Applying Meta-Analytic-Predictive Priors with the R Bayesian Evidence Synthesis Tools

Sebastian Weber
Novartis Pharma AG

Yue Li
Novartis Pharma AG

John W. Seaman III
Novartis Pharma AG

Tomoyuki Kakizume
Novartis Pharma K.K.

Heinz Schmidli [iD]
Novartis Pharma AG

Abstract

Use of historical data in clinical trial design and analysis has shown various advantages such as reduction of number of subjects and increase of study power. The meta-analytic-predictive (MAP) approach accounts with a hierarchical model for between-trial heterogeneity in order to derive an informative prior from historical data. In this paper, we introduce the package **RBesT** (R Bayesian evidence synthesis tools) which implements the MAP approach with normal (known sampling standard deviation), binomial and Poisson endpoints. The hierarchical MAP model is evaluated by Markov chain Monte Carlo (MCMC). The MCMC samples representing the MAP prior are approximated with parametric mixture densities which are obtained with the expectation maximization algorithm. The parametric mixture density representation facilitates easy communication of the MAP prior and enables fast and accurate analytical procedures to evaluate properties of trial designs with informative MAP priors. The paper first introduces the framework of robust Bayesian evidence synthesis in this setting and then explains how **RBesT** facilitates the derivation and evaluation of an informative MAP prior from historical control data. In addition we describe how the meta-analytic framework relates to further applications including probability of success calculations.

Keywords: Bayesian inference, clinical trial, extrapolation, historical control, operating characteristics, prior, probability of success, robust analysis.

1. Introduction

More efficient clinical trials are of great demand in drug development for all stakeholders such as pharmaceutical companies, regulatory agencies, health-care organizations and, most

importantly, for patients. Use of historical data for quantitative trial design has become more and more attractive for the same reason (Wandel, Schmidli, and Neuenschwander 2015; Neuenschwander and Schmidli 2020). Using historical data can reduce the size of the control group, leading to a smaller size of clinical trials which are more ethical and shorten study duration. Therefore, studies utilizing historical data may speed up informative decision making and eventually make better medicines available to patients sooner.

Borrowing information from historical studies has always been a part of the design of clinical trials. For example, the definition of a patient population in a new clinical study compared to previous similar studies, or how much of a clinically relevant treatment effect to expect compared to the control treatment. Contributions from statisticians in a more quantitative manner started more than 40 years ago, by Pocock (1976). Since then, the relevant statistical approaches have been developed by many, mostly in a Bayesian framework (Chen and Ibrahim 2000; Spiegelhalter, Abrams, and Myles 2004; Neuenschwander, Capkun-Niggli, Branson, and Spiegelhalter 2010; Hobbs, Sargent, and Carlin 2012).

In this paper, we focus on robust meta-analytic-predictive (MAP) priors (Spiegelhalter *et al.* 2004; Neuenschwander *et al.* 2010; Schmidli, Gsteiger, Roychoudhury, O'Hagan, Spiegelhalter, and Neuenschwander 2014), which summarize the historical information as an informative prior to be used in a Bayesian analysis for the study of interest. The MAP prior is derived using a random effects meta-analysis model, also referred to as hierarchical model. The meta-analysis accounts for between-trial heterogeneity and leads to a discounting of the historical information. As in any meta-analytic approach, it is important to first examine the characteristics of the historical trials with clinical inputs. These include quantitative descriptions of trial population such as subject demographics, baseline characteristics and qualitative features such as concomitant medications. This helps to ensure that selected historical controls are comparable to those in the new trial, such that one of the key assumptions of exchangeability between the trials holds. Secondly, the assumption for the between-trial heterogeneity needs to be reasonable as the number of trials is commonly small to borrow from and considerable uncertainty needs to be accounted for adequately in the analysis. Then the MAP prior can be derived from a random-effects meta-analysis of historical data via Markov chain Monte Carlo (MCMC) algorithms as the analysis is commonly not tractable analytically. The predictive distribution of the effect parameter for a new study then forms the informative prior for the study of interest. In many cases, the goal is to replace some sample size of the planned control group with information from the MAP prior. The effective sample size of the MAP prior can then be used as a convenient measure of informativeness for the potential sample size savings of the actual trial. However, since in practice we can never rule out the unexpected, a robustification of the MAP prior (Schmidli *et al.* 2014) is advisable in order to ensure sensible inference whenever actual trial data and MAP prior strongly deviate from one another. Application of the MAP approach to incorporate historical data in early phase clinical trials has been more widely accepted, not only by statisticians but also in medical societies (Baeten, Baraliakos, Braun, Sieper, Emery, Van Der Heijde, McInnes, Van Laar, Landewé, Wordsworth, Wollenhaupt, Kellner, Paramarta, Wei, Brachat, Bek, Laurent, Li, Wang, Bertolino, Gsteiger, Wright, and Hueber 2013).

While there exists a number of R (R Core Team 2021) packages which perform meta-analysis, these target the conduct of the meta-analysis for the purpose of synthesizing the available historical information alone. However, these lack the aspect of using the combined historical information for the purpose of a trial design and analysis. For example, **netmeta** implements a

frequentist network meta-analysis approach (Rücker 2012; Rücker, Krahn, König, Efthimiou, and Schwarzer 2021), wheras **bayesmeta** conducts a Bayesian random-effects meta-analysis (Röver 2020), **metafor** a Bayesian mixed-effects meta-analysis (including meta-regression) (Viechtbauer 2010), and **MetaStan** performs a meta-analysis with random-effects for primarily binary endpoints using a Bayesian approach (Günhan, Röver, and Friede 2020; Günhan 2020). There are also R packages developed for network meta-analysis, which seeks to combine the historical information for multiple treatments, i.e., **gemtc** uses **JAGS** (Plummer 2003) for arm-based network meta-analysis (Van Valkenhoef and Kuiper 2021), **pcnetmeta** is based on contrast-based network meta-analysis (Lin, Zhang, Hodges, and Chu 2017) and **nmaINLA** employs the integrated nested Laplace approximations (Günhan, Friede, and Held 2018).

The R package **RBesT** for R Bayesian evidence synthesis tools (Weber 2021) is available from the Comprehensive R Archive Network (CRAN) at `https://CRAN.R-project.org/package=RBesT`. In this package, we implement the MAP approach tailored for the design and analysis of clinical trials as conducted during (usually early phases of) drug development. The focus on trial design leads to a number of considerations specific to this application. For example, the limited sample sizes involved may not warrant that the often used normal approximation holds well enough such that in addition to the common normal endpoint a binomial and a Poisson endpoint are supported by **RBesT**. As the package facilitates trial design and clinical trial protocol writing, a key step, which makes **RBesT** unique, is that the derived MAP prior can be easily documented in parametric form and, most importantly, the implications of actually using an informative prior for the trial design can be readily evaluated. Thus, in **RBesT** the meta-analysis is the starting point and large emphasis is on facilitating trial design and supporting compilation of the statistical analysis plan which pre-specifies the (primary) trial analysis.

In order to document and easily pre-specify an informative MAP prior for a trial, **RBesT** offers functionality to approximate the posterior MCMC sample representing the MAP prior with a parametric mixture distribution. The parametric mixture distribution is arbitrarily exact and is estimated using the expectation maximization (EM) algorithm. Representing the MAP prior in parametric form is key to design clinical trials. For one, the exact statistical analysis of the trial results are pre-specified ahead of time of trial analysis and, for two, the parametric representation allows seamless evaluation of the design in question. For each endpoint specific conjugate mixture densities are estimated. This allows the use of accurate and fast analytical results. **RBesT** supports trial design evaluation of one- or two-sample situations with flexible success criteria. In addition to standard frequentist operating characteristics, the probability of success or derivation of critical values is supported. All of these subsequent steps after MAP prior derivation operate in **RBesT** on the parametric mixture densities such that a large fraction of the package is devoted to evaluation of mixture densities.

The organization of this paper is as follows. We will first give an overview of the package key functionality and describe the details of one motivating example from a real clinical study in Section 2. In Section 3, the theoretical background of the MAP approach and the robustification of MAP will be explained, followed by how these approaches can be applied with **RBesT** to the example in Section 4. Finally we close the paper with summary and discussions in Section 5.

2. Historical control in clinical trials

One of the main use-cases of **RBesT** is the use of historical information in the analysis of clinical trials. The goal is to reduce the trial (usually the control group) sample size while maintaining its target power under the assumed true effect size of an alternative hypothesis. The **RBesT** package facilitates the (i) prior derivation using MCMC, (ii) parametric (mixture) approximation of the MAP prior and finally (iii) evaluation of the clinical trial design.

Prior derivation. Once relevant data for the prior derivation have been compiled, the first step is to use the `gMAP` function which performs a random-effects meta-analysis using MCMC implemented by **rstan** (Stan Development Team 2020). The syntax follows standard R modeling conventions and the returned analysis object can be modified for sensitivity analyses with the `update` command or queried with the generic modeling functions. In particular, the `plot` function creates graphical model diagnostics like a forest plot augmented with model estimates. As the forest plot is a key summary of the data making up the prior a dedicated `forest_plot` function is available which allows for various customization of the forest plot.

Parametric mixture approximation. After running the MCMC analysis with `gMAP`, the next step is to find an accurate parametric distributional representation. The package uses the EM algorithm implemented by the `mixfit` function which can be applied to 'gMAP' analysis objects or MCMC samples directly. The EM procedure requires specification of the number of mixture components to be used. To ease this step even further, the package also provides an `automixfit` function which selects the number of mixture components automatically based on information criteria measures. The resulting parametric mixture distribution objects generated from `mixfit` also support the `plot` method which produces a graphical check of the fitted mixture distribution by a comparison to a histogram of the MCMC sample.

Trial design evaluation. It is first recommended to robustify the parametric MAP prior with the `robustify` function. This function adds a weakly-informative prior component to the mixture derived from the previous steps. The so-obtained robust MAP prior is intended for use in a pre-specified trial analysis. Its informativeness can be assessed with the effective sample size with the `ess` command. At the design stage of the trial the frequentist operating characteristics or probability of success of the trial are of interest in order to document the properties of the statistical analysis as planned in the statistical analysis plan of the trial. The trial evaluation is supported in **RBesT** for one- or two-sample designs. In a first step, the statistical success criterion of the trial is specified with the `decision1S` (`decision2S`) function which return objects representing the decision function. These decision functions are then required by the functions `oc1S` (`oc2S`), `pos1S` (`pos2S`) and `decision1S_boundary` (`decision2S_boundary`) to specify the trial design (defined by prior, sample size and decision function) for the purpose of evaluating the frequentist operating characteristics, the probability of success or the critical decision boundaries of the trial design, respectively.

Ankylosing spondylitis example. As an example we will use a study in ankylosing spondylitis comparing the test treatment secukinumab with placebo (Baeten *et al.* 2013). The primary efficacy endpoint was a binary responder analysis for the percentage of patients with a 20% response according to the Assessment of SpondyloArthritis international Society

Study	1	2	3	4	5	6	7	8	Total
Patients	107	44	51	39	139	20	78	35	513
Responders	23	12	19	9	39	6	9	10	127

Table 1: Historical data of placebo response rates used in the example ankylosing spondylitis study. The data set is available in the `AS` data frame as part of the **RBesT** package.

criteria for improvement (ASAS20) at week 6. Eight historical trials, totaling 513 patients as shown in Table 1, were used to derive the MAP prior for the control arm.

This trial was conducted using the MAP approach (before the availability of **RBesT**). The trial used a $\text{Beta}(11, 32)$ prior for the control arm and performed a $4 : 1$ randomization ratio of active vs. control patients. The final trial compared $n_1 = 24$ patients on active treatment out of which $r_1 = 15$ responded vs. $n_2 = 6$ patients on control out of which $r_2 = 1$ responded and declared success based on meeting the success criterion defined as requiring that the response rate in the active group, θ_1, exceeds the response rate in the control group, θ_2, with a probability of at least 95% such that $\mathsf{P}(\delta = \theta_1 - \theta_2 > 0|y) > 0.95$ holds. This example is discussed with greater focus on the statistical aspects in the vignette "Getting started with **RBesT** (binary)" included in package **RBesT**.

3. Bayesian evidence synthesis and prediction

Important decisions should arguably be evidence based, especially in medicine (Eddy 1990; Wandel and Roychoudhury 2016). For example, decisions regarding design and analysis of clinical trials are important for trial sponsors, patients, physicians and policy makers. To support such decisions, relevant sources of information should be collected, appropriately synthesized through meta-analytic approaches, and used to make predictions on the planned target trial. We use the term meta-analytic-predictive (MAP) approach (Neuenschwander *et al.* 2010) to denote the synthesis of evidence from various sources, and the prediction/extrapolation to the target. Although the MAP approach is useful in a broad range of applications, we consider here the medical setting (European Medicines Agency 2013).

Methodology for Bayesian evidence synthesis and prediction is well developed. Textbooks on this topic include Stangl and Berry (2000), Spiegelhalter *et al.* (2004), Welton, Sutton, Cooper, Abrams, and Ades (2012), and Dias, Ades, Welton, Jansen, and Sutton (2018). Robust hierarchical models play a key role here, as explained in the following Sections 3.1 and 3.2. Applications of the MAP approach are very diverse, and we briefly discuss some settings in Section 3.3.

3.1. Meta-analytic-predictive methodology

Figure 1 schematically depicts the MAP approach for evidence synthesis and prediction. Suppose that a sponsor plans a new clinical trial (the target, labeled by the star symbol). This trial will generate data $Y_\star$, to be described by a statistical model $p(Y_\star|\theta_\star)$, with parameters $\theta_\star$. Usually, several relevant sources of information will be available, e.g., clinical trials in the same or similar patient population, and with (partly) the same treatments. Each source of information consists of data Y_j, modeled by $p(Y_j|\theta_j)$, with corresponding parameters θ_j, $j = 1, \ldots, J$. To borrow strength from the source information, a model is required that links
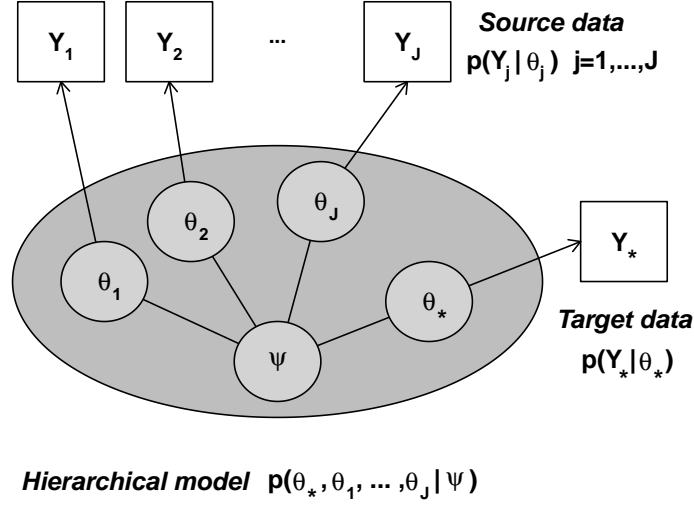
Figure 1: MAP approach to evidence synthesis and prediction.

the parameters from both source and target: $p(\theta_\star, \theta_1, \ldots, \theta_J | \Psi)$, with hyper-parameters Ψ. Such hierarchical models are very natural and convenient for the synthesis of the evidence and the prediction to the target (Spiegelhalter *et al.* 2004). Within the Bayesian framework, a prior for the hyper-parameters $p(\Psi)$ is needed, which will be specific to the considered setting.

At the planning stage of the target trial, the data $Y_\star$ are not available. Hence the posterior distribution of the parameters $p(\theta_\star, \theta_1, \ldots, \theta_J, \Psi | Y_1, \ldots, Y_J)$ is based on the source data only. The marginal posterior for the target parameter $p_{\mathrm{MAP}}(\theta_\star) = p(\theta_\star | Y_1, \ldots, Y_J)$ is the prior information for the target, called the MAP prior in the following.

Once the target data $Y_\star$ are available, the posterior for $\theta_\star$ is $p(\theta_\star | Y_\star) \propto p(Y_\star | \theta_\star) p_{\mathrm{MAP}}(\theta_\star)$. Exactly the same posterior could also be obtained through a meta-analytic-combined (MAC) approach, from $p(\theta_\star, \theta_1, \ldots, \theta_J, \Psi | Y_\star, Y_1, \ldots, Y_J)$ (Schmidli *et al.* 2014).

An analytical derivation of the MAP prior is typically not possible, and hence MCMC methods have to be used (Gelman, Stern, Carlin, Dunson, Vehtari, and Rubin 2013). These generate a large sample from the posterior distribution of the parameters, including $\theta_\star$. However, an approximate analytical description of the MAP prior facilitates communication and use with standard software. Mixtures of normal distributions generally provide such an analytical approximation (West 1993). When conjugate priors exist, mixtures of conjugate priors may be used (Dallal and Hall 1983; Diaconis and Ylvisaker 1985). These are preferable, as they allow analytical posterior calculations for the target trial in simple settings (O'Hagan and Forster 2004; Schmidli *et al.* 2014).

3.2. Robustness to prior-data conflict

The MAP approach provides the prior for the target parameters as $p_{\mathrm{MAP}}(\theta_\star)$. Occasionally, this MAP prior may turn out to be in conflict with the emerging target data $Y_\star$, despite great care in the selection of relevant sources and in the specification of the model. The behavior of the posterior distribution to prior-data conflicts is governed by the tails of the prior (O'Hagan and Pericchi 2012). For example, conjugate priors are not heavy-tailed, and consequently the posterior will always be a compromise between prior and data. However, MAP priors are

typically heavy-tailed, and are essentially discarded in case of prior-data conflict. This is a desirable feature in most settings.

Although the MAP prior should be robust to prior-data conflict by accounting for heterogeneity, a faster reaction to prior-data conflicts may be achieved by adding a weakly-informative mixture component $p_V(\theta_\star)$ (Schmidli *et al.* 2014; Neuenschwander, Wandel, Roychoudhury, and Bailey 2016). The robustified MAP prior is:

$$p_{\mathrm{MAPr}}(\theta_\star) = (1 - w)\, p_{\mathrm{MAP}}(\theta_\star) + w\, p_V(\theta_\star), \tag{1}$$

where w may be interpreted as the prior probability that the source information is not relevant for the target, expressing some degree of skepticism towards borrowing strength. The choice of the weakly-informative prior $p_V(\theta_\star)$ varies by endpoint and application. For binary endpoints a Jeffrey's prior, a neutral prior (Kerman 2011) or a uniform prior on the response scale can be used, for example. In the case of other endpoints, like a normal, a unit information prior (Kass and Wasserman 1995) may be suitable.

3.3. Applications

Methodology and diverse applications of the MAP approach in medicine are reviewed in Wandel *et al.* (2015), Schmidli, Gsteiger, and Neuenschwander (2021), and Neuenschwander and Schmidli (2020). Specific applications include comparison of several treatments through a network MAP approach (Schmidli, Wandel, and Neuenschwander 2013), the design and analysis of non-inferiority and biosimilar clinical trials (Gamalo-Siebers *et al.* 2016; Mielke, Schmidli, and Jones 2018), and the use of external data in adaptive clinical trials (Gsponer, Gerber, Bornkamp, Ohlssen, Vandemeulebroecke, and Schmidli 2014; Mütze, Schmidli, and Friede 2018). In the following, four common applications are briefly described.

Random-effects meta-analysis. Random-effects meta-analyses of clinical trials are very common in medicine (Higgins and Green 2008). These typically synthesize the evidence on the comparative effectiveness of two interventions in patients with a specific disease. Higgins, Thompson, and Spiegelhalter (2009) emphasize that both the overall effect size and the prediction for the true effect in a new trial are important for decision makers.

Sources of information are often J randomized clinical trials comparing a test and a control treatment, with a continuous clinical endpoint. Data available from the jth source trial are taken from publications, and are usually the estimated effect Y_j with standard error se_j (taken as exactly known). These data are modeled as $Y_j \sim \mathrm{Normal}(\theta_j, se_j^2)$. The parameters are linked through a model: $\theta_\star, \theta_1, \ldots, \theta_J \sim \mathrm{Normal}(\mu, \tau^2)$, with overall effect size μ and between-trial standard deviation τ. The parameter $\theta_\star$ denotes the true effect in a new trial. Priors for the hyper-parameters $\Psi = (\mu, \tau)$ are, e.g., a weakly informative Normal prior for μ and a Half-Normal prior for τ. In case with few trials (i.e., $J < 5$), an appropriate choice of the prior for τ is crucial (Gelman 2006; Friede, Röver, Wandel, and Neuenschwander 2017,B). After having specified the priors, **RBesT** may be used to obtain a sample from the posterior distribution of the parameters, and to graphically and numerically summarize these.

Evaluation of probability of success. Clinical trials aiming to show the superiority of a test treatment over a control treatment are often analyzed using a frequentist approach.

The trial is considered a success, if a statistically significant treatment effect is observed, with one-sided significance level $\alpha = 0.025$. The sample size of the trial is chosen such that a power of, e.g., 80% is achieved, conditional on a specific treatment effect $\theta_\star$. However, the power does not provide the unconditional probability of success (or assurance), as it ignores the uncertainty on the treatment effect. If relevant source data on the treatment effect are available, the uncertainty on $\theta_\star$ is captured by the MAP prior $p_{\mathrm{MAP}}(\theta_\star)$.

The probability of success (PoS) is the prior expectation of the power, averaged over the MAP prior (O'Hagan, Stevens, and Campbell 2005).

$$\mathrm{PoS} = \int \mathrm{CP}(\theta_\star)\, p_{\mathrm{MAP}}(\theta_\star)\, d\theta_\star \;, \tag{2}$$

where $\mathrm{CP}(\theta_\star)$ is the conditional power function, i.e., the probability of success conditional on an assumed true treatment effect. From a MCMC sample $\theta_\star^{(1)}, \ldots, \theta_\star^{(M)}$ of the MAP prior the PoS may be calculated as $1/M \sum_m \mathrm{CP}(\theta_\star^{(m)})$. Alternatively, the MAP prior may be approximated by a mixture of normal priors with **RBesT**, and PoS can be evaluated by numerical integration.

PoS evaluations are also relevant for decision makers at interim analyses of clinical trials. If the PoS (or predictive power) at interim is low, the trial may be stopped early to avoid unnecessary exposure of patients to ineffective treatments and to save resources (Spiegelhalter, Freedman, and Blackburn 1986; Schmidli, Bretz, and Racine-Poon 2007; Neuenschwander, Roychoudhury, and Schmidli 2016). For these interim analyses, the MAP prior is updated with the interim data.

Extrapolation. New treatments are typically first investigated in adult patients, before starting clinical trials in children. For many diseases and treatments, a similar effect may be expected for children and adults, using a possibly modified children version of the treatment (e.g., dosing based on body weight). Borrowing strength from the available adult trials should therefore always be considered (Food and Drug Administration (FDA) for Devices 2015; European Medicines Agency 2018). The MAP approach may be used for extrapolation (Wandel, Neuenschwander, Röver, and Friede 2017; Röver, Wandel, and Friede 2019; Schmidli, Häring, Thomas, Cassidy, Weber, and Bretz 2020), although alternative methods are also available (Gamalo-Siebers *et al.* 2017; Wadsworth, Hampson, and Jaki 2018).

The source data are usually J randomized clinical trials in adults comparing test and control treatment. These can be summarized with a random-effects meta-analysis as described above, which provides the MAP prior for the treatment effect in a new trial in adults $p_{\mathrm{MAP}}(\theta_\star)$. In some settings essentially the same treatment effect in adults and children may be expected, based on a scientific understanding of the disease and the mode-of-action of the treatment. Hence, the MAP prior for adults may also be used for a new trial in children. Skepticism on the relevance of the adult data may be expressed by robustifying the MAP prior (Section 3.2). In simple settings, **RBesT** can be used to derive the MAP prior, robustify it, evaluate frequentist operating characteristics of the trial in children, and finally obtain the posterior distribution of the treatment effect in children, once results from the children trial are available.

Historical controls. In many disease areas, multiple randomized controlled trials (RCTs) have been conducted, with the same control treatment (e.g., placebo) but different test treatments. When planning to investigate a new test treatment in an RCT, the question arose

whether one could borrow strength from the historical control data (Viele *et al.* 2014). In this setting, the sources of information are the control data from J trials. For a clinical endpoint, the control mean from the jth trial may be modeled as $Y_j \sim \text{Normal}(\theta_j, se_j^2)$, with true control mean θ_j, and standard error se_j (taken as exactly known). A model is used to provide the link to the true control mean in the new trial $\theta_\star$ as: $\theta_\star, \theta_1, \ldots, \theta_J \sim \text{Normal}(\mu, \tau^2)$. With appropriate priors for the hyper-parameters, the MAP prior $p_{\text{MAP}}(\theta_\star)$ is derived, and used as the informative prior for the control group in the new trial. Again, it is often advisable to robustify the MAP prior in case of some doubt on the relevance of the historical control information (Section 3.2). **RBesT** may be used for MAP prior derivation, evaluation of frequentist operating characteristics and the final analysis. An example data set is described in Section 2 and in Section 4 we present how **RBesT** facilitates the use of historical control information in clinical trials.

Use of the MAP approach in historical control settings has also been described for data modeled by the one-parameter exponential family (Schmidli *et al.* 2014), count data (Gsteiger, Neuenschwander, Mercier, and Schmidli 2013), recurrent event data (Holzhauer, Wang, and Schmidli 2018), time-to-event data (Holzhauer 2017) and variance data (Schmidli, Neuenschwander, and Friede 2017).

4. Application

In the following the use of **RBesT** is explained using as example the study in the indication ankylosing spondylitis as introduced in Section 2. The **RBesT** package facilitates the (i) prior derivation using MCMC, (ii) parametric (mixture) approximation of the MAP prior and finally (iii) evaluation of the clinical trial design.

4.1. Prior derivation

The statistical models implemented in the package follow the standard generalized linear regression modeling conventions and are implemented with the `gMAP` function mostly analogous as in the R `glm` command of the **stats** package. The supported sampling distributions are normal (with known sampling standard deviation σ), binomial and Poisson. These use the canonical link functions of the identity, logit and log link, respectively. The `gMAP` function call for the secukinumab trial is:

```
R> library("RBesT")
R> AS

    study   n   r
1 Study 1 107 23
2 Study 2  44 12
3 Study 3  51 19
4 Study 4  39  9
5 Study 5 139 39
6 Study 6  20  6
7 Study 7  78  9
8 Study 8  35 10
```

```
R> set.seed(35667)
R> map_mcmc <- gMAP(cbind(r, n-r) ~ 1 | study, family = binomial, data = AS,
+   tau.dist = "HalfNormal", tau.prior = 1, beta.prior = cbind(0, 2))
```

The first argument is the `formula` argument which specifies a two-column matrix as response for a binary response and contains in the first column the number of responders r and in the second column the number of non-responders $n - r$. The response is modeled by the terms on the right of the ~ operator. In this example a single intercept term is used here, but additional covariates can be specified using standard R formulae syntax. The last element of the formula is the grouping factor, which is separated by a vertical bar | from the model for the response. The grouping factor defines what constitutes a trial in the data set. If no grouping factor is specified, then each row in the input data set is interpreted as a group. The next argument is the `family` argument which specifies the sampling distribution. It is strongly recommended to use the `data` argument and pass a '`data.frame`' to it where all data for the model is stored (otherwise the environment will be searched for the respective columns). Finally, the priors of the model are specified. As the between-trial heterogeneity parameter τ is of particular importance for a meta-analysis, the `gMAP` function allows the user to choose the distributional class for the τ parameter with the `tau.dist` argument whereas the regression coefficients β are restricted to Normal priors. The arguments `tau.prior` and `beta.prior` both take a two column argument with the convention that each row corresponds to the respective parameter and the columns correspond to respective parameters of the prior distribution. Whenever a prior distribution for τ only needs a single parameter, a vector can be given as well, which is the case for the HalfNormal distribution used here with a standard deviation of 1. For the `beta.prior`, the first and second column correspond to the means and standard deviations of the normal prior distributions, respectively.

Internally the `gMAP` function uses **Stan** via the **rstan** package for sampling from the posterior distribution. The diagnostics of the MCMC sampler as provided by **Stan** are automatically inspected and potential issues are reported with a warning. The `gMAP` function returns an S3 object, which can then be further processed by standard R modeling functions. The default `print` method

```
R> map_mcmc

Generalized Meta Analytic Predictive Prior Analysis

Call:  gMAP(formula = cbind(r, n - r) ~ 1 | study, family = binomial,
    data = AS, tau.dist = "HalfNormal", tau.prior = 1, beta.prior = cbind(0,
        2))

Exchangeability tau strata: 1
Prediction tau stratum     : 1
Maximal Rhat               : 1

Between-trial heterogeneity of tau prediction stratum
  mean     sd   2.5%    50%  97.5%
0.3770 0.2120 0.0426 0.3510 0.8610
```
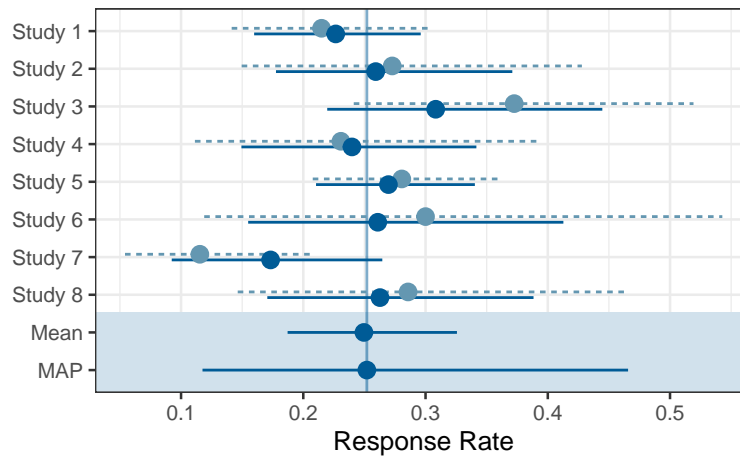
Figure 2: Recommended diagnostic plot for `gMAP` analyses which is a forest plot augmented with meta-analytic shrinkage estimates θ_j per trial. Shown are the per-trial point estimates (light dot) and the 95% frequentist confidence intervals (dashed line) and the model derived median (dark point) and the 95% credible interval of the meta-analytic model. In addition the model derived typical parameter estimate μ and the MAP estimate $\theta_\star$ is shown.

```
MAP Prior MCMC sample
  mean     sd   2.5%    50%  97.5%
0.2620 0.0869 0.1180 0.2520 0.4660
```

shows a short summary of the `gMAP` analysis. The derived MAP prior corresponds to the intercept-only model of the fitted statistical model (relevant when using covariates) and is given on the response scale (here: response rate) by default as opposed to the link scale (here: log-odds). More information of the model estimates is available with the `summary` method. Importantly, the `plot` method provides key graphical summaries of the `gMAP` analysis:

```
R> model_plots <- plot(map_mcmc, size = 0.5)
R> names(model_plots)

[1] "densityThetaStar"     "densityThetaStarLink" "forest_model"
```

The graphical summaries are returned as plain R list of **ggplot2** (Wickham 2016) plots, which can be plotted by printing a given element of the list. The plot in Figure 2 can be obtained with:

```
R> model_plots$forest_model
```

The density estimate plot on the response or link scale of the MAP prior shows each fitted chain (by default 4 chains are used) separately. The overlaid display by chain allows to assess graphically the MCMC convergence, since each chain must have sampled the same posterior resulting in numerically very similar densities. As key diagnostic plot we recommend to inspect the `forest_model` plot as shown in Figure 2. The plot gives a graphical summary of the MAP analysis and can serve as a fast check for coding errors.
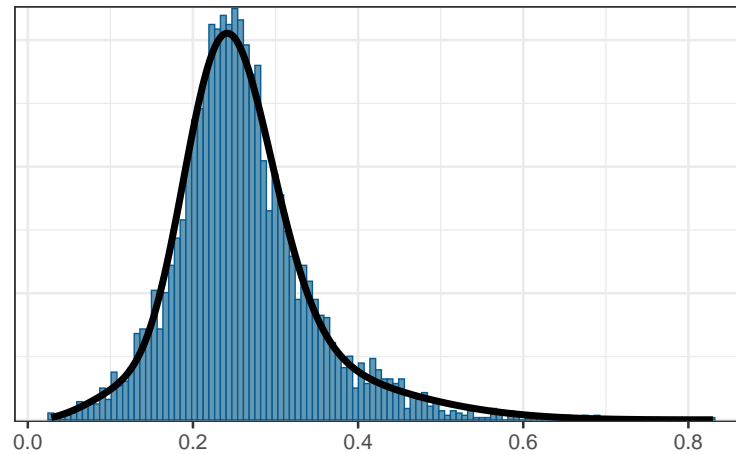
Figure 3: Recommended diagnostic plot for parametric approximation of the MCMC represented MAP prior. The plot overlays the MCMC histogram of the MAP prior as obtained from `gMAP` and the respective parametric mixture approximation as determined from `automixfit` or `mixfit`.

4.2. Parametric mixture prior derivation

The MAP prior, represented numerically with a large MCMC simulation sample, can be approximated with a parametric representation with the `automixfit` function. This function fits a parametric mixture representation using the EM algorithm. When calling this function with a 'gMAP' analysis object, the EM algorithm is run on the MAP prior MCMC values on the response scale (as opposed to the transformed link scale). The distributional class of the fitted mixture densities depends on the sampling distribution of the `family` argument. For the `gaussian`, `binomial` and `poisson` family a mixture of normal, beta and gamma distributions are fitted with the EM algorithm, respectively. These choices allow **RBesT** to take advantage of analytical results as their likelihood-prior combinations are conjugate to one another.

The EM algorithm requires a pre-specified number of components, which is automatically chosen by `automixfit` through the use of the Akaike information criterion (AIC). The function fits parametric mixture models of increasing complexity with up to four components and then selects the one with the lowest AIC. The output below shows the log-likelihood results for the selected mixture model, as well as the mixture components of the beta mixture. Mixtures are represented in **RBesT** with a set of K mixture components and each being defined by a triplet $(w, a, b)_k$ which corresponds to the weight w of the component, the first standard parameter a and the second standard parameter b of the respective density. Please refer to the overview in Table 2 for further details.

```
R> map <- automixfit(map_mcmc)
R> map


EM for Beta Mixture Model
Log-Likelihood = 4446.925


Univariate beta mixture
```

```
Mixture Components:
  comp1     comp2
w  0.641628  0.358372
a 17.472900  3.497193
b 52.774167  8.669930
```

To consider the quality of the EM fit we recommend to compare the MCMC sample with the parametric mixture density in a graphical manner. In Figure 3 the output of the `plot` method is shown for the generated `mix` plot, which overlays the fitted mixture density marginal with a histogram of the MCMC sample. In rare cases the response scale can be inadequate to compare the parametric mixture density appropriately (for example, if the response rate is very small or very large). To obtain Figure 3:

```
R> em_diagnostic <- plot(map)
R> em_diagnostic$mix + ggtitle(NULL)
```

Additional plots which are not shown can be obtained using:

```
R> em_diagnostic$mixdens
R> em_diagnostic_link <- plot(map, link = "logit")
R> em_diagnostic_link$mix
```

with the second plot resulting in the same plot as in Figure 3 except that the plot is on the logit scale.

Once the user has derived a parametric mixture representation for the MAP prior, the **RBesT** package provides additional functions to further investigate the fit as indicated by the overview in Table 2. In the following we discuss the key functions needed in the context of informative prior derivation from historical data.

As the goal is to reduce the required sample size in a future trial, the informative MAP prior enables unequal randomization of active vs. control. In the ankylosing spondylitis example a 4:1 randomization ratio was chosen. The `ess` function provides approximations to the effective sample size of a given prior for various methods. The effective sample size of the MAP prior gives a rough guide by how much the sample size can be reduced when using the respective frequentist power calculation as a reference, for example. For the ankylosing spondylitis example, the overall 513 historical control patients in 8 different trials constitute

```
R> ess(map)
```

```
[1] 35.53539
```

patients worth of information on the used control patients for a future trial. The considerable reduction in information content reflects the inherent heterogeneity between the trials.

Instead of using the MAP prior directly in a new study, we recommend to robustify the prior with a weakly-informative component (Equation 1 of Section 3.2) as follows:

```
R> map_robust <- robustify(map, weight = 0.2, mean = 0.5)
```

Function name	Use	Notes
c(w, a, b)	Defines a component of a mixture distribution. Specifies for each component its weight, the first and second standard parameter.	Normal densities use mean and standard deviation, beta densities use α and β, gamma densities use shape and rate.
mn2norm mn2beta mn2gamma	Maps to the standard parametrization given mean and number of observations.	
ms2beta ms2gamma	Maps to the standard parametrization given mean and standard deviation.	
mixnorm mixbeta mixgamma	Create supported base mixture distribution objects.	For normal densities the attribute sigma sets the known sampling standard deviation. For gamma densities a likelihood attribute sets the likelihood the density is intended to be used with.
mixcombine	Combines mixture distributions.	
(d/p/q/r)mix	Density/distribution/quantile/random number generation functions.	The quantile function uses numerical search routines.
robustify	Adds a weakly-informative component to a mixture.	
(d/p/q/r)mixdiff	Common distribution functions for the difference of two mixture distributions.	Uses convolution theorem with numerical integration while for normal mixtures analytically exact results are used.
ess	Calculates the effective sample size of a mixture distribution.	Methods implemented are "elir" (default), "moment" and "morita", please refer to Appendix B.
postmix	Calculates posterior mixture distribution.	Uses analytically exact results.
preddist	Returns the predictive mixture distribution.	
decision1S_boundary decision2S_boundary	Calculates critical decision boundary.	Please refer to Appendix C.
oc1S oc2S	Calculates conditional power.	Please refer to Appendix C.
pos1S pos2S	Calculates probability of success.	Please refer to Appendix C.
summary	Produces descriptive statistics of mixture.	
plot	Produces diagnostic plots.	

Table 2: Overview on functions in **RBesT** defined for parametric mixtures.

RBesT offers many functions which facilitate to explore the implications of an informative prior. For example, the predictive distribution of the data in a new trial can easily be derived with the `preddist` command. While the MAP prior is the predictive distribution of the mean parameter for a new trial and accounts for parameter uncertainty and between-trial heterogeneity, the predictive distribution of data accounts in addition for sampling uncertainty. For beta mixture densities the respective predictive distribution is the beta-binomial mixture distribution. This can be used to illustrate possible outcomes in a future trial or to calculate the Bayes factor of observed data with respect to the prior.

Moreover, **RBesT** provides the `postmix` command which updates a prior with the data as observed and computes analytically the posterior mixture distribution. For two-sample cases we are often interested in the difference distribution of two densities (representing a treatment difference). In **RBesT** the difference distribution of mixtures of the same class is supported through the use of the convolution theorem which allows for an accurate evaluation. Table 2 summarizes all functions available for parametric mixtures supported in **RBesT**. Further details are available in the help files for each function.

4.3. Trial design evaluation

Once a parametric mixture MAP prior has been derived, it is crucial to understand its influence on the statistical analysis of a clinical trial. In the context of the historical control example the goal is commonly to reduce the control group sample size while maintaining adequate power for trial success under the alternative hypothesis which assumes some true effect size. The main focus here is the evaluation of the (frequentist) operating characteristics of the trial design with respect to achieving trial success.

RBesT follows a modular, step-wise approach for design evaluation. First, a success criterion is defined, then the design of the trial is specified and finally the desired evaluation of the trial can be conducted for possible scenarios of interest. The success criteria supported are restricted to one-sided criteria which are referred to as decision functions. These can be set up for one-sample (`decision1S`) and two-sample (`decision2S`) situations. In **RBesT** the decision functions can be set up to require that multiple critical probability thresholds and quantiles for the difference distribution have to be met in the two-sample case while these thresholds are applied directly to the posterior density in the one-sample case. This enables evaluation of so-called dual criterion designs (Roychoudhury, Scheuer, and Neuenschwander 2018) which evaluate a statistical confidence criterion (high confidence in a positive/negative difference) and a minimally observed treatment difference (observed median difference being greater/smaller than some value). For the ankylosing spondylitis trial, success was declared whenever the posterior treatment difference is positive with a probability which exceeds 95%. In **RBesT** this is expressed by first defining the decision function as used in the trial:

```
R> success <- decision2S(0.95, 0, lower.tail = FALSE)
```

An alternative which demands in addition to see at least a median difference of at least 10% is defined by

```
R> success_dual <- decision2S(c(0.95, 0.5), c(0, 0.1), lower.tail = FALSE)
```

For convenience, we can print decision function objects:

```
R> success
```

```
2 sample decision function
Conditions for acceptance:
P(x1 - x2 > 0) > 0.95
Link: identity
```

```
R> success_dual
```

```
2 sample decision function
Conditions for acceptance:
P(x1 - x2 > 0) > 0.95
P(x1 - x2 > 0.1) > 0.5
Link: identity
```

The returned object represents the decision function and is a binary function taking two arguments. The arguments need to be two mixture densities of x1 and x2 (those will usually be the mixture *posteriors*) and the decision function acts as indicator function returning 1 whenever the condition for success on the difference distribution of the two mixture densities is fulfilled and 0 otherwise. Optionally, the success criterion allows for transformation of the input mixture distributions prior to forming the difference distribution using the link argument. The log and logit link enable relative risk and log odds-ratio success criteria to be evaluated with **RBesT**.

The next step is to define the design of the trial using the operating characteristics function oc2S (or oc1S for a one-sample case). The design of the trial includes the priors for each arm, the total sample size per arm and the decision function for trial success. The aim of using informative priors on the control response rate is to reduce the sample size needed for control. As such, it is useful to start with the respective frequentist power calculation to find an initial sample size needed when not using prior information. For the ankylosing spondylitis example the assumed true response rates for control were 25% and for active 60%, which results in a per-arm sample size of at least 24 when targeting a power of 80%, using function power.prop.test from package **stats**:

```
R> power.prop.test(p1 = 0.25, p2 = 0.6, power = 0.8, sig.level = 0.05,
+    alternative = "one.sided")

     Two-sample comparison of proportions power calculation

              n = 23.59764
             p1 = 0.25
             p2 = 0.6
      sig.level = 0.05
          power = 0.8
    alternative = one.sided

NOTE: n is number in *each* group
```

Thus, 24 patients per arm are required to reach 80% for these assumptions and hence the MAP prior with an effective sample size of 36 (25 for the robust MAP prior) represents a considerable amount of information. Therefore, the sample size of the control group can be reduced by a sizable fraction in this case. A 4:1 randomization ratio between active and control treatment was chosen for the trial. To now evaluate the impact a MAP prior has on the operating characteristics compared to the respective robust design, one can determine a prior for the treatment response rate in trial with a-priori median 25% response rate matched to the a-priori assumption of the control response rate using

```
R> treat_prior <- mixbeta(c(1, 1/2, 1))
```

and explore the different trial designs with

```
R> design_nonrobust <- oc2S(treat_prior, map, 24, 6, success)
R> design_robust <- oc2S(treat_prior, map_robust, 24, 6, success)
```

The `oc2S` function returns a binary function which is finally used to calculate the frequency of trial success as a function of an assumed truth for each arm. For the binomial sampling distribution, the function takes assumed true response rates θ_1 and θ_2. Whenever these two are set to the same numerical value $\theta = \theta_1 = \theta_2$, the scenario of no treatment difference is evaluated, which corresponds to a type-I error in a respective frequentist trial analysis which assumes as null hypothesis a difference of 0 between active and control. Setting θ_2 to a plausible control response rate and varying θ_1 as a function of the difference $\delta = \theta_1 - \theta_2$ yields the power of the trial. The functions take vectors of equal length as arguments so that we can evaluate the type-I error rates over large parameter ranges as:

```
R> theta <- c(0.25, 0.5, 0.75)
R> round(design_nonrobust(theta, theta), 3)

[1] 0.020 0.323 0.775

R> round(design_robust(theta, theta), 3)

[1] 0.018 0.190 0.173
```

Here we see that the Bayesian design (with informative priors) does not control the type-I error and that the error rates are conditional on the response rates being equal to their respective values in each treatment arm. However, the 97.5% quantile of the MAP prior is approximately 0.48 such that response rates greater than this would be very unusual. The power of the trial under the assumed true response rates in each treatment group is:

```
R> round(design_nonrobust(0.25 + 0.35, 0.25), 3)

[1] 0.871

R> round(design_robust(0.25 + 0.35, 0.25), 3)

[1] 0.77
```

Given that **RBesT** is fast and accurate for these calculations, it is recommended to use graphical plots in addition to tables to visualize the (error) rates of interest as demonstrated in the vignettes of **RBesT**.

4.4. Probability of success for interim trial analysis

The operating characteristics evaluate the frequency at which trial success is declared while assuming known true values for the response rates in each treatment arm. However, in practice it can be of interest at an interim analysis, when only a fraction of the total sample size is collected, to evaluate the frequency of trial success whenever the response rates are subject to uncertainty. Should the chances for success be too low at the interim analysis, then the trial can be stopped early for futility. While this has not been applied in the ankylosing spondylitis Phase II trial, we here assume that an interim analysis is conducted with 12 patients in the active treatment arm and 3 in the control arm. The observed number of responders are assumed to be 5 and 1 for the active treatment and control arm, respectively.

At the interim analysis the accrued data updates the priors. **RBesT** calculates this analytically using the `postmix` function:

```
R> interim_treat <- postmix(treat_prior, r = 5, n = 12)
R> interim_control <- postmix(map_robust, r = 1, n = 3)
```

The probability of success calculation now follows the same conventions as before in that first the design is specified:

```
R> pos_robust <- pos2S(interim_treat, interim_control, 24 - 12, 6 - 3,
+     success)
```

Note that the specification of the probability of success uses the *interim state* of the accrued knowledge as prior and the *remaining* sample size. The returned object is again a binary function accepting two arguments which need to be mixture distributions representing the densities of the per treatment arm response rates and hence the exact value of the response rates are treated as unknown. For the situation of an interim analysis, the state of knowledge about the per treatment arm response rate are merely the interim posteriors and as such the probability of success is in this example:

```
R> pos_robust(interim_treat, interim_control)
```

```
[1] 0.203874
```

However, if one would like to calculate the probability for success at the interim with the non-robust MAP prior for the control response rate (while still using the robust prior for the analysis), then the calculation is:

```
R> interim_control_nonrobust <- postmix(map, r = 1, n = 3)
R> pos_robust(interim_treat, interim_control_nonrobust)
```

```
[1] 0.2113132
```

While the probability of success calculation uses for the calculation of the trial success the robust prior for the control arm, the knowledge at the interim about the control response rate uses the plain non-robust MAP prior as basis. This reflects that the trial success depends on the analysis prior while at the interim we may have additional information available (not part of the final trial analysis) to calculate the probability of success. The **RBesT** package contains two further vignettes about the probability of success to detail this functionality.

4.5. Trial analysis

Once the final data of the trial is available, the **RBesT** package can be used to conduct the final analysis. For the ankylosing spondylitis trial 15 responders were reported out of 24 patients in the active treatment arm and 1 responder out of 6 in the control arm. This leads to a successful trial result as demonstrated by the following analysis. First the final posterior per treatment arm is obtained:

```
R> final_treat <- postmix(treat_prior, r = 15, n = 24)
R> final_control <- postmix(map_robust, r = 1, n = 6)
```

Then the success indicator function is applied to the posteriors by treatment arm:

```
R> success(final_treat, final_control)
```

```
[1] 1
```

The output of 1 indicates that the previously defined success criteria are met. The success indicator function simply thresholds the difference probability density distribution using the defined success criteria. Alternatively one may calculate directly the probability for a positive value of the difference between the control and treatment response rate posterior as:

```
R> pmixdiff(final_treat, final_control, 0, lower.tail = FALSE)
```

```
[1] 0.9948083
```

Internally, all calculations with the mixture distributions use analytical results whenever possible and always avoid the use of Monte Carlo simulations. This warrants fast and accurate results which do not depend (numerically) on the choice of a random seed. For the trial evaluation functions the key quantity calculated is the critical decision boundary determined by the success criterion and the trial design. At the critical outcome boundary the success criterion changes its value between 0 and 1. While in the one-sample case this corresponds to a single value, it is a function of the outcome in the second sample in the two-sample case. As the decision boundary can be useful for other applications, the boundary can be obtained with the functions `decision2S_boundary` (`decision1S_boundary`). These are also useful when communicating various data scenarios and their respective decisions to non-statisticians like clinical teams. For more details please refer to the appendix for their calculation and the vignettes of **RBesT** contain example visualizations of the boundaries.

5. Summary

In this paper, we introduced the **RBesT** package which implements the MAP approach via MCMC sampling algorithms for a number of common sampling and prior distributions. Incorporating historical data in clinical studies has various advantages, such as reducing the number of subjects randomized to a control arm or getting more precise information for decision making. Incorporation of historical data should lead to more ethical and efficient clinical trials.

The MAP approach is a hierarchical modeling method allowing heterogeneity between historical trials, which can incorporate historical data in a meta-analytic framework. The **RBesT** package allows for easier implementation of MAP priors. After selection of appropriate historical information, **RBesT** facilitates the prior derivation using MCMC, the parametric (mixture) approximation of the MAP prior, the evaluation of the clinical trial design and finally the analysis of the trial data.

Computational details

The **RBesT** package depends on packages **Rcpp** (Eddelbuettel and François 2011; Eddelbuettel 2013; Eddelbuettel and Balamuta 2018), **RcppParallel** (Allaire, François, Ushey, Vandenbrouck, Geelnard, and Intel 2021), Stan (Carpenter *et al.* 2017), **rstan** (Stan Development Team 2020), **rstantools** (Gabry, Goodrich, and Lysy 2020), **assertthat** (Wickham 2019), **mvtnorm** (Genz *et al.* 2021; Genz and Bretz 2009), **Formula** (Zeileis and Croissant 2010), **checkmate** (Lang 2017), **bayesplot** (Gabry and Mahr 2021; Gabry, Simpson, Vehtari, Betancourt, and Gelman 2019), **ggplot2** (Wickham 2016), **dplyr** (Wickham, François, Henry, and Müller 2021), **BH** (Eddelbuettel, Emerson, and Kane 2020), and **RcppEigen** (Bates and Eddelbuettel 2013). All R packages are available from CRAN at `https://CRAN.R-project.org/`.

Acknowledgments

References

Allaire JJ, François R, Ushey K, Vandenbrouck G, Geelnard M, Intel (2021). ***RcppParallel: Parallel Programming Tools for Rcpp***. R package version 5.1.4, URL `https://CRAN.R-project.org/package=RcppParallel`.

Baeten D, Baraliakos X, Braun J, Sieper J, Emery P, Van Der Heijde D, McInnes I, Van Laar JM, Landewé R, Wordsworth P, Wollenhaupt J, Kellner H, Paramarta J, Wei J, Brachat A, Bek S, Laurent D, Li Y, Wang YA, Bertolino AP, Gsteiger S, Wright AM, Hueber W (2013). "Anti-Interleukin-17A Monoclonal Antibody Secukinumab in Treatment of Ankylosing Spondylitis: A Randomised, Double-Blind, Placebo-Controlled Trial." *The Lancet*, **382**(9906), 1705–1713. `doi:10.1016/s0140-6736(13)61134-4`.

Bates D, Eddelbuettel D (2013). "Fast and Elegant Numerical Linear Algebra Using the **RcppEigen** Package." *Journal of Statistical Software*, **52**(5), 1–24. `doi:10.18637/jss.v052.i05`.

Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A (2017). "Stan: A Probabilistic Programming Language." *Journal of Statistical Software*, **76**(1), 1–32. `doi:10.18637/jss.v076.i01`.

Chen MH, Ibrahim JG (2000). "Power Prior Distributions for Regression Models." *Statistical Science*, **15**(1), 46–60. `doi:10.1214/ss/1009212673`.

Dallal S, Hall W (1983). "Approximating Priors by Mixtures of Natural Conjugate Priors." *Journal of the Royal Statistical Society B*, **45**(2), 278–286. `doi:10.1111/j.2517-6161.1983.tb01251.x`.

Diaconis P, Ylvisaker D (1985). "Quantifying Prior Opinion." In JM Bernardo, MH DeGroot, DV Lindley, AFM Smith (eds.), *Bayesian Statistics 2: Proceedings of the Second Valencia International Meeting*, pp. 133–148.

Dias S, Ades AE, Welton NJ, Jansen JP, Sutton AJ (2018). *Network Meta-Analysis for Decision-Making.* John Wiley & Sons. `doi:10.1002/9781118951651`.

Eddelbuettel D (2013). *Seamless R and C++ Integration with **Rcpp**.* Springer-Verlag, New York. `doi:10.1007/978-1-4614-6868-4`.

Eddelbuettel D, Balamuta JJ (2018). "Extending R with C++: A Brief Introduction to **Rcpp**." *The American Statistician*, **72**(1), 28–36. `doi:10.1080/00031305.2017.1375990`.

Eddelbuettel D, Emerson JW, Kane MJ (2020). **BH: Boost C++ Header Files.** R package version 1.72.0-3, URL `https://CRAN.R-project.org/package=BH`.

Eddelbuettel D, François R (2011). "**Rcpp**: Seamless R and C++ Integration." *Journal of Statistical Software*, **40**(8), 1–18. `doi:10.18637/jss.v040.i08`.

Eddy DM (1990). "Practice Policies: Where Do They Come From?" *JAMA: The Journal of the American Medical Association*, **263**(9), 1265–1275. `doi:10.1001/jama.1990.03440090103036`.

European Medicines Agency (2013). "Concept Paper on Extrapolation of Efficacy and Safety in Medicine Development." See URL.

European Medicines Agency (2018). "Reflection Paper on the Use of Extrapolation in the Development of Medicines for Paediatrics." See URL.

Food and Drug Administration (FDA) for Devices (2015). "Leveraging Existing Clinical Data for Extrapolation to Pediatric Uses of Medical Devices. Draft Guidance for Industry and Food and Drug Administration Staff." See URL.

Friede T, Röver C, Wandel S, Neuenschwander B (2017). "Meta-Analysis of Two Studies in the Presence of Heterogeneity with Applications in Rare Diseases." *Biometrical Journal*, **59**(4), 658–671. `doi:10.1002/bimj.201500236`.

Friede T, Röver C, Wandel S, Neuenschwander B (2017B). "Meta-Analysis of Few Small Studies in Orphan Diseases." *Research Synthesis Methods*, **8**(1), 79–91. `doi:10.1002/jrsm.1217`.

Gabry J, Goodrich B, Lysy M (2020). **rstantools**: *Tools for Developing* R *Packages Interfacing with* Stan. R package version 2.1.1, URL `https://CRAN.R-project.org/package=rstantools`.

Gabry J, Mahr T (2021). "**bayesplot**: Plotting for Bayesian Models." R package version 1.8.1, URL `https://CRAN.R-project.org/package=bayesplot`.

Gabry J, Simpson D, Vehtari A, Betancourt M, Gelman A (2019). "Visualization in Bayesian Workflow." *Journal of the Royal Statistical Society A*, **182**(2), 389–402. `doi:10.1111/rssa.12378`.

Gamalo-Siebers M, Gao A, Lakshminarayanan M, Liu G, Natanegara F, Railkar R, Schmidli H, Song G (2016). "Bayesian Methods for the Design and Analysis of Noninferiority Trials." *Journal of Biopharmaceutical Statistics*, **26**(5), 823–841. `doi:10.1080/10543406.2015.1074920`.

Gamalo-Siebers M, Savic J, Basu C, Zhao X, Gopalakrishnan M, Gao A, Song G, Baygani S, Thompson L, Xia HA, *et al.* (2017). "Statistical Modeling for Bayesian Extrapolation of Adult Clinical Trial Information in Pediatric Drug Evaluation." *Pharmaceutical Statistics*, **16**(4), 232–249. `doi:10.1002/pst.1807`.

Gelman A (2006). "Prior Distributions for Variance Parameters in Hierarchical Models." *Bayesian Analysis*, **1**(3), 515–534. `doi:10.1214/06-ba117a`.

Gelman A, Stern HS, Carlin JB, Dunson DB, Vehtari A, Rubin DB (2013). *Bayesian Data Analysis*. Chapman & Hall/CRC. `doi:10.1201/b16018`.

Genz A, Bretz F (2009). *Computation of Multivariate Normal and t Probabilities*. Lecture Notes in Statistics. Springer-Verlag, Heidelberg. `doi:10.1007/978-3-642-01689-9`.

Genz A, Bretz F, Miwa T, Mi X, Leisch F, Scheipl F, Hothorn T (2021). **mvtnorm**: *Multivariate Normal and t Distributions*. R package version 1.1-3, URL `https://CRAN.R-project.org/package=mvtnorm`.

Gsponer T, Gerber F, Bornkamp B, Ohlssen D, Vandemeulebroecke M, Schmidli H (2014). "A Practical Guide to Bayesian Group Sequential Designs." *Pharmaceutical Statistics*, **13**(1), 71–80. `doi:10.1002/pst.1593`.

Gsteiger S, Neuenschwander B, Mercier F, Schmidli H (2013). "Using Historical Control Information for the Design and Analysis of Clinical Trials with Overdispersed Count Data." *Statistics in Medicine*, **32**(21), 3609–3622. `doi:10.1002/sim.5851`.

Günhan BK (2020). **MetaStan**: *Bayesian Meta-Analysis via* **Stan**. R package version 0.2.0, URL `https://CRAN.R-project.org/package=MetaStan`.

Günhan BK, Friede T, Held L (2018). "A Design-by-Treatment Interaction Model for Network Meta-Analysis and Meta-Regression with Integrated Nested Laplace Approximations." *Research Synthesis Methods*, **9**(2), 179–194. `doi:10.1002/jrsm.1285`.

Günhan BK, Röver C, Friede T (2020). "Random-Effects Meta-Analysis of Few Studies Involving Rare Events." *Research Synthesis Methods*, **11**(1), 74–90. `doi:10.1002/jrsm.1370`.

Higgins JPT, Green ST (2008). *Cochrane Handbook for Systematic Reviews of Interventions.* John Wiley & Sons. `doi:10.1002/9780470712184`.

Higgins JPT, Thompson SG, Spiegelhalter DJ (2009). "A Re-Evaluation of Random-Effects Meta-Analysis." *Journal of the Royal Statistical Society A*, **172**(1), 137–159. `doi:10.1111/j.1467-985x.2008.00552.x`.

Hobbs BP, Sargent DJ, Carlin BP (2012). "Commensurate Priors for Incorporating Historical Information in Clinical Trials Using General and Generalized Linear Models." *Bayesian Analysis*, **7**(3), 639–674. `doi:10.1214/12-ba722`.

Holzhauer B (2017). "Meta-Analysis of Aggregate Data on Medical Events." *Statistics in Medicine*, **36**(5), 723–737. `doi:10.1002/sim.7181`.

Holzhauer B, Wang C, Schmidli H (2018). "Evidence Synthesis from Aggregate Recurrent Event Data for Clinical Trial Design and Analysis." *Statistics in Medicine*, **37**(6), 867–882. `doi:10.1002/sim.7549`.

Kass RE, Wasserman L (1995). "A Reference Bayesian Test for Nested Hypotheses and Its Relationship to the Schwarz Criterion." *Journal of the American Statistical Association*, **90**(431), 928–934. `doi:10.1080/01621459.1995.10476592`.

Kerman J (2011). "Neutral Noninformative and Informative Conjugate Beta and Gamma Prior Distributions." *Electronic Journal of Statistics*, **5**, 1450–1470. `doi:10.1214/11-ejs648`.

Lang M (2017). "**checkmate**: Fast Argument Checks for Defensive R Programming." *The R Journal*, **9**(1), 437–445. `doi:10.32614/rj-2017-028`.

Lin L, Zhang J, Hodges JS, Chu H (2017). "Performing Arm-Based Network Meta-Analysis in R with the **pcnetmeta** Package." *Journal of Statistical Software*, **80**(5), 1–25. `doi:10.18637/jss.v080.i05`.

Ma Z, Leijon A (2009). "Beta Mixture Models and the Application to Image Classification." In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pp. 2045–2048. IEEE. `doi:10.1109/ICIP.2009.5414043`.

Mielke J, Schmidli H, Jones B (2018). "Incorporating Historical Information in Biosimilar Trials: Challenges and a Hybrid Bayesian-Frequentist Approach." *Biometrical Journal*, **60**(3), 564–582. `doi:10.1002/bimj.201700152`.

Morita S, Thall PF, Müller P (2008). "Determining the Effective Sample Size of a Parametric Prior." *Biometrics*, **64**(2), 595–602. `doi:10.1111/j.1541-0420.2007.00888.x`.

Mütze T, Schmidli H, Friede T (2018). "Sample Size Re-Estimation Incorporating Prior Information on a Nuisance Parameter." *Pharmaceutical Statistics*, **17**(2), 126–143. `doi:10.1002/pst.1837`.

Neuenschwander B, Capkun-Niggli G, Branson M, Spiegelhalter DJ (2010). "Summarizing Historical Information on Controls in Clinical Trials." *Clinical Trials*, **7**(1), 5–18. `doi:10.1177/1740774509356002`.

Neuenschwander B, Roychoudhury S, Schmidli H (2016). "On the Use of Co-Data in Clinical Trials." *Statistics in Biopharmaceutical Research*, **8**(3), 345–354. `doi:10.1080/19466315.2016.1174149`.

Neuenschwander B, Schmidli H (2020). "Use of Historical Data." In E Lesaffre, G Baio, B Boulanger (eds.), *Bayesian Methods in Pharmaceutical Research.* Chapman & Hall/CRC, Boca Raton.

Neuenschwander B, Wandel S, Roychoudhury S, Bailey S (2016). "Robust Exchangeability Designs for Early Phase Clinical Trials with Multiple Strata." *Pharmaceutical Statistics*, **15**(2), 123–134. `doi:10.1002/pst.1730`.

Neuenschwander B, Weber S, Schmidli H, O'Hagan A (2020). "Predictively Consistent Prior Effective Sample Sizes." *Biometrics*, **76**(2), 578–587. `doi:10.1111/biom.13252`.

O'Hagan A, Forster JJ (2004). *Kendall's Advanced Theory of Statistics*, volume 2B. 1st edition. Arnold.

O'Hagan A, Pericchi L (2012). "Bayesian Heavy-Tailed Models and Conflict Resolution: A Review." *Brazilian Journal of Probability and Statistics*, **26**(4), 372–401. `doi:10.1214/11-bjps164`.

O'Hagan A, Stevens JW, Campbell MJ (2005). "Assurance in Clinical Trial Design." *Pharmaceutical Statistics*, **4**(3), 187–201. `doi:10.1002/pst.175`.

Peel DJ, McLachlan GJ (2000). "Robust Mixture Modelling Using the t Distribution." *Statics and Computing*, **10**(4), 339–348. `doi:10.1023/a:1008981510081`.

Plummer M (2003). "**JAGS**: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling." In K Hornik, F Leisch, A Zeileis (eds.), *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003).* Technische Universität Wien, Vienna, Austria. URL `https://www.R-project.org/conferences/DSC-2003/Proceedings/Plummer.pdf`.

Pocock SJ (1976). "The Combination of Randomized and Historical Controls in Clinical Trials." *Journal of Chronic Diseases*, **29**(3), 175–188. `doi:10.1016/0021-9681(76)90044-8`.

R Core Team (2021). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. URL `https://www.R-project.org/`.

Röver C (2020). "Bayesian Random-Effects Meta-Analysis Using the **bayesmeta** R package." *Journal of Statistical Software*, **93**(6), 1–51. `doi:10.18637/jss.v093.i06`.

Röver C, Wandel S, Friede T (2019). "Model Averaging for Robust Extrapolation in Evidence Synthesis." *Statistics in Medicine*, **38**(4), 674–694. `doi:10.1002/sim.7991`.

Roychoudhury S, Scheuer N, Neuenschwander B (2018). "Beyond p-Values: A Phase II Dual-Criterion Design with Statistical Significance and Clinical Relevance." *Clinical Trials*, **15**(5), 452–461. `doi:10.1177/1740774518770661`.

Rücker G (2012). "Network Meta-Analysis, Electrical Networks and Graph Theory." *Research Synthesis Methods*, **4**(3), 312–324. `doi:10.1002/jrsm.1058`.

Rücker G, Krahn U, König J, Efthimiou O, Schwarzer G (2021). **netmeta***: Network Meta-Analysis Using Frequentist Methods*. R package version 2.0-1, URL `https://CRAN.R-project.org/package=netmeta`.

Schmidli H, Bretz F, 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**(27), 4925–4938. `doi:10.1002/sim.2957`.

Schmidli H, Gsteiger S, Neuenschwander B (2021). "Incorporating Historical Data into Randomized Clinical Trials." In K Kim, F Bretz, YK Cheung, LV Hampson (eds.), *Handbook of Statistical Methods for Randomized Controlled Trials*. Chapman & Hall/CRC, Boca Raton.

Schmidli H, Gsteiger S, Roychoudhury S, O'Hagan A, Spiegelhalter DJ, Neuenschwander B (2014). "Robust Meta-Analytic-Predictive Priors in Clinical Trials with Historical Control Information." *Biometrics*, **70**(4), 1023–32. `doi:10.1111/biom.12242`.

Schmidli H, Häring DA, Thomas M, Cassidy A, Weber S, Bretz F (2020). "Beyond Randomized Clinical Trials: Use of External Controls." *Clinical Pharmacology & Therapeutics*, **107**(4), 806–816. `doi:10.1002/cpt.1723`.

Schmidli H, Neuenschwander B, Friede T (2017). "Meta-Analytic-Predictive Use of Historical Variance Data for the Design and Analysis of Clinical Trials." *Computational Statistics & Data Analysis*, **113**, 100–110. `doi:10.1016/j.csda.2016.08.007`.

Schmidli H, Wandel S, Neuenschwander B (2013). "The Network Meta-Analytic-Predictive Approach to Non-Inferiority Trials." *Statistical Methods in Medical Research*, **22**(2), 219–240. `doi:10.1177/0962280211432512`.

Spiegelhalter DJ, Abrams KR, Myles JP (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. John Wiley & Sons. `doi:10.1002/0470092602`.

Spiegelhalter DJ, Freedman LS, Blackburn PR (1986). "Monitoring Clinical Trials: Conditional or Predictive Power?" *Controlled Clinical Trials*, **7**(1), 8–17. `doi:10.1016/0197-2456(86)90003-6`.

Stan Development Team (2020). "**rstan**: The R Interface to **Stan**." R package version 2.21.2, URL `https://CRAN.R-project.org/package=rstan`.

Stangl D, Berry DA (2000). *Meta-Analysis in Medicine and Health Policy*. Chapman & Hall/CRC.

Van Valkenhoef G, Kuiper J (2021). **gemtc***: Network Meta-Analysis Using Bayesian Methods*. R package version 1.0-1, URL `https://CRAN.R-project.org/package=gemtc`.

Viechtbauer W (2010). "Conducting Meta-Analyses in R with the **metafor** Package." *Journal of Statistical Software*, **36**(3), 1–48. `doi:10.18637/jss.v036.i03`.

Viele K, Berry S, Neuenschwander B, Amzal B, Chen F, Enas N, Hobbs B, Ibrahim J, Kinnersley N, Lindborg S, Micallef S, Roychoudhury S, Thompson L (2014). "Use of Historical Control Data for Assessing Treatment Effects in Clinical Trials." *Pharmaceutical Statistics*, **13**(1), 41–54. `doi:10.1002/pst.1589`.

Wadsworth I, Hampson LV, Jaki T (2018). "Extrapolation of Efficacy and Other Data to Support the Development of New Medicines for Children: A Systematic Review of Methods." *Statistical Methods in Medical Research*, **27**(2), 398–413. `doi:10.1177/0962280216631359`.

Wandel S, Neuenschwander B, Röver C, Friede T (2017). "Using Phase II Data for the Analysis of Phase III Studies: An Application in Rare Diseases." *Clinical Trials*, **14**(3), 277–285. `doi:10.1177/1740774517699409`.

Wandel S, Roychoudhury S (2016). "Designing and Analysing Clinical Trials in Mental Health: An Evidence Synthesis Approach." *Evidence-Based Mental Health*, **19**(4), 114–117. `doi: 10.1136/eb-2016-102491`.

Wandel S, Schmidli H, Neuenschwander B (2015). "Use of Historical Data." In S George (ed.), *Cancer Clinical Trials: Current and Controversial Issues in Design and Analysis*. Chapman & Hall/CRC.

Weber S (2021). **RBesT**: *R Bayesian Evidence Synthesis Tools*. R package version 1.6-3, URL `https://CRAN.R-project.org/package=RBesT`.

Welton NJ, Sutton AJ, Cooper N, Abrams KR, Ades AE (2012). *Evidence Synthesis for Decision Making in Healthcare*. Statistics in Practice. John Wiley & Sons. `doi:10.1002/9781119942986`.

West M (1993). "Approximating Posterior Distributions by Mixtures." *Journal of the Royal Statistical Society B*, **55**(2), 409–422. `doi:10.1111/j.2517-6161.1993.tb01911.x`.

Wickham H (2016). **ggplot2**: *Elegant Graphics for Data Analysis*. Springer-Verlag. `doi: 10.1007/978-3-319-24277-4`.

Wickham H (2019). **assertthat**: *Easy Pre and Post Assertions*. R package version 0.2.1, URL `https://CRAN.R-project.org/package=assertthat`.

Wickham H, François R, Henry L, Müller K (2021). **dplyr**: *A Grammar of Data Manipulation*. R package version 1.0.7, URL `https://CRAN.R-project.org/package=dplyr`.

Zeileis A, Croissant Y (2010). "Extended Model Formulas in R: Multiple Parts and Multiple Responses." *Journal of Statistical Software*, **34**(1), 1–13. `doi:10.18637/jss.v034.i01`.

A. Parametric mixture inference

In **RBesT** the EM algorithm is used to find parametric mixture approximations to the numerical representation of the MAP prior. Thus, we consider in this section as data $\boldsymbol{Y}$ the MCMC sample representing the MAP prior (or any other MCMC sample). Direct application of maximum likelihood is numerically problematic, since the log-likelihood for a mixture prior, $\log p(\boldsymbol{Y}|\boldsymbol{w}, \boldsymbol{a}, \boldsymbol{b}) = \sum_{i=1}^{N} \log \sum_{k=1}^{K} w_k \, p_k(Y_i|a_k, b_k)$, involves the sum over the log of the component densities $w_k \, p_k(Y_i|a_k, b_k)$. The inner summation is on the natural scale and is required as we do not know the component k which a data item Y_i is a member of. However, extending the observed data ($\boldsymbol{Y}$), also referred to as incomplete data, to the so-called complete data ($\boldsymbol{Y}, \boldsymbol{Z}$) leads to a numerically stable problem. The unobserved data $\boldsymbol{Z}$ is defined as the latent component indicator such that $Z_{i,k}$ is 1 whenever data item i is part of component k and 0 otherwise. The extended problem is related to the original through marginalization, $p(\boldsymbol{Y}|\boldsymbol{w}, \boldsymbol{a}, \boldsymbol{b}) = E_{\boldsymbol{Z}}[p(\boldsymbol{Y}, \boldsymbol{Z}|\boldsymbol{w}, \boldsymbol{a}, \boldsymbol{b})]$. However, the extended log-likelihood factorizes in the usual way

$$\log p(\boldsymbol{Y}, \boldsymbol{Z}|\boldsymbol{w}, \boldsymbol{a}, \boldsymbol{b}) = \sum_{i=1}^{N} \sum_{k=1}^{K} Z_{i,k} \left[\log(w_k) + \log p_k(Y_i|a_k, b_k)\right].$$

The EM algorithm begins with a fixed number of mixture components K and an initial guess of all parameters. The initial guess of the parameters is achieved with the k nearest neighbors (knn) algorithm in **RBesT** with the exception of the normal mixture case as detailed below. The parameter vector $(\boldsymbol{w}, \boldsymbol{a}, \boldsymbol{b})$ is then updated iteratively. The nth iteration of the EM consists of first performing the *expectation* (E) step and then the *maximization* (M) step which updates the parameter vector for the next iteration. These EM steps are then repeated until convergence to a maximum or saddle-point of the log-likelihood. While it is guaranteed that in each iteration the log-likelihood is always increased, the EM algorithm may only find a local saddle-point rather than a global extremum. Moreover, applying the EM algorithm in cases with an unbounded log-likelihood can also cause issues (as it can occur at the boundaries of a Beta distribution with any of the shape parameters being smaller than 1).

E-step. The E-step calculates the posterior probability for the latent indicators $p(\boldsymbol{Z}|\boldsymbol{Y}, \boldsymbol{w}, \boldsymbol{a}, \boldsymbol{b})$ in order to compute the expected posterior mean weights

$$E[Z_{i,k}|\boldsymbol{Y}, \boldsymbol{w}, \boldsymbol{a}, \boldsymbol{b}] = \frac{w_k \, p_k(Y_i|a_k, b_k)}{\sum_k w_k \, p_k(Y_i|a_k, b_k)} = \gamma(Z_{i,k}).$$

The expression $\gamma(Z_{i,k})$ is often referred to as the responsibility of mixture component k for data item i. The overall responsibility of mixture component k

$$N_k = \sum_{i=1}^{N} \gamma(Z_{i,k})$$

can be interpreted as the number of data items belonging to mixture component k and hence the overall fraction of data points assigned to a specific component k is

$$\pi_k = \frac{N_k}{N}.$$

Finally, the E-step computes the expectation of the complete log-likelihood with respect to $\mathbf{Z}$ conditional on the parameter vector of the current nth iteration and the data $\mathbf{Y}$,

$$E_{\mathbf{Z}|\mathbf{Y},\mathbf{w}^{(n)},\mathbf{a}^{(n)},\mathbf{b}^{(n)}}[\log p(\mathbf{Y},\mathbf{Z}|\mathbf{w},\mathbf{a},\mathbf{b})] = \sum_{i=1}^{N}\sum_{k=1}^{K}\gamma(Z_{i,k})\left[\log(w_k) + \log p_k(Y_i|a_k,b_k)\right]$$

$$= Q(\mathbf{w},\mathbf{a},\mathbf{b}|\mathbf{w}^{(n)},\mathbf{a}^{(n)},\mathbf{b}^{(n)}).$$

The dependence of $Q(\mathbf{w},\mathbf{a},\mathbf{b}|\mathbf{w}^{(n)},\mathbf{a}^{(n)},\mathbf{b}^{(n)})$ on the parameters in the nth iteration ($\mathbf{w}^{(n)}$, $\mathbf{a}^{(n)}$ and $\mathbf{b}^{(n)}$) is through $\gamma(Z_{i,k})$.

M-step.   The M-step then proceeds by finding a new parameter vector through maximization

$$(\mathbf{w}^{(n+1)},\mathbf{a}^{(n+1)},\mathbf{b}^{(n+1)}) = \underset{\mathbf{w},\mathbf{a},\mathbf{b}}{\arg\max}\, Q(\mathbf{w},\mathbf{a},\mathbf{b}|\mathbf{w}^{(n)},\mathbf{a}^{(n)},\mathbf{b}^{(n)}).$$

The updated weights are constrained to sum to one. Maximization with this constraint is achieved through Lagrange multipliers and leads to the solution

$$w_k^{(n+1)} = \pi_k.$$

To find the maximum with respect to the parameters of each component k, we take the gradient $(\partial_{a_k},\partial_{b_k})$ of $Q(\mathbf{w},\mathbf{a},\mathbf{b}|\mathbf{w}^{(n)},\mathbf{a}^{(n)},\mathbf{b}^{(n)})$ and equate these to zero.

Normal mixtures.   For normal mixtures **RBesT** implements internally a multi-variate normal EM, but only exposes the uni-variate functionality at the moment. Empirically we observed that the normal EM algorithm is easily trapped into local extrema which is caused by the commonly heavy tailed distributions of MAP priors. For this reason, we initialize the normal EM with the result of a Student-t EM procedure as described in Peel and McLachlan (2000). The Student-t EM is itself initialized with k nearest neighbors. The maximization equations can be solved analytically in the normal mixture case (note that each data item i is a vector $\mathbf{Y}_i$):

$$\boldsymbol{\mu}_k = \frac{1}{N_k}\sum_{i=1}^{N}\gamma(Z_{i,k})\,\mathbf{Y}_i$$

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k}\sum_{i=1}^{N}\gamma(Z_{i,k})\left(\mathbf{Y}_i - \boldsymbol{\mu}_k\right)\left(\mathbf{Y}_i - \boldsymbol{\mu}_k\right)^{\top}.$$

The analytical solution is a weighted mean and covariance estimate with the weight for each data item $\mathbf{Y}_i$ equal to $\gamma(Z_{i,k})/N_k$.

Beta mixtures.   For beta mixture distributions the M-step requires solving the joint equation system of (see also Ma and Leijon 2009):

$$\psi(a_k) - \psi(a_k + b_k) = \frac{1}{N_k}\sum_{i=1}^{N}\gamma(Z_{i,k})\log(Y_i)$$

$$\psi(b_k) - \psi(a_k + b_k) = \frac{1}{N_k}\sum_{i=1}^{N}\gamma(Z_{i,k})\log(1 - Y_i).$$

Here $\psi(x)$ is the digamma function defined as $\partial_x \log(\Gamma(x))$. This equation system is solved simultaneously for a_k and b_k through numerical minimization.

Gamma mixtures. With gamma mixtures the algebraic equation system

$$\psi(a_k) - \log(b_k) = \frac{1}{N_k} \sum_{i=1}^{N} \gamma(Y_{i,k}) \log(Y_i)$$

$$\frac{a_k}{b_k} = \frac{1}{N_k} \sum_{i=1}^{N} \gamma(Z_{i,k}) Y_i$$

is obtained for the M-step. This system can be reduced to a single equation, which is again solved through numerical minimization.

B. Effective sample size

The effective sample size is an approximate measure for the number of observations a prior is equivalent to. In the setting of conjugate likelihood-prior pairs without mixtures the standard parameters can be cast into an effective sample size. For mixture priors the conjugacy of the likelihood-prior pair is preserved and analytical results still apply with some modifications. However, in the mixture prior case there is no more a clear relationship between the standard parameters and the effective sample size. **RBesT** implements three approaches to calculate the effective sample size for mixture priors:

`"elir"` : The expected local information ratio has been proposed in Neuenschwander, Weber, Schmidli, and O'Hagan (2020) and is a predictively consistent effective sample size measure. The predictive consistency requires that the effective sample size of the prior is equal to the average effective sample size of the respective posterior of this prior after simulation of m samples from the predictive prior distribution and subtracting m from the averaged posterior effective sample size. The method is neither liberal nor conservative and is the default method in **RBesT**.

`"morita"` : The method from Morita, Thall, and Müller (2008) evaluates the curvature of the prior at a reference point (mode, median or mode) and compares this against the expected curvature of a posterior of a variance inflated prior which is updated with m samples from the prior predictive of the prior. The m is chosen to minimize the difference in curvature. Since the Morita method evaluates the prior at a single point, the approach can be sensitive to the curvature at this point and has been observed to report relatively large effective sample sizes when used with mixture priors.

`"moment"` : The moment based approach matches the mean and the variance of a given prior to its respective canonical prior from which the effective sample size can be obtained directly from the standard parameters. This approach has been found empirically to report very conservative (low) effective sample sizes.

The key expressions involved in the effective sample size calculations are the prior information

$$i(p(\theta)) = -\partial_\theta^2 \log(p(\theta))$$

Sampling distribution $p(Y\|\theta)$	Fisher information $i_F(\theta)$	Prior density $p_k(\theta\|a,b)$	Prior information $i(p_k(\theta))$
$\text{Normal}(Y\|\theta,\sigma^2)$	σ^{-2}	$\text{Normal}(\theta\|m,s^2)$	s^{-2}
$\text{Binomial}(Y\|\theta,n)$	$\frac{n}{\theta(1-\theta)}$	$\text{Beta}(\theta\|a,b)$	$\frac{a-1}{\theta^2}+\frac{b-1}{(1-\theta)^2}$
$\text{Poisson}(Y\|\theta)$	θ^{-1}	$\text{Gamma}(\theta\|a,b)$	$\frac{a-1}{\theta^2}$
$\text{Exp}(Y\|\theta)^{\dagger}$	θ^{-2}	$\text{Gamma}(\theta\|a,b)$	$\frac{a-1}{\theta^2}$

Table 3: Overview on supported conjugate likelihood-prior pairs supported in **RBesT**. †Note that for the exponential sampling distribution only the effective sample size calculations are supported as of **RBesT** 1.6-2 (the predictive distribution is not available for this likelihood-prior pair).

and the Fisher information

$$i_F(\theta) = E_{Y_1|\theta}\left[i(p(Y_1|\theta))\right] = -E_{Y_1|\theta}\left[\partial_\theta^2 \log(p(Y_1|\theta)))\right].$$

The Fisher information is derived from the sampling distribution, but the second derivative of the log mixture prior density with K components, defined as

$$p(\theta|\boldsymbol{w},\boldsymbol{a},\boldsymbol{b}) = \sum_{k=1}^{K} w_k\, p_k(\theta|a_k,b_k),$$

needs some considerations for a numerically stable evaluation. It is key to evaluate the log density instead of the density on the natural scale whenever possible. We will in the following suppress the weights and standard parameters of $p(\theta)$ and $p_k(\theta)$ for simplicity. Using the equality $\partial_\theta p(\theta) = p(\theta)\,\partial_\theta \log(p(\theta))$ one finds that the prior information for a mixture is

$$i(p(\theta)) = \frac{1}{p(\theta)^2}\left[\sum_{k=1}^{K} w_k\, p_k(\theta)\,\partial_\theta \log p_k(\theta)\right]^2$$
$$-\frac{1}{p(\theta)}\sum_{k=1}^{K} w_k\, p_k(\theta)\left[(\partial_\theta \log p_k(\theta))^2 + \partial_\theta^2 \log p_k(\theta)\right]. \quad (3)$$

Table 3 lists all the main expressions required for the supported conjugate likelihood-prior pairs in **RBesT**.

C. Informative prior evaluation

In **RBesT** one-sided decision functions with multiple criteria are supported for one- and two-sample cases. The decision functions are indicator functions through thresholding density distributions such that critical quantiles must exceed predefined probability thresholds. Denoting with $H(x)$ the Heaviside step function, which is 0 for $x \leq 0$ and 1 otherwise, the decision functions are defined as

$$D(p(\theta)) = \prod_i H(\mathsf{P}(\theta \leq q_{i,c}) - p_{i,c}) \qquad \text{one sample,}$$

$$D(p_1(\theta_1),p_2(\theta_2)) = \prod_i H(\mathsf{P}(\theta_1 - \theta_2 \leq q_{i,c}) - p_{i,c}) \qquad \text{two sample.}$$

In the two-sample case the difference distribution of $p_1(\theta_1)$ and $p_2(\theta_2)$ is thresholded.

Critical decision boundary. With the design of a trial the priors and the sample size per sample are defined and these determine the critical decision boundary of the decision function,

$$D_c = \sup_{y}\{D(p(\theta|y)) = 1\} = y_c \qquad \text{one sample,}$$

$$D_{c,1}(y_2) = \sup_{y_1}\{D(p_1(\theta_1|y_1), p_2(\theta_2|y_2)) = 1\} \qquad \text{two sample.}$$

While for the one-sample cases the decision boundary is a single value, $D_c = y_c$, the decision boundary for the two-sample case is a function of the outcome in one of the samples (by convention the second sample), $D_{c,1}(y_2)$.

Conditional power. The critical decision boundary is used in **RBesT** to simplify the conditional power calculation in the following manner,

$$\mathrm{CP}(\theta) = \int D(p(\theta'|y))\, p(y|\theta)\, dy = \int_{-\infty}^{y_c} p(y|\theta)\, dy$$

$$= \mathsf{P}(y \le y_c|\theta) \qquad \text{one sample,}$$

$$\mathrm{CP}(\theta_1, \theta_2) = \iint D(p_1(\theta_1'|y_1), p_2(\theta_2'|y_2))\, p_1(y_1|\theta_1)\, p_2(y_2|\theta_2)\, dy_1\, dy_2$$

$$= \int P_1(y_1 \le D_{c,1}(y_2)|\theta_1)\, p_2(y_2|\theta_2)\, dy_2 \qquad \text{two sample.}$$

Therefore, the conditional power simplifies in the one-sample case to evaluating the cumulative distribution function corresponding to the sampling distribution and in the two-sample case the integration is simplified to a one dimensional integral instead of a two dimensional one. For the case of a binomial sampling distribution the calculations are carried out exactly. With normal and Poisson endpoints the respective integrals are evaluated with quadrature integration on a domain which corresponds to $1 - \epsilon$ ($\epsilon = 10^{-6}$ by default) of probability mass.

Probability of success. The probability of success as defined in Equation 2 of Section 3.3 results in a double (triple) integral for the one-(two-)sample case. To simplify this, the prior predictive distribution of the prior $p(\theta)$,

$$p(y) = \int p(y|\theta)\, p(\theta)\, d\theta,$$

is used. The prior predictive distribution is available in analytic form and allows to re-arrange the evaluation of the Equation 2 as

$$\mathrm{PoS} = \int \mathrm{CP}(\theta)\, p(\theta)\, d\theta = \iint D(p(\theta'|y))\, p(y|\theta)\, p(\theta)\, d\theta\, dy = \int D(p(\theta'|y))\, p(y)\, dy.$$

This re-arrangement also holds for the two-sample case. This leads to the same calculations as previously for the operating characteristics with the only difference in that the sampling distribution, $p(y|\theta)$, is replaced by the predictive distribution of the prior, $p(y)$. For the two-sample case the integration is performed using numerical integration while for the one-sample case the cumulative distribution function of the predictive is evaluated.

Affiliation:

Sebastian Weber
Novartis Pharma AG
4002 Basel, Switzerland
E-mail: sebastian.weber@novartis.com