Biometrics 70, 1023–1032
December 2014

DOI: 10.1111/biom.12242

Robust Meta-Analytic-Predictive Priors in Clinical Trials with Historical Control Information

Heinz Schmidli,[1],* Sandro Gsteiger,[2] Satrajit Roychoudhury,[3] Anthony O'Hagan,[4] David Spiegelhalter,[5] and Beat Neuenschwander[6]

[1]Statistical Methodology, Development, Novartis Pharma AG, Basel, Switzerland
[2]Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland
[3]Statistical Methodology, Oncology, Novartis Pharmaceuticals Corporation, East Hanover, New Jersey, U.S.A.
[4]Department of Probability and Statistics, University of Sheffield, Sheffield, U.K.
[5]Statistical Laboratory, University of Cambridge, Cambridge, U.K.
[6]Statistical Methodology, Oncology, Novartis Pharma AG, Basel, Switzerland
**email:* heinz.schmidli@novartis.com

Summary. Historical information is always relevant for clinical trial design. Additionally, if incorporated in the analysis of a new trial, historical data allow to reduce the number of subjects. This decreases costs and trial duration, facilitates recruitment, and may be more ethical. Yet, under prior-data conflict, a too optimistic use of historical data may be inappropriate. We address this challenge by deriving a Bayesian meta-analytic-predictive prior from historical data, which is then combined with the new data. This prospective approach is equivalent to a meta-analytic-combined analysis of historical and new data if parameters are exchangeable across trials. The prospective Bayesian version requires a good approximation of the meta-analytic-predictive prior, which is not available analytically. We propose two- or three-component mixtures of standard priors, which allow for good approximations and, for the one-parameter exponential family, straightforward posterior calculations. Moreover, since one of the mixture components is usually vague, mixture priors will often be heavy-tailed and therefore robust. Further robustness and a more rapid reaction to prior-data conflicts can be achieved by adding an extra weakly-informative mixture component. Use of historical prior information is particularly attractive for adaptive trials, as the randomization ratio can then be changed in case of prior-data conflict. Both frequentist operating characteristics and posterior summaries for various data scenarios show that these designs have desirable properties. We illustrate the methodology for a phase II proof-of-concept trial with historical controls from four studies. Robust meta-analytic-predictive priors alleviate prior-data conflicts - they should encourage better and more frequent use of historical data in clinical trials.

Key words: Adaptive design; Adaptive randomization; Bayesian inference; Clinical trials; Exponential family; Meta-analysis; Mixture distribution; Robustness.

1. Introduction

Randomized controlled clinical trials are the most appropriate way to investigate a new test treatment in clinical research. These trials are usually evaluated based solely on study data. However, historical trials in the same patient population have often been carried out with the same control, for example, placebo. In such settings, it seems attractive to use the historical control data in the design and analysis of the new trial (Berry, 2006), as fewer patients can then be randomized to control. This lowers both cost and trial duration, facilitates recruitment, and may be more ethical in some situations.

Clinical trials with historical controls are used in earlier phases of drug development (Neuenschwander et al., 2010; Hueber et al., 2012; Trippa, Rosner, and Müller, 2012; Baeten et al., 2013; DiScala, Kerman, and Neuenschwander, 2013; Gsteiger et al., 2013), occasionally in phase III trials (Schmidli, Bretz, and Racine, 2007; French et al., 2012), and also in special areas such as medical devices (FDA, 2010a), serious conditions (FDA, 2013) and pediatric studies (Berry, 1989). In addition, proper use of historical information is critical for non-inferiority trials (FDA, 2010b; Schmidli, Wandel, and Neuenschwander, 2013). Clinical trials where the control arm is entirely replaced by historical information are popular in phase II oncology (Simon, 1989), but may lead to biases due to lack of randomization and blinding.

In a seminal article, Pocock (1976) suggested a blinded, randomized and controlled design for the new trial, where historical and concurrent controls are combined in the analysis. He also proposed conditions for the relevance of historical trials, and provided an approach for combining historical and concurrent controls. Several similar methods for combining historical and current data have been proposed since then, including Dempster, Selwyn, and Weeks (1983), Ryan (1993), Ibrahim and Chen (2000), Spiegelhalter, Abrams, and Myles (2004), Neuenschwander et al. (2010), Hobbs et al. (2011), and Hobbs, Sargent, and Carlin (2012). A common feature of the approaches is that they discount historical data to account for between-trial heterogeneity.

A major concern with the use of historical controls is prior-data conflict. Despite careful selection of historical trials, past information may not be relevant for the new trial due to unanticipated differences in study design, conduct or patient population (ICH, 2000). Irrespective of prior-data conflict, conjugate Bayesian analyses result in a pre-defined compromise (Fuquene, Cook, and Pericchi, 2009). For example, with normal endpoints and conjugate priors, the posterior mean is a weighted average of prior mean and data mean. For well-conducted clinical trials, prior-data conflict should be resolved by strong discounting of prior information. Dawid (1973) and O'Hagan (1979) showed that the relative weights of the tails of prior and sample distribution determine how conflict is resolved. Heavy-tailed priors will eventually be discarded with increasing prior-data conflict. Considerable progress on Bayesian heavy-tailed modeling and conflict resolution has been achieved in the past 40 years (O'Hagan and Pericchi, 2012). Of particular interest are mixture priors of conjugate distributions as these can be concisely characterized and provide tractable analyses (O'Hagan and Forster, 2004). These properties are important in biometric practice, as prior information needs to be specified in clinical trial protocols and medical publications to facilitate review by ethical committees, regulatory authorities and referees. The tractability of the analysis lowers the hurdle for implementation, and allows a very fast evaluation of operating characteristics.

We consider here the use of historical controls in a meta-analytic framework. In particular, we focus on the Bayesian version with a robust prior (derived from historical data) for the control arm. The robust prior is a mixture prior with two components. The first component, derived from historical data, is a meta-analytic-predictive (MAP) prior (Spiegelhalter et al., 2004; Neuenschwander et al., 2010), which already provides better robustness than a simple conjugate prior. However, the MAP prior is not available in analytical form. To allow for a concise description of the prior and tractable posterior analysis we approximate the MAP prior by a mixture of conjugate priors, with the Kullback–Leibler divergence as a measure of discrepancy. Any prior can be well described by such a mixture (Dalal and Hall, 1983; Diaconis and Ylvisaker, 1985), so that no relevant information is lost in this process. The second (weakly informative) component of the mixture prior ensures further robustness against prior-data conflict.

When designing a clinical trial, the number of patients allocated to control and test treatment needs to be specified. If historical prior information is used, it is important to know the prior effective sample size, that is, the equivalent number of patients corresponding to the prior information. While the prior effective sample size is well-defined for conjugate priors, this is more difficult in general (Morita, Thall, and Müller, 2008).

If the control data and the robust prior are in clear conflict, the prior information will essentially be discarded in the posterior analysis. This may result in inconclusive trial results, as not enough control information may then be available. Adaptive trials can minimize this risk (Hobbs, Carlin, and Sargent, 2013), since they allow to change features of the design based on interim analyses (Berry et al., 2010). We will discuss a two-stage adaptive design, where more patients are randomized to control in the second stage, if interim results suggest prior-data conflict.

In Section 2, we first introduce the retrospective and prospective meta-analytic approach and show that they are equivalent. We then focus on meta-analytic-predictive priors, describe how they can be approximated by mixtures of conjugate priors, and how further robustness can be achieved by adding an extra weakly-informative component to the prior. Finally, the approximate prior effective sample size is discussed. Section 3 describes a two-stage adaptive design with binary endpoints, and provides corresponding frequentist operating characteristics. In Section 4, we illustrate the methodology and its practical implementation for a phase II proof-of-concept trial. The article concludes with a discussion.

2. Methods

2.1. *Meta-Analytic Approaches to Incorporate Historical Data*

We consider a new clinical trial comparing a test treatment with a control, and where relevant historical data on the control group are available, which one aims to incorporate in the analysis of the new trial. For the new trial, data and parameter are denoted by $X_\star$ and $\phi_\star$ for the test and by $Y_\star$ and $\psi_\star$ for the control treatment, respectively. For ease of presentation we will assume that no nuisance parameters are present, and that no relevant prior information on the test parameter $\phi_\star$ is available.

Denote the control data and parameters of the H historical trials by $Y_{\mathcal{H}} = (Y_1, \ldots, Y_H)$ and $\psi_{\mathcal{H}} = (\psi_1, \ldots, \psi_H)$, where $\mathcal{H} = \{1, \ldots, H\}$. The structure of the control data suggest a hierarchical model:

$$Y_h|\psi_h \sim F(\psi_h; n_h)\ ,\ \ \psi_h|\eta \sim G(\eta)\ ,\ \ \eta \sim P, \tag{1}$$

where $h \in \mathcal{H}_\star = \{1, \ldots, H, \star\}$, $n_1, \ldots, n_H, n_\star$ are the sample sizes of the trials, and F, G, P are the sampling, exchangeability (random-effects), and hyper-prior distribution, respectively. The framework could be extended to accommodate study level covariates, so that trials are then partially exchangeable.

Inference for the control parameter $\psi_\star$ in the new trial will be based on direct $(Y_\star)$ and indirect evidence $(Y_{\mathcal{H}})$. This can be done in two ways:

(i) by a *meta-analytic-combined (MAC)* approach. At the end of the trial, a meta-analysis of all the control data will be performed, providing the inference for $\psi_\star$ given all the data by $p(\psi_\star|Y_\star, Y_{\mathcal{H}})$,

(ii) or, by a two-step approach. At the design stage of the current trial, a *meta-analytic-predictive (MAP)* prior $p(\psi_\star|Y_{\mathcal{H}})$ is derived from the historical control data. Then, at the end of the trial, the MAP prior is combined with the current control data $Y_\star$ via Bayes' theorem: $p(\psi_\star|Y_\star, Y_{\mathcal{H}}) \propto p(Y_\star|\psi_\star)p(\psi_\star|Y_{\mathcal{H}})$.

The MAC and MAP analyses are equivalent for model (1), as the data are conditionally independent, and do not involve the hyper-parameters (see Appendix). However, at the planning stage of the new trial, only historical control data are available. This makes the use of the two-step MAP approach particularly attractive, as then the prior information on the

Table 1

Sampling model, conjugate prior and posterior, marginal probability f (up to proportionality; see Section 2.3) and transformed parameter $\theta = g(\psi)$ for some members of the one-parameter exponential family with sufficient statistics n and $\bar{y}$. The nuisance parameter is assumed to be known for the Negative Binomial and Normal distribution.

Model $F(\psi)$	Prior and posterior	f	θ

Bernoulli(ψ)	Beta(a, b)
	Beta($a + n\bar{y}, b + n - n\bar{y}$)
		$\frac{B(a+n\bar{y},b+n-n\bar{y})}{B(a,b)}$	$\log\{\psi/(1-\psi)\}$

NegBin(ψ, r)	Beta(a, b)
	Beta($a + nr, b + n\bar{y}$)
		$\frac{B(a+nr,b+n\bar{y})}{B(a,b)}$	$\log\{\psi/(1-\psi)\}$

Poisson(ψ)	Gamma(a, b)
	Gamma($a + n\bar{y}, b + n$)
		$\frac{\Gamma(a+n\bar{y})/(b+n)^{a+n\bar{y}}}{\Gamma(a/b^a}$	$\log(\psi)$

Exponential(ψ)	Gamma(a, b)
	Gamma($a + n, b + n\bar{y}$)
		$\frac{\Gamma(a+n)/(b+n\bar{y})^{a+n}}{\Gamma(a/b^a}$	$\log(\psi)$

Normal(ψ, s^2)	Normal($m_0, s^2/n_0$) Normal$\left(\frac{n_0 m_0 + n\bar{y}}{n_0 + n}, s^2/n_0 + s^2/n\right)$	$\frac{exp\{-0.5(\bar{y}-m_0)^2/(s^2/n_0+s^2/n)\}}{\sqrt{s^2/n_0+s^2/n}}$	ψ

control can be quantified in advance and influence the design. For example, fewer patients may then be randomized to control. With prior $p(\phi_\star, \psi_\star | Y_{\mathcal{H}}) = p(\phi_\star) p(\psi_\star | Y_{\mathcal{H}})$, the posterior is given by $p(\phi_\star, \psi_\star | X_\star, Y_\star, Y_{\mathcal{H}}) = p(\phi_\star | X_\star) p(\psi_\star | Y_\star, Y_{\mathcal{H}})$, from which the posterior for the treatment effect can be derived.

2.2. *Meta-Analytic-Predictive (MAP) Priors*

We consider the planning of a new clinical trial, to compare a test treatment with a control, where $n_\star$ patients are randomized to the control group. The control data $Y_\star$ are assumed to follow a distribution from the regular one-parameter exponential family $Y_\star \sim F(\psi_\star; n_\star)$. For example, if the endpoint is binary, the number of responders $Y_\star$ follows a binomial distribution with parameter $\psi_\star$; see Table 1 for other common endpoints.

We suppose that several historical controlled clinical trials in the same patient population are available, and that the control data Y_h in trial h with n_h patients is distributed as

$$Y_h \sim F(\psi_h; n_h), h = 1, \ldots, H. \tag{2}$$

The similarity of new and historical trials is expressed by exchangeable parameters $\theta = g(\psi)$ (see Table 1)

$$\theta_\star, \theta_1, \ldots, \theta_H \sim \text{Normal}(\mu, \tau^2) \tag{3}$$

with population mean μ and between-trial standard deviation τ. For binary endpoints, the logit-transformation $g(\psi) = log\{\psi/(1-\psi)\}$ is commonly used. Although we use a normal distribution here, a more flexible random-effects distribution could also be considered.

Finally, a prior distribution for the hyper-parameters is chosen as $p(\mu, \tau) = p(\mu)p(\tau)$. For the population mean μ a vague prior is used here, as data are sufficiently informative. More care is needed for the between-trial standard deviation τ, especially with few historical trials. Spiegelhalter et al. (2004), Gelman (2006), and Polson and Scott (2012) discuss several priors and recommend Half-T distributions, which include the Half-Normal and Half-Cauchy as special cases. In the following, we will use Half-Normal priors, with standard deviations chosen such that unrealistically large values of τ have small probability. Sensitivity analyses with various realistic priors for τ can be useful, especially if fewer than four historical studies are available (Gelman, 2006; Hobbs et al., 2012).

Based on model (2), (3) and the hyper-prior, the MAP prior distribution $p_H(\psi_\star) = p(\psi_\star | Y_1, \ldots, Y_H)$ for the control parameter in the new trial can be derived. Markov Chain Monte Carlo (MCMC) can be used to generate a sample $\theta_\star^{(1)}, \ldots, \theta_\star^{(M)}$, and then $\psi_\star^{(i)} = g^{-1}(\theta_\star^{(i)})$.

A kernel-density estimate from the MCMC sample or a Rao-Blackwellized density estimate (Gelfand and Smith, 1990; Web Appendix A) can be used to describe the MAP prior. However, there are practical disadvantages both in the communication of such density estimates and their use in the analysis of a new trial, as these density estimates have a very large number of parameters. A parsimonious and convenient alternative approximation are mixtures of conjugate priors. For example in the binary case, this mixture prior is

$$\hat{p}_H(\psi_\star) = \sum_{k=1}^{K} w_k \text{Beta}(\psi_\star | a_k, b_k) \tag{4}$$

with positive weights w_k summing up to one.

Dalal and Hill (1983) as well as Diaconis and Ylvisaker (1985) have shown that any prior can be closely approximated in this way. The number of components K, the mixture weights and the hyper-parameters for each component must be specified, such that the approximate prior is close to $p_H(\psi_\star)$. We consider here the Kullback–Leibler divergence, as this is arguably the most appropriate measure in pure inference problems (Bernardo and Smith, 1994; O'Hagan and Forster, 2004). The Kullback–Leibler divergence from the exact prior $p_H(\psi_\star)$ to the approximate prior $\hat{p}_H(\psi_\star)$ is

$$KL(p_H(\psi_\star), \hat{p}_H(\psi_\star)) = \int \log\{p_H(\psi_\star)\} p_H(\psi_\star) \mathrm{d}\psi_\star - \int \log\{\hat{p}_H(\psi_\star)\} p_H(\psi_\star) \mathrm{d}\psi_\star. \tag{5}$$

The best approximation to the exact prior is obtained by choosing weights w_k and hyper-parameters of the conjugate priors such that the second term on the right in (5) is maximal, which requires numerical optimization. It should be noted that a Monte-Carlo estimate of the integral is given by $1/M\sum_{i=1}^{M}\log\{\hat{p}_H(\psi_\star^{(i)})\}$, using the MCMC sample from the posterior distribution. This term is formally identical to the log-likelihood of the MCMC sample with mixture model $\widehat{p}_H(\psi_\star)$. Hence, the weights and the hyper-parameters can be obtained as maximum-likelihood (ML) estimates. The choice of the number of components can be based on both numerical and graphical methods, and is further discussed in Section 4. In our context, the prior is typically unimodal, and two to three components are usually sufficient to adequately approximate the exact prior.

2.3. *Robust MAP Prior*

When considering the use of historical controls in a new trial, a careful selection of the historical trials is necessary to render the exchangeability assumption for the control parameters plausible. Pocock (1976) proposed criteria for the selection process. Nevertheless, one has to acknowledge the possibility of prior-data conflict. Hence we consider a robust version of the MAP prior of Section 2.1, as

$$\widehat{p}_{HR}(\psi_\star) = (1-w_R)\widehat{p}_H(\psi_\star) + w_R\ p_V(\psi_\star)\ , \tag{6}$$

where $\widehat{p}_H(\psi_\star)$ is the approximated MAP prior, $p_V(\psi_\star)$ is a vague conjugate prior, and w_R is the prior probability that the new trial differs systematically from the historical trials. The choice of w_R could be based on the degree of confidence of the clinical trial team in the relevance of the historical data. Posterior summaries for several data scenarios or operating characteristics may also be considered when selecting the weight given to the vague component of the prior. The choice of w_R in (6) will determine how quickly historical information is discounted with increasing prior-data conflict. This second component must be proper if one would like to interpret w_R as a probability, and hence we use here weakly informative priors. For binary endpoints, either Jeffreys' or the uniform prior could be used, or, for other endpoints, unit information priors (Kass and Wasserman, 1995).

Since the robust prior $\widehat{p}_{HR}(\psi_\star)$ is again a mixture of conjugate priors, the posterior $\widehat{p}_{HR}(\psi_\star|y_\star)$ is also a mixture of conjugate posteriors, with updated mixture weights (Bernardo and Smith, 1994). For binary endpoints, the robust prior is

$$\widehat{p}_{HR}(\psi_\star) = (1-w_R)\sum_k w_k\mathrm{Beta}(\psi_\star|a_k,b_k) + w_R\mathrm{Beta}(\psi_\star|a_0,b_0), \tag{7}$$

and the posterior is given by

$$\begin{aligned}\widehat{p}_{HR}(\psi_\star|y_\star) = {} & (1-\tilde{w}_R)\sum_k \tilde{w}_k\mathrm{Beta}(\psi_\star|a_k+y_\star, b_k+n_\star-y_\star)\\ & +\tilde{w}_R\mathrm{Beta}(\psi_\star|a_0+y_\star, b_0+n_\star-y_\star),\end{aligned} \tag{8}$$

where $\tilde{w}_R \propto w_R f_0/\{w_R f_0 + (1-w_R)\sum_k w_k f_k\}$ and $\tilde{w}_k \propto w_k f_k/\{\sum_k w_k f_k\}$, and $f_k = B(a_k+y_\star, b_k+n_\star-y_\star)/B(a_k,b_k)$ is proportional to the marginal probability, for k=0,…,K, where $B(a,b)$ is the Beta function. The corresponding marginal probabilities f for other endpoints (disregarding proportionality constants) are shown in Table 1.

2.4. *Effective Sample Size of the Robust MAP Prior*

When designing a new clinical trial, fewer patients can be randomized to control by borrowing strength from historical information. However, the prior effective sample size (ESS) needs then to be quantified. For conjugate priors, the ESS is easily obtained for the exponential family (Bernardo and Smith, 1994), for example, for binary endpoints $ESS = a+b$ with a Beta(a,b) prior. For non-conjugate priors, normal approximations can be used (Morita et al., 2008; Neuenschwander et al., 2010; Morita, Thall, and Müller, 2012). Here we apply the methodology by Morita et al. (2008). The ESS is the sample size such that the expected information of the posterior under a non-informative prior is the same as the information of the informative prior $p(\psi_\star)$, where the information is evaluated at the mode $\tilde{\psi}_\star$ of the informative prior. The information of the prior is given by $I = \mathrm{d}^2\log\ p(\psi_\star)/\mathrm{d}\ \psi_\star^2\ |_{\psi_\star=\tilde{\psi}_\star}$. The expected information for the posterior with sample size m under a non-informative prior is $EI_0(m) = \int\{\mathrm{d}^2\log\ p_0(\psi_\star|y_\star)/\mathrm{d}\ \psi_\star^2\ |_{\psi_\star=\tilde{\psi}_\star}\}p(y_\star)\mathrm{d}y_\star$, where $p(y_\star)$ is the prior predictive distribution with respect to the informative prior $p(\psi_\star)$. The ESS is then the largest m such that $EI_0(m) < I$. For a conjugate prior, the ESS is the usual prior effective sample size.

3. Adaptive Design

We consider a clinical trial comparing a test treatment with a control, using a vague prior for the test treatment and a robust MAP prior for the control. Fewer patients are here randomized to control, as the robust MAP prior approximately corresponds to ESS control patients (see Section 2.4). However in case of prior-data conflict, the robust MAP prior will be discounted, and hence the information available at the end of the trial may then not be sufficient for decision making; see Web Appendix B. Use of an adaptive design can reduce this risk as described by Hobbs et al. (2013), and we use a similar approach in the following.

We propose a two-stage adaptive design, where m and n are the desired effective sample sizes at the end of the trial for test treatment and control, respectively (e.g., $m = n = 40$). The number of patients in the two stages are then

Stage 1: m_I in test treatment and n_I in control (e.g., $m_I = 20$, $n_I = 15$);
Stage 2: $(m - m_I)$ in test treatment and $max(n - ESS_I, n_{min})$ in control (e.g., $n_{min} = 5$).

Here ESS_I is the posterior effective sample size based on the first stage control data with a robust MAP prior. If the interim control data and the robust MAP prior are consistent, then $ESS_I \approx ESS + n_I$. However, if prior and data are in clear conflict, then $ESS_I \approx n_I$. The effective sample size at interim ESS_I determines how many patients will be randomized to control in the second stage (at least n_{min}). At the

Table 2

Type I error and power (%) under different control rates $\psi_\star$ and treatment effects δ for four priors: Beta$(4, 16)$ (Beta), $0.5 \times$ Beta$(4, 16) + 0.5 \times$ Beta$(1, 1)$ (Mix50) , $0.9 \times$ Beta$(4, 16) + 0.1 \times$ Beta$(1, 1)$ (Mix90) , Beta$(1, 1)$ (Unif). Also shown is the expected sample size in the control group for the two mixture priors.

	Treatment effect (δ=0)				Treatment effect (δ=0.3)				Expected sample size control	
Control rate ($\psi_\star$)	Mix50	Mix90	Beta	Unif	Mix50	Mix90	Beta	Unif	Mix50	Mix90
0.1	0.6	0.1	0.0	1.8	92.0	81.4	81.6	89.7	27.6	20.0
0.2	2.5	1.5	1.6	2.3	88.4	85.7	87.8	82.1	25.5	20.3
0.3	3.9	5.5	6.1	2.4	83.0	88.4	93.4	79.5	28.5	21.2
0.4	4.2	10.4	13.7	2.6	76.7	86.8	97.9	79.5	33.5	23.2
0.5	3.4	12.3	26.0	2.8	77.5	85.4	99.6	81.9	37.4	26.9
0.6	3.0	9.5	44.4	2.6	86.4	89.7	100.0	89.8	38.9	31.8

end of the trial, data from both stages are used in the final analysis.

In a clinical trial setting, frequentist operating characteristics are typically used for evaluating adaptive trial designs (Berry et al., 2010). Computations are very fast, as the posteriors can be calculated analytically when priors are mixtures of conjugate priors; see (8). For illustration, we investigate a setting typical for proof-of-concept trials with binary endpoint, where $m = n = 40$, $m_I = 20$, $n_I = 15$, and $n_{min} = 5$. We compare frequentist properties for a uniform prior on the test treatment parameter $\phi_\star$, and the following priors on the control parameter $\psi_\star$:

(i) a simple conjugate prior: Beta(4, 16);
(ii) a two-component mixture prior (weight 0.9): $0.9 \times$ Beta$(4, 16) + 0.1 \times$ Beta$(1, 1)$;
(iii) a two-component mixture prior (weight 0.5): $0.5 \times$ Beta$(4, 16) + 0.5 \times$ Beta$(1, 1)$;
(iv) the uniform prior: Beta(1, 1).

Prior (i), a simple conjugate prior, is the method of choice in many Bayesian applications with binary data. This prior could arise from an approximation of the MAP prior by a single Beta distribution. Mixture priors (ii) and (iii) could arise in two situations: as a robust version of prior (i), or, from a two-component mixture approximation of a MAP prior, where one component is usually weakly-informative (small Beta parameters) due to the heavy tails of the MAP prior. The prior effective sample sizes (Section 2.4) for the four priors are 20, 18, 11, and 2. The 95% prior probability intervals for the response rates are (0.06,0.40), (0.06,0.75), (0.04,0.95), and (0.025,0.975), respectively.

Table 2 shows Type-I error and power properties of the two-stage adaptive design, where study success will be declared if $P(\delta > 0|\text{data}) > 0.975$, for treatment effect $\delta = \phi_\star - \psi_\star$. In contrast to the informative conjugate prior (i), the mixture priors do not lead to an excessive increase of Type-I error. Also, the two-stage adaptive design with informative mixture priors results in considerable savings in sample size compared to a design using vague priors. For example, if historical and current control data are consistent, then on average 25–50% of the 40 control patients can be replaced by historical information. It should be noted that the operating characteristics depend upon the direction of bias induced by incorporating the historical information.

The frequentist properties of the Bayesian point estimate (posterior mean) for the control at the end of the adaptive two-stage trial are also of interest. Figure 1 shows the root mean squared error (rMSE) and bias for the three informative priors and the uniform prior. For the conjugate informative prior (i), the number of control patients is always 20, while this number varies between 20 and 40 for the two mixture priors (ii) and (iii). Hence for the uniform prior (iv), we consider a fixed design with $n = 20$ or $n = 40$ control patients for comparison. For rMSE the conjugate prior (i) and the

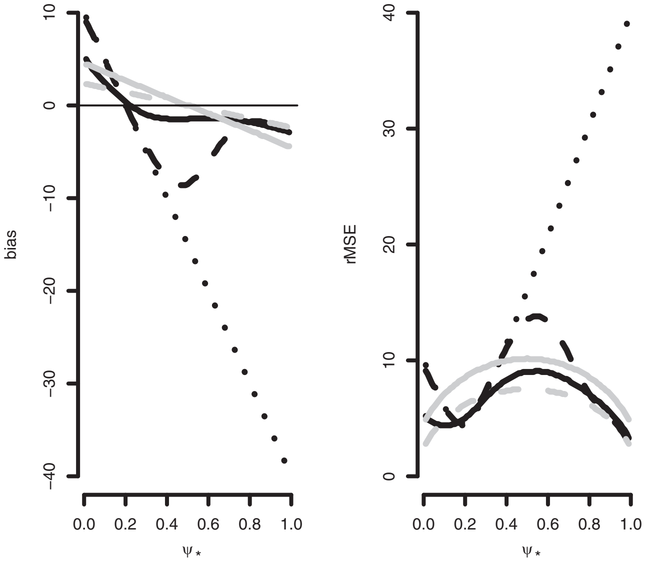

Figure 1. Bias (in %) and root mean-squared error (rMSE) of the posterior mean of the control response rate $\psi_\star$ for (i) Beta(4,16) prior (black dotted line), (ii) 0.9× Beta(4,16) + 0.1× Beta(1,1) (black dashed line), and (iii) 0.5× Beta(4,16) + 0.5× Beta(1,1) (black solid line). Also shown are rMSE and bias for a uniform prior with $n = 20$ (gray solid line) or $n = 40$ (gray dashed line).

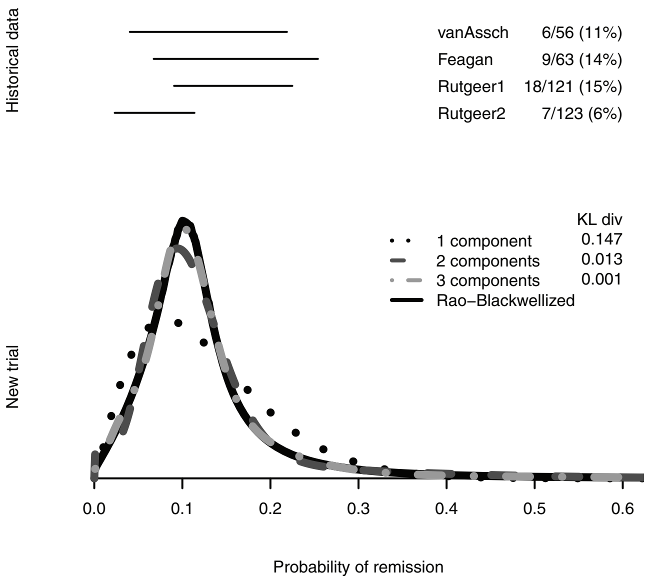

Figure 2. Observed placebo remission rates from four historical ulcerative colitis trials with 95% intervals, MAP prior for the rate in a new trial (Rao-Blackwellized density estimate), and one-, two-, and three-component Beta mixtures with corresponding Kullback–Leibler divergence (KL div).

mixture priors (ii) and (iii) offer gains compared to the uniform prior with $n = 40$ control patients, if the true response rate $\psi_\star$ is in the range of prior support (roughly 0.10–0.35). These gains are achieved although the designs with informative priors (i)–(iii) include considerably less than 40 control patients (see Table 2). The gains compared to the uniform prior with $n = 20$ control patients are even more impressive, if prior and data are consistent. In case of prior-data conflict ($\psi_\star > 0.4$), the negative impact on rMSE for prior (i) is clearly much stronger compared to the robust versions (ii) and (iii). For very clear prior-data conflict ($\psi_\star > 0.8$), the rMSE of the robust priors (ii) and (iii) is slightly higher than for the uniform prior with $n = 40$ control patients, but the corresponding adaptive designs also use slightly less control patients. Similar results can be seen for bias, with a much better behavior for the mixture priors compared to prior (i).

In summary, the example suggests that an adaptive two-stage design with mixture priors containing a weakly-informative component has good robustness properties. In a specific application a decision about what is acceptable with regard to frequentist metrics is needed. This includes careful judgment about how likely scenarios of prior-data conflict are.

4. Clinical Trial Example

To illustrate the methodology introduced in Section 2, we consider a proof-of-concept study in ulcerative colitis (Neuenschwander et al., 2010). The primary outcome, remission after 8 weeks of treatment, is binary. Four relevant historical placebo-controlled trials with a total of 363 placebo patients were identified, with remission rates in placebo ranging from 5.7% to 14.9% (Figure 2).

Based on the historical placebo data, the MAP prior for the remission rate in a new trial was derived, using a weakly informative Half-Normal prior with a standard deviation of 1 for the between-trial standard deviation τ, which puts approximately 5% probability for values of τ greater than 2. A value of $\tau = 2$ corresponds to very large between-trial variability on the log-odds-scale, and would essentially lead to no borrowing from the historical data (Spiegelhalter et al., 2004). Sensitivity analyses are provided in Web Appendix C. Figure 2 shows the MAP prior (Rao-Blackwellized density estimate, Web Appendix A) from an MCMC sample of size $M = 100,000$, obtained with WinBUGS (Lunn et al., 2000).

The MAP prior is now approximated by a mixture of conjugate priors as in (4). Fitting a single Beta density to this sample by ML does obviously not lead to a satisfactory approximation ($a_1 = 2.3$ and $b_1 = 16.0$). A two component mixture of Beta densities considerably improves the fit ($a_1 = 6.2, b_1 = 50.8$; $a_2 = 1.0, b_2 = 4.7$; $w_1 = 0.77$). The mixture of three Beta densities

$$\widehat{p}_H(\psi_\star) = 0.53 \times \mathrm{Beta}(2.5, 19.1) + 0.38 \times \mathrm{Beta}(14.6, 120.2) + 0.08 \times \mathrm{Beta}(0.9, 2.8) \qquad (9)$$

fits the density estimate almost perfectly. The Kullback–Leibler divergence (5) quickly drops towards zero with increasing numbers of components (Figure 2).

The first component in (9) is close to the ML estimate of the single Beta fit and has a weight of more than fifty percent. The second component corresponds to a highly informative peak centered around the same value as the first component (prior mean of 11%). Finally, the third component of the mixture (mean 24%) is only worth about four patients and has a low prior weight of eight percent. Since this third component is not very informative, the prior $\widehat{p}_H(\psi_\star)$ is already fairly robust. If desired, further robustification can be achieved as described in equation (7) by adding a Beta(1, 1) component with, for example, weight $w_R = 0.1$ to the mixture (9), which leads to

$$\widehat{p}_{HR}(\psi_\star) = 0.48 \times \mathrm{Beta}(2.5, 19.1) + 0.34 \times \mathrm{Beta}(14.6, 120.2) + 0.07 \times \mathrm{Beta}(0.9, 2.8) + 0.1 \times \mathrm{Beta}(1, 1). \qquad (10)$$

We consider now a new trial with $n_\star = 20$ placebo patients. From Section 2.4, the ESS for the two component mixtures without and with robustification are 47 and 37, respectively. The ESS of the three component mixture (9) is 81, and 63 for its robust version (10).

Figure 3 shows the posterior mean and posterior standard deviation of the two and three component mixtures and their robustified versions, for all possible values of the observed response rate $y_\star/20$. For good agreement between new and historical data ($y_\star/20 \leq 4/20$, say), the posterior means and standard deviations are very similar. For the range of potential prior-data conflict (5/20 to 10/20), uncertainty is increased for the mixture priors. Finally, under clear prior-data conflict ($\geq$11/20), the informative parts in the mixture distributions are strongly discounted, and the posterior means are close to the observed rates. Figure 3 also shows the simple conjugate Beta(2.3, 16) prior for which posterior inference is clearly not robust; see also Web Appendix D.

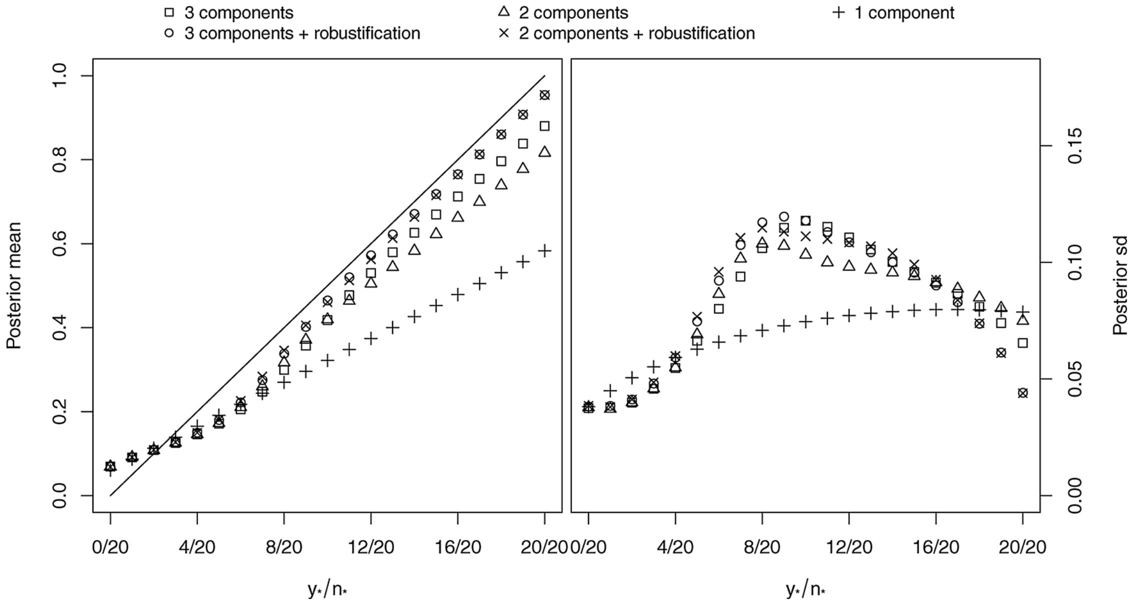

Figure 3. Posterior means and standard deviations (SD) versus observed placebo remission rates for all possible number of responders $y_\star$ with $n_\star = 20$ placebo patients, for different mixture priors approximating the MAP prior, and their robustifications.

Table 3 compares various characteristics of the three component mixture with and without robustification. Although the inference is similar for the two priors, the robustified version discounts the conflicting prior information more quickly.

Table 3 also shows the posterior ESS, which is very similar for the two priors. If prior and data are consistent, the posterior ESS is approximately the sum of prior ESS and new sample size $n_\star$. For scenarios of clear conflict, however, the posterior ESS decreases to $n_\star = 20$. Therefore, the actual posterior effective sample size depends on the outcome of the new trial in non-adaptive designs. This risk can be reduced with a two-stage adaptive design as discussed in Section 3. For example if the results shown here would correspond to the first stage, the patients allocated to control in a second stage could be increased if necessary.

The judgment whether a historical data prior and the new data are compatible can be based on the prior predictive tail area probabilities (Box, 1980) given by $\mathrm{P}(Y \geq y_\star) = \int \mathrm{P}(Y \geq y_\star|\psi_\star)p_H(\psi_\star)\mathrm{d}\psi_\star$. Table 3 shows one-sided tail area probabilities (lower or upper, whichever is smaller) calculated for $\widehat{p}_H$ and $\widehat{p}_{HR}$.

Table 3

Prior and posterior summaries for hypothetical trial outcomes $y_\star/n_\star$, for the three component mixture prior $\widehat{p}_H$, and its robustified version $\widehat{p}_{HR}$. The last column gives the prior predictive tail probability (see text).

	w_1	w_2	w_3	w_4	Mean	2.5%	97.5%	ESS	Prior pred. prob. (%)
Prior $\widehat{p}_H$	0.53	0.38	0.08		0.12	0.02	0.35	81	
Posterior for $y_\star/n_\star$									
0/20	0.62	0.30	0.08		0.07	0.01	0.15	78	14.9
2/20	0.50	0.46	0.04		0.11	0.04	0.20	110	59.6
5/20	0.59	0.31	0.11		0.17	0.08	0.33	74	13.7
10/20	0.25	0.01	0.74		0.42	0.20	0.64	14	1.5
15/20	0.004	0.00	0.996		0.67	0.47	0.84	24	0.3
Prior $\widehat{p}_{HR}$	0.48	0.34	0.07	0.10	0.16	0.02	0.76	63	
Posterior for $y_\star/n_\star$									
0/20	0.60	0.29	0.08	0.03	0.07	0.01	0.15	76	13.9
2/20	0.49	0.45	0.04	0.02	0.11	0.04	0.21	108	55.1
5/20	0.54	0.28	0.10	0.08	0.18	0.08	0.37	69	20.0
10/20	0.11	0.00	0.32	0.56	0.46	0.23	0.69	20	6.6
15/20	0.00	0.00	0.16	0.84	0.72	0.51	0.88	22	3.1

5. Discussion

We proposed a meta-analytic-predictive (MAP) approach to derive the prior for the control in the new trial from the historical control data. Heavy-tailed MAP priors imply a degree of robustness against prior-data conflicts which is absent in simple conjugate analyses. Further robustness and a more rapid adaptation to prior-data conflict can be obtained by adding a weakly informative component. This additional component acknowledges the possibility to be mistaken, and hence is in line with Cromwell's rule (Lindley, 2006): despite a careful selection of the historical trials, the new trial may have some unsuspected features which make it non-exchangeable to the historical trials.

To approximate the MAP prior, we have used mixtures of conjugate distributions. Alternative approaches to robust analyses could be considered, for example, the use of t-distributions.

We have used weakly informative priors for the vague component of the mixture prior. With such proper priors, the weight given to the vague component can be interpreted as a probability, which would not be the case for improper "flat" priors. Similar issues with the use of improper priors occur in Bayesian testing with Bayes factors (Kass and Wasserman, 1995).

We have assumed that the parameters for the historical trials are exchangeable. The model could be extended to allow for the possibility of conflict between the historical controls by using a mixture distribution with a weakly informative component as the random-effects distribution. However, if there is clear evidence of conflict between historical controls, then use of this historical information would make the new trial less credible, and hence may perhaps best be avoided. If there is conflict between historical controls and a normal random-effects distribution is nevertheless assumed, then the predictive distribution for the new trial will widen, and hence be less informative, so that the error is on the safe side. Alternatively, nonparametric Bayesian approaches could be considered to model partial exchangeability of trials (Leon-Novelo et al., 2012; Müller and Mitra, 2013).

If historical and concurrent control data are in clear conflict, the prior will essentially be discarded, if the MAP prior is robust. This may result in inconclusive trial results, as not enough control information may then be available. Adaptive designs allow to increase number of controls based on interim data, and hence reduce this risk (see also Hobbs et al., 2013).

We considered the case where several historical trials with control information are available. In this setting, a meta-analytic approach seems particularly appropriate, as the between-trial variability τ can then be assessed based on the data. For the case where only one historical study is available, the challenging problem of a Bayesian meta-analysis with only two studies (one historical, and one new) has to be addressed, in particular the choice of an appropriate informative prior distribution for τ (Gelman, 2006). A good judgment about plausible values of τ is important here, and may be obtained by elicitation from experts. Alternatively, a prior on τ may be derived by considering information from related patient populations or similar diseases (Higgins and Whitehead, 1996; Turner et al., 2012). Hobbs et al. (2011) suggested the use of a commensurate prior to specifically address the case of one historical trial. However, for this case, commensurate and MAP priors are the same (Hobbs et al., 2012). The power prior (Ibrahim and Chen, 2000) has also been used to include prior information from one historical trial. Typically the power parameter for the historical data is set to a fixed value, which approximately corresponds to fixing of the between-trial variability τ in the meta-analytic model (Chen and Ibrahim, 2006). Although inference on the power parameter is possible, a proper normalization is required (Neuenschwander, Branson, and Spiegelhalter, 2009).

In this article, we described our approach for the common situation where historical information on the control treatment is available, but not on the test treatment. If historical data for the latter are available, a robust MAP prior can then also be used. In more complex settings, meta-regression (Witte et al., 2011) or network meta-analytic models (Schmidli et al., 2013) allow to derive prior information for both control and test treatment.

We focused our discussion on endpoints from the one-parameter exponential family. Extensions to other distributions including nuisance parameters are possible, although more involved (see Web Appendix E). Nevertheless the key element of our proposal can be applied quite generally, namely the addition of a weakly informative component to the historical prior, inspired by De Groot who always carried an epsilon of probability for surprise in his pocket (Parmigiani and Inoue, 2009).

6. Supplementary Materials

Web Appendices A–E referenced in Sections 2–5 and code with example data are available with this paper at the *Biometrics* website on Wiley Online Library.

Acknowledgements

We would like to thank the editor Jeremy M. G. Taylor, an associate editor, and two referees for their helpful and constructive comments.

References

Baeten, D., Baraliakos, X., Braun, J., Sieper, J., Emery, P., van der Heijde, D., McInnes, I., van Laar, J., Landewe, R., Wordsworth, P., Wollenhaupt, J., Kellner, H., Paramarta, J., Wei, J., Brachat, A., Bek, S., Laurent, D., Li, Y., Wang, Y., Bertolino, A., Gsteiger, S., Wright, A. M., and Hueber, W. (2013). Anti-interleukin-17A monoclonal antibody secukinumab in ankylosing spondylitis: A randomized, double-blind, placebo-controlled trial. *The Lancet* **382**, 1705–1713.

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory.* Chichester: Wiley.

Berry, D. A. (1989). Comment: Ethics and ECMO. *Statistical Science* **4**, 306–310.

Berry, D. A. (2006). Bayesian clinical trials. *Nature Review Drug Discovery* **5**, 27–36.

Berry, S. M., Carlin, B. P., Lee, J. J., and Müller, P. (2010). *Bayesian Adaptive Methods for Clinical Trials.* Boca Raton: Chapman and Hall.

Box, G. E. P. (1980). Sampling and Bayes inference in scientific modeling and robustness. *Journal of the Royal Statistical Society, Series A* **143**, 383–430.

Chen, M. H. and Ibrahim, J. G. (2006). The relationship between the power prior and hierarchical models. *Bayesian Analysis* **1**, 551–574.

Dallal, S. and Hall, W. (1983). Approximating priors by mixtures of natural conjugate priors. *Journal of the Royal Statistical Society, Series B* **45**, 278–286.

Dawid, A. P. (1973). Posterior expectations for large observations. *Biometrika* **60**, 664–667.

Dempster, A., Selwyn, M., and Weeks, B. (1983). Combining historical and randomized controls for assessing trends in proportions. *Journal of the American Statistical Association* **78**, 221–227.

Diaconis, P. and Ylvisaker, D. (1985). Quantifying prior opinion. In: *Bayesian Statistics 2*, J. M. Bernardo, M. H. DeGroot, D. V. Lindley, A. F. M. Smith (eds), 133–156. Netherlands: Elsevier.

DiScala, L., Kerman, J., and Neuenschwander, B. (2013). Collection, synthesis, and interpretation of evidence: A proof-of-concept study in COPD. *Statistics in Medicine* **32**, 1621–1634.

FDA (2010a). Guidance for the use of Bayesian statistics in medical device clinical trials. www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071072.htm (accessed Jan 2014).

FDA (2010b). Non-inferiority clinical trials: Guidance for industry (draft). www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM202140.pdf (accessed Jan 2014).

FDA (2013). Expedited Programs for Serious Conditions—Drugs and Biologics (draft). www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM358301.pdf (accessed Jan 2014).

French, J. A., Temkin, N. R., Shneker, B. F., Hammer, A. E., Caldwell, P. T., and Messenheimer, J. A. (2012). Lamotrigine XR conversion to monotherapy: First study using a historical control group. *Neurotherapeutics* **9**, 176–184.

Fuquene, J. A., Cook, D., and Pericchi, L. R. (2009). A case for robust Bayesian priors with applications to clinical trials. *Bayesian Analysis* **4**, 817–846.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398–409.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1**, 515–533.

Gsteiger, S., Neuenschwander B., Mercier, F., and Schmidli, H. (2013). Using historical control information for the design and analysis of clinical trials with over-dispersed count data. *Statistics in Medicine* **32**, 3609–3622.

Higgins, J. P. and Whitehead, A. (1996). Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine* **15**, 2733–2749.

Hobbs, B. P., Carlin, B. P., Mandrekar, S. J., and Sargent, D. J. (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics* **67**, 1047–1056.

Hobbs, B. P., Carlin, B. P., and Sargent, D. J. (2013). Adaptive adjustment of the randomization ratio using historical control data. *Clinical Trials* **10**, 430–440.

Hobbs, B. P., Sargent, D. J., and Carlin, B. P. (2012). Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Analysis* **7**, 639–674.

Hueber, W., Sands, B. E., Lewitzky, S., Vandemeulebroecke, M., Reinisch, W., Higgins, P. D., Wehkamp, J., Feagan, B. G., Yao, M. D., Karczewski, M., Karczewski, J., Pezous, N., Bek, S., Bruin, G., Mellgard, B., Berger, C., Londei, M., Bertolino, A. P., Tougas, G., and Travis, S.P. (2012). Secukinumab, a human anti-IL-17A monoclonal antibody, for moderate to severe Crohn's disease: Unexpected results of a randomised, double-blind placebo-controlled trial. *Gut* **61**, 1693–1700.

ICH (2000). E10: Choice of control group in clinical trials. www.ich.org/LOB/media/MEDIA486.pdf (accessed Jan 2014).

Ibrahim, J. G. and Chen, M. H. (2000). Power prior distributions for regression models. *Statistical Science* **15**, 46–60.

Kass, R. E. and Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association* **90**, 928–934.

Leon-Novelo, L. G., Bekele, B. N., Müller, P., Quintana, F., and Wathen, K. (2012). Borrowing strength with nonexchangeable priors over subpopulations. *Biometrics* **68**, 550–558.

Lindley, D. V. (2006). *Understanding Uncertainty*. Chichester: Wiley.

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS—A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing* **10**, 325–337.

Morita, S., Thall, P. F., and Müller, P. (2008). Determining the effective sample size of a parametric prior. *Biometrics* **64**, 595–602.

Morita, S., Thall, P. F., and Müller, P. (2012). Prior effective sample size in conditionally independent hierarchical models. *Bayesian Analysis* **7**, 591–614.

Müller, P. and Mitra, R. (2013). Bayesian nonparametric inference—Why and how. *Bayesian Analysis* **8**, 269–302.

Neuenschwander, B., Branson, M., and Spiegelhalter, D. J. (2009). A note on the power prior. *Statistics in Medicine* **28**, 3562–3566.

Neuenschwander, B., Capkun-Niggli, G., Branson, M., and Spiegelhalter, D. J. (2010). Summarizing historical information on controls in clinical trials. *Clinical Trials* **7**, 5–18.

O'Hagan, A. (1979). On outlier rejection phenomena in Bayes inference. *Journal of the Royal Statistical Society, Series B* **41**, 358–367.

O'Hagan, A. and Forster, J. (2004). *Bayesian Inference, Kendall's Advanced Theory of Statistics*, Volume 2B. Chichester: Wiley.

O'Hagan, A. and Pericchi, L. (2012). Bayesian heavy-tailed models and conflict resolution: A review. *Brazilian Journal of Probability and Statistics* **26**, 372–401.

Parmigiani, G. and Inoue, L. (2009). *Decision Theory: Principles and Approaches*. Chichester: Wiley.

Pocock, S. J. (1976). The combination of randomized and historical controls in clinical trials. *Journal of Chronic Diseases* **29**, 175–188.

Polson, N. G. and Scott, J. G. (2012). On the Half-Cauchy prior for a global scale parameter. *Bayesian Analysis* **7**, 887–902.

Ryan, L. (1993). Using historical controls in the analysis of developmental toxicity data. *Biometrics* **49**, 1126–1135.

Schmidli, H., Bretz, F., and Racine-Poon, A. (2007). Bayesian predictive power for interim adaptation in seamless phase II/III trials where the endpoint is survival up to some specified timepoint. *Statistics in Medicine* **26**, 4925–4938.

Schmidli, H., Wandel, S., and Neuenschwander, B. (2013). The network meta-analytic-predictive approach to non-inferiority trials. *Statistical Methods in Medical Research* **22**, 219–240.

Simon, R. (1989). Optimal two-stage designs for phase II clinical trials. *Controlled Clinical Trials* **10**, 1–10.

Spiegelhalter, D. J., Abrams, K. R., and Myles J. P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Chichester: Wiley.

Trippa, L., Rosner, G. L., and Müller, P. (2012). Bayesian enrichment strategies for randomized discontinuation trials (with discussion). *Biometrics* **68**, 203–225.

Turner, R. M., Davey, J., Clarke, M. J., Thompson, S. G., and Higgins, J.P. (2012). Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. *International Journal of Epidemiology* **41**, 818–827.

Witte, S., Schmidli, H., O'Hagan, A. and Racine, A. (2011). Designing a non-inferiority study in kidney transplantation: A case study. *Pharmaceutical Statistics* **10**, 427–432.

Received July 2013. *Revised June* 2014. *Accepted July* 2014.

Appendix

It is shown that the meta-analytic-predictive (MAP) and the meta-analytic-combined (MAC) approaches are equivalent for the model specified by (1) in Section 2.1. Equivalence of MAC and MAP follows from

$$\begin{aligned} p(\psi_\star | Y_\star, Y_\mathcal{H}) &\propto p(\psi_\star, \psi_\mathcal{H} | Y_\star, Y_\mathcal{H}) \\ &\propto p(Y_\star, Y_\mathcal{H} | \psi_\star, \psi_\mathcal{H}) \times p(\psi_\star, \psi_\mathcal{H}) \\ &= p(Y_\star | \psi_\star) \times p(Y_\mathcal{H} | \psi_\mathcal{H}) \times p(\psi_\star, \psi_\mathcal{H}) \\ &\propto p(Y_\star | \psi_\star) \times p(\psi_\star, \psi_\mathcal{H} | Y_\mathcal{H}) \\ &\propto p(Y_\star | \psi_\star) \times p(\psi_\star | Y_\mathcal{H}), \end{aligned}$$

where $\propto$ denotes proportionality with regard to $\psi_\star$.