then becomes

$$\pi_{\text{MPP}}(\theta_C, \underline{\boldsymbol{\delta}} \,|\, \underline{\boldsymbol{y}}_{\boldsymbol{0}}, \underline{\boldsymbol{n}}_{\boldsymbol{0}}) \propto \frac{\theta_C^{\sum \delta_j y_{0j} + \alpha_\theta - 1} (1-\theta_C)^{\sum \delta_j (n_{0j} - y_{0j}) + \beta_\theta - 1} \prod_{j=1}^{K} p(\delta_j)}{B\left(\sum \delta_j y_{0j} + \alpha_\theta, \sum \delta_j (n_{0j} - y_{0j}) + \beta_\theta\right)}. \tag{14}$$

The independence power prior also enjoys some robustness property in that it can ignore a single (or more than one) historical study if too discrepant from the current controls. Thus, in contrast to the robustified MAP, not the whole set of historical controls will be used or neglected based on its similarity to the current controls.

5.2 | Dependent power parameters

For comparable historical and current control patients that satisfy Pocock's criteria, it seems reasonable to desire related weight parameters for the different historical data. Hence, in this study, we consider the MPP with dependent weight parameters in a hierarchical Bayesian framework by assuming the same parent distribution for δ_j, ie,

$$\delta_j \sim \text{Beta}(\alpha_\delta, \beta_\delta), (\, j = 1, 2, \, \ldots, K). \tag{15}$$

The hyperparameters α_δ and β_δ control the likely degree of borrowing from the historical data.[11] For this study, they are also helpful in the computation of the robustified mixture MPP of Section 5.3. These hyperparameters are reparameterized to the mean μ_δ and variance σ_δ^2 of the beta distribution as $\mu_\delta = \frac{\alpha_\delta}{\alpha_\delta + \beta_\delta}$ and variance $\sigma_\delta^2 = \frac{\mu_\delta(1-\mu_\delta)}{\alpha_\delta + \beta_\delta + 1}$. The DMPP can be given by considering dependent distributions for the weight parameters as

$$\pi_{\text{DMPP}}(\theta_C, \underline{\boldsymbol{\delta}}, \mu_\delta, \sigma_\delta^2 \,|\, \underline{\boldsymbol{D}}_{\boldsymbol{0}}) \propto \frac{\left[\prod_{j=1}^{K} L(\theta_C \,|\, D_{0j})^{\delta_j} p\left(\delta_j \,|\, \mu_\delta, \sigma_\delta^2\right)\right] p(\mu_\delta) p\left(\sigma_\delta^2\right) p(\theta_C)}{C(\underline{\boldsymbol{\delta}})}. \tag{16}$$

Including the current data, the posterior distribution of the DMPP is

$$p_{\text{DMPP}}\left(\theta_C, \Delta, \underline{\boldsymbol{\delta}}, \mu_\delta, \sigma_\delta^2 \,|\, \underline{\boldsymbol{D}}_{\boldsymbol{0}}, D\right) \propto L(\theta_C, \Delta \,|\, D)\, \pi_{\text{DMPP}}\left(\theta_C, \underline{\boldsymbol{\delta}}, \mu_\delta, \sigma_\delta^2 \,|\, \underline{\boldsymbol{D}}_{\boldsymbol{0}}\right). \tag{17}$$

For dichotomous outcome data, with the scaling constant $C(\underline{\boldsymbol{\delta}})$ as in (14), the MPP with dependent weight parameters can be given by (see Appendix A.2.2)

$$\pi_{\text{DMPP}}(\theta_C, \underline{\boldsymbol{\delta}} \,|\, \underline{\boldsymbol{y}}_{\boldsymbol{0}}, \underline{\boldsymbol{n}}_{\boldsymbol{0}}) \propto \frac{\theta_C^{\sum \delta_j y_{0j} + \alpha_\theta - 1} (1-\theta_C)^{\sum \delta_j (n_{0j} - y_{0j}) + \beta_\theta - 1} \prod_{j=1}^{K} p\left(\delta_j \,|\, \mu_\delta, \sigma_\delta^2\right) p(\mu_\delta) p\left(\sigma_\delta^2\right)}{B\left(\sum \delta_j y_{0j} + \alpha_\theta, \sum \delta_j (n_{0j} - y_{0j}) + \beta_\theta\right)}. \tag{18}$$

Comparing Equations (14) and (18), we see that only the prior for δ_j is different. The effect of the hierarchical prior on δ_j is different than for the MAP. While the MAP assumes that the θ's are similar, but not necessarily the same, now, we assume that the powers are not too different. A motivation for this prior is that Pocock's criteria should guarantee the similarity of the historical controls.

5.3 | The robust DMPP

A robust version of the DMPP can be developed to more effectively account for the possibility of prior-data conflict. This time, we aim to downweight the historical data when there is a conflict between the historical and current data through the distribution of the weight parameters. This may be achieved with a mixture prior on the weight parameters having the above dependent prior as one component and a component concentrated at zero. Robustness to the DMPP can be applied in two ways: (1) by giving each historical trial an individual mixing proportion or (2) by giving the same mixing proportion to all historical trials simultaneously. If we denote the hierarchical prior in (18) by $p(\delta_j \,|\, \mu_\delta, \sigma_\delta^2)$, then the robustified dependent prior to each individual weight parameter is given by

$$\delta_j \sim (1 - w_R) * p\left(\delta_j \,|\, \mu_\delta, \sigma_\delta^2\right) + w_R * p_R(\delta_j), \tag{19}$$

whereas the second robustified version of the dependent prior is given by

$$\underline{\boldsymbol{\delta}} \sim (1 - w_R) * p\left(\underline{\boldsymbol{\delta}} \,|\, \mu_\delta, \sigma_\delta^2\right) + w_R * p_R(\underline{\boldsymbol{\delta}}), \tag{20}$$

with $p_R(\underline{\boldsymbol{\delta}})$ the vector of spike components at zero.

For computational reasons, we chose a spike distribution concentrated closely around zero having a small variance. We used a half-normal distribution with variance parameter $\frac{\sigma_\delta^2}{\kappa_P}$ with κ_P relatively large and σ_δ^2 the variance of the slab part; see Figure A1. This component allows the historical data to be largely discarded when there is a prior-data conflict.

The proportion w_R of the robust component is fixed depending on the relevance of the historical data. As for the robust MAP, we have taken here $w_R = 0.1$. Alternatively, one could assume a prior for w_R. In Appendix A.4, we show that assuming a robust DMPP with a fixed $w_R = 0.5$ is equivalent to assuming a uniform prior on [0,1] for w_R. Because of this result, we will not consider the case of a stochastic mixing proportion.

Hence, the robust DMPP is inspired by the robust MAP but implemented differently.

6 | ANALYSIS OF THE REAL LIFE DATA SETS

We have applied versions of the MAP and the MPP to incorporate the control data of the two historical trials HOVON 29 and HOVON 42 for the analysis of the HOVON 42A data and the three historical trials Van Assche, Feagan, and Rutgeerts-1 for the analysis of the Rutgeerts-2 data. The MPP methods include the MPP with independently distributed weight parameters ("MPP Ind"), with dependently distributed weight parameters (DMPP) and the robustified version of the DMPP with robustness on each individual component ("Robust DMPP 1") or globally ("Robust DMPP 2"). In addition, we applied the MAP and Robust MAP as well as a "Current data" analysis (a Bayesian analysis of the current trial only, ie, without historical data) and a "Pooled data" analysis (a pooled Bayesian analysis that includes the data of all trials without accounting for between-trial heterogeneity).

6.1 | Settings

We assume a Beta(1, 1) prior for θ_C in the MPP methods, but also for the θ_{C_j} in the "Pooled data" analysis and for θ_{CC} in both the "Current data" and the "Pooled data" analyses. For the hyperparameters μ and τ^2 in the MAP approach, which are expressed on a log-odds scale, we assumed a $N(0, 10^6)$ and a $HN(0, 1)$ half-normal prior, respectively. In all methods, a vague $N(0, 10^6)$ prior is assumed for the treatment effect Δ on the original scale.

In the "Robust MAP," we set $\kappa_M = 10$ to obtain a larger variance for the weakly informative component π_R with 10% attributed to the vague component of the mixture prior. For the "MPP Ind," we assumed a Beta(1, 1) prior for each δ_j. In the DMPP, the hyperparameters μ_δ and σ_δ^2 are assumed to have a $U(0, 1)$ and a $IG(0.01, 0.01)$ prior, respectively. For the robust components $p_R(\delta_j)$ and $p_R(\underline{\boldsymbol{\delta}})$ in the "Robust DMPP 1" and "Robust DMPP 2," respectively, we used $\kappa_P = 25$ to obtain a spike distribution as done by George and McCulloch.[27]

The computations involve Markov chain Monte Carlo (MCMC) computations. These were done using the JAGS software[28] in combination with R statistical software.[29] For all methods, a single chain was initiated and 50 000 MCMC iterations were run after 50 000 burn-in iterations. Convergence was assessed using Geweke's diagnostic. Codes are available in the Online Appendix.

6.2 | Results

6.2.1 | HOVON data set

In Table 3, the posterior mean of the treatment effect Δ (difference in CR rates of treatment arm minus control arm) in the HOVON 42A trial is given for the above defined methods that take the historical trials (HOVON 29 and HOVON 42) into account. The first observation is that all estimated treatment effects are quite small and do not differ much between the approaches. The posterior mean of Δ using the "Current data" analysis is 1.13%. When the historical controls are pooled with the data of the HOVON 42A trial, the posterior mean of Δ becomes −0.73%. The posterior means of Δ for the MPP and the MAP methods lie between the above two percentages. Also, the posterior SDs of Δ for the Bayesian methods lie between the "Current data" and the "Pooled data" posterior SD. This shows that all borrowing methods show some discounting of the historical data. The posterior means of Δ for the DMPP method is negative and closer to the posterior mean of the "Pooled data" analysis. Also, the posterior SDs of Δ using the DMPP are smaller as compared to the posterior SDs of other Bayesian methods, which indicates that the DMPP method borrows more information from the historical controls. Nonetheless, for all methods, the 95% credible interval (CI) of Δ includes zero. This shows that the investigational treatment (G-CSF priming) in HOVON 42A has no significant effect on the CR rate of the AML patients.

TABLE 3 HOVON acute myeloid leukemia trial: the posterior distribution of the treatment effect (in %) in the HOVON 42A trial using different methods for including historical data

Methods	Mean	Median	SD	95% CI
Current data	1.13	1.11	3.30	(−5.24, 7.60)
Pooled data	−0.73	−0.67	2.52	(−5.88, 4.07)
MAP	0.33	0.28	3.07	(−5.73, 6.49)
Robust MAP	0.32	0.31	3.10	(−5.65, 6.53)
MPP Ind	−0.22	−0.14	2.75	(−5.96, 5.00)
DMPP	−0.28	−0.25	2.63	(−5.57, 4.75)
Robust DMPP 1	−0.17	−0.10	2.74	(−5.69, 5.02)
Robust DMPP 2	−0.21	−0.12	2.67	(−5.47, 4.96)

Abbreviations: CI, credible interval; DMPP, dependent modified power prior; MAP, meta-analytic predictive; MPP, modified power prior; SD, standard deviation.

TABLE 4 Proof-of-concept trial: the posterior distribution of the treatment effect (in %) in the Rutgeerts-2 trial using different methods for including historical data

Methods	Mean	Median	SD	95% CI
Current data	27.74	27.72	4.83	(18.56, 37.26)
Pooled data	22.92	22.80	4.56	(14.18, 32.07)
MAP	26.38	26.38	4.93	(16.67, 36.12)
Robust MAP	26.51	26.52	5.00	(16.83, 36.39)
MPP Ind	24.45	24.42	4.80	(15.41, 33.97)
DMPP	24.38	24.43	4.76	(15.16, 33.98)
Robust DMPP 1	24.77	24.71	4.56	(15.44, 33.86)
Robust DMPP 2	24.87	24.82	4.67	(15.93, 34.62)

Abbreviations: CI, credible interval; DMPP, dependent modified power prior; MAP, meta-analytic predictive; MPP, modified power prior; SD, standard deviation.

The posterior median of τ using the MAP method is 0.251 (95% CI: 0.017, 1.204) on the log-odds scale (Table A1). This value shows that the variability among the HOVON trials (namely HOVON 29, 42 and 42A) is moderate according to criteria given by Neuenschwander et al,[7] see also section 7.1. of this paper.

6.2.2 | Proof-of-concept study

The posterior distribution of the treatment effect in the Rutgeerts-2 study (Application 2) is presented in Table 4. The 95% CIs of Δ based on all methods do not include the value 0, which shows that there is a significant effect of 5 mg of Infliximab on the remission rate of UC patients at week 8. The posterior mean of Δ based on the "Current data" analysis is 27.74%, whereas the posterior mean of Δ becomes 22.92% if the control data from the three historical trials are pooled with the data of the Rutgeerts-2 trial. As for the HOVON data, the posterior means of Δ for all Bayesian methods lie between the "Current data" and the "Pooled data" posterior mean. In this application, the SD of Δ for the MAP methods exceeds that of the "Current data" posterior SD, showing that there was no substantial gain of information from the historical trials.

The posterior median of τ was 0.451 (95% CI: 0.025,1.435) on the log-odds scale (Table A2), which means that the variability among the proof-of-concept trials is high. The posterior means of the weight parameters are closer to each other and lower in value for the robust DMPP than the posterior means for the MPP Ind method.

7 | SIMULATION STUDY

7.1 | Design and settings of the simulation study

Simulation studies were performed to compare the performance of the different borrowing approaches discussed above. We considered different scenarios for incorporating compatible and noncompatible historical trials into the analysis of a current trial.

For the compatible scenarios, we considered $K = 3$ and 5 historical trials with 100 and 150 patients per arm in each trial. Dichotomous outcomes were generated according to a Bernoulli distribution for the historical controls and for both treatment and control arms of the current trial. The probability of success for the ith trial p_i, where $i = 1, \ldots, K+1$ with $1, \ldots, K$ for the historical trials and $K+1$ for the current trial, was computed as follows:

$$p_i = 1/(1+\exp(-Z_i)), Z_i = \beta_0 + \beta_1 T_i + \epsilon_i \quad \text{with}\, \epsilon_i \sim N\left(0, \tau_Z^2\right),$$

where T_i is a binary covariate equal to 0 for the control arm and value 1 for the treatment arm. τ_Z^2 is the variance of the trial-specific effect on the log-odds scale that varies with the scenario, but the model parameters of the historical controls and the current control data were kept the same in the simulation study.

In practice, the between-trial heterogeneity τ_Z^2 often lies between 0.01 and 0.25 on the log-odds scale; see the works of Neuenschwander et al,[7] Spiegelhalter et al,[9] and Veroniki et al.[30] We considered in our simulation study $\tau_Z^2 = 0$, 0.01, 0.04, and 0.16 for no, low, moderate, and high between-trial heterogeneity, respectively. Inspired by the HOVON studies, we set the CR rate for AML patients to 72% for the control treatment[31] and a treatment effect of 13% so that the response rate for the investigational arm was 85%. Accordingly, the baseline log-odds is set to $\beta_0 = \log\{0.72/(1-0.72)\} = 0.944$.

The log-odds scale is convenient for specifying the variation between trials, but for interpretation purposes, the treatment effect is expressed as $\Delta = \theta_T - \theta_{CC}$, where θ_T represents the parameter for the treatment and θ_{CC} for the control arms of the current trial. Hence, we expressed β_1 in terms of Δ, ie, $\beta_1 = \log\{\frac{(0.72+\Delta)/(1-(0.72+\Delta))}{0.72/(1-0.72)}\}$. This helps to perform the simulation study with different settings for the treatment effect Δ: (I) with treatment effect ($\Delta = 0.13$) and (II) without treatment effect ($\Delta = 0$). Results of the simulations based on these scenarios of compatibility are presented in Tables 5, 6, and 7 in Section 7.2.

We performed simulations for two additional scenarios for the case of incorporating incompatible historical trials into the analysis of a current trial by considering three historical trials with 150 patients per arm in each trial. For the first scenario, we let one of the historical control groups differ from all current and historical control groups. In the second scenario, the current control group is taken to be different from the historical controls. For these scenarios of incompatibility, the baseline CR rate for the control treatment is also set to be 72%. In the first scenario, one of the historical control groups has a 30% lower response rate than the other, homogeneous control groups, whereas in the second scenario, the 30% lower response rate applies to the current control group. Such situations are classically referred to a prior-data conflict.[3] That is,

$$Z_i = \beta_0 + \beta_1 T_i - 1.2 \quad \text{for}\, i = 1, \ldots, K.$$

Simulation results based on these two scenarios of incompatibility are presented in Table 8 in Section 7.2.

We simulated 1000 data sets for each scenario and setting. For some scenarios, evaluations were performed based on 6000 simulation runs but they gave almost identical results as with 1000 simulation runs. The methods were compared using frequentist measures like the type I error rate (no treatment effect) and statistical power (treatment effect of 0.13). To obtain a fair comparison of methods that incorporates a trade-off between the power and the type I error rate, we calculated a calibrated version. For this calibrated power, the rejection region was based on that equal-tail CI, which yields approximately an observed type I error rate of 5% in the simulations. Note that this calibration is performed for the above simulations that are based on given sample sizes and true response rates in the control groups. We also computed the precision and root mean square deviation (RMSD) of the posterior mean of the treatment effect Δ.

7.2 | Results of the simulation study

The type I error rate and the statistical power of the methods are reported in Tables 5 and 6, respectively. All MPP methods yielded higher statistical power than the MAP prior and its robust version. The DMPP method yielded the highest power of all borrowing methods. This method offered approximately 12% more power than the Current data approach (ie, an uninformative prior), but this is at the cost of an inflated (about 10%) type I error rate with moderate or high between-trial heterogeneity. The MAP methods produced an estimated type I error rate close to 5% in all scenarios and settings.

TABLE 5 The type I error rate of the treatment effect in the simulation study based on 1000 simulated data sets

Number of Historical Trials (H)	Number of Patients (N)	Method	Between-Trial Heterogeneity: No Heterogeneity	Low Heterogeneity	Moderate Heterogeneity	High Heterogeneity
$H = 3$	$N = 100$	Current data	0.050	0.051	0.045	0.062
		Pooled data	0.051	0.073	0.106	0.240
		MAP	0.039	0.050	0.043	0.060
		Robust MAP	0.039	0.050	0.043	0.062
		MPP Ind	0.039	0.054	0.048	0.099
		DMPP	0.040	0.055	0.057	0.109
		Robust DMPP 1	0.041	0.057	0.054	0.100
		Robust DMPP 2	0.035	0.052	0.054	0.097
	$N = 150$	Current data	0.045	0.048	0.054	0.051
		Pooled data	0.044	0.078	0.128	0.302
		MAP	0.036	0.041	0.051	0.058
		Robust MAP	0.036	0.043	0.049	0.057
		MPP Ind	0.035	0.043	0.061	0.091
		DMPP	0.036	0.051	0.076	0.122
		Robust DMPP 1	0.040	0.049	0.064	0.106
		Robust DMPP 2	0.040	0.047	0.062	0.090
$H = 5$	$N = 100$	Current data	0.054	0.043	0.046	0.056
		Pooled data	0.051	0.076	0.115	0.255
		MAP	0.043	0.042	0.052	0.052
		Robust MAP	0.043	0.042	0.050	0.050
		MPP Ind	0.045	0.053	0.073	0.126
		DMPP	0.041	0.057	0.074	0.120
		Robust DMPP 1	0.043	0.043	0.068	0.111
		Robust DMPP 2	0.044	0.056	0.064	0.065
	$N = 150$	Current data	0.047	0.043	0.045	0.044
		Pooled data	0.051	0.074	0.139	0.304
		MAP	0.048	0.046	0.050	0.044
		Robust MAP	0.046	0.045	0.048	0.048
		MPP Ind	0.050	0.049	0.080	0.132
		DMPP	0.048	0.050	0.084	0.138
		Robust DMPP 1	0.048	0.048	0.071	0.119
		Robust DMPP 2	0.043	0.049	0.074	0.105

Abbreviations: DMPP, dependent modified power prior; MAP, meta-analytic predictive; MPP, modified power prior.

Based on the calibrated power in Table 7, the MPP methods outperformed the MAP methods for homogeneous data and for low and moderate between-trial heterogeneities, whereas the MAP method had the best results for high heterogeneity. The DMPP method produced higher calibrated power than the "MPP Ind" method, especially for homogeneous data and lower between-trial heterogeneity. For high between-trial heterogeneity, the robust version of DMPP yielded higher calibrated power than the other MPP methods. In the calculation of the calibrated power, we observed that the average rejection region for the MPP methods with moderate and high between-trial heterogeneity is approximately 0.04 and 0.02, respectively. This value is lower than 0.05 because the MPP methods yielded inflated type I error rates for 95% equal-tail CIs. The average rejection region for the MAP methods is about 0.05 because these methods could maintain the nominal type I error rate of 5%.

The power of the methods considerably increases with the number of patients in the trials, but the increase in power with respect to the number of historical studies is small. The "Pooled data" analysis had the highest calibrated power of all methods with low between-trial heterogeneity but performed poorly for the moderate and high heterogeneity settings.

The average RMSDs show how much the different methods benefit from incorporating the historical data (Table A3). Compared to the other methods, the DMPP yielded the smallest RMSDs, even smaller than the "Pooled data" analysis for the high between-trial heterogeneity. In all scenarios and settings, the four MPP methods achieved lower RMSDs than the MAP methods, and the "Current data" analysis yielded the largest RMSDs. The RMSDs decreased with the number of patients per trial and the number of historical trials.

For all methods, increasing the number of patients per trial decreases the SD of the estimated treatment effect (Table A4). Except for the "Current data" analysis, increasing the number of historical studies has the advantage of

TABLE 6 The power of the treatment effect in the simulation study based on 1000 simulated data sets

Number of Historical Trials (H)	Number of Patients (N)	Method	Between-Trial Heterogeneity: No Heterogeneity	Low Heterogeneity	Moderate Heterogeneity	High Heterogeneity
$H = 3$	$N = 100$	Current data	0.621	0.612	0.617	0.589
		Pooled data	0.786	0.786	0.754	0.708
		MAP	0.696	0.677	0.656	0.605
		Robust MAP	0.688	0.672	0.645	0.607
		MPP Ind	0.728	0.716	0.694	0.667
		DMPP	0.754	0.748	0.727	0.699
		Robust DMPP 1	0.742	0.730	0.713	0.679
		Robust DMPP 2	0.734	0.721	0.701	0.666
	$N = 150$	Current data	0.789	0.785	0.779	0.778
		Pooled data	0.935	0.936	0.912	0.848
		MAP	0.853	0.861	0.832	0.795
		Robust MAP	0.852	0.856	0.826	0.794
		MPP Ind	0.901	0.904	0.884	0.855
		DMPP	0.914	0.916	0.896	0.860
		Robust DMPP 1	0.898	0.905	0.887	0.851
		Robust DMPP 2	0.899	0.905	0.875	0.824
$H = 5$	$N = 100$	Current data	0.610	0.614	0.598	0.623
		Pooled data	0.817	0.819	0.791	0.727
		MAP	0.720	0.706	0.673	0.641
		Robust MAP	0.714	0.704	0.662	0.637
		MPP Ind	0.761	0.768	0.745	0.691
		DMPP	0.772	0.789	0.761	0.711
		Robust DMPP 1	0.771	0.776	0.757	0.710
		Robust DMPP 2	0.774	0.762	0.761	0.709
	$N = 150$	Current data	0.779	0.764	0.765	0.751
		Pooled data	0.946	0.927	0.900	0.835
		MAP	0.879	0.867	0.835	0.789
		Robust MAP	0.879	0.866	0.834	0.785
		MPP Ind	0.915	0.906	0.876	0.853
		DMPP	0.919	0.909	0.889	0.850
		Robust DMPP 1	0.919	0.910	0.885	0.863
		Robust DMPP 2	0.910	0.902	0.874	0.848

Abbreviations: DMPP, dependent modified power prior; MAP, meta-analytic predictive; MPP, modified power prior.

increasing the precision of the estimate. This shows that these different methods borrow a considerable amount of information from the historical data. The SDs of the MPP methods and the MAP methods lie in between the SDs of the "Pooled data" and "Current data" analysis. These estimates are lowest for the DMPP method in all scenarios and settings.

In Table 8, the calibrated power computed for the two scenarios of incompatible historical studies is presented. When one of the historical control groups deviates from the other control groups (scenario 1), the "Robust DMPP 1" produced better calibrated power than the other methods. However, with a prior-data conflict between all historical control groups and the current control group, the "Robust DMPP 2" and the MAP methods yielded better power. In that scenario, the "Robust DMPP 2" method gave the lowest posterior mean (standard deviation) of the weight parameters (0.02(3.71E-6)) of all MPP methods and, thus, shows the strongest downweighting of the incompatible historical data.

8 | DISCUSSION

The MPP has become an established method for including historical data. However, previous applications of this method either included data from only a single historical study or naively pooled the data of the historical studies. This study evaluated the extension of MPP methods to account for multiple historical trials in the analysis of a current trial, with different possible priors for the study-specific weight parameters. For the inclusion of historical controls in the analysis of current clinical trial, the evaluation of Pocock's criteria for the comparability of the historical and current control patients is central. Accordingly, the power prior approach assumes the same parameter for the historical controls and the current

TABLE 7 The calibrated power of the treatment effect in the simulation study based on 1000 simulated data sets

Number of Historical Trials (H)	Number of Patients (N)	Method	Between-Trial Heterogeneity: No Heterogeneity	Low Heterogeneity	Moderate Heterogeneity	High Heterogeneity
$H = 3$	$N = 100$	Current data	0.621	0.612	0.641	0.556
		Pooled data	0.783	0.727	0.648	0.332
		MAP	0.731	0.677	0.674	0.586
		Robust MAP	0.727	0.672	0.663	0.570
		MPP Ind	0.776	0.704	0.700	0.523
		DMPP	0.791	0.717	0.706	0.543
		Robust DMPP 1	0.781	0.711	0.688	0.533
		Robust DMPP 2	0.771	0.713	0.688	0.532
	$N = 150$	Current data	0.795	0.799	0.777	0.774
		Pooled data	0.938	0.906	0.814	0.425
		MAP	0.885	0.882	0.820	0.780
		Robust MAP	0.884	0.869	0.827	0.779
		MPP Ind	0.918	0.915	0.872	0.729
		DMPP	0.928	0.915	0.868	0.736
		Robust DMPP 1	0.915	0.906	0.869	0.745
		Robust DMPP 2	0.917	0.905	0.858	0.740
$H = 5$	$N = 100$	Current data	0.581	0.647	0.613	0.607
		Pooled data	0.816	0.751	0.638	0.363
		MAP	0.733	0.726	0.659	0.624
		Robust MAP	0.730	0.736	0.662	0.631
		MPP Ind	0.780	0.747	0.678	0.534
		DMPP	0.802	0.770	0.689	0.556
		Robust DMPP 1	0.767	0.791	0.708	0.581
		Robust DMPP 2	0.785	0.752	0.690	0.603
	$N = 150$	Current data	0.785	0.779	0.781	0.769
		Pooled data	0.945	0.907	0.757	0.521
		MAP	0.887	0.872	0.835	0.796
		Robust MAP	0.885	0.868	0.840	0.795
		MPP Ind	0.911	0.907	0.831	0.670
		DMPP	0.922	0.909	0.847	0.698
		Robust DMPP 1	0.894	0.897	0.856	0.749
		Robust DMPP 2	0.920	0.902	0.836	0.733

Abbreviations: DMPP, dependent modified power prior; MAP, meta-analytic predictive; MPP, modified power prior.

TABLE 8 Calibrated power of the treatment effect for noncompatible scenarios for three historical trials with 150 patients per arm in each trial based on 1000 simulated data sets

Method	Noncompatible One Historical Control	Noncompatible Current Control
MAP	0.812	0.851
Robust MAP	0.810	0.850
MPP Ind	0.899	0.771
DMPP	0.896	0.770
Robust DMPP 1	0.911	0.758
Robust DMPP 2	0.863	0.851

Abbreviations: DMPP, dependent modified power prior; MAP, meta-analytic predictive; MPP, modified power prior.

controls, albeit with a lower weight for the historical data in the analysis. Based on real data sets and simulation study, the DMPP method, which assumes dependent weight parameters for the different historical studies, borrows more historical information than the other methods. For homogeneous controls and lower between-trial heterogeneities this method outperforms the other methods in terms of statistical power.

Despite the fact that the MPP methods produce better power than the MAP approach, they resulted in inflated type I error rates for higher between-trial heterogeneities. As studied by the same research group,[17] the MAP approach is able to control the type I error rate to approximately 5% in all scenarios and settings. For the trade-off between the type I error rate and the power, in this study, the calibrated power was computed by fixing the type I error rate in the simulations to 5%. Based on this criterion, the MAP methods seem to perform better for incorporating comparable historical controls with high between-trial heterogeneity. However, the robust versions of the DMPP that protect against prior-data conflict improve the power for incorporating heterogeneous and noncompatible historical trials.

In the HOVON application, for all methods considered, the treatment of G-CSF priming had no significant effect on the response rate of the AML patients in HOVON 42A. This adds to previous studies by Sung et al[32] and Löwenberg et al,[33] which suggested no benefit of CSF priming on response rates in patients receiving chemotherapy for AML. In the analysis of HOVON 42A, only two historical trials that satisfy Pocock's comparability criteria are incorporated, namely, HOVON 29 and HOVON 42. The heterogeneity among these HOVON trials was estimated to be moderate. The DMPP method gained more information from HOVON 29 and HOVON 42 than the other methods, as it estimated a lower SD of the treatment effect, and the posterior mean of the treatment effect was closer to the posterior mean of the "Pooled data" analysis.

In the second application, all methods yielded a significant effect of 5 mg of Infliximab on the remission rate of UC patients at week 8. A review of Gisbert et al[34] and Rutgeerts et al[20] suggested that Infliximab is more effective than placebo for UC. Due to high between-trial heterogeneity among the proof-of-concept trials, all borrowing methods, especially the MAP methods, gained little information from the three historical trials, namely, Van Assche, Feagan, and Rutgeerts-1, for the analysis of the Rutgeerts-2 data. In the two real-life examples and in the simulation study, the MAP methods tended to borrow less information than the MPP methods.

The main advantage of the borrowing methods considered in this paper is the increase in power compared to the analysis without historical data. This advantage should be weighed against the main disadvantages of including historical data, namely, the potential increase in the type I error rate and the additional effort required to implement these methods. The inflation in type I error depends on the level of heterogeneity, with effectively no increase for the no and low heterogeneity scenarios, a small increase to 6% to 8% for moderate heterogeneity, and a considerable increase for the high heterogeneity scenarios. In the more explorative settings where these borrowing methods are likely to be applied, the small increases in type I error rate in the moderate heterogeneity setting could well be acceptable to practitioners, whereas the type I error rate inflation of the high heterogeneity setting will generally not be acceptable. Because the data typically provide limited information on the level of heterogeneity, the borrowing methods can be used without adjustment only if the possibility of a high level of heterogeneity can be ruled out *a priori*, through a strict application of comparability criteria. For example, if historical trials are chosen in line with Pocock's criteria, the most likely scenario in practice will be incorporating almost homogeneous historical trials and certainly not data with large heterogeneity. For that case, the DMPP would be an appropriate method.

In case a strict control of the nominal type I error rate is required, the inflation of the type I error rate could be addressed by adjusting (lowering) the significance level as in the calculation of the calibrated power. This adjusted significance level can be obtained from the rejection region observed based on equal-tail CIs that yield approximately a type I error rate of 5%. In our results, an adjusted significance level of 0.03 or 0.04 was sufficient to maintain nominal type I error rates in all scenarios with at most moderate heterogeneity, whereas a significance level of 0.01 or 0.02 was needed for the high heterogeneity setting. An advantage of the adjustment based on calibrated power is that it provides a single number by which to compare the potential performance of different methods. But, it has the limitation that it is probably not feasible in real life and not easily calculated in practical applications, since the true response for control and level of heterogeneity will not be known exactly. The level of heterogeneity is difficult to estimate and there would likely be differences in the value of the adjusted significance level (and thus the calibrated power) between different levels of heterogeneity. This problem could conceivably be mitigated by ruling out the highest level of heterogeneity between studies through a strict application of comparability criteria or by applying more conservative priors for the weight parameters.

Sampling from the posterior distribution in the power prior approach is computationally difficult due to the challenging implementation of the scaling constant with which the posterior is multiplied to satisfy the likelihood principle. For a binary endpoint with a binomial distribution, the integration of the scaling constant can be implemented analytically. However, further studies on models with multiple parameters for incorporating several historical studies using the power prior approach are needed. In this study, the MPP methods were performed using MCMCpack package in R, whereas the "Current data," the "Pooled data" analysis, the MAP and its robust version were performed using another program, ie, Jags.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the VLIR JU-IUC project and the BOF Bilateral Cooperation for the financial support to the first author for his research visits. For the simulations, we used the infrastructure of the Flemish Supercomputer Center (VSC), funded by the Hercules Foundation and the EWI Department of the Flemish Government. The authors thank the Dutch-Belgian Hemato-Oncology Cooperative Group (HOVON) and Professor Bob Löwenberg, Department of Hematology, Erasmus University Medical Center, Rotterdam, The Netherlands, for kindly providing the data set and for interesting discussions. This research was financially supported by the Jimma University Inter-University Cooperation (IUC-JU) and the BOF-BILA of UHasselt.

COMPETING INTERESTS

The authors declare that they have no competing interests.

ORCID

Akalu Banbeta http://orcid.org/0000-0002-4963-5264
Joost van Rosmalen http://orcid.org/0000-0002-9187-244X

REFERENCES

1. Kendall JM. Designing a research project: randomized controlled trials and their principles. *Emerg Med J*. 2003;20(2):164-168.
2. Campbell G. Guidance for industry and FDA staff: guidance for the use of Bayesian statistics in medical device clinical trials. https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071121.pdf. 2010. Accessed August 26, 2017.
3. Schmidli H, Gsteiger S, Roychoudhury S, O'Hagan A, Spiegelhalter D, Neuenschwander B. Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*. 2014;70(4):1023-1032.
4. LeBlond D. FDA Bayesian statistics guidance for medical device clinical trials-application to process validation. *J Validation Technol*. 2010;16(4):24.
5. Guideline IH. Choice of control group in clinical trials e10. *Choice*. 2000;E:10. https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E10/Step4/E10_Guideline.pdf. Accessed July 9, 2017.
6. Pocock SJ. The combination of randomized and historical controls in clinical trials. *J Chronic Dis*. 1976;29(3):175-188.
7. Neuenschwander B, Capkun-Niggli G, Branson M, Spiegelhalter DJ. Summarizing historical information on controls in clinical trials. *Clin Trials*. 2010;7(1):5-18.
8. Ryan L. Using historical controls in the analysis of developmental toxicity data. *Biometrics*. 1993;49(4):1126-1135.
9. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Vol 13. England, UK: John Wiley & Sons; 2004;48.
10. Ibrahim JG, Chen MH. Power prior distributions for regression models. *Stat Sci*. 2000;15(1):46-60.
11. Hobbs BP, Carlin BP, Mandrekar SJ, Sargent DJ. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics*. 2011;67(3):1047-1056.
12. Hobbs BP, Sargent DJ, Carlin BP. Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Anal*. 2012;7(3):639-674.
13. Evans M, Moshonov H. Checking for prior-data conflict. *Bayesian Anal*. 2006;1(4):893-914.
14. Duan Y, Ye K, Smith EP. Evaluating water quality using power priors to incorporate historical information. *Environmetrics*. 2006;17(1):95-106.
15. De Greef GE, Van Putten WL, Boogaerts M, et al. Criteria for defining a complete remission in acute myeloid leukaemia revisited. An analysis of patients treated in HOVON-SAKK co-operative group studies. *Brit J Haematol*. 2005;128(2):184-191.
16. Larson RA, Löwenberg B, Connor RF. Remission criteria in acute myeloid leukemia and monitoring for residual disease. http://www.uptodate.com/contents/remission-criteria-in-acute-myeloid-leukemia-and-monitoring-for-residual-disease. 2012. Accessed June 10, 2017.
17. van Rosmalen J, Dejardin D, van Norden Y, Löwenberg B, Lesaffre E. Including historical data in the analysis of clinical trials: is it worth the effort? *Stat Methods Med Res*. 2017;27(10):3167-3182.
18. Van Assche G, Sandborn W, Feagan BG, et al. Daclizumab, a humanised monoclonal antibody to the interleukin 2 receptor (CD25), for the treatment of moderately to severely active ulcerative colitis: a randomised, double blind, placebo controlled, dose ranging trial. *Gut*. 2006;55(11):1568-1574.
19. Feagan BG, Greenberg GR, Wild G, et al. Treatment of ulcerative colitis with a humanized antibody to the $\alpha_4\beta_7$ integrin. *N Engl J Med*. 2005;352(24):2499-2507.

20. Rutgeerts P, Sandborn WJ, Feagan BG, et al. Infliximab for induction and maintenance therapy for ulcerative colitis. *N Engl J Med*. 2005;353(23):2462-2476.
21. Chen MH, Ibrahim JG, Shao QM. Power prior distributions for generalized linear models. *J Stat Plan Infer*. 2000;84(1-2):121-137.
22. Birnbaum A. On the foundations of statistical inference. *J Amer Stat Assoc*. 1996;57(298):269-306.
23. Neuenschwander B, Branson M, Spiegelhalter DJ. A note on the power prior. *Statist Med*. 2009;28(28):3562-3566.
24. Gelman A, Meng XL. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Stat Sci*. 1998;13(2):163-185.
25. Friel N, Pettitt AN. Marginal likelihood estimation via power posteriors. *J Royal Stat Soc Ser B*. 2008;70(3):589-607.
26. Dejardin D. The use of modified power prior to partially take into account historical data: the case of survival analysis [PhD thesis]. Leuven, Belgium: KU Leuven; 2013.
27. George EI, McCulloch RE. Variable selection via Gibbs sampling. *Amer Stat Assoc*. 1993;88(423):881-889.
28. Plummer M. Jags: a program for analysis of Bayesian graphical models using Gibbs sampling; In: Proceedings of the 3rd International Workshop on Distributed Statistical Computing; 2003; Wien, Vienna.
29. Team RC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.
30. Veroniki AA, Jackson D, Viechtbauer W, et al. Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Res Synth Methods*. 2016;7(1):55-79.
31. Lee JH, Joo YD, Kim H, et al. A randomized trial comparing standard versus high-dose daunorubicin induction in patients with acute myeloid. *Blood*. 2011;118(14):3832-3841.
32. Sung L, Alibhai SM, Beyene J, et al. Hematopoietic colony-stimulating factor priming does not influence survival in acute myeloid leukemia: a meta-analysis of randomized trials. *Leukemia*. 2009;23(4):811-813.
33. Löwenberg B, Van Putten W, Theobald M, et al. Effect of priming with granulocyte colony-stimulating factor on the outcome of chemotherapy for acute myeloid leukemia. *N Engl J Med*. 2003;349(8):743-752.
34. Gisbert JP, Gonzalez-Lama Y, Mate J. Systematic review: infliximab therapy in ulcerative colitis. *Aliment Pharm Ther*. 2007;25(1):19-37.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Banbeta A, van Rosmalen J, Dejardin D, Lesaffre E. Modified power prior with multiple historical trials for binary endpoints. *Statistics in Medicine*. 2019;38:1147–1169. https://doi.org/10.1002/sim.8019

APPENDIX

In this supplementary material, the proof that the MPP methods satisfy the likelihood principle is given, as well as a derivation of the posteriors of the MPP methods for multiple historical trials with independently and dependently distributed weight parameters for binary end points. Plots of the spike-and-slab distribution, results based on real-life data, and simulation studies are also presented.

A.1 | The likelihood principle

Let θ_C be the model parameter and $\underline{\boldsymbol{\delta}} = \{\delta_1, \ldots, \delta_K\}$ denote the total set of the weight parameters for the K historical data set $\underline{\boldsymbol{D}}_{\mathbf{0}}$, the joint power prior for multiple historical data that has been proposed by Ibrahim and Chen[10] is given as

$$\pi(\theta_C, \underline{\boldsymbol{\delta}} \,|\, \underline{\boldsymbol{D}}_{\mathbf{0}}) \propto \left[\prod_{j=1}^{K} L(\theta_C \,|\, D_{0j})^{\delta_j} p(\delta_j)\right] p(\theta_C). \tag{A1}$$

If each likelihood function, $L(\theta_C \,|\, D_{0j})$ is multiplied by a constant $\kappa_j, j = 1, \ldots, K$, the joint prior distribution of $(\theta_C, \underline{\boldsymbol{\delta}})$ becomes

$$\pi(\theta_C, \underline{\boldsymbol{\delta}} \,|\, \underline{\boldsymbol{D}}_{\mathbf{0}}) \propto \left[\prod_{j=1}^{K} \kappa_j^{\delta_j}\right] \left[\prod_{j=1}^{K} L(\theta_C \,|\, D_{0j})^{\delta_j} p(\delta_j)\right] p(\theta_C). \tag{A2}$$

Hence, the joint prior of $(\theta_C, \underline{\boldsymbol{\delta}})$ and consequently the posterior will be changed by a factor of $\prod_{j=1}^{K} \kappa_j^{\delta_j}$. This violates the likelihood principle that multiplying the likelihood function by a constant term should not affect the posterior distribution. However, this problem is solved in the MPP due to the scaling constant multiplied with the joint power prior.

$$\begin{aligned}\pi(\theta_C, \underline{\boldsymbol{\delta}} \mid \underline{\boldsymbol{D_0}}) &\propto \frac{\left[\prod_{j=1}^{K} \kappa_j^{\delta_j}\right]\left[\prod_{j=1}^{K} L(\theta_C \mid D_{0j})^{\delta_j} p(\delta_j)\right] p(\theta_C)}{\left[\prod_{j=1}^{K} \kappa_j^{\delta_j}\right]\int_{\theta_C}\left[\prod_{j=1}^{K} L(\theta_C \mid D_{0j})^{\delta_j}\right) p(\theta_C) d\theta_C} \\ &\propto \frac{\left[\prod_{j=1}^{K} L(\theta_C \mid D_{0j})^{\delta_j} p(\delta_j)\right] p(\theta_C)}{\int_{\theta_C}\left[\prod_{j=1}^{K} L(\theta_C \mid D_{0j})^{\delta_j}\right) p(\theta_C) d\theta_C}. \end{aligned} \tag{A3}$$

This means that, if each likelihood function $L(\theta_C \mid D_{0j})$ is multiplied by a constant κ_j, the MPP will not be changed.

A.2 | The MPP for multiple historical trials

A.2.1 | With independent weight parameters

The MPP for multiple historical studies is defined as

$$\pi_{\mathrm{MPP}}(\theta_C, \underline{\boldsymbol{\delta}} \mid \underline{\boldsymbol{D_0}}) \propto \frac{\left[\prod_{j=1}^{K} L(\theta_C \mid D_{0j})^{\delta_j} p(\delta_j)\right] p(\theta_C)}{C(\underline{\boldsymbol{\delta}})}. \tag{A4}$$

For dichotomous historical data with sample sizes $n_{01}, \ldots, n_{0K}$ and numbers of successes $y_{01}, \ldots, y_{0K}$, the scaling constant $C(\underline{\delta})$ used in (A4) can be computed analytically as

$$\begin{aligned} C(\underline{\delta}) &= \int_{\theta_C}\left[\prod_{j=1}^{K} L(\theta_C \mid D_{0j})^{\delta_j}\right] p(\theta_C) d\theta_C = \int_{\theta_C}\left[\prod_{j=1}^{K} L(\theta_C \mid y_{0j}, n_{0j})^{\delta_j}\right] p(\theta_C) d\theta_C \\ &= \frac{\prod_{j=1}^{K}\binom{n_{0j}}{y_{0j}}^{\delta_j}}{B(\alpha_\theta, \beta_\theta)} \int_{\theta_C} \theta_C^{\sum \delta_j y_{0j} + \alpha_\theta - 1}(1-\theta_C)^{\sum \delta_j (n_{0j} - y_{0j}) + \beta_\theta - 1} d\theta_C \\ &= \frac{\prod_{j=1}^{K}\binom{n_{0j}}{y_{0j}}^{\delta_j}}{B(\alpha_\theta, \beta_\theta)} B\left(\sum \delta_j y_{0j} + \alpha_\theta, \sum \delta_j (n_{0j} - y_{0j}) + \beta_\theta\right). \end{aligned} \tag{A5}$$

The numerator in (A4) can be computed as

$$\begin{aligned} &\left[\prod_{j=1}^{K}\binom{n_{0j}}{y_{0j}}^{\delta_j} \theta_C^{\delta_j y_{0j}}(1-\theta_C)^{\delta_j (n_{0j} - y_{0j})} p(\delta_j)\right] \frac{\theta_C^{\alpha_\theta - 1}(1-\theta_C)^{\beta_\theta - 1}}{B(\alpha_\theta, \beta_\theta)} \\ &= \frac{\left[\prod_{j=1}^{K}\binom{n_{0j}}{y_{0j}}^{\delta_j}\right]}{B(\alpha_\theta, \beta_\theta)} \theta_C^{\sum \delta_j y_{0j} + \alpha_\theta - 1}(1-\theta_C)^{\sum \delta_j (n_{0j} - y_{0j}) + \beta_\theta - 1} \prod_{j=1}^{K} p(\delta_j). \end{aligned} \tag{A6}$$

Substituting (A5) and (A6) in (A4), the MPP for multiple historical studies with a binary endpoint can be computed as

$$\frac{\left[\prod_{j=1}^{K}\binom{n_{0j}}{y_{0j}}^{\delta_j}\right]}{B(\alpha_\theta,\beta_\theta)}\theta_C^{\sum\delta_j y_{0j}+\alpha_\theta-1}(1-\theta_C)^{\sum\delta_j(n_{0j}-y_{0j})+\beta_\theta-1}\prod_{j=1}^{K}p(\delta_j)}{\frac{\left[\prod_{j=1}^{K}\binom{n_{0j}}{y_{0j}}^{\delta_j}\right]}{B(\alpha_\theta,\beta_\theta)}B\left(\sum\delta_j y_{0j}+\alpha_\theta,\sum\delta_j(n_{0j}-y_{0j})+\beta_\theta\right)} =$$

$$\frac{\theta_C^{\sum\delta_j y_{0j}+\alpha_\theta-1}(1-\theta_C)^{\sum\delta_j(n_{0j}-y_{0j})+\beta_\theta-1}\prod_{j=1}^{K}p(\delta_j)}{B\left(\sum\delta_j y_{0j}+\alpha_\theta,\sum\delta_j(n_{0j}-y_{0j})+\beta_\theta\right)}. \tag{A7}$$

A.2.2 | With dependent weight parameters

The DMPP is defined as

$$\pi(\theta_C,\underline{\delta},\mu,\sigma\,|\,\underline{D}_{0j}) \propto \frac{\left[\prod_{j=1}^{K}L(\theta_C\,|\,D_{0j})^{\delta_j}p(\delta_j\,|\,\mu,\sigma)\right]p(\mu)p(\sigma)p(\theta_C)}{C(\underline{\delta})}. \tag{A8}$$

With the scaling constant $C(\underline{\delta})$ as in (A5), the MPP for the dichotomous data $\pi(\theta_C,\underline{\delta},\mu,\sigma\,|\,\underline{y}_{0j},\underline{n}_{0j})$ can be computed as

$$\frac{\left[\prod_{j=1}^{K}\binom{n_{0j}}{y_{0j}}^{\delta_j}\theta_C^{\delta_j y_{0j}}(1-\theta_C)^{\delta_j(n_{0j}-y_{0j})}p(\delta_j\,|\,\mu,\sigma)\right]p(\mu)p(\sigma)\frac{\theta_C^{\alpha_\theta-1}(1-\theta_C)^{\beta_\theta-1}}{B(\alpha_\theta,\beta_\theta)}}{\frac{\left[\prod_{j=1}^{K}\binom{n_{0j}}{y_{0j}}^{\delta_j}\right]}{B(\alpha_\theta,\beta_\theta)}B\left(\sum\delta_j y_{0j}+\alpha_\theta,\sum\delta_j(n_{0j}-y_{0j})+\beta_\theta\right)} =$$

$$\frac{\frac{\left[\prod_{j=1}^{K}\binom{n_{0j}}{y_{0j}}^{\delta_j}\right]}{B(\alpha_\theta,\beta_\theta)}\theta_C^{\sum\delta_j y_{0j}+\alpha_\theta-1}(1-\theta_C)^{\sum\delta_j(n_{0j}-y_{0j})+\beta_\theta-1}\prod_{j=1}^{K}p(\delta_j\,|\,\mu,\sigma)p(\mu)p(\sigma)}{\frac{\left[\prod_{j=1}^{K}\binom{n_{0j}}{y_{0j}}^{\delta_j}\right]}{B(\alpha_\theta,\beta_\theta)}B\left(\sum\delta_j y_{0j}+\alpha_\theta,\sum\delta_j(n_{0j}-y_{0j})+\beta_\theta\right)} =$$

$$\frac{\theta_C^{\sum\delta_j y_{0j}+\alpha_\theta-1}(1-\theta_C)^{\sum\delta_j(n_{0j}-y_{0j})+\beta_\theta-1}\prod_{j=1}^{K}p(\delta_j\,|\,\mu,\sigma)p(\mu)p(\sigma)}{B\left(\sum\delta_j y_{0j}+\alpha_\theta,\sum\delta_j(n_{0j}-y_{0j})+\beta_\theta\right)}. \tag{A9}$$

Let us have n samples for each of the treatment and the control arms of a new trial with x and y number of successes and parameters θ_T and θ_C, respectively. The treatment effect is defined as $\Delta = \theta_T - \theta_C$. We can put θ_T as $\theta_C + \Delta$ and we can perform the analysis to compare both arms using θ_C and Δ. Hence, the posterior distribution after incorporating the current data can be computed as

$$p(\theta_C,\Delta,\underline{\delta},\mu,\sigma\,|\,\underline{y}_{0j},\underline{n}_{0j},y,x,n) \propto L(\theta_C,\Delta\,|\,y,x,n)\pi(\theta_C,\underline{\delta},\mu,\sigma\,|\,\underline{y}_{0j},\underline{n}_{0j}) \propto$$

$$(\theta_C+\Delta)^x(1-(\theta_C+\Delta))^{n-x}\times$$

$$\frac{\theta_C^{\sum\delta_j y_{0j}+y+\alpha_\theta-1}(1-\theta_C)^{\sum\delta_j(n_{0j}-y_{0j})+(n-y)+\beta_\theta-1}\prod_{j=1}^{K}p(\delta_j\,|\,\mu,\sigma)p(\mu)p(\sigma)p(\Delta)}{B\left(\sum\delta_j y_{0j}+\alpha_\theta,\sum\delta_j(n_{0j}-y_{0j})+\beta_\theta\right)}. \tag{A10}$$

A.3 | Spike-and-slab prior for the weight parameter

To achieve the Robust DMPP, a mixture prior on the weight parameters having the dependent prior Beta$(\alpha_\delta,\beta_\delta)$ as a slab component and a spike component at zero can be applied. For the spike component, we can use a half-normal distribution that has the same height as a spike distribution with a smaller variance $\frac{\sigma_\delta^2}{\kappa_P}$ with κ_P relatively large, like $\kappa_P = 25$ and σ_δ^2 the variance of the slab part (see the work of George and McCulloch[27]). As an example, let a weight parameter δ have a mixture prior with a slab component Beta$(10,10)$. This distribution has a mean $\mu_\delta = \frac{\alpha_\delta}{\alpha_\delta+\beta_\delta} = 0.5$ and a variance

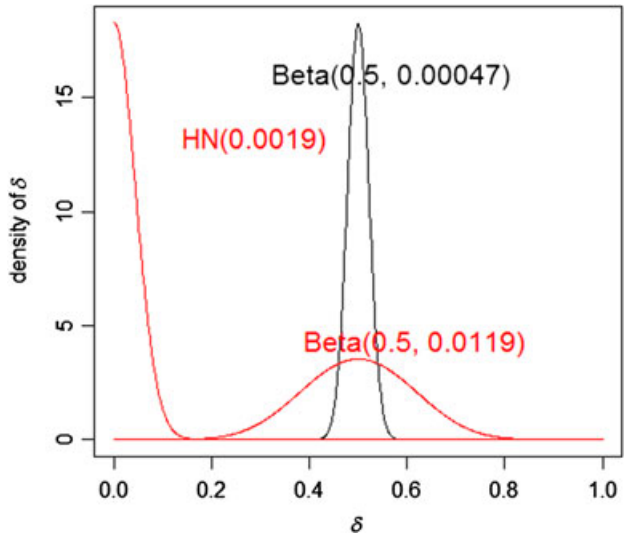

FIGURE A1 Examples of a spike-and-slab distribution and a half-normal distribution for the weight parameter in the robust dependent modified power prior [Colour figure can be viewed at wileyonlinelibrary.com]

$\sigma_\delta^2 = \frac{\mu_\delta(1-\mu_\delta)}{\alpha_\delta+\beta_\delta+1} = 0.0119$. The spike distribution can be formed with variance $\frac{\sigma_\delta^2}{25} = 0.00047$. A half-normal distribution with variance $\frac{\sigma_\delta^2}{6.25} = 0.0019$ has the same height as this spike distribution. Figure A1 demonstrates the spike-and-slab plot of this particular example. It is always true that a beta distribution with variance $\frac{\sigma_\delta^2}{25}$ has the same height as a half-normal distribution with variance $\frac{\sigma_\delta^2}{6.25}$. As a result, in this study, we used a half-normal distribution with variance $\frac{\sigma_\delta^2}{6.25}$ as a spike distribution in the mixture prior, where σ_δ^2 is the variance of the slab part Beta$(\alpha_\delta, \beta_\delta)$.

A.4 | Random proportion for a robust component in the mixture prior for the distribution of the weight parameter

The following is the proof that shows how the power prior with a mixture prior for the weight parameter with a random w_R can be equivalent to a fixed $w_R = 0.5$. For simplicity, the proof is based on a single historical trial. Note that the power prior of θ_C for a current study is defined as

$$\pi(\theta_C, \delta \mid D_0) \propto \frac{L(\theta_C \mid D_0)^\delta p(\theta_C) p(\delta)}{C(\delta)}. \tag{A11}$$

We let a mixture prior for the weight parameter δ to form the robust version of the power prior through the weight parameter.

$$\delta \sim (1 - w_R) * \text{Beta}(\alpha_\delta, \beta_\delta) + w_R * p_R(\delta), \tag{A12}$$

where $p_R(\delta)$ is the robust component and w_R is the proportion of this component. If we fix the value of the proportion w_R to, eg, 0.1 or 0.5, the formulation in (A11) will not be changed. Simply, it can be formulated conditional on the proportion w_R as

$$\pi(\theta_C, \delta \mid D_0, w_R) \propto \frac{L(\theta_C \mid D_0)^\delta p(\theta_C) p(\delta \mid w_R)}{C(\delta)}. \tag{A13}$$

However, if a random proportion having a Beta$(1, 1)$ distribution is assumed for w_R, then the power prior in (A13) can be formulated as

$$\pi(\theta_C, \delta, w_R \mid D_0) \propto L(\theta_C \mid D_0)^\delta p(\theta_C) p(\delta \mid w_R) p(w_R). \tag{A14}$$

The joint posterior distribution of (θ_c, δ) can given by

$$
\begin{aligned}
p(\theta_c, \delta) &= \int_0^1 p(\theta_c, \delta, w_R) dw_R = \int_0^1 L(\theta_C \mid y_0, n_0)^{\delta} p(\theta_C) p(\delta \mid w_R) p(w_R) dw_R \\
&= \int_0^1 \binom{n_0}{y_0}^{\delta} \theta_C^{\delta y_0} (1-\theta_C)^{\delta(n_0-y_0)} \frac{\theta_C^{\alpha_\theta - 1}(1-\theta_C)^{\beta_\theta - 1}}{B(\alpha_\theta, \beta_\theta)} p(\delta \mid w_R) p(w_R) dw_R \\
&= \frac{\binom{n_0}{y_0}^{\delta}}{B(\alpha_\theta, \beta_\theta)} \theta_C^{\delta y_0 + \alpha_\theta - 1} (1-\theta_C)^{\delta(n_0-y_0)+\beta_\theta - 1} \int_0^1 p(\delta \mid w_R) p(w_R) dw_R. \qquad \text{(A15)}
\end{aligned}
$$

Let us call the constant part $c = \frac{\binom{n_0}{y_0}^{\delta}}{B(\alpha_\theta, \beta_\theta)} \theta_C^{\delta y_0 + \alpha_\theta - 1} (1-\theta_C)^{\delta(n_0-y_0)+\beta_\theta - 1}$. Consider the distribution of the weight parameter in (A12) and assume that the proportion w_R has a Beta(α_w, β_w) distribution. Then, Equation (A15) will be

$$
\begin{aligned}
&= c \int_0^1 p(\delta / w_R) p(w_R) dw_R \\
&= c \int_0^1 [(1-w_R) * \text{Beta}(\alpha_\delta, \beta_\delta) + w_R * p_R(\delta)] \text{Beta}(\alpha_w, \beta_w) dw_R \\
&= c \Bigg[\int_0^1 \text{Beta}(\alpha_\delta, \beta_\delta) \text{Beta}(\alpha_w, \beta_w) dw_R - \int_0^1 w_R * \text{Beta}(\alpha_\delta, \beta_\delta) \text{Beta}(\alpha_w, \beta_w) dw_R \\
&\quad + \int_0^1 w_R * p_R(\delta) \text{Beta}(\alpha_w, \beta_w) dw \Bigg] \\
&= c \Bigg[\text{Beta}(\alpha_\delta, \beta_\delta) \int_0^1 \text{Beta}(\alpha_w, \beta_w) dw_R - \text{Beta}(\alpha_\delta, \beta_\delta) \int_0^1 w_R * \frac{w_R^{\alpha_w - 1}(1-w_R)^{\beta_w - 1}}{B(\alpha_w, \beta_w)} dw_R \\
&\quad + p_R(\delta) \int_0^1 w_R * \frac{w_R^{\alpha_w - 1}(1-w_R)^{\beta_w - 1}}{B(\alpha_w, \beta_w)} dw_R \Bigg] \\
&= c \left[\text{Beta}(\alpha_\delta, \beta_\delta) \left(1 - \frac{B(\alpha_w + 1, \beta_w)}{B(\alpha_w, \beta_w)} \right) + p_R(\delta) \frac{B(\alpha_w + 1, \beta_w)}{B(\alpha_w, \beta_w)} \right]
\end{aligned}
\qquad \text{(A16)}
$$

Setting $w_R = \frac{B(\alpha_w+1, \beta_w)}{B(\alpha_w, \beta_w)}$ shows that the power prior methods with a random and a fixed proportion for the robust component are equivalent.

A.5 | Additional outputs

Tables A1 and A2 present the posterior distributions of the parameters for the amount of borrowing information from historical trials using the MPP and MAP methods into the analysis of a current trial based on the HOVON data and the proof-of-concept trial. In Tables A3 and A4, the RMSD and SD of the treatment effect in the simulation studies for the different methods are presented.

TABLE A1 HOVON AML trial: the posterior distributions of parameters for the amount of borrowing from the historical data using the MPP and the MAP methods

Methods	Mean	Median	SD	95% CI
MAP				
τ	0.342	0.251	0.315	(0.015, 1.217)
Robust MAP				
τ	0.336	0.241	0.315	(0.016, 1.203)
MPP Ind				
δ_1	0.476	0.451	0.282	(0.027, 0.969)
δ_2	0.549	0.579	0.276	(0.045, 0.978)
DMPP				
δ_1	0.515	0.512	0.179	(0.179, 0.852)
δ_2	0.518	0.518	0.176	(0.184, 0.847)
Robust DMPP 1				
δ_1	0.451	0.452	0.193	(0.014, 0.838)
δ_2	0.496	0.465	0.175	(0.150, 0.847)
Robust DMPP 2				
δ_1	0.473	0.468	0.212	(0.036, 0.891)
δ_2	0.479	0.469	0.212	(0.033, 0.887)

Abbreviations: AML, acute myeloid leukemia; CI, credible interval; DMPP, dependent modified power prior; MAP, meta-analytic predictive; MPP, modified power prior; SD, standard deviation.

TABLE A2 Proof-of-concept trial: the posterior distributions of parameters for the amount of borrowing from the historical data using the MPP and the MAP methods

Methods	Mean	Median	SD	95% CI
MAP				
τ	0.520	0.450	0.366	(0.025, 1.435)
Robust MAP				
τ	0.513	0.436	0.513	(0.035, 1.460)
MPP Ind				
δ_1	0.535	0.553	0.285	(0.032, 0.972)
δ_2	0.441	0.414	0.286	(0.019, 0.962)
δ_3	0.383	0.327	0.277	(0.013, 0.950)
DMPP				
δ_1	0.441	0.421	0.202	(0.097, 0.861)
δ_2	0.438	0.419	0.205	(0.094, 0.860)
δ_3	0.434	0.411	0.205	(0.087, 0.860)
Robust DMPP 1				
δ_1	0.414	0.403	0.231	(0.016, 0.882)
δ_2	0.399	0.392	0.234	(0.017, 0.862)
δ_3	0.326	0.311	0.240	(0.004, 0.839)
Robust DMPP 2				
δ_1	0.440	0.433	0.211	(0.036, 0.856)
δ_2	0.435	0.429	0.209	(0.041, 0.844)
δ_3	0.432	0.427	0.211	(0.034, 0.847)

Abbreviations: CI, credible interval; DMPP, dependent modified power prior; MAP, meta-analytic predictive; MPP, modified power prior; SD, standard deviation.

TABLE A3 The average root mean square deviation (95% confidence interval) of the estimated treatment effect in the simulation study

Number of Historical Trials (H)	Number of Patients (N)	Method	Between-Trial Heterogeneity			
			No Heterogeneity	Low Heterogeneity	Moderate Heterogeneity	High Heterogeneity
$H = 3$	$N = 100$	Current data	0.076 (0.075-0.078)	0.077 (0.076-0.079)	0.078 (0.076-0.079)	0.081 (0.079-0.082)
		Pooled data	0.057 (0.056-0.058)	0.058 (0.057-0.059)	0.061 (0.059-0.062)	0.069 (0.067-0.071)
		MAP	0.066 (0.065-0.067)	0.067 (0.066-0.069)	0.070 (0.068-0.071)	0.075 (0.074-0.077)
		Robust MAP	0.066 (0.065-0.067)	0.068 (0.067-0.069)	0.070 (0.069-0.071)	0.075 (0.074-0.077)
		MPP Ind	0.062 (0.060-0.063)	0.063 (0.061-0.064)	0.064 (0.063-0.065)	0.068 (0.067-0.069)
		DMPP	0.060 (0.059-0.061)	0.061 (0.060-0.062)	0.063 (0.061-0.064)	0.067 (0.066-0.068)
		Robust DMPP 1	0.062 (0.060-0.063)	0.063 (0.061-0.064)	0.063 (0.061-0.064)	0.068 (0.067-0.070)
		Robust DMPP 2	0.062 (0.061-0.063)	0.063 (0.062-0.064)	0.065 (0.063-0.066)	0.070 (0.069-0.072)
	$N = 150$	Current data	0.063 (0.062-0.064)	0.063 (0.062-0.064)	0.064 (0.063-0.066)	0.066 (0.065-0.067)
		Pooled data	0.046 (0.046-0.047)	0.047 (0.046-0.048)	0.049 (0.048-0.050)	0.059 (0.057-0.061)
		MAP	0.055 (0.054-0.056)	0.055 (0.054-0.056)	0.058 (0.056-0.059)	0.062 (0.060-0.063)
		Robust MAP	0.055 (0.054-0.056)	0.055 (0.054-0.056)	0.058 (0.057-0.059)	0.062 (0.061-0.063)
		MPP Ind	0.051 (0.050-0.052)	0.051 (0.050-0.051)	0.052 (0.051-0.053)	0.055 (0.054-0.057)
		DMPP	0.049 (0.049-0.050)	0.049 (0.049-0.050)	0.051 (0.050-0.052)	0.055 (0.054-0.057)
		Robust DMPP 1	0.051 (0.050-0.052)	0.051 (0.050-0.051)	0.052 (0.051-0.053)	0.056 (0.055-0.057)
		Robust DMPP 2	0.051 (0.050-0.052)	0.051 (0.050-0.052)	0.053 (0.052-0.054)	0.058 (0.057-0.060)
$H = 5$	$N = 100$	Current data	0.077 (0.076-0.079)	0.077 (0.076-0.079)	0.078 (0.077-0.080)	0.079 (0.078-0.081)
		Pooled data	0.055 (0.054-0.056)	0.055 (0.054-0.057)	0.058 (0.057-0.060)	0.068 (0.066-0.070)
		MAP	0.064 (0.062-0.065)	0.064 (0.063-0.065)	0.067 (0.066-0.068)	0.073 (0.071-0.074)
		Robust MAP	0.064 (0.063-0.065)	0.064 (0.063-0.065)	0.068 (0.066-0.069)	0.073 (0.072-0.075)
		MPP Ind	0.059 (0.058-0.060)	0.059 (0.058-0.060)	0.061 (0.060-0.062)	0.066 (0.064-0.067)
		DMPP	0.059 (0.058-0.060)	0.059 (0.058-0.060)	0.061 (0.059-0.062)	0.066 (0.064-0.067)
		Robust DMPP 1	0.059 (0.058-0.061)	0.059 (0.058-0.060)	0.061 (0.060-0.063)	0.066 (0.065-0.067)
		Robust DMPP 2	0.060 (0.058-0.061)	0.060 (0.059-0.062)	0.062 (0.061-0.063)	0.069 (0.067-0.070)
	$N = 150$	Current data	0.063 (0.062-0.064)	0.063 (0.062-0.064)	0.064 (0.062-0.065)	0.065 (0.064-0.066)
		Pooled data	0.044 (0.043-0.045)	0.045 (0.044-0.046)	0.049 (0.048-0.050)	0.058 (0.056-0.060)
		MAP	0.051 (0.050-0.052)	0.053 (0.052-0.054)	0.055 (0.054-0.056)	0.060 (0.059-0.061)
		Robust MAP	0.051 (0.051-0.052)	0.053 (0.052-0.054)	0.056 (0.055-0.057)	0.060 (0.059-0.062)
		MPP Ind	0.047 (0.046-0.048)	0.048 (0.047-0.049)	0.050 (0.049-0.051)	0.054 (0.053-0.055)
		DMPP	0.047 (0.046-0.048)	0.048 (0.047-0.049)	0.050 (0.049-0.051)	0.055 (0.053-0.056)
		Robust DMPP 1	0.048 (0.047-0.048)	0.049 (0.048-0.049)	0.051 (0.050-0.052)	0.054 (0.053-0.055)
		Robust DMPP 2	0.048 (0.047-0.049)	0.049 (0.048-0.050)	0.052 (0.051-0.053)	0.057 (0.056-0.058)

Abbreviations: DMPP, dependent modified power prior; MAP, meta-analytic predictive; MPP, modified power prior.

TABLE A4 The average standard error (95% confidence interval) of the estimated treatment effect in the simulation study

Number of Historical Trials (H)	Number of Patients (N)	Method	Between-Trial Heterogeneity: No Heterogeneity	Low Heterogeneity	Moderate Heterogeneity	High Heterogeneity
$H = 3$	$N = 100$	Current data	0.057 (0.057-0.057)	0.057 (0.057-0.057)	0.057 (0.057-0.057)	0.057 (0.056-0.057)
		Pooled data	0.042 (0.042-0.042)	0.042 (0.042-0.042)	0.042 (0.042-0.042)	0.042 (0.042-0.043)
		MAP	0.050 (0.050-0.051)	0.051 (0.051-0.051)	0.052 (0.052-0.053)	0.054 (0.054-0.055)
		Robust MAP	0.051 (0.050-0.051)	0.051 (0.051-0.051)	0.052 (0.052-0.053)	0.055 (0.054-0.055)
		MPP Ind	0.047 (0.047-0.047)	0.047 (0.047-0.048)	0.048 (0.048-0.048)	0.050 (0.050-0.050)
		DMPP	0.046 (0.046-0.046)	0.046 (0.046-0.046)	0.046 (0.046-0.047)	0.048 (0.048-0.049)
		Robust DMPP 1	0.047 (0.047-0.047)	0.047 (0.047-0.047)	0.048 (0.047-0.048)	0.050 (0.049-0.050)
		Robust DMPP 2	0.047 (0.047-0.047)	0.047 (0.047-0.048)	0.048 (0.048-0.048)	0.050 (0.050-0.051)
	$N = 150$	Current data	0.046 (0.046-0.047)	0.046 (0.046-0.046)	0.046 (0.046-0.046)	0.046 (0.046-0.046)
		Pooled data	0.034 (0.034-0.034)	0.034 (0.034-0.034)	0.034 (0.034-0.034)	0.034 (0.034-0.034)
		MAP	0.041 (0.041-0.042)	0.042 (0.042-0.042)	0.043 (0.043-0.043)	0.044 (0.044-0.045)
		Robust MAP	0.042 (0.041-0.042)	0.042 (0.042-0.042)	0.043 (0.043-0.043)	0.044 (0.044-0.045)
		MPP Ind	0.037 (0.037-0.037)	0.039 (0.039-0.039)	0.039 (0.039-0.040)	0.041 (0.041-0.041)
		DMPP	0.038 (0.037-0.038)	0.037 (0.037-0.038)	0.038 (0.038-0.038)	0.040 (0.039-0.040)
		Robust DMPP 1	0.038 (0.038-0.038)	0.038 (0.038-0.039)	0.039 (0.039-0.039)	0.041 (0.040-0.041)
		Robust DMPP 2	0.038 (0.038-0.039)	0.038 (0.038-0.039)	0.040 (0.039-0.040)	0.041 (0.041-0.042)
$H = 5$	$N = 100$	Current data	0.057 (0.057-0.057)	0.057 (0.056-0.057)	0.057 (0.056-0.057)	0.056 (0.056-0.057)
		Pooled data	0.040 (0.040-0.040)	0.040 (0.040-0.040)	0.040 (0.040-0.040)	0.040 (0.040-0.040)
		MAP	0.048 (0.047-0.048)	0.048 (0.048-0.049)	0.050 (0.050-0.051)	0.054 (0.053-0.054)
		Robust MAP	0.048 (0.048-0.048)	0.048 (0.048-0.049)	0.051 (0.050-0.051)	0.054 (0.053-0.054)
		MPP Ind	0.044 (0.044-0.044)	0.044 (0.044-0.044)	0.044 (0.044-0.045)	0.046 (0.046-0.046)
		DMPP	0.044 (0.044-0.044)	0.044 (0.043-0.044)	0.044 (0.044-0.045)	0.046 (0.045-0.046)
		Robust DMPP 1	0.044 (0.044-0.045)	0.044 (0.044-0.045)	0.045 (0.045-0.045)	0.047 (0.046-0.047)
		Robust DMPP 2	0.045 (0.044-0.045)	0.045 (0.045-0.045)	0.045 (0.045-0.046)	0.048 (0.047-0.048)
	$N = 150$	Current data	0.047 (0.046-0.047)	0.047 (0.046-0.047)	0.046 (0.046-0.047)	0.046 (0.046-0.046)
		Pooled data	0.033 (0.033-0.033)	0.033 (0.033-0.033)	0.033 (0.033-0.033)	0.033 (0.032-0.033)
		MAP	0.039 (0.039-0.039)	0.040 (0.040-0.040)	0.042 (0.042-0.042)	0.044 (0.044-0.044)
		Robust MAP	0.039 (0.039-0.040)	0.040 (0.040-0.040)	0.042 (0.042-0.042)	0.044 (0.044-0.045)
		MPP Ind	0.036 (0.036-0.036)	0.036 (0.036-0.036)	0.037 (0.036-0.037)	0.038 (0.038-0.038)
		DMPP	0.036 (0.036-0.036)	0.036 (0.036-0.036)	0.036 (0.036-0.037)	0.038 (0.037-0.038)
		Robust DMPP 1	0.036 (0.036-0.037)	0.037 (0.036-0.037)	0.037 (0.037-0.037)	0.039 (0.039-0.039)
		Robust DMPP 2	0.037 (0.036-0.037)	0.037 (0.037-0.037)	0.038 (0.038-0.038)	0.040 (0.039-0.040)

Abbreviations: DMPP, dependent modified power prior; MAP, meta-analytic predictive; MPP, modified power prior.