PRACTICAL APPLICATION

Bayesian Meta-Analysis of Health State Utility Values: A Tutorial with a Practical Application in Heart Failure

Joseph Alvin Ramos Santos[1] · Robert Grant[2,3] · Gian Luca Di Tanna[1,4]

Abstract

Researchers incorporate health state utility values as inputs to inform economic models. However, for a particular health state or condition, multiple utility values derived from different studies typically exist and a single study is often insufficient to represent the best available source of utility needed to inform policy decisions. The purpose of this paper is to provide an introductory guidance for conducting Bayesian meta-analysis of health state utility values to generate a single parameter input for economic evaluation, using R. The tutorial is illustrated using data from a systematic review of health state utilities of patients with heart failure, with 21 studies that reported utilities measured using the EuroQol-5D (EQ-5D). Explanations, key considerations and suggested readings are provided for each step of the tutorial, adhering to a clear workflow for conducting Bayesian meta-analysis: (1) setting-up the data; (2) employing methods to impute missing standard deviations; (3) defining the priors; (4) fitting the model; (5) diagnosing model convergence; (6) interpreting the results; and (7) performing sensitivity analyses. The posterior distributions for the pooled effect size (i.e. mean health state utility) and between-study heterogeneity are discussed and interpreted in light of the data, priors and models used. We hope that this tutorial will foster interest in Bayesian methods and their applications in the meta-analysis of utilities.

Key Points for Decision Makers

The Bayesian methods' ability to incorporate prior information, handle heterogeneity between studies explicitly and provide intuitive probabilistic interpretations, make it a powerful framework for pooling utility data.

This tutorial provides step-by-step guidance on how to conduct a Bayesian meta-analysis to pool health state utility values using a ready-to-use R script.

✉ Gian Luca Di Tanna
   gianluca.ditanna@supsi.ch

1   Department of Business Economics, Health and Social Care (DEASS), University of Applied Sciences and Arts of Southern Switzerland (SUPSI), Manno, Ticino, Switzerland

2   BayesCamp Ltd, Winchester, UK

3   Kingston University, London, UK

4   Department of Clinical Research (DCR), University of Bern, Bern, Switzerland

1 Introduction

Utility refers to a measure placed by an individual on quality of life that is commonly associated with different health states [1]. It is measured by a value between one (representing full health) and zero (representing death), although it may take a value less than zero for extreme health conditions viewed as worse than death [1]. Estimates of quality-adjusted life years (QALYs) are directly informed by both health state utility values and the length of time spent in the health states of interest. In the context of cost-effectiveness analysis and health technology assessment (HTA), utility values and QALYs are used to quantify the effect of health interventions or technologies on an individual's quality of life, and offer a standardized measure to compare alternative interventions and allocate resources efficiently [2]. Thus, utility holds significant importance in economic evaluations.

Researchers employ utility values as part of the inputs to inform economic models. However, for a particular health state or condition, multiple utility values derived from different studies typically exist, making the selection of the most suitable utility value challenging. This complexity arises due to the differences in study characteristics, for example, in terms of the health-utility instrument used, the timing

and frequency of collection, the variables collected, and the heterogeneity of the study population [3]. Additionally, the lack of methodological harmonization in utility measurement methods, along with the deliberate selection of a utility value to be incorporated into decision-support economic models, can contribute to discrepancies in utility outcomes and add to the complexity in selecting the appropriate utility value. For these reasons, a single study is often insufficient to represent the best available source of utility values needed for informing a policy decision [4], and researchers rely on systematic literature reviews of utilities to assess and harmonize disparities in the estimated values across studies [4–6].

Meta-analysis of utilities has been recommended to generate a single parameter input for economic evaluation within a clinical domain or for a specific health state or condition [4]. The studies conducted so far on meta-analytic approaches for pooling utility values have shown that such methods can generate reliable utility estimates suitable for incorporation into economic models [4, 7, 8]. Petrou et al. [4] described three approaches (i.e. fixed-effect meta-analysis, random-effects meta-analysis and mixed-effects meta-regression) for combining utility data from various studies, but cautioned about possible methodological issues associated with the application of these approaches [4]. Hatswell et al. [7] conducted a comparison between the frequentist and Bayesian meta-regression models for the synthesis of utility data. The authors found comparable outcomes between the two approaches, and recommended the Bayesian analysis as the preferred approach, due to its capacity to incorporate prior information into the analysis, making it possible to utilize utility values identified from previous reviews [7]. A subsequent study by Hatswell [8] introduced the use of the Bayesian power prior [9] to adjust prior knowledge according to perceived relevance of the study source. The use of this prior produced comparable results with random-effects meta-analysis, but not with fixed-effect meta-analysis, which yielded very narrow confidence intervals. The authors noted clear benefits associated with the method, and suggested various avenues for development [8]. Overall, these studies offered valuable guidance for conducting meta-analysis of utilities, contributing to the progress of evidence synthesis in HTA to inform health policy decisions.

Bayesian meta-analysis involves the combination of information from multiple studies, while incorporating prior knowledge or beliefs about the parameters of interest, to generate updated estimates and quantify uncertainty. As such, it provides a powerful framework for the synthesis of health state utility values. Firstly, the inclusion of prior knowledge into the analysis is relevant in the case of utility values, given the increasing number of primary studies and systematic reviews of utility values across various health states and population groups [4]. The findings from these studies can serve as valuable prior information in the

meta-analysis. Secondly, Bayesian meta-analysis handles the degree of heterogeneity among studies more explicitly, which is crucial since utility values can be highly variable across studies [3]. The integration of prior information for the effect size and between-study heterogeneity helps in pooling diverse sources of data to produce more robust and precise estimates [10]. Lastly, Bayesian meta-analysis generates posterior distributions that encompass a range of possible values for the parameters of interest, allowing for better assessment of uncertainties around estimates and enabling direct calculation of posterior probabilities [11]. This facilitates clear interpretation of results and allows researchers to make informed inferences about plausible values for utility.

Effectively applying Bayesian inference to real-world problems requires a blend of statistical and programming skills, domain expertise and an understanding of the decision-making process within data analysis [12]. Together, these components form a sophisticated Bayesian workflow, which encompasses several tasks, including pre-specifying the analysis plan, incorporating diverse priors, providing scientific rationale for the priors and comparing different models [13]. Against the above backdrop, this tutorial aims to provide step-by-step guidance on how to perform Bayesian meta-analysis of utilities using R, with the view to empowering meta-analysts to utilize statistical modelling more effectively, enhancing confidence in the inferences and decisions derived, and going beyond the simple pooling of results. To this aim, we have provided codes to aid practitioners in understanding and applying Bayesian methods using data from a systematic review of health state utilities of patients with heart failure [14].

2 Data and Software

2.1 Summary of the Systematic Review

2.1.1 Systematic Review Methods

The data used in this tutorial has been reproduced with permission from the study authors [14]. The objective of the systematic review was to identify and summarize utility values of patients with heart failure. The search strategy included a peer-reviewed database search from their start date until June 2019, supplemented by a grey literature search including HTA websites and by relevant publications from a parallel review on cost-effectiveness models for pharmacological interventions in heart failure led by the same primary author [15]. Studies were included if they reported health state utility values for adults aged 18 years and above with heart failure, regardless of study design. Details on the study design, the instrument used to elicit utility, the value set used to produce utility values, the health state to which

the utility data was reported (i.e. chronic heart failure, hospitalized, and other acute heart failure) and the utility value and its measure of variability were extracted from the eligible studies. Studies that had a sample size of ≥ 100 were included in the calculation of the interquartile limits (25th and 75th percentile) for health states and heart failure subgroups. Meta-analysis was not carried out.

2.1.2 Findings from the Systematic Review

The review identified 161 publications with primary utility data of patients with heart failure elicited from 142 studies. The studies varied in design and study population. The EuroQol-5D (EQ-5D) (3L or 5L) was the most common instrument used to elicit utility ($n = 104$) although several studies did not specify which version was used ($n = 37$). The majority of publications did not report the value set used to calculate utility ($n = 88$), although the UK value set was the most commonly reported ($n = 33$). Utility values were reported in 128 publications for chronic heart failure, 39 publications for hospitalized patients with heart failure and three for other acute heart failure. Of the publications that reported EQ-5D utility values and met the criteria for calculating the interquartile limits, the calculated limits for chronic heart failure ($n = 35$) were 0.64–0.72, with a trend of decreasing utility with increasing disease severity. The limits for hospitalized patients with heart failure were 0.54–0.63 during hospital admission ($n = 4$) and 0.64–0.73 at hospital discharge ($n = 6$).

2.2 Description of the Dataset Used in the Meta-Analysis

For the purpose of this tutorial, the Bayesian meta-analysis only includes studies that reported heart failure utility values using EQ-5D (either 3L or 5L). This was an attempt to establish a reasonable degree of comparability between the studies and utility values included in the meta-analysis, to enhance the validity of the results. However, it is important to note that there were still some variations in the design (e.g. randomized controlled trials, non-randomized trial, observational), diagnosis or health state, and source of data (e.g. some were obtained from conference abstracts) across the studies included in the meta-analysis.

For studies with more than two treatment arms and those that reported utility values by subgroups (e.g. by the New York Heart Association class), where appropriate and possible, groups were combined to produce a single weighted average of utility values. For studies with multiple time points and intervention studies, only baseline data were utilized to avoid the introduction of any confounding effect of the intervention into the analysis [4]. For studies that reported multiple utility values derived from different value sets, the utility value based on the UK value set was preferred [4]. The final dataset includes 21 studies, of which six did not report a measure of variability (e.g. standard deviation) (Table 1). The dataset and R Script (hereinafter referred to as the "script") can be found in the Online Supplementary Material (see the electronic supplementary material).

2.3 Setting-Up R and RStudio

The open-source R software [35], along with its popular integrated development environment RStudio [36], are used in this tutorial. Several other open-source software and packages are available for carrying out Bayesian meta-analysis, which allow researchers to implement and customize models according to their needs and preferences. Some examples include JAGS [37], BUGS [38] and JASP [39].

Many introductory courses and workshops on data manipulation, analysis and visualization for R and RStudio are available online. Throughout the tutorial code, existing packages and user-defined functions in R are used. Packages can be easily installed using the R function 'install.packages()'.

3 An example of Bayesian Meta-Analysis of Health State Utilities in R

The tutorial adheres to the following structure: (1) set-up the data in R; (2) employ methods to impute missing standard deviations; (3) define the priors; (4) fit the model; (5) diagnose model convergence; (6) interpret the results; and (7) perform sensitivity analyses.

3.1 Set-Up the Data in R

All the packages needed to run our model are loaded using 'library()' (**Box 1**). The *brms* (*Bayesian regression models using Stan*) package is used for the implementation of the Bayesian meta-analysis [40]. Stan is a probabilistic programming language for statistical modelling [41], and *brms* extends the functionality of Stan by offering an interface (similar to the traditional regression modelling syntax in R) to fit Bayesian models. The *mfp* and *mice* packages are utilized to impute missing standard deviations through fractional polynomial regression and multiple imputation using chained equations, respectively [42, 43]. Plots are generated either through the built-in functions in the *brms* package or through the *bayesplot* [44], *shinystan* [45], *tidybayes* [46] or the *ggplot2* [47] packages.

Table 1 Characteristics of studies included in the meta-analysis

First author, year	Type of EQ-5D	Study design	Population/health state	N	Mean utility	SD
Adamson 2015 [16]	EQ-5D-3L	RCT	HF patients with CardioMEMS HF system	550	0.71	–
Allemann 2018 [17]	EQ-5D-3L	Observational	HF patients with implantable cardioverter defibrillator	1515	0.79	–
Clark 2015 [18]	EQ-5D-3L	RCT	Chronic HF patients	139	0.52	0.26
Delgado 2014 [19]	Unspecified	Observational	Chronic HF patients	369	0.76	0.27
García-Pérez 2012 [20]	Unspecified	Observational	HF patients	176	0.53	–
González-Guerrero 2018 [21]	EQ-5D-3L	RCT	Disease management program for geriatric HF patients	117	0.44	0.36
Hansson 2016 [22]	EQ-5D-3L	Non-randomized interventional	Chronic HF patients hospitalized for worsening HF	248	0.58	0.32
Hong 2018 [2]	EQ-5D-5L	Observational	Stable chronic HF patients	100	0.87	0.09
Hwang 2019 [23]	Unspecified	RCT	Chronic HF patients	53	0.71	0.21
Jackson 2018 [24]	EQ-5D-3L	Observational	HF patients	931	0.80	0.20
Krotneva 2016 [25]	Unspecified	RCT	HF patients	4199	0.68	–
Lee 2016 [26]	EQ-5D-3L	Observational	HF patients	140	0.74	0.25
Lewis 2014 [27]	EQ-5D-3L	RCT	Myocardial infarction patients with evidence of HF	597	0.70	0.29
Mantis 2018 [28]	Unspecified	Non-randomized interventional	HF patients	52	0.69	0.10
McNeil 2012[a]	Unspecified	Observational	HF patients who had ventricular assist device Heart-Mate II implant	29	0.76	0.21
Srinonprasert 2019 [29]	EQ-5D-3L	Observational	Non-valvular atrial fibrillation patients with HF as complication	875	0.77	–
Sullivan 2006 [30]	EQ-5D-3L	Observational	Chronic HF patients	284	0.64	–
Teng 2018 [31]	EQ-5D-3L	RCT	HF patients	84	0.80	0.14
Van Spall 2019 [32]	EQ-5D-5L	RCT	Hospitalized HF patients	606	0.70	0.24
Whitty 2013 [33]	Unspecified	Observational	Hospitalized chronic HF patients	91	0.75	0.20
Zanaboni 2013 [34]	Unspecified	RCT	Chronic HF patients with implantable defibrillators	180	0.76	0.21

EQ-5D EuroQol-5D, *HF* heart failure, *N* number of observations, *RCT* randomized controlled trial, *SD* standard deviation

[a]Link unavailable. Abstract based on patients who had ventricular assist device Heart-Mate II implant as a bride-to-transplant or destination therapy

Box 1 Set-up the data in R

```
# load the packages

library(brms)
library(mfp)
library(mice)
library(bayesplot)
library(shinystan)
library(tidybayes)
library(ggplot2)

# load the data

dat <- read.csv("Utilities.csv")
```

We load the dataset using the 'read.csv()' function. The dataset contains four variables: *studyid,* which refers to the first author and year of publication; *n*, the study sample size; *utility*, the reported or calculated mean health state utility per study; and *sd*, the standard deviation.

3.2 Employ Methods to Impute Standard Deviations

Commonly, meta-analyses exclude studies that lack a measure of variability. However, there are several approaches to deal with missing standard deviations in meta-analysis [48], and the choice of imputation method usually depends on the nature of the outcome variable or data structures. For the purpose of demonstration, two approaches are shown in this tutorial to impute missing standard deviations of utility values. The first method involves fitting a regression model using fractional polynomials based on the methods of Royston and Altman [49]. This method has been applied in a previous meta-analysis of utilities of chronic kidney disease

patients [50]. The second method involves multiple imputation using chained equations. The first approach is applied in the main analysis, while the second approach was used in the sensitivity analysis.

To fit a fractional polynomial regression model of the observed standard deviations against the utility estimates, the *mfp* function from the *mfp* package is used [42] (**Box 2**). The advantage of using fractional polynomial regression is that it allows for non-linear modelling of relationships between variables [49]. Briefly, the *mfp* package works by fitting several models with different combinations of fractional polynomial transformations of the predictor variable, and then selects the model that best fits the data through a stepwise model selection approach. In our model, *sd* is the dependent variable, while *utility* is the predictor. The term 'fp(utility)' tells R that we want to investigate different fractional polynomials of *utility*. The results are stored in an object called *sd.model1*.

Box 2 Employ methods to impute missing standard deviations

```
# fit a fractional polynomial regression model

sd.model1 <-  mfp(formula = sd ~ fp(utility), data = dat)

# examine the results

sd.model1

# extract the coefficients of the best model

coef <- coef(sd.model1)

# generate the equation and impute missing standard deviations

dat$sd.imp1 <- ifelse(is.na(dat$sd), (coef[1] + coef[2]*dat$utility^1),
                      dat$sd)

# calculate the standard error

dat$se.imp1 <- dat$sd.imp1/sqrt(dat$n)
```

The output shows that given our data, the best model for predicting the standard deviation of a utility estimate is a simple linear model with the following equation: sd = $0.558 - 0.474 \times$ utility. We also calculate the standard error of each utility estimate since this is required as an input for fitting the Bayesian meta-analysis model using the *brms* package in the subsequent step.

3.3 Define the Priors

One of the advantages of employing Bayesian methods lies in their capacity to integrate prior knowledge or beliefs into the analytical process, and combine this prior knowledge with the observed data to update parameter estimates.

Prior distributions are specified for each parameter in the model, and in the context of Bayesian meta-analysis, this means that we can directly model our assumptions about two parameters of interest: (1) the effect size (i.e. mean health state utility) and (2) the between-study heterogeneity *tau*. The inclusion of prior information for these parameters can help improve the precision of the estimates, particularly when dealing with a limited number of studies or highly variable data, by shrinking estimates towards more plausible values [51].

In general, priors can be classified into three types: flat, informative and weakly informative. Flat priors are typically used when one wants to input as little information as possible about the parameters of interest, thus assigning equal

probability to all possible parameter values. In contrast, informative priors incorporate specific prior knowledge, for instance, from previous research or literature, which can influence the plausibility of some parameter values. Lastly, weakly informative priors fall between flat and informative priors, and provide some information to guide the analysis without strongly influencing the results. These are usually employed when there is some prior knowledge about the parameters of interest, but not strong enough to justify a more constrained prior [11]. While Bayesian meta-analysis offers flexibility in including priors in the model, it is important to carry out sensitivity analysis to assess if specifying different prior information affect the results.

We define the priors using the 'prior()' function from the *brms* package (**Box 3**). This function takes two arguments, the prior distribution and the class. For illustration purposes, we use a Normal prior centred at 0.5 with a standard deviation of 0.05 for the mean health state utility, but other priors can be used [7, 8]. We set the class as *intercept* since it is a fixed population-level effect. For the between-study heterogeneity, we use a half-Cauchy prior to restrict values to positive numbers (since standard deviations cannot be negative), with a peak of 0 and a scale of 0.5. We set the class as *sd* since it is a measure of variability. The priors are saved into an object called *priors.model1*.

Box 3 Define the priors and perform prior predictive check

```
# define the priors

priors.model1 <- c(prior(normal(0.5,0.05), class = Intercept),
                    prior(cauchy(0,0.5), class = sd))

# conduct prior predictive check

fit.prior <- brm(formula = utility | se(se.imp1) ~ 1 + (1|studyid),
                 data = dat,
                 prior = priors.model1,
                 iter = 4000,
                 sample_prior = "only")

# examine the plot

pp_check(fit.prior, ndraws = 50)
```
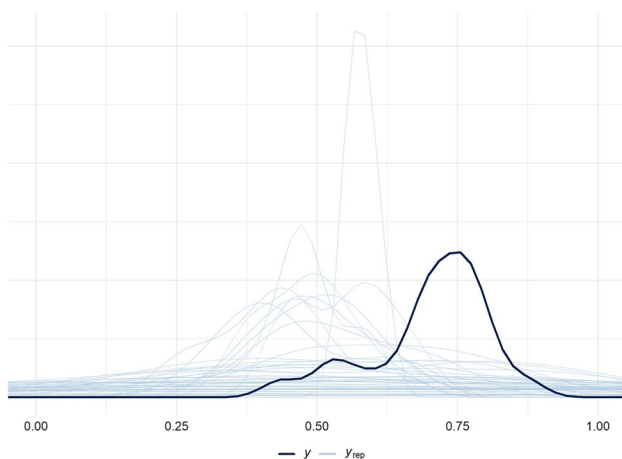

Fig. 1 Prior predictive check. The black line represents the observed data, while the blue lines represent the simulated data based on the prior distributions specified for the model parameters.

The adequacy of the priors can be checked by performing prior predictive checks [44]. A prior predictive check involves generating simulated data based on the chosen prior distribution and comparing it to the observed data. In *brms*, this is carried out by fitting the model (the arguments of the *brm* function are explained in the next section) and including the 'sample_prior = "only"' argument [40]. The 'pp_check()' function is used to plot the simulated data points and the observed data. In our example, the prior predictive check showed that the simulated data aligns with our expectation (that the utility value is around 0.5 and has a standard deviation drawn from the half-Cauchy prior) and covers the range of the observed data (Fig. 1). It is to note that if the simulated data from the prior predictive distribution consistently diverge from the observed data, it implies that the suggested priors are at odds with the data and likely need to be reconsidered.

3.4 Fit the Model

The *brm* function from the *brms* package is used to fit the Bayesian meta-analysis model. The *brms* package uses the No-U-Turn sampler (NUTS) to find and draw samples from the posterior distribution [40, 52]. The NUTS algorithm is considered better than traditional Markov Chain Monte Carlo (MCMC) in terms of efficiency, adaptability and scalability [11, 52]. The NUTS algorithm is packaged into Stan [11], which the *brms* package applies for fitting Bayesian multilevel models.
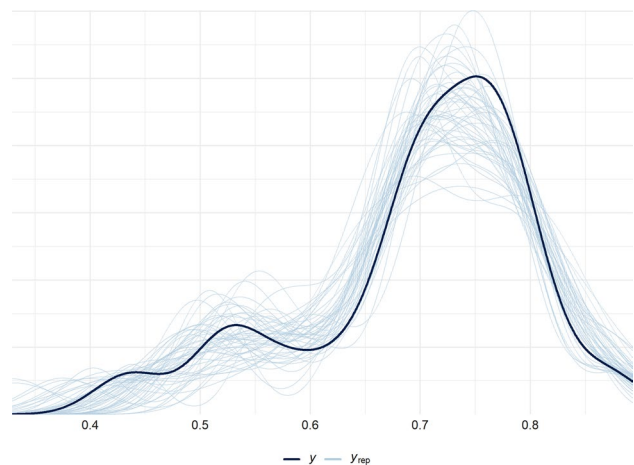
For our example, we define the model for our meta-analysis by specifying the *formula*, *data*, *prior* and *iter* (**Box 4**). The *formula* argument follows the standard regression notation, with some modifications since we are doing a meta-analysis. The part 'utility | se(se.imp1) ~ 1' indicates that our outcome is the utility value weighted according to the standard error of each study and that we do not have any predictors in the model. However, if one wishes to perform a meta-regression to account for factors that could potentially influence the pooled utility value (for instance, the year of study, study design, or the instrument used to elicit utility), then the syntax would be replaced by 'utility | se(se.imp1) ~ covariates'. It is worthy to note that inclusion of covariates into the model requires specifying priors for those parameters as well. The part '+ (1 | studyid)' indicates that the utility values are assumed to be nested within studies and as such we want to use a random-effects model. We specify our

Fig. 2 Posterior predictive check. The black line represents the observed data, while the blue lines show the simulated data based on the posterior distribution of utility (which takes into account both the observed data and the prior distributions of the model parameters).

dataset in the *data* argument, the priors for the effect size and between-study heterogeneity in the *prior* argument (which we have already set in the previous step), and the number of iterations per chain in the *iter* argument. By default, the *brm* function runs four chains. We save the fitted model into an object called *fit.model1*.

Box 4 Fit the model

```
fit.model1 <- brm(formula = utility | se(se.imp1) ~ 1 + (1|studyid),
                  data = dat,
                  prior = priors.model1,
                  iter = 4000)
```

3.5 Diagnose Model Convergence

Several tools are available to evaluate model convergence. By default, the *brm* output offers two convergence metrics: the Gelman–Rubin convergence diagnostic (i.e. Rhat) [53] and the number of effective sample size (i.e. bulk_ESS and tail_ESS). Rhat serves as a numerical summary for evaluating convergence. In practical applications, many researchers employ a threshold value greater than 1.1 to indicate non-convergence [54]. The effective sample size refers to the number of independent samples from the posterior distribution after taking into account autocorrelation of chains [11]. A low effective sample size indicates high autocorrelation, which means that the sequential samples are closely related to the previous one, rendering the chains inefficient.

As a rough guide, both bulk_ESS and tail_ESS should be at least 100 per chain to be able to consider the estimates reliable [55]. Additionally, graphical diagnostics can also be used to assess model convergence, for example, using a trace plot and a posterior predictive check plot. If the model has converged well, we can expect a trace plot with a stable path and good mixing, and a posterior predictive check plot where the density of the generated effect size aligns with the observed data.

We evaluate the Rhat, bulk_ESS and tail_ESS using the 'summary()' function (**Box 5**). The 'plot()' function from the *brms* package displays both the density and trace plots for the parameters of interest (in our case, the effect size and between-study heterogeneity), whereas the 'mcmc_trace()' function from the *bayesplot* package displays just the trace

plot. The 'pp_check()' function is used to display the posterior predictive check plot. This function works by drawing samples of model parameters from the posterior distribution and generating simulated data points that match the structure of the observed data. Lastly, the 'launch_shinystan()' function from the *shinystan* package opens an interactive window where we can further examine diagnostic plots and assess the performance of the model.

that the simulated effect sizes aligned with the observed effect size, particularly at the tails of the distribution (Fig. 2).

3.6 Interpret the Results

Guidance about interpreting the results from Bayesian analysis is available [11, 56]. In our example, we interpret the results by looking at the pooled effect size and between-study heterogeneity in the summary output (**Box 6**). The pooled

Box 5 Diagnose model convergence

```
# examine rhat and ess
summary(fit.model1)
# create a trace plot
plot(fit.model1, variable = c("b_Intercept","sd_studyid__Intercept"))
mcmc_trace(fit.model1, pars = c("b_Intercept","sd_studyid__Intercept"))
# conduct posterior predictive checks
pp_check(fit.model1, ndraws = 50)
# further checks through shinystan
launch_shinystan(fit.model1)
```

Model diagnostics showed that our model achieved convergence. No parameter had an Rhat above 1.1 or bulk_ESS and tail_ESS of less than 400 (100 × 4 chains) for both effect size and between-study heterogeneity parameters. There were no divergent transitions recorded. The trace plot showed stationarity and good mixing (see Figure S1 in the electronic supplementary material). The posterior predictive check plot showed
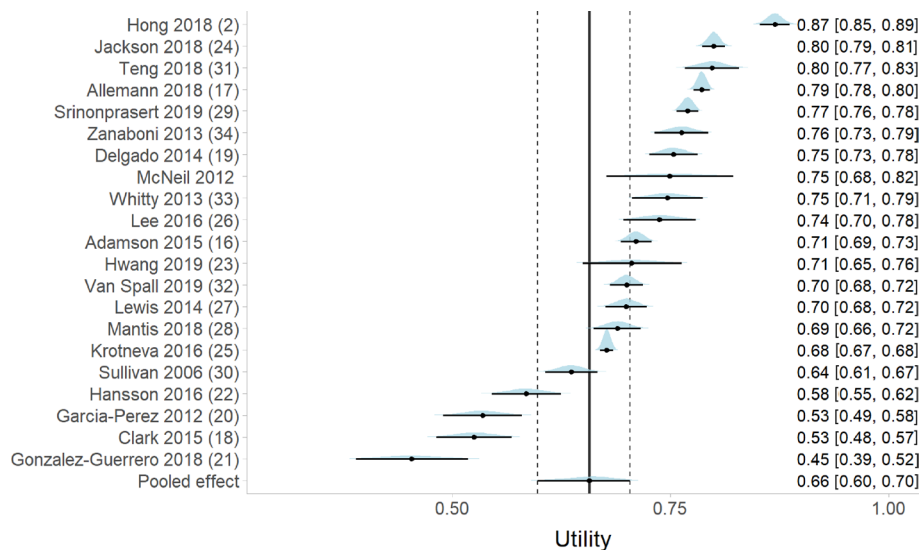
effect size, in our case the pooled utility value, is 0.66 with a 95% credible interval (CrI) of 0.60–0.70, given the data, priors and model used. The between-study heterogeneity *tau* is 0.12 (95% CrI 0.08–0.18). Since we fitted a random-effects model under the assumption that each study has its unique effect size, we can also look at the study-specific effect sizes (by summing up the pooled effect size and the deviations from each study) using the 'ranef()' function.

Fig. 3 Forest plot showing the posterior distribution of effect sizes for each study and the pooled effect size

Study	Utility [95% CrI]
Hong 2018 (2)	0.87 [0.85, 0.89]
Jackson 2018 (24)	0.80 [0.79, 0.81]
Teng 2018 (31)	0.80 [0.77, 0.83]
Allemann 2018 (17)	0.79 [0.78, 0.80]
Srinonprasert 2019 (29)	0.77 [0.76, 0.78]
Zanaboni 2013 (34)	0.76 [0.73, 0.79]
Delgado 2014 (19)	0.75 [0.73, 0.78]
McNeil 2012	0.75 [0.68, 0.82]
Whitty 2013 (33)	0.75 [0.71, 0.79]
Lee 2016 (26)	0.74 [0.70, 0.78]
Adamson 2015 (16)	0.71 [0.69, 0.73]
Hwang 2019 (23)	0.71 [0.65, 0.76]
Van Spall 2019 (32)	0.70 [0.68, 0.72]
Lewis 2014 (27)	0.70 [0.68, 0.72]
Mantis 2018 (28)	0.69 [0.66, 0.72]
Krotneva 2016 (25)	0.68 [0.67, 0.68]
Sullivan 2006 (30)	0.64 [0.61, 0.67]
Hansson 2016 (22)	0.58 [0.55, 0.62]
Garcia-Perez 2012 (20)	0.53 [0.49, 0.58]
Clark 2015 (18)	0.53 [0.48, 0.57]
Gonzalez-Guerrero 2018 (21)	0.45 [0.39, 0.52]
Pooled effect	0.66 [0.60, 0.70]

Utility

Box 6 Interpret the results

```
# examine the pooled effect size and between-study heterogeneity

summary(fit.model1)

# examine the deviation of each study from the overall effect size

ranef(fit.model1)
```

Bayesian meta-analysis naturally provides the posterior distribution of the pooled effect, which we can examine and use to make explicit probability statements regarding our parameters of interest [10]. We can extract the parameters of interest from the fitted model using the 'posterior_samples()' function, and then perform manual calculations or use the 'ecdf()' function to calculate posterior probabilities (**Box 7**). The *ecdf* function takes a value or set of values as input and returns the cumulative probabilities associated with those values. In our example, the probability that the pooled utility value is less than 0.80 is 100.00%, while the probability that it is less than 0.70 is 96.06%. Figure S2 displays the cumulative posterior distribution plot, which shows the cumulative probability associated with values less than or equal to each utility value on the *x*-axis (see the electronic supplementary material).

Box 7 Calculate posterior probabilities

```
# extract parameters of interest from the fitted model

model1.post <- posterior_samples(fit.model1, pars = "b_Intercept")

# rename the parameter

names(model1.post) <- "utility"

# explore posterior probabilities (manual)

(sum(model1.post$utility < 0.80) / length(model1.post$utility))*100

(sum(model1.post$utility < 0.70) / length(model1.post$utility))*100

# explore posterior probabilities (using ecdf)

model1.ecdf <- ecdf(model1.post$utility)

model1.ecdf(c(0.80, 0.70))
```

Table 2 Summary of results and convergence diagnostics

Model	Utility value	95% CrI	Rhat	Bulk_ESS	Tail_ESS
Model 1: with imputed SDs using fractional polynomial regression	0.66	0.60–0.70	1.00	850	1240
Model 2: with more informative priors	0.66	0.61–0.70	1.00	835	1653
Model 3: with imputed SDs using *mice*	0.66	0.60–0.70	1.00	3665	5583
Model 4: excluding studies with missing SDs	0.63	0.55–0.70	1.00	862	1247

CrI credible interval, *ESS* effective sample size, *mice* multiple imputation using chained equations, *Rhat* Gelman–Rubin convergence diagnostic, *SD* standard deviation

Lastly, we can also generate a forest plot by following the step-by-step guide from the *tidybayes* package [46] (Fig. 3). Note that this requires installation of other packages. The full code is provided in the script.

3.7 Perform Sensitivity Analysis

For this tutorial, we explored the results of using a more informative prior for the mean health state utility, employing multiple imputation using chained equations to impute missing standard deviations, and excluding studies with missing standard deviations. For the first sensitivity analysis, a Normal prior centred at 0.6 with a standard deviation of 0.03 is used (**Box 8**). This is regarded as being more informative than the prior used in the main analysis since it is above the midpoint of possible utility values with a smaller standard deviation. Given that mean utility

is skewed and has a lower limit) or truncated log-normal distribution reversed to have an upper limit. Using a beta prior distribution might also be an appropriate choice, as it aligns well with the above constraint and could potentially reflect the uncertainty around utility values better [57]. For example, in the script, this could be implemented by using 'prior(beta(1,1), class = Intercept)' rather than 'prior(normal(0.5,0.05)'. Lastly, if the analysis conceives of the possibility of negative utilities, has relatively small number of observations and anticipates results around zero, it may be necessary to define a lower boundary to the utilities. The prior for the between-study heterogeneity can also be changed, but is kept the same as the main analysis in this example for simplicity. Guidance on the selection of prior distributions for the between-study heterogeneity parameter is available [58, 59].

Box 8 Sensitivity analysis 1: using a more informative prior

```
# define a more informative prior for the intercept

priors.model2 <- c(prior(normal(0.6,0.03), class = Intercept),
                   prior(cauchy(0,0.5), class = sd))

# re-fit the model

fit.model2 <- brm(formula = utility | se(se.imp1) ~ 1 + (1|studyid),
                  data = dat,
                  prior = priors.model2,
                  iter = 4000)
```

values are generally bounded between 0 and 1 (although, at the individual participant level, they may occasionally take negative values), it is important to choose a prior distribution that will not violate these bounds. When the number of observations is large, the likelihood will be more important than the prior, and the posterior distribution will approach a normal distribution. In such scenarios, the normal distribution is a justifiable choice for the prior, and is easy for non-statistical collaborators to understand. However, when utility values are likely to be close to the boundaries of possible values (0 or 1), skewness will be introduced that only an extremely large number of observations will overcome. In these cases, a different prior distribution is justified. Options include a truncated normal distribution, a truncated log-normal distribution (which

For the second sensitivity analysis, we use the 'mice()' function from the *mice* package. Multiple imputation using chained equations involves generating several datasets with imputed values for the missing data. Within each dataset, missing values are filled in one at a time, using the observed values and other variables in the dataset. The process iterates, refining the imputed values based on the results of the previous round, until all missing data have been imputed. In our example, we generate five sets of imputed standard errors (**Box 9**). We use the *brm_multiple* function (rather than the *brm* function) from the *brms* package to fit the model since it is compatible with fitting multiple imputed datasets generated by *mice* and pooling the posterior distributions of those imputed datasets [40].

Box 9 Sensitivity analysis 2: using multiple imputation using chained equations

```
# calculate the standard error for available studies

dat$se <- dat$sd/sqrt(dat$n)

# use mice function to generate 5 sets of imputed SEs

dat.mice <- mice(dat[,c("studyid","utility","se")], m = 5,
                 print = FALSE)

# re-fit the model

fit.model3 <- brm_multiple(formula = utility | se(se) ~ 1 + (1|studyid),
                           data = dat.mice,
                           prior = priors.model1,
                           iter = 4000)
```

For the last sensitivity analysis, only studies with available standard deviations were included (**Box 10**). The results are summarized in Table 2. The sensitivity analyses conducted did not materially change the pooled utility value and the 95% CrI. The model excluding the studies with missing SDs (model 4) generated less precise estimates compared to the other models, while the model with more informative priors for the effect size (model 2) yielded slightly narrower CrIs. All models achieved convergence. Model 3 has the highest bulk_ESS and tail_ESS since five imputed datasets were used and pooled to produce the estimates.

3.8 Comparison to Frequentist Approach

Random-effects meta-analysis using the DerSimonian and Laird method were also carried out using the *metafor* package [60] to compare the results between the frequentist and Bayesian approaches. The codes for the frequentist meta-analysis are included in the script. The three frequentist models (i.e. with imputed SDs using fractional polynomial regression, with imputed SDs using *mice*, and excluding studies with missing SDs) produced identical pooled utility values (0.70), which were slightly higher than the values from their Bayesian counterparts (Table 3). As expected, the confidence interval widths were narrower than the CrIs,

Box 10 Sensitivity analysis 3: excluding studies with missing SDs

```
fit.model4 <- brm(formula = utility | se(se) ~ 1 + (1|studyid),
                  data = dat,
                  prior = priors.model1,
                  iter = 4000)
```

Table 3 Comparison between Bayesian and frequentist meta-analysis results

Model	Bayesian		Frequentist	
	Mean utility (95% CrI)	Tau (95% CrI)	Pooled utility (95% CI)	Tau
Model 1: with imputed SDs using fractional polynomial regression	0.66 (0.60–0.70)	0.12 (0.08–0.18)	0.70 (0.67–0.73)	0.07
Model 2: with imputed SDs using *mice*	0.66 (0.60–0.70)	0.12 (0.08–0.18)	0.70 (0.66–0.74)[a]	D1: 0.10 D2: 0.10 D3: 0.09 D4: 0.09 D5: 0.09
Model 3: excluding studies with missing SDs	0.63 (0.55–0.70)	0.15 (0.09–0.24)	0.71 (0.66–0.75)	0.09

CI confidence interval, *CrI* credible interval, *D* dataset, *mice* multiple imputation using chained equations, *SD* standard deviation

[a]Similar across the 5 imputed datasets

and the taus were lower in the frequentist models than in the Bayesian models. Lastly, all *p* values produced from the frequentist models were statistically significant ($p < 0.05$).

4 Discussion

This paper outlines the fundamental steps in conducting Bayesian meta-analysis of utilities in R. By providing an illustrative example with data and codes, the paper highlights the applicability of Bayesian modelling in synthesizing utility values.

The tutorial benefits from following a clear workflow for conducting Bayesian meta-analysis. This workflow is adaptable to a wide range of statistical problems, and encourages the clear and transparent communication of assumptions, prior beliefs, and data, to increase rigor and replicability in research. The tutorial also benefits from using the *brms* package [40]—a powerful and versatile tool for fitting Bayesian models using Stan. The *brms* package greatly simplifies the model specification process since it follows the coding language in other widely used R packages (e.g. the *lme4* package [62]). This makes Bayesian modelling more accessible and approachable to individuals without a deep understanding of Stan, and also eases the transition of individuals who are already familiar with R to Bayesian modelling. Nonetheless, users have the flexibility to select from a range of available software and packages for conducting Bayesian meta-analysis based on their preferences and specific requirements.

The sensitivity analysis was deemed as an insightful exercise, showcasing some approaches that can be implemented to deal with missing standard deviations. This is essential in the context of economic evaluations, given that reporting of measures of variability around utility values is poor [8], despite it being promoted as good practice by HTA agencies [63]. It has been shown that imputing missing standard deviations in meta-analyses is generally better than excluding studies [48, 64]. It is worth noting that the methods presented here (i.e. fractional polynomial regression and multiple imputation using chained equations) are both executed prior to model fitting, meaning that missing data are filled-in before running the model. In contrast, imputation can be built into the Bayesian meta-analysis rather than a two-step process, which is considered as a superior approach since imputation is integrated into the model fitting process [48, 65]. However, it is computationally demanding and requires programming in the underlying Stan software, and falls outside the scope of this tutorial. This method allows extra flexibility by allowing the inclusion of prior information and uncertainties related to the missing data, updating these uncertainties through the sharing of information within the hierarchical structure of the model, and ultimately generating a posterior distribution for each missing data point [11].

The sensitivity analysis also showed that using a more informative prior did not change the results, suggesting that the data had more influence on the analysis than the prior in our example. In Bayesian analysis, the consideration of priors is a crucial aspect since they can impact the final results and conclusions drawn from the analyses. It is therefore essential to carefully select and specify priors for each parameter of interest, and perform sensitivity analyses to check how different initial states of the model can affect the estimates [11].

The comparison of Bayesian and frequentist meta-analytic approaches produced roughly similar results. Yet, the interpretation between these two approaches differs. For the Bayesian approach, the 95% CrI is easier to interpret, such that it indicates the 95% probability that the pooled utility value lies between the lower and upper limits of the interval. On the contrary, interpreting the 95% confidence interval presents more challenge, as it implies conducting the analysis repeatedly with the assumption that 95% of the generated confidence intervals will contain the true value. The significant *p* value from the frequentist models is a useless test in the context of pooling utility values since it does not provide meaningful information in terms of interpretation of findings. The posterior distribution, on the other hand, can be used and explored to estimate parameters, calculate posterior probabilities and quantify uncertainties. In terms of the tau, the frequentist models appeared to underestimate the level of heterogeneity between studies. Although not to a great extent in our example, frequentist meta-analysis had been shown to perform poorly when there is a high between-study heterogeneity and small number of studies included in the meta-analysis [66]. These issues are addressed in the Bayesian model by inclusion of priors for tau and effect size, which helps in handling the uncertainty around these parameters. Thus, while the frequentist approach is easier to implement in practice (in terms of speed and simplicity, as it could be implemented with a few lines of codes), the Bayesian approach offers several advantages that makes it suitable for the meta-analysis of health state utility values.

5 Conclusion

In conclusion, Bayesian method offers several advantages when conducting meta-analysis of utility values. Its ability to incorporate prior information, handle heterogeneity between studies explicitly and provide intuitive probabilistic interpretations make it a valuable tool for synthesizing utility data. In this tutorial, we provided a pooled utility value (and its CrI) for patients with heart failure, which can be used as an input for economic evaluations. We hope that this fosters an interest in Bayesian methods and their applications in meta-analysis of utilities.

Declarations

References

1. Tosh JC, Longworth LJ, George E. Utility values in National Institute for Health and Clinical Excellence (NICE) technology appraisals. Value Health. 2011;14(1):102–9.
2. Hong SH, Lee JY, Park SK, Nam JH, Song HJ, Park SY, et al. The utility of 5 hypothetical health states in heart failure using Time Trade-Off (TTO) and EQ-5D-5L in Korea. Clin Drug Investig. 2018;38(8):727–36.
3. Wolowacz SE, Briggs A, Belozeroff V, Clarke P, Doward L, Goeree R, et al. Estimating health-state utility for economic models in clinical studies: an ISPOR good research practices task force report. Value Health. 2016;19(6):704–19.
4. Petrou S, Kwon J, Madan J. A Practical guide to conducting a systematic review and meta-analysis of health state utility values. Pharmacoeconomics. 2018;36(9):1043–61.
5. Papaioannou D, Brazier J, Paisley S. Systematic searching and selection of health state utility values from the literature. Value Health. 2013;16(4):686–95.
6. Ara R, Brazier J, Peasgood T, Paisley S. The identification, review and synthesis of health state utility values from the literature. Pharmacoeconomics. 2017;35(Suppl 1):43–55.
7. Hatswell AJ, Burns D, Baio G, Wadelin F. Frequentist and Bayesian meta-regression of health state utilities for multiple myeloma incorporating systematic review and analysis of individual patient data. Health Econ. 2019;28(5):653–65.
8. Hatswell AJ. Incorporating prior beliefs into meta-analyses of health-state utility values using the bayesian power prior. Value Health. 2023;26(9):1389–97.
9. Chen M-H, Ibrahim JG. Power prior distributions for regression models. Stat Sci. 2000;15(1):46–60, 15.
10. Spiegelhalter DJ. Incorporating Bayesian ideas into health-care evaluation. Stat Sci. 2004;19(1):156–74.
11. McElreath R. Statistical Rethinking: A Bayesian Course with Examples in R and Stan. Dominici F, Faraway J, Tanner M, Zidek J, editors. 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742: CRC Press: Taylor & Francis Group; 2016.
12. Gelman A, Vehtari A, Simpson D, Margossian CC, Carpenter B, Yao Y, et al. Bayesian Workflow. 2020. arXiv:2011.01808.
13. Goligher EC, Harhay MO. What is the point of Bayesian analysis? Am J Respir Crit Care Med. 2024;209(5):485–7.
14. Di Tanna GL, Urbich M, Wirtz HS, Potrata B, Heisen M, Bennison C, et al. Health state utilities of patients with heart failure: a systematic literature review. Pharmacoeconomics. 2021;39(2):211–29.
15. Di Tanna GL, Bychenkova A, O'Neill F, Wirtz HS, Miller P, Hartaigh BÓ, et al. Evaluating cost-effectiveness models for pharmacologic interventions in adults with heart failure: a systematic literature review. Pharmacoeconomics. 2019;37(3):359–89.
16. Adamson PB, Bharmi R, Bauman J, Dalal N, Martinson M, Abraham WT. Cost effectiveness assessment of pulmonary artery pressure monitoring for heart failure management (AB36-01). Paper presented at: Heart Rhythm Society 36th Annual Scientific Sessions, May 13–15, 2015; Boston, MA. 2015.
17. Allemann H, Strömberg A, Thylén I. Perceived social support in persons with heart failure living with an implantable cardioverter defibrillator: a cross-sectional explorative study. J Cardiovasc Nurs. 2018;33(6):E1-e8.
18. Clark AL, Johnson M, Fairhurst C, Torgerson D, Cockayne S, Rodgers S, et al. Does home oxygen therapy (HOT) in addition to standard care reduce disease severity and improve symptoms in people with chronic heart failure? A randomised trial of home oxygen therapy for patients with chronic heart failure. Health Technol Assess. 2015;19(75):1–120.
19. Delgado JF, Oliva J, Llano M, Pascual-Figal D, Grillo JJ, Comín-Colet J, et al. Health care and nonhealth care costs in the treatment of patients with symptomatic chronic heart failure in Spain. Rev Esp Cardiol (Engl Ed). 2014;67(8):643–50.
20. García-Pérez L, Linertová R, Pinilla-Domínguez P, Dávila-Ramos M, Copca-Álvarez A, Ruiz-Hernández JJ, Díaz-Escofet M, Escobar A. EQ-5D utilities in patients hospitalised with heart failure in Canary Islands. PCV99 2012.
21. González-Guerrero JL, Hernández-Mocholi MA, Ribera-Casado JM, García-Mayolín N, Alonso-Fernández T, Gusi N. Cost-effectiveness of a follow-up program for older patients with heart failure: a randomized controlled trial. Eur Geriatr Med. 2018;9(4):523–32.
22. Hansson E, Ekman I, Swedberg K, Wolf A, Dudas K, Ehlers L, et al. Person-centred care for patients with chronic heart failure - a cost-utility analysis. Eur J Cardiovasc Nurs. 2016;15(4):276–84.
23. Hwang R, Morris NR, Mandrusiak A, Bruning J, Peters R, Korczyk D, et al. Cost-Utility analysis of home-based telerehabilitation compared with centre-based rehabilitation in patients with heart failure. Heart Lung Circ. 2019;28(12):1795–803.
24. Jackson JD, Cotton SE, Bruce Wirta S, Proenca CC, Zhang M, Lahoz R, et al. Burden of heart failure on caregivers in China: results from a cross-sectional survey. Drug Des Devel Ther. 2018;12:1669–78.
25. Krotneva S, Kansal AR, Zheng Y, Patel HK, Kielhorn A, Böhm M, et al. Abstract 16738: estimation of decrements

of utility associated with hospitalizations in a population with heart failure from the Systolic Heart Failure Treatment with the If Inhibitor Ivabradine Trial (SHIFT). Circulation. 2016;134(suppl_1):A16738.

26. Lee JY, Lee E. Assessment of utility for heart failure using Visual Analogue Scale (VAS), Time-Trade off (TTO) and Euroqol-5 Dimension (EQ-5D) in the Korean General Population. Value Health. 2016;19(7):A868–9.

27. Lewis EF, Li Y, Pfeffer MA, Solomon SD, Weinfurt KP, Velazquez EJ, et al. Impact of cardiovascular events on change in quality of life and utilities in patients after myocardial infarction: a VALIANT study (valsartan in acute myocardial infarction). JACC Heart Fail. 2014;2(2):159–65.

28. Mantis C, Anadiotis A, Patsilinakos S. Impact of sacubitril/valsartan on functional exercise capacity and quality of life in patients with heart failure with reduced ejection fraction. Eur J Prev Cardiol. 2018;25(S73).

29. Srinonprasert V, Ratanasumawong K, Thongsri T, Dutsadeevettakul S, Jittham P, Wiwatworapan W, et al. Factors associated with low health-related quality of life among younger and older Thai patients with non-valvular atrial fibrillation. Qual Life Res. 2019;28(8):2091–8.

30. Sullivan PW, Ghushchyan V. Preference-based EQ-5D index scores for chronic conditions in the United States. Med Decis Making. 2006;26(4):410–20.

31. Teng HC, Yeh ML, Wang MH. Walking with controlled breathing improves exercise tolerance, anxiety, and quality of life in heart failure patients: a randomized controlled trial. Eur J Cardiovasc Nurs. 2018;17(8):717–27.

32. Van Spall HGC, Lee SF, Xie F, Oz UE, Perez R, Mitoff PR, et al. Effect of patient-centered transitional care services on clinical outcomes in patients hospitalized for heart failure: the PACT-HF randomized clinical trial. JAMA. 2019;321(8):753–61.

33. Whitty JA, Stewart S, Carrington MJ, Calderone A, Marwick T, Horowitz JD, et al. Patient preferences and willingness-to-pay for a home or clinic based program of chronic heart failure management: findings from the Which? trial. PLoS ONE. 2013;8(3): e58347.

34. Zanaboni P, Landolina M, Marzegalli M, Lunati M, Perego GB, Guenzati G, et al. Cost-utility analysis of the EVOLVO study on remote monitoring for heart failure patients with implantable defibrillators: randomized controlled trial. J Med Internet Res. 2013;15(5): e106.

35. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2023.

36. Posit Team. RStudio: integrated development environment for R. Boston: Posit Software, PBC; 2023.

37. Plummer M. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. In: Proceedings of the 3rd international workshop on distributed statistical computing (DSC 2003), Vienna, 20–22 March 2003. p. 1–10.

38. Lunn D, Spiegelhalter D, Thomas A, Best N. The BUGS project: evolution, critique and future directions. Stat Med. 2009;28(25):3049–67.

39. JASP Team. JASP (Version 0.18.3)[Computer software]. 2024. https://jasp-stats.org/.

40. Bürkner P-C. Bayesian item response modeling in R with brms and Stan. J Stat Softw. 2021;100(5):1–54.

41. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: a probabilistic programming language. J Stat Softw. 2017;76(1):1–32.

42. Ambler G, Benner A. mfp: Multivariable fractional polynomials. R package version 1.5.4. 2023.

43. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. J Stat Softw. 2011;45(3):1–67.

44. Gabry J, Simpson D, Vehtari A, Betancourt M, Gelman A. Visualization in Bayesian workflow. J R Stat Soc Ser A Stat Soc. 2019;182(2):389–402.

45. Gabry J, Veen D. shinystan: Interactive visual and numerical diagnostics and posterior analysis for Bayesian models. R package version 2.6.0 ed2022.

46. Matthew Kay. tidybayes: Tidy data and Geoms for Bayesian models. R package version 3.0.6 ed2023.

47. Wickham H. ggplot2: Elegant graphics for data analysis. New York: Springer; 2016.

48. Weir CJ, Butcher I, Assi V, Lewis SC, Murray GD, Langhorne P, et al. Dealing with missing standard deviation and mean values in meta-analysis of continuous outcomes: a systematic review. BMC Med Res Methodol. 2018;18(1):25.

49. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. J R Stat Soc Ser C (Appl Stat). 1994;43(3):429–67.

50. Wyld M, Morton RL, Hayen A, Howard K, Webster AC. A systematic review and meta-analysis of utility-based quality of life in chronic kidney disease treatments. PLoS Med. 2012;9(9): e1001307.

51. Röver C, Friede T. Dynamically borrowing strength from another study through shrinkage estimation. Stat Methods Med Res. 2020;29(1):293–308.

52. Hoffman M, Gelman A. The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. J Mach Learn Res. 2014;15:1593–623.

53. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Stat Sci. 1992;7(4):457–72.

54. Gelman A, Carlin J, Stern H, Dunson D, Vehtari A, Rubin D. Bayesian data analysis third edition (with errors fixed as of 13 February 2020). 2020. http://www.stat.columbia.edu/~gelman/book/BDA3.pdf. Accessed Dec 2023.

55. Vehtari A, Gelman A, Simpson D, Carpenter B, Burkner PC. Rank-normalization, folding, and localization: an improved R for assessing convergence of MCMC. Bayesian Anal. 2021;16(2):667–718.

56. van de Schoot R, Kaplan D, Denissen J, Asendorpf JB, Neyer FJ, van Aken MAG. A gentle introduction to Bayesian analysis: applications to developmental research. Child Dev. 2014;85(3):842–60.

57. Blythe R, White N, Kularatna S, McPhail S, Barnett A. A Bayesian approach for incorporating the EQ-5D visual analog scale when estimating the health-related quality of life. Value Health. 2022;25(9):1575–81.

58. Turner RM, Jackson D, Wei Y, Thompson SG, Higgins JP. Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. Stat Med. 2015;34(6):984–98.

59. Röver C, Bender R, Dias S, Schmid CH, Schmidli H, Sturtz S, et al. On weakly informative prior distributions for the heterogeneity parameter in Bayesian random-effects meta-analysis. Res Synth Methods. 2021;12(4):448–74.

60. Viechtbauer W. Conducting meta-analyses in R with the metafor Package. J Stat Softw. 2010;36(3):1–48.

61. Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JP. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. Int J Epidemiol. 2012;41(3):818–27.

62. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67(1):1–48.

63. Husereau D, Drummond M, Augustovski F, de Bekker-Grob E, Briggs AH, Carswell C, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) 2022 explanation and elaboration: a report of the ISPOR CHEERS II good practices task force. Value Health. 2022;25(1):10–31.

64. Furukawa TA, Barbui C, Cipriani A, Brambilla P, Watanabe N. Imputing missing standard deviations in meta-analyses can provide accurate results. J Clin Epidemiol. 2006;59(1):7–10.

65. Halme AS, Tannenbaum C. Performance of a Bayesian approach for imputing missing data on the SF-12 health-related quality-of-life measure. Value Health. 2018;21(12):1406–12.

66. Seide SE, Röver C, Friede T. Likelihood-based random-effects meta-analysis with few studies: empirical and simulation studies. BMC Med Res Methodol. 2019;19(1):16.