



beanz: An R Package for Bayesian Analysis of Heterogeneous Treatment Effects with a Graphical User Interface

Chenguang Wang
Johns Hopkins University

Thomas A. Louis
Johns Hopkins University

Nicholas C. Henderson
Johns Hopkins University

Carlos O. Weiss
Mercy Health Physician Partners

Ravi Varadhan
Johns Hopkins University

Abstract

In patient-centered outcomes research, it is essential to assess the heterogeneity of treatment effects (HTE) when making health care decisions for an individual patient or a group of patients. Nevertheless, it remains challenging to evaluate HTE based on information collected from clinical studies that are often designed and conducted to evaluate the efficacy of a treatment for the overall population. The Bayesian framework offers a principled and flexible approach to estimate and compare treatment effects across subgroups of patients defined by their characteristics. In this paper, we describe the package **beanz** which facilitates the conduct of Bayesian analysis of HTE by allowing users to explore a wide range of Bayesian HTE analysis models and produce posterior inferences about HTE. The package **beanz** also provides a web-based graphical user interface (GUI) for users to conduct the Bayesian analysis of HTE in an interactive and user-friendly manner. With the GUI feature, package **beanz** can also be used by analysts not familiar with the R environment. We demonstrate package **beanz** using data from a randomized controlled trial on angiotensin converting enzyme inhibitor for treating congestive heart failure ($N = 2569$).

Keywords: Bayesian analysis, HTE, GUI, patient-centered outcomes research, R, **shiny**, Stan, subgroup analysis, web-based Bayesian analysis.

1. Introduction

Clinical trials typically evaluate the efficacy of a medical intervention based on its overall treatment effect, the average benefit or harm the enrolled patients will achieve. However,

individuals respond to medical interventions in different ways. Even for a treatment that is effective on average, some individuals may derive substantial benefit; some little benefit; others could actually be harmed. Thus, understanding this heterogeneity of treatment effect (HTE) is critical in evaluating how well a treatment can be expected to work for an individual patient. Such understanding is especially important in patient-centered outcomes research (PCOR) that aims to help patients and stakeholders make informed personalized health care decisions (Varadhan, Segal, Boyd, Wu, and Weiss 2013).

Although the importance of understanding HTE is obvious, reliable identification of HTE is challenging. Ideally, we would like to be able to estimate individual treatment effects (ITE). However, ITE estimation is not feasible because it is impossible to observe the outcomes of both the treatment and control for the same individual. This is the fundamental problem of causal inference (Holland 1986). Consequently, subgroup analysis is commonly used to evaluate HTE, but it is challenging due to the need to specify the groups and control for multiplicity in a low-information context.

The Bayesian approach provides a principled way to formulate treatment effect heterogeneity across patient characteristics. The approach includes a prior distribution that allows expression of optimism, skepticism or agnosticism concerning the differences in treatment effect. The Bayesian approach also allows a great deal of flexibility in making inferences on a variety of questions such as “what is the probability that women derive more benefit from the treatment than men?” A frequentist approach, on the other hand, would ask: “what is the probability of observing a difference in treatment effect between men and women, if in truth there were no difference?” This way of framing the question implies a premature dichotomization: the treatment effect is either the same in men and women or it is not, based on arbitrary conventions (e.g., the significance level is 0.05). While the Bayesian approach is principled and flexible, there are several challenges in conducting Bayesian HTE analysis, including the need to specify prior distributions, little guidance, and the lack of easy-to-use software to implement the HTE models.

There are a few software packages for subgroup analysis available for the statistical software environment R (R Core Team 2018) on the Comprehensive R Archive Network (CRAN), most of which implement frequentist methods. Package **quint** (Dusseldorp, Doove, and van Mechelen 2015) applies a tree-based clustering method for inducing subgroups that are involved in qualitative treatment-subgroup interactions. Package **subgroup** (Schou and Marschner 2014) considers order statistics-based measures to evaluate the magnitude and the nature of the variation in treatment effects and provides non-inferential visual aids. Package **FINDIt** (Egami, Ratkovic, and Imai 2018) applies machine learning methods to identify subsets of the population who benefit or are harmed by a treatment and to select the most or least efficacious treatment from many alternative treatments. Package **DSBayes** (Varadhan and Yao 2014) implements the Bayesian Dixon and Simon model (Dixon and Simon 1991) for subgroup analysis. To the best of our knowledge, there is no statistical software package for comprehensive Bayesian HTE analysis, let alone one with a graphical user interface (GUI).

To this end, we describe the R package **beanz** (Wang, Varadhan, and Trustees of Columbia University 2018) that facilitates conducting Bayesian HTE analysis. The package gives users the ability to explore a wide range of Bayesian HTE analysis models and obtain posterior inferences related to HTE. Specifically, we develop a web-based GUI for package **beanz** that allows users to apply functions in package **beanz** in an interactive and user-friendly manner. With the GUI feature, package **beanz** can also be used by analysts not familiar with the R

environment. Package **beanz** is available from the Comprehensive R Archive Network (CRAN) at <https://CRAN.R-project.org/package=beanz>.

In this paper, we use the data from the SOLVD trial (SOLVD Investigators *et al.* 1991) for the demonstration of package **beanz**. SOLVD was a randomized, double-blind, placebo-controlled study evaluating the efficacy of the drug Enalapril compared to placebo on mortality and hospitalization in patients with congestive heart failure (CHF). From 1986 to 1989, a total of 2569 patients with CHF and ejection fraction ≤ 0.35 were enrolled and randomly assigned to the Enalapril arm or the placebo arm in a 1:1 randomization ratio. At the end of the study, 735 patients had died or were hospitalized in the placebo group as compared with 613 in the Enalapril group. The overall treatment effect on mortality was statistically significant with log-rank test p value < 0.0001 . The SOLVD protocol pre-specified examination of subgroup effects for clinically important baseline factors. For the current demonstration we focus on baseline covariates sodium level, usage of vasodilators other than angiotensin-converting-enzyme inhibitors, and ejection fraction.

We emphasize that subgroup analysis are mainly exploratory, unless they were pre-specified in the study protocol at the design stage (Assmann, Pocock, Enos, and Kasten 2000; Pocock, Assmann, Enos, and Kasten 2002; Jones, Ohlssen, Neuenschwander, Racine, and Branson 2011). Moreover, although the software makes it easy to conduct sophisticated Bayesian modeling for subgroup analysis, there are intricate issues involved in such modeling. Chiefly, specifying the prior distribution for parameters in the model is challenging, and there is no automatic method that will handle all situations. In Henderson, Louis, Wang, and Varadhan (2016), we discussed the issue of prior selection and sensitivity analysis for Bayesian subgroup analysis in greater detail.

The remainder of the paper is organized as follows. In Section 2, we introduce the Bayesian HTE analysis models implemented in package **beanz**. In Section 3, we demonstrate the **beanz** package, used in the R interactive mode, with a comprehensive example using data from SOLVD. In Section 4, we describe the details of the GUI of package **beanz**, including its architecture and user manual. We revisit the SOLVD example to illustrate the GUI of package **beanz** in Section 5. Section 6 is devoted to the discussion.

2. Probability models

2.1. Notation

Consider a randomized two-arm clinical trial. For patient i , let Y_i denote the response and Z_i denote treatment arm assignment ($Z_i \in \{0, 1\}$). For subgroup analysis, assume there are P baseline covariates of interest, X_1, X_2, \dots, X_P , that are binary, ordinal with numerical values, or nominal. If a covariate is ordinal with character labels, then we assume the covariate can be re-coded with numerical values. For $j = 1, \dots, P$, assume $X_j \in \{x_{j,1}, \dots, x_{j,n_j}\}$ with $n_j \geq 2$. If X_j is binary or ordinal with numerical values, $x_{j,k}$ ($k = 1, \dots, n_j$) denotes the value of the k -th level of X_j . If X_j is a nominal variable, $x_{j,k}$ denotes the vector of dummy variables with dimension $n_j - 1$ that corresponds to the k -th level of X_j . For example, if X_j is patients' race with three levels: "white", "black" and "other", we may have $x_{j,1} = (0, 0)$ for "white", $x_{j,2} = (0, 1)$ for "black" and $x_{j,3} = (1, 0)$ for "other".

Let

$$\Omega = \left\{ (X_1, \dots, X_P) : X_j \in \{x_{j,1}, \dots, x_{j,n_j}\} \text{ for } j = 1, \dots, P \right\}$$

denote the collection of subgroups defined by X_1, \dots, X_P .

Let $G_i \in \Omega$ denote the subgroup for patient i , and $\mathbf{X}_g = (X_{g,1}, \dots, X_{g,P})$ be the covariates for subgroup $G = g$. Let θ_g^* denote the treatment effect in subgroup $G = g$. Note that θ_g^* may be measured on different scales. For example, if the response is continuous, θ_g^* may be measured as the mean difference by

$$\theta_g^* = \mathbb{E}(Y|Z = 1, G = g) - \mathbb{E}(Y|Z = 0, G = g).$$

If the response is binary, θ_g^* may be measured as the odds ratio by

$$\theta_g^* = \frac{\mathbb{P}(Y = 1|Z = 1, G = g) \mathbb{P}(Y = 0|Z = 0, G = g)}{\mathbb{P}(Y = 0|Z = 1, G = g) \mathbb{P}(Y = 1|Z = 0, G = g)}.$$

If the response is time-to-event, θ_g^* may be measured as the hazard ratio by

$$\theta_g^* = \frac{\lambda_{1,g}}{\lambda_{0,g}},$$

where $\lambda_{z,g}$ is the hazard rate of arm $Z = z$ in subgroup $G = g$.

We define $\theta_g = h(\theta_g^*)$ where h is a link function that allows the support of θ_g to be unrestricted. For example, we may let h be the identity link function for the mean difference of continuous responses and let h be the log link function for the odds ratio of binary and hazard ratio of time-to-event responses.

Lastly, we let $\hat{\theta}_g$ be the estimated θ_g with $\hat{\sigma}_g^2$ the estimated variance associated with $\hat{\theta}_g$.

2.2. Bayesian models

We consider the subgroup modeling strategy and selection of Bayesian hierarchical models suggested by Jones *et al.* (2011). First, assume that the estimated treatment effect $\hat{\theta}_g$ in subgroup $G = g$ approximately follows a normal distribution,

$$\hat{\theta}_g | \theta_g, \sigma_g^2 \sim N(\theta_g, \sigma_g^2). \quad (1)$$

Next, we consider assigning an informative prior to σ_g . For example, we can let

$$\log \sigma_g | \hat{\sigma}_g, \Delta \sim N(\log \hat{\sigma}_g, \Delta). \quad (2)$$

Alternatively, we may let

$$\log \sigma_g | \hat{\sigma}_g, \Delta \sim \text{Unif}(\log \hat{\sigma}_g - \Delta, \log \hat{\sigma}_g + \Delta). \quad (3)$$

In both cases, Δ is a non-negative parameter specified by the user. These priors are ‘‘centered’’ at the point estimate of the variance $\hat{\sigma}_g$ corresponding to $\hat{\theta}_g$. They account for the uncertainty in $\hat{\sigma}_g$ by means of Δ . Note that in Jones *et al.* (2011), $\hat{\sigma}_g$ was assumed known, which is equivalent to setting $\Delta = 0$, and assuming $\hat{\theta}_g | \theta_g \sim N(\theta_g, \hat{\sigma}_g^2)$.

Next, we consider a set of models together with the priors for parameters in θ_g . These models are listed in the following.

No subgroup effect model:

$$\begin{aligned}\theta_g &= \mu, \\ \mu &\sim N(0, B),\end{aligned}\tag{4}$$

where B is large in relation to the magnitude of the treatment effect size so that the prior for μ is essentially non-informative.

Note. This model assumes that patients in all the subgroups are common with similar characteristics. That is, all the subgroups are statistically identical with regard to the treatment effect and there is no subgroup effect. Information about treatment effects can be directly combined from all subgroups for inference.

Full stratification model:

$$\begin{aligned}\theta_g &= \mu_g, \\ \mu_g &\sim N(0, B).\end{aligned}\tag{5}$$

Note. The subgroups are fully distinguished from each other with regard to the treatment effect. There is no information about treatment effects shared between any subgroups.

Simple regression model:

$$\begin{aligned}\theta_g &= \mu + \sum_{j=1}^P X_{g,j}^\top \gamma_j, \\ \mu &\sim N(0, B), \\ \gamma_j &\sim N(0, \mathbf{1}C), \quad j = 1, \dots, P,\end{aligned}\tag{6}$$

where C is large in relation to the magnitude of γ so that the priors for γ are essentially non-informative.

Note. When X_j is nominal, $X_{g,j}$ and γ_j are vectors with dimension $n_j - 1$ and $\mathbf{1}$ is the $(n_j - 1) \times (n_j - 1)$ identity matrix. Otherwise, $X_{g,j}$ and γ_j are scalars and $\mathbf{1}$ is equal to the scalar 1. Unless otherwise specified, the same notation and modeling strategy also apply to the following models: *Dixon and Simon model* (7), *simple regression and shrinkage model* (9) and *extended Dixon and Simon model* (10).

The model introduces a first-order, linear regression structure for θ_g . This model takes into account the information that the subgroups are formulated based on the set of baseline covariates X_1, \dots, X_P . The coefficients γ_j 's are shared among subgroups, i.e., these coefficients are assumed to arise from the same distribution. Information about treatment effects are shared between subgroups with similar baseline covariates through these coefficients.

Dixon and Simon model (Dixon and Simon 1991):

$$\begin{aligned}\theta_g &= \mu + \sum_{j=1}^P X_{g,j}^\top \gamma_j, \\ \mu &\sim N(0, B), \\ \gamma_j &\sim N(0, \mathbf{1}\omega^2), \\ \omega &\sim \text{Half-}N(D),\end{aligned}\tag{7}$$

where Half- N denotes the Half-normal distribution with variance D . By definition, when $\omega' \sim N(0, D)$, $\omega = |\omega'|$ follows a Half- $N(D)$ distribution.

Note. This model assumes that the coefficients γ_j 's are shared across the subgroups, which allows information about the treatment effect to be shared between the subgroups. This model has exactly the same structure as the *simple regression model* (6). However, the *simple regression model* assumes that all the γ_j 's are independent, whereas this model allows them to be shared across the subgroups, leading to shrinkage of the subgroup-specific treatment effects towards each other.

Basic shrinkage model:

$$\begin{aligned}\theta_g &= \mu + \phi_g, \\ \mu &\sim N(0, B), \\ \phi_g &\sim N(0, \omega^2), \\ \omega &\sim \text{Half-}N(D).\end{aligned}\tag{8}$$

Note. Here, ϕ_g denotes a random subgroup effect for subgroup $G = g$ and, directly estimated subgroup treatment effects are shrunk towards the overall mean μ . This approach assumes all subgroups are similar with regards to the treatment effect.

Simple regression and shrinkage model:

$$\begin{aligned}\theta_g &= \mu + \sum_{j=1}^P X_{g,j}^\top \gamma_j + \phi_g, \\ \mu &\sim N(0, B), \\ \gamma_j &\sim N(0, \mathbf{1}C), \quad j = 1, \dots, P, \\ \phi_g &\sim N(0, \omega^2), \\ \omega &\sim \text{Half-}N(D).\end{aligned}\tag{9}$$

Note. This model combines basic regression with shrinkage, with a linear regression structure and a random effect term. Direct estimates are shrunk towards the regression surface.

Extended Dixon and Simon model (Jones *et al.* 2011):

$$\begin{aligned}\theta_g &= \mu + \sum_{k=1}^P \sum_{j \in \xi^{(k)}} \mathbf{X}_{\xi^{(k)},j}^\top \gamma_j^{(k)}, \\ \mu &\sim N(0, B), \\ \gamma_j^{(k)} &\sim N(0, \mathbf{1}\omega_k^2), \quad k = 1, \dots, P, \quad j \in \xi^{(k)}, \\ \omega_k &\sim \text{Half-}N(D),\end{aligned}\tag{10}$$

where $\xi^{(k)}$ denotes the set of k -th order interaction terms among X_1, \dots, X_P and $\mathbf{X}_{\xi^{(k)},j}$ denotes the j -th element in $\xi^{(k)}$. The dimension of $\gamma_j^{(k)}$ and $\mathbf{1}$ is the same as $\mathbf{X}_{\xi^{(k)},j}$.

Note. This approach extends the *Dixon and Simon model* (7) by introducing the higher-order interactions and sharing the interaction effects of the same order. The model may be highly sensitive to the choices of priors for ω_k when k is large.

2.3. Inference

Treatment effect evaluation and treatment effect comparison among subgroups are based on the posterior distribution $p(\theta_g | \hat{\theta}_g, \hat{\sigma}_g^2, \Delta)$.

3. The beanz package

3.1. Installation

The **beanz** package is available from CRAN at <http://CRAN.R-project.org/package=beanz>. To install and load package **beanz**, type the following in R:

```
R> install.packages("beanz")
R> library("beanz")
```

The major functions provided in package **beanz** are listed in Table 1.

3.2. Data format

There are two types of data structures that package **beanz** recognizes:

- *Patient level raw data.* Each row represents a patient with covariates that define the subgroup to which the patient belongs, treatment indicator and outcome. The outcome can be binary, continuous, or time to event.
- *Summary treatment effect data.* Each row represents a subgroup with covariates that define the subgroup, estimated treatment effect in the subgroup and variance for the estimation.

Package **beanz** provides dataset `solvd.sub` from the SOLVD trial as an example of *patient level raw data*. Table 2 reports the frequency of the covariates in `solvd.sub`.

Function	Description
<code>bzGetSubgrpRaw</code>	Get subgroup treatment effect estimation and variance for subject level data.
<code>bzCallStan</code>	Call Bayesian models for HTE analysis.
<code>bzSummary</code> <code>bzPlot</code> <code>bzForest</code>	Present the posterior subgroup treatment effects using table, density plots and forest plots.
<code>bzSummaryComp</code> <code>bzPlotComp</code> <code>bzForestComp</code>	Present the difference in the posterior treatment effects between subgroups using table, density plots and forest plots.
<code>bzPredSubgrp</code> <code>bzRptTbl</code>	Get the predictive distribution of the subgroup treatment effects. Overall summary of Bayesian HTE analysis.
<code>bzShiny</code>	Run package beanz using web-based GUI.

Table 1: Major functions in package **beanz**.

Covariates	Level	No. (%) of patients		
		Enalapril	Placebo	
Sodium (mmol/liter)	≤ 141	0	914 (72)	907 (72)
	> 141	1	355 (28)	359 (28)
Vasodilator	No	0	581 (46)	545 (43)
	Yes	1	688 (54)	721 (57)
Ejection fraction (%)	≤ 29	0	879 (69)	865 (68)
	> 29	1	390 (31)	401 (32)

Table 2: Baseline covariates frequencies in dataset `solvd.sub`.

```
R> head(solvd.sub)
```

```

  trt      y  censor lvef sodium any.vasodilator.use
1  0  803.01      1    0     1                      0
2  0  829.01      0    1     1                      1
3  0 1577.01      0    0     0                      1
4  1 1476.01      0    0     0                      0
5  0  224.01      1    0     1                      0
6  1 1168.01      0    1     0                      1

```

If *patient level raw data* is provided, package **beanz** provides function `bzGetSubgrpRaw` for estimating subgroup effects for each subgroup. The return value from `bzGetSubgrpRaw` is a data frame with the format of *summary treatment effect data*.

For continuous and binary response data, `bzGetSubgrpRaw` calls `glm` with model

$$g(Y|Z = z, G = g) = \mu_g + \theta_g z, \quad (11)$$

where $g(\cdot)$ is the identity or logit link function. For time to event data, package **beanz** calls the `coxph` function in package **survival** (Therneau 2017) with proportional hazard model

$$\lambda(t|Z = z, G = g) = \lambda_0(t) \exp(\theta_g z). \quad (12)$$

The model parameter estimation $\hat{\theta}_g$ and the associated variance $\hat{\sigma}_g^2$ are further used in the Bayesian HTE analysis.

```
R> subgrp.effect <- bzGetSubgrpRaw(solvd.sub, var.resp = "y",
+   var.trt = "trt", var.cov = c("lvef", "sodium", "any.vasodilator.use"),
+   var.censor = "censor", resptype = "survival")
R> subgrp.effect
```

```

Subgroup lvef sodium any.vasodilator.use Estimate Variance N
1         1    0     0                0 -0.37783038 0.01212786 562
2         2    0     0                1 -0.34655336 0.01004499 695
3         3    0     1                0 -0.79235451 0.03939983 237
4         4    0     1                1 -0.39334304 0.02969421 250
5         5    1     0                0  0.06776454 0.04629163 223
6         6    1     0                1 -0.23655764 0.02400353 341
7         7    1     1                0  0.15435495 0.10365396 104
8         8    1     1                1  0.05947290 0.07761840 123

```

3.3. Bayesian analysis

For Bayesian inference, the function `bzCallStan` calls the `sampling` function in package `rstan` (Carpenter, Gelman, Hoffman, Lee, Goodrich, Betancourt, Brubaker, Guo, Li, and Riddell 2017) to draw MCMC samples for Bayesian models introduced in Section 2. The parameter `mdls` allows the following modeling options: "nse" for *no subgroup effect model*, "fs" for *full stratification model*, "sr" for *simple regression model*, "ds" for *Dixon and Simon model*, "bs" for *basic shrinkage model*, "srs" for *simple regression with shrinkage model* and "eds" for *extended Dixon and Simon model*. Note that the results obtained are platform-dependent and will thus vary slightly for different platforms.

Priors of the parameters in the Bayesian models are passed to the function `bzCallStan` by parameter `par.pri` as a vector. The return value is a list object of class 'beanz.stan' containing the name of the Bayesian HTE model, raw `rstan` results, the posterior sample matrix, the posterior samples of the subgroup treatment effects (θ_g), the deviance information criterion (DIC, Spiegelhalter, Best, Carlin, and Van Der Linde 2002) and the leave-one-out cross-validation information criteria (LOOIC, Vehtari, Gelman, and Gabry 2017).

We consider the *simple regression model* in the following example:

```
R> var.cov <- c("lvef", "sodium", "any.vasodilator.use")
R> var.estvar <- c("Estimate", "Variance")
R> rst.sr <- bzCallStan(mdls = "sr", dat.sub = subgrp.effect,
+   var.estvar = var.estvar, var.cov = var.cov, par.pri = c(B = 1000,
+   C = 1000), chains = 4, iter = 4000, warmup = 2000)
```

```
SAMPLING FOR MODEL 'sr' NOW (CHAIN 1).
```

```
Gradient evaluation took 3.1e-05 seconds
1000 transitions using 10 leapfrog steps per transition would take 0.31 seconds.
Adjust your expectations accordingly!
```

```
Chain 1, Iteration:    1 / 4000 [  0%] (Warmup)
Chain 1, Iteration:   400 / 4000 [ 10%] (Warmup)
Chain 1, Iteration:   800 / 4000 [ 20%] (Warmup)
Chain 1, Iteration:  1200 / 4000 [ 30%] (Warmup)
Chain 1, Iteration:  1600 / 4000 [ 40%] (Warmup)
Chain 1, Iteration:  2000 / 4000 [ 50%] (Warmup)
Chain 1, Iteration:  2001 / 4000 [ 50%] (Sampling)
Chain 1, Iteration:  2400 / 4000 [ 60%] (Sampling)
Chain 1, Iteration:  2800 / 4000 [ 70%] (Sampling)
Chain 1, Iteration:  3200 / 4000 [ 80%] (Sampling)
Chain 1, Iteration:  3600 / 4000 [ 90%] (Sampling)
Chain 1, Iteration:  4000 / 4000 [100%] (Sampling)
```

```
Elapsed Time: 0.183723 seconds (Warm-up)
              0.10574 seconds (Sampling)
              0.289463 seconds (Total)
```

```
[... ]
```

The function `bzSummary` creates a data frame with the summary of the posterior subgroup treatment effects. The summary statistics include the posterior mean, standard deviation (SD), 2.5% (Q025), 25% (Q25), Median, 75% (Q75), 97.5% (Q975) quantiles of the treatment effect in each selected subgroup and the posterior probability that the subgroup treatment effect is less than a given cut off value (`ProbLT0`), which can be specified by parameter `cut`.

The following example shows the posterior subgroup treatment effect summary table for the *simple regression model*.

```
R> tbl.sub <- bzSummary(rst.sr, digits = 2, cut = 0)
R> tbl.sub
```

	Subgroup	Mean	SD	Q025	Q25	Median	Q75	Q975	ProbLT0
1	Subgroup 1	-0.4	0.09	-0.59	-0.46	-0.4	-0.34	-0.21	1
2	Subgroup 2	-0.38	0.09	-0.56	-0.44	-0.38	-0.32	-0.21	1
3	Subgroup 3	-0.49	0.13	-0.74	-0.57	-0.48	-0.4	-0.24	1
4	Subgroup 4	-0.47	0.12	-0.71	-0.55	-0.47	-0.38	-0.22	1
5	Subgroup 5	-0.06	0.13	-0.33	-0.15	-0.06	0.03	0.2	0.68
6	Subgroup 6	-0.04	0.12	-0.28	-0.12	-0.04	0.04	0.19	0.64
7	Subgroup 7	-0.15	0.16	-0.47	-0.26	-0.14	-0.04	0.17	0.82
8	Subgroup 8	-0.13	0.15	-0.41	-0.23	-0.13	-0.03	0.16	0.81

The package **beanz** also provides functions `bzPlot` and `bzForest` to present the posterior densities and forest plots (Lewis and Clarke 2001), respectively, for the posterior subgroup treatment effects (Figure 1).

```
R> bzPlot(rst.sr)
R> bzForest(rst.sr)
```

The **beanz** package provides functions for treatment effect comparisons between subgroup. The function `bzSummaryComp` creates a data frame with the summary of the posterior subgroup treatment effect differences between subgroups. The parameter `sel.grps` allows to select specific subgroups for the comparison.

```
R> tbl.sub <- bzSummaryComp(rst.sr, sel.grps = 1:3, digits = 2, cut = 0)
R> tbl.sub
```

	Comparison	Mean	SD	Q025	Q25	Median	Q75	Q975	ProbLT0
1	Subgroup 2-1	0.02	0.13	-0.23	-0.07	0.02	0.11	0.27	0.44
2	Subgroup 3-1	-0.09	0.16	-0.4	-0.19	-0.09	0.02	0.22	0.71
3	Subgroup 3-2	-0.1	0.16	-0.41	-0.21	-0.1	0	0.2	0.75

The functions `bzPlotComp` and `bzForestComp` present the posterior densities and forest plots, respectively, for the posterior subgroup treatment effect differences (Figure 2).

```
R> bzPlotComp(rst.sr, sel.grps = 1:3)
R> bzForestComp(rst.sr, sel.grps = 1:3)
```

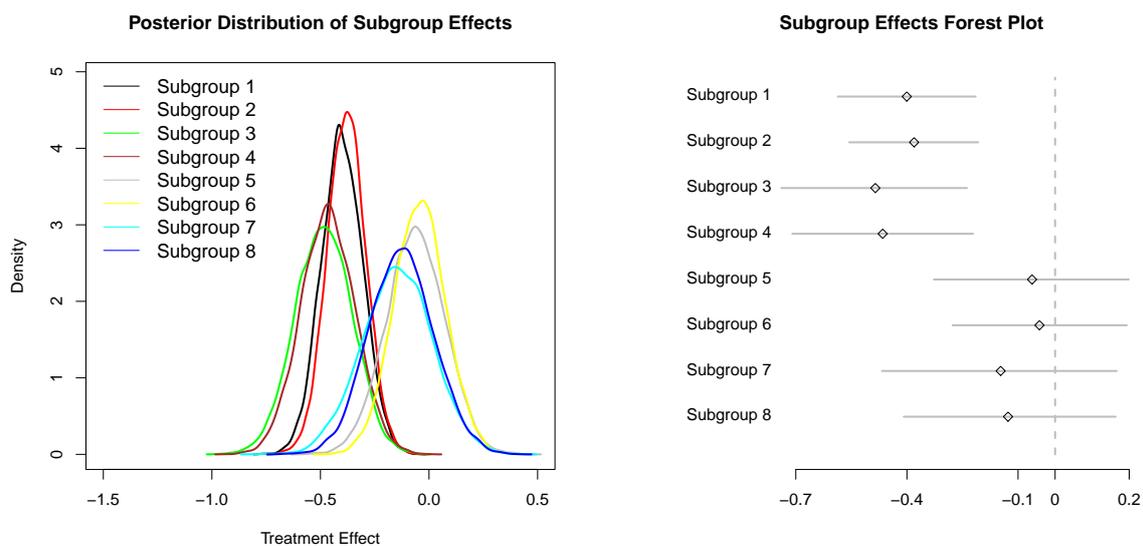


Figure 1: Posterior density and forest plot of the subgroup treatment effect from the *simple regression model*.

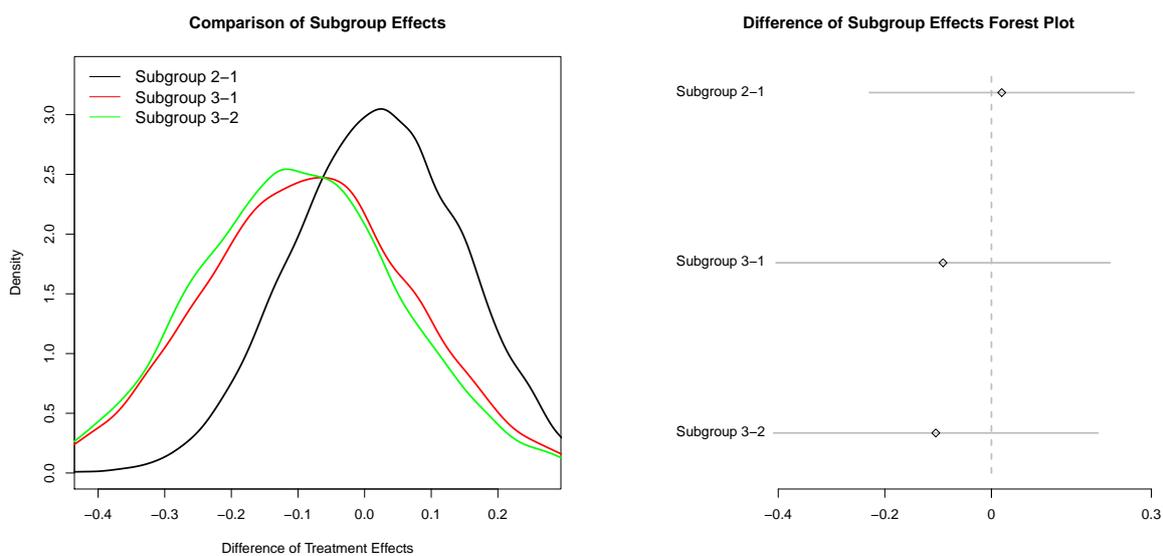


Figure 2: Posterior density and forest plot of the subgroup treatment effect differences of subgroups 1-3 from the *simple regression model*.

3.4. Model selection

We consider the *basic shrinkage* model and compare its LOOIC with the *simple regression model*.

```
R> rst.bs <- bzCallStan(mdl = "bs", dat.sub = subgrp.effect,
+   var.estvar = var.estvar, var.cov = var.cov,
+   par.pri = c(B = 1000, D = 1), chains = 4, iter = 4000, warmup = 2000)
```

```
R> rst.bs$looic$looic
```

```
[1] 3.426074
```

```
R> rst.sr$looic$looic
```

```
[1] 4.530313
```

Thus, the LOOIC is smaller for the *basic shrinkage model*, which indicates that the *basic shrinkage model* may fit the data better than the *simple regression model*. Further comparison between the models can be done following the suggestions in [Vehtari et al. \(2017\)](#).

The package **beanz** provides function `bzRptTbl` to generate the summary posterior subgroup treatment effect table from the model with the smallest LOOIC:

```
R> lst.rst <- list(sr = rst.sr, bs = rst.bs)
R> tbl.summary <- bzRptTbl(lst.rst, dat.sub = subgrp.effect,
+   var.cov = c("lvef", "sodium", "any.vasodilator.use"))
R> tbl.summary
```

	Model	Subgroup	lvef	sodium	any.vasodilator.use	Mean
Subgroup 1	Basic shrinkage	1	0	0	0	-0.352
Subgroup 2	Basic shrinkage	2	0	0	1	-0.333
Subgroup 3	Basic shrinkage	3	0	1	0	-0.518
Subgroup 4	Basic shrinkage	4	0	1	1	-0.345
Subgroup 5	Basic shrinkage	5	1	0	0	-0.141
Subgroup 6	Basic shrinkage	6	1	0	1	-0.267
Subgroup 7	Basic shrinkage	7	1	1	0	-0.167
Subgroup 8	Basic shrinkage	8	1	1	1	-0.180
	SD	Prob < 0				
Subgroup 1	0.096	1.000				
Subgroup 2	0.086	1.000				
Subgroup 3	0.185	1.000				
Subgroup 4	0.130	0.994				
Subgroup 5	0.183	0.780				
Subgroup 6	0.122	0.978				
Subgroup 7	0.215	0.793				
Subgroup 8	0.191	0.832				

3.5. Gail-Simon qualitative interaction test

Qualitative interaction is a type of treatment effect heterogeneity, where one or more of the subgroups have a treatment effect in the opposite direction relative to the overall treatment effect. It is critically important to detect such interactions, if they are present. Package **beanz** provides the frequentist Gail-Simon qualitative interaction test ([Gail and Simon 1985](#)) for examining qualitative interactions by function `bzGailSimon`. The Gail-Simon test examines the null hypothesis that all the subgroup treatment contrasts lie within the positive orthant

$O^+ = \{G_g : \theta_g \geq d\}$ or in the negative orthant, $O^- = \{G_g : \theta_g \leq -d\}$, where $d \geq 0$ denotes the threshold for clinically meaningful qualitative interaction.

The Gail-Simon test is to reject the null of no qualitative interaction if both of the following conditions are met:

$$\sum_g \mathbf{I}(\hat{\theta}_g > d) (\hat{\theta}_g - d)^2 / \hat{\sigma}_g^2 > c, \quad (13)$$

$$\sum_g \mathbf{I}(\hat{\theta}_g < -d) (\hat{\theta}_g + d)^2 / \hat{\sigma}_g^2 > c. \quad (14)$$

The critical value c is calculated such that when $\Omega \subset O^-$ or $\Omega \subset O^+$, the probability that both Equations 13 and 14 hold is no greater than the significance level α .

```
R> gs.pval <- with(subgrp.effect, bzGailSimon(Estimate, sqrt(Variance)))
R> gs.pval
```

```
[1] 0.9191656
```

The Gail-Simon test p value is 0.9192, which shows that there is no significant qualitative interactions in the SOLVD data.

4. The beanz GUI

4.1. Overall architecture

The **beanz** GUI is web-based and developed in R using the **shiny** (Chang, Cheng, Allaire, Xie, and McPherson 2017) web application framework. Figure 3 shows the architecture of the **beanz** GUI.

The GUI can be accessed within R using the function **bzShiny**:

```
R> bzShiny()
```

The function **bzShiny** calls **runApp** function in package **shiny**. To stop the GUI application, one may click the **Exit** button on the main title banner.

Alternatively, an online version of the **beanz** GUI is available at <https://olssol.shinyapps.io/beanz/>.

4.2. Tabpanel description

The main tabpanels that can be accessed through the menubar in the **beanz** GUI include **Start**, **Upload Data**, **Subgroup Specification**, **Configuration**, **Bayesian Analysis**, **Toolbox** and **Report**. The details of each panel are given as follows:

Start panel:

The **Start** panel serves as an introduction page for the software. The sections on this panel include:

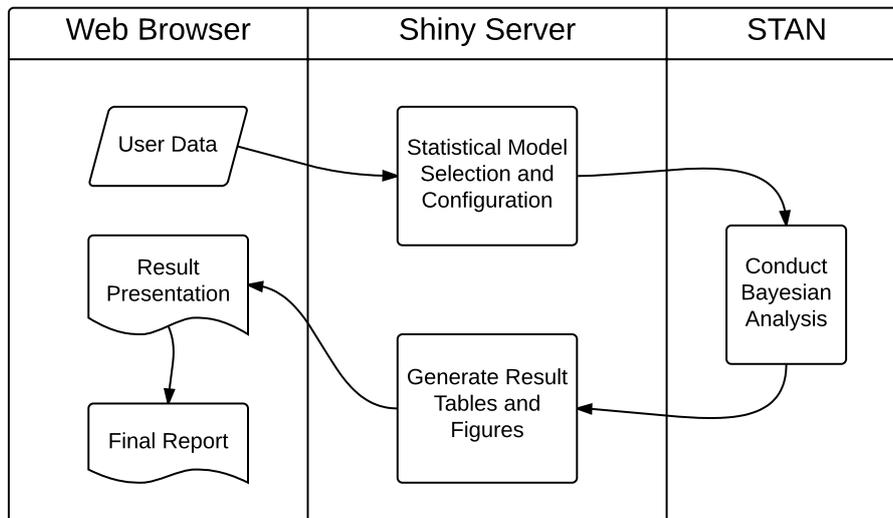


Figure 3: Architecture of the **beanz** GUI.

- **What does **beanz** do?**
Presents the background introduction to package **beanz** and the basic steps of using the software.
- **What does **beanz** need?**
Explains the different formats of the data files that package **beanz** accepts.
- **What does **beanz** provide?**
Introduces the results generated with an package **beanz** analysis.
- **Warnings**
Warnings that emphasize that subgroup analysis are mostly exploratory and sensitivity analysis are necessary to evaluate the robustness of the findings with regards to different prior choices.

Upload Data panel:

The Upload Data panel provides an interface for users to upload the data to be analyzed. The sections and items within each section on this panel include:

- **Upload Data**
 - Choose File** Clicking the `Browse...` button will load local data files in `csv` or `plain text` format.
 - Data Format** *Summary treatment effect data or patient level raw data.*
 - Separator** Field separating character.

Quote	Quoting character.
Other	There are two additional options: <i>Header.</i> Checkbox indicating if the first line of the file contains the names of the columns. <i>Show Data.</i> Checkbox indicating whether to present the uploaded data in the Review Data section on this panel.

- Try An Example
Clicking the button will load the example SOLVD Patient level raw data dataset. Details of the example dataset are also given in this section.
- Review Data
Presents the uploaded dataset in a table view.

Subgroup Specification panel:

The Subgroup Specification panel is designed to identify the columns from the uploaded dataset for formulating subgroups. This panel is only available after a data file has been successfully uploaded. The sections and items within each section on this panel include:

- Select Variables

Treatment Effect Estimates	Dropdown list for specifying the column for treatment effect estimation. Only available for <i>summary treatment effect data</i> .
Treatment Effect Variance	Dropdown list for specifying the column for the variance associated with the treatment effect estimation. Only available for <i>summary treatment effect data</i> .
Treatment	Dropdown list for specifying the column for treatment. Only available for <i>patient level raw data</i> .
Response	Dropdown list for specifying the column for response. Only available for <i>patient level raw data</i> .
Censoring	Dropdown list for specifying the column for censoring. Only available for <i>patient level raw data</i> with response that is <i>time to event with treatment effect measured by log hazard ratio</i> .
Type of Response	Radio buttons for specifying the type of response data. The types include <i>continuous</i> , <i>binary with treatment effect measured by log odds ratio</i> and <i>time to event with treatment effect measured by log hazard ratio</i> . Only available for <i>patient level raw data</i> .
Covariates	Checkboxes for specifying columns for the covariates. The covariates should be <i>binary</i> , <i>ordinal with numerical values</i> , or <i>nominal</i> . It is strongly recommended that the user specifies numerical values for ordinal variables. Otherwise, ordinal variables with character levels are converted to integers starting from 1 by the software.

Nominal Covariates Checkboxes for specifying columns that are nominal covariates. Only available if there are categorical covariates that contain more than two levels. Ordinal and nominal covariates are modeled differently in the *simple regression model*, *simple regression shrinkage model*, *Dixon and Simon model* and *extended Dixon and Simon model*. See Section 2.2 for more details.

Get Subgroups Button for generating subgroups based on the column specifications. A specification validation check will be first conducted. If there are errors in the specifications, error messages will be shown in this section.

- **Subgroups**

Provides a table view of the subgroups. Each row represents a subgroup and the columns include covariates, treatment effect estimation and the variance associated with the treatment effect estimation.

Configuration panel:

The Configuration panel contains options for Bayesian model priors, MCMC sampling and results presentation. The sections and items within each section on this panel include:

- **Statistical Models and Priors**

Checkboxes for selecting the Bayesian models to include in the analysis. Note that the *no subgroup effect model* is always included.

For each model, numeric inputs are provided for specifying the parameters in the priors. See Section 2 for details of the priors.

- **Prior Of Variance**

Prior of log SD Choose *normal distribution* or *uniform distribution* as the prior distribution for $\log \sigma$. See Section 2 for details.

Uncertainty of log SD Numeric input for specifying parameter Δ .

- **MCMC Parameters**

Number of iterations Stan parameter specifying how many iterations including burn-in for posterior sampling.

Number of burn-in Stan parameter specifying how many burn-in for posterior sampling.

Number of thinning Stan parameter specifying the period for saving posterior samples.

Number of Chains Stan parameter specifying the number of MCMC chains for sampling.

Random seed	Stan parameter for random number generation.
Number of cores	Stan parameter for the number of cores for parallel computation.
Rhat Warning	Threshold for generating warnings of problematic convergence based on the Gelman and Rubin potential scale reduction statistic Rhat (Gelman and Rubin 1992) in the Stan posterior samples.
Initial Step-size	Stan parameters that affect the MCMC convergence.
Target Metropolis Acceptance Rate	Stan parameters that affect the MCMC convergence.
• Display Parameters	
Cut off for treatment effects	The subgroup treatment effect result tables will present the probability that the subgroup effect is less than this cut off value.
Cut off for comparison	The treatment effect comparison between subgroups result tables will present the probability that the subgroup treatment effect difference is less than this cut off value.
Digits	Integer indicating the number of decimal places in the result tables.
Maximum subgroups for comparison plots	When the number of subgroups is large, the treatment effect comparison density plots will not be presented. This integer specifies the maximum number of subgroups for the plot to display.
Organize results	Different ways to present the analysis results: <i>By results.</i> In the results, each row represents a statistical model. The columns include treatment effect, subgroup comparison and Stan diagnosis. <i>By model.</i> In the results, each column represents a statistical model. The rows include treatment effect, subgroup comparison and Stan diagnosis.
Transformation	Checkbox indicating if the treatment effect should be measured by log-odds ratio or odds ratio. Only effective for <i>patient level raw data</i> and <i>time to event with treatment effect measured by log hazard ratio</i> .
Reference	Checkbox indicating whether the results from the no subgroup effect model should be included in the result tables and figures for the other models as a reference.
Select subgroups to display	A table for selecting a subset of subgroups to be included in the result presentation. The selected subgroups are highlighted in the table.

Bayesian Analysis panel:

The Bayesian Analysis panel allows users to conduct the Bayesian analysis by clicking the `Conduct Bayesian Analysis` button and presents the analysis results following the display configurations on the Configuration panel. This panel is only available after subgroups have been successfully specified.

The sections and items within each section on this panel include:

- **Effect**

Table	Presents the posterior mean, standard deviation, 2.5%, 25%, 50%, 75%, 97.5% quantiles of the treatment effect in each selected subgroup (see <code>Select subgroups to display</code>) and the posterior probability that the subgroup treatment effect is less than a given cut off value (see <code>Cut off for treatment effects</code>).
Density	Presents the posterior density plots of the subgroup treatment effects for the selected subgroups.
Forest plot	Presents the posterior forest plots of the subgroup treatment effects for the selected subgroups.
Predictive plot	Presents the posterior predictive distribution of the median, standard deviation, minimum and maximum of the subgroup treatment effects and compare them to the observed statistics.

- **Comparison**

Table	Presents the posterior mean, standard deviation, 2.5%, 25%, 50%, 75%, 97.5% quantiles of the treatment effect difference in each pairwise comparison of the selected subgroups and the posterior probability that the subgroup treatment effect difference is less than a given cut off value (see <code>Cut off for comparison</code>).
Density	Presents the posterior density plots of the treatment effect difference for each pairwise comparison of the selected subgroups.
Forest plot	Presents the posterior forest plots of the treatment effect difference for each pairwise comparison of the selected subgroups.

- **Stan Diagnosis**

Raw output	Presents the outputs from Stan.
Information criteria	Presents the deviance information criterion (DIC) and the leave-one-out cross-validation information criterion (LOOIC).
Trace plot	Presents trace plots of the posterior samples for evaluating the convergence of the Monte-Carlo sampling chains.

Rhat Presents the distribution of the Rhat statistics and lists the parameters with possible convergence issues based on the Rhat warning threshold (see **Rhat warning**).

Toolbox panel:

The **Toolbox** panel provides tools that implement frequentist methods for the evaluation of HTE. The sections on this panel include:

- **ANOINT**

By clicking the **Conduct Analysis** button, package **beanz** conducts frequentist subgroup analysis using the **anoimt** package and presents the outputs generated in R in this section. The details of package **anoimt** can be found in [Varadhan and Kovalchik \(2015\)](#).

This section is only available for *patient level raw data* and when subgroups have been successfully formulated.

- **Gail-Simon Test for Qualitative Interactions**

Provides the p value for the Gail-Simon test for qualitative interactions. See Section 3.5 for details.

Clinically meaningful threshold allows the user to specify the threshold for defining the positive and negative orthants.

Report panel:

The **Report** panel presents a summary of the Bayesian analysis results and provides a link for downloading a final report. This panel is only available after the Bayesian analysis has been successfully conducted. The sections on this panel include:

- **Summary**

Presents a summary result table for the selected subgroups. The table contains columns of covariates, treatment effect posterior mean and standard deviation, and the posterior probability that the treatment effect is smaller than a cut off value. The results are generated by the Bayesian model with the smallest LOOIC with the given set of priors.

- **Download the analysis report**

The final report can be downloaded by clicking the **Download** button. The available document formats for the report include PDF, HTML and Word.

5. Demonstration of beanz GUI

We demonstrate the **beanz** GUI using the SOLVD clinical trial data in this section.

Step 1. We first upload the patient-level SOLVD data file to package **beanz** from the **Upload Data** panel (Figure 4). One can also load the data from package **beanz** by clicking the **Try it** button.

Upload Data

Please upload data file.

Choose File: No file selected

Data Format:

- Summary treatment effect data
- Subject level raw data

Separator:

- Comma
- Semicolon
- Tab
- Space

Quote:

- None
- Double Quote
- Single Quote

Other:

- Header
- Show Data

Review Data

Show entries

Search:

trt	y	censor	lvef	sodium	any_vasodilator.use
0	803.0100098	1	0	1	0
0	829.0100098	0	1	1	1
0	1577.01001	0	0	0	1
1	1476.01001	0	0	0	0
0	224.0099945	1	0	1	0
1	1168.01001	0	1	0	1
1	1483.01001	0	0	1	0
0	973.0100098	0	0	0	1
0	961.0100098	1	0	1	1
1	1449.01001	0	0	0	0
1	752.0100098	1	0	1	1
1	586.0100098	1	0	0	0
1	618.0100098	1	0	1	1

Figure 4: Upload data.

Select Variables For Subject Level Raw Data

Specify variables for defining subgroups. If a covariate has more than two levels in the data, please specify whether the covariate is nominal or ordinal. Ordinal covariates are assumed to have a linear trend across levels in its effect on the outcome.

Treatment:

Response:

Censoring:

Covariates:

- lvef
- sodium
- any_vasodilator.use

Type of response:

- Continuous
- Binary with treatment effect measured by log odds ratio
- Time to event with treatment effect measured by log hazard ratio

Subgroup specification is valid.

Subgroups

Search:

Subgroup	lvef	sodium	any_vasodilator.use	Estimate	Variance	N
1	0	0	0	-0.377830384551549	0.0121278643460031	562
2	0	0	1	-0.346553356437156	0.0100449942640191	695
3	0	1	0	-0.792354513684684	0.0393998309129795	237
4	0	1	1	-0.393343043259766	0.0296942137491711	250
5	1	0	0	0.0677645444041309	0.0462916293607645	223
6	1	0	1	-0.236557643587652	0.0240035268423054	341
7	1	1	0	0.154354951524952	0.103653960793714	104
8	1	1	1	0.059472906058487	0.077618400960085	123

Showing 1 to 8 of 8 entries

A PCORI Project • Team

Figure 5: Subgroup specification.

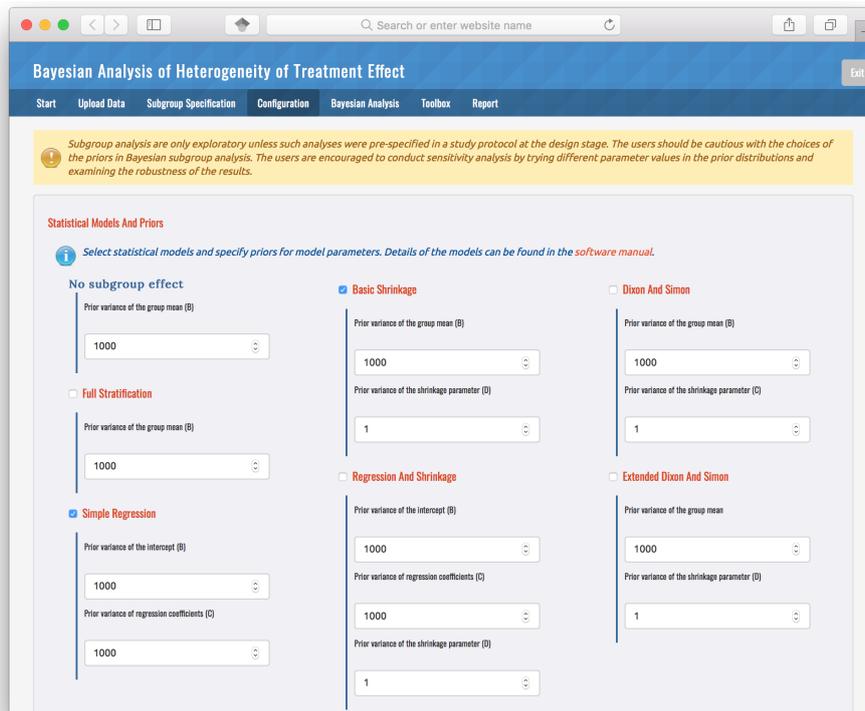


Figure 6: Configuration of model selection.

Step 2. Next, on the **Subgroup Specification** panel we first select the **Type** of response to be time-to-event. We then specify the columns for **Treatment**, **Response** and **Censoring** and then include the baseline covariates sodium, vasodilator and LVEF indicators. By clicking the **Get Subgroups** button, package **beanz** checks the validity of the subgroup specification, estimates the treatment effect that is measured by log hazard ratio for each subgroup and reports them in the **Subgroups** section on this panel (Figure 5).

We can see that there are a total of 8 subgroups. The subgroup with ejection fraction $> 29\%$, sodium > 141 and no vasodilator (subgroup 7) shows the least Enalapril drug effect with $\theta_7 = 0.15$.

Step 3. After the subgroups are formulated, on the **Configuration** panel, for this demonstration, we select the **Simple regression** and the **Basic shrinkage** models with their default priors ($B = 1000$, $C = 1000$, $D = 1$) for the Bayesian analysis (Figure 6). We use the default MCMC parameters: 4000 iterations, 2000 burn-ins and 2 thinning iterations. We choose the normal distribution with $\Delta = 0$ for $\log \sigma_g$. We require 4 MCMC chains and set the target Metropolis acceptance rate to be 0.95. Lastly, we set the random seed for Stan to be 1000. For reporting results, we set the cut off value for subgroup treatment effects to be 0 to report the posterior probability that the log hazard ratio is smaller than 0. We use the default value 0 for the cut off value for subgroup treatment effect differences. We also set the format of the result tables with default value 3 digits. We choose the **By model** option that organizes results with each column presenting a model. Lastly, we choose the **Display no subgroup effect outcome** option to present the results from

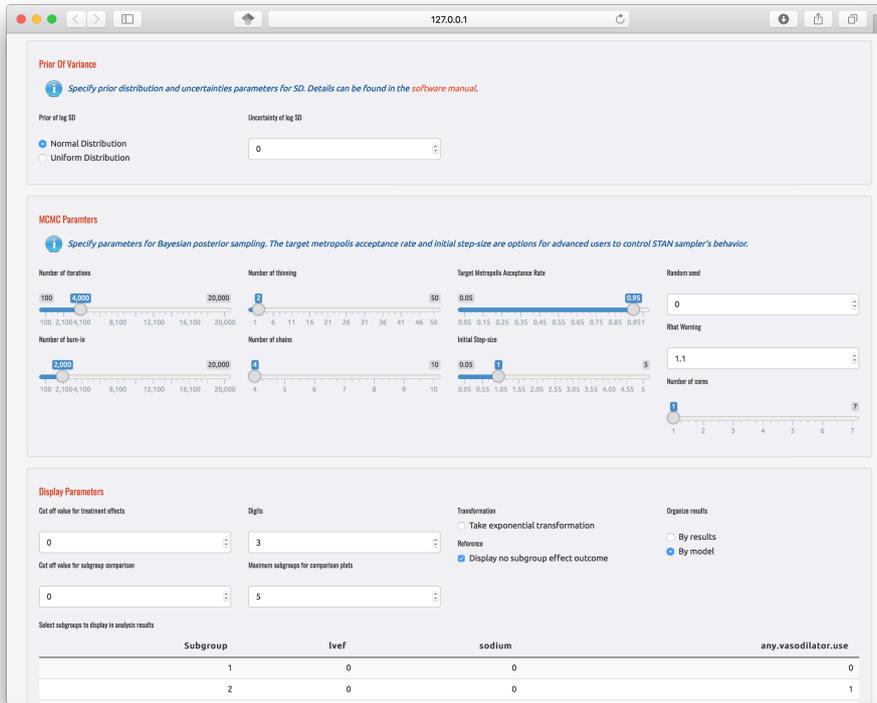


Figure 7: Configuration of MCMC parameters.

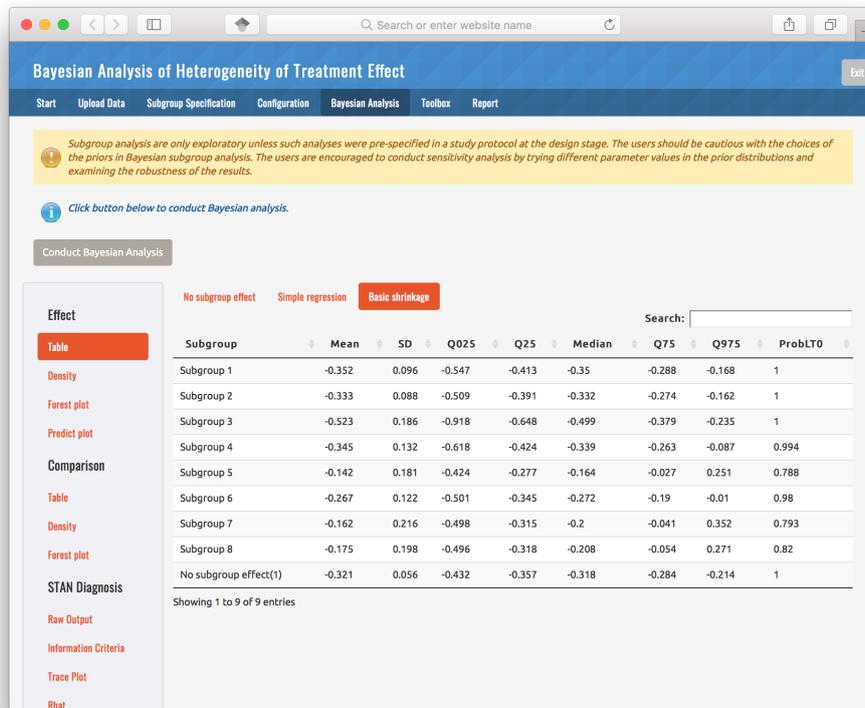


Figure 8: Bayesian HTE analysis: basic shrinkage model subgroup treatment effects.

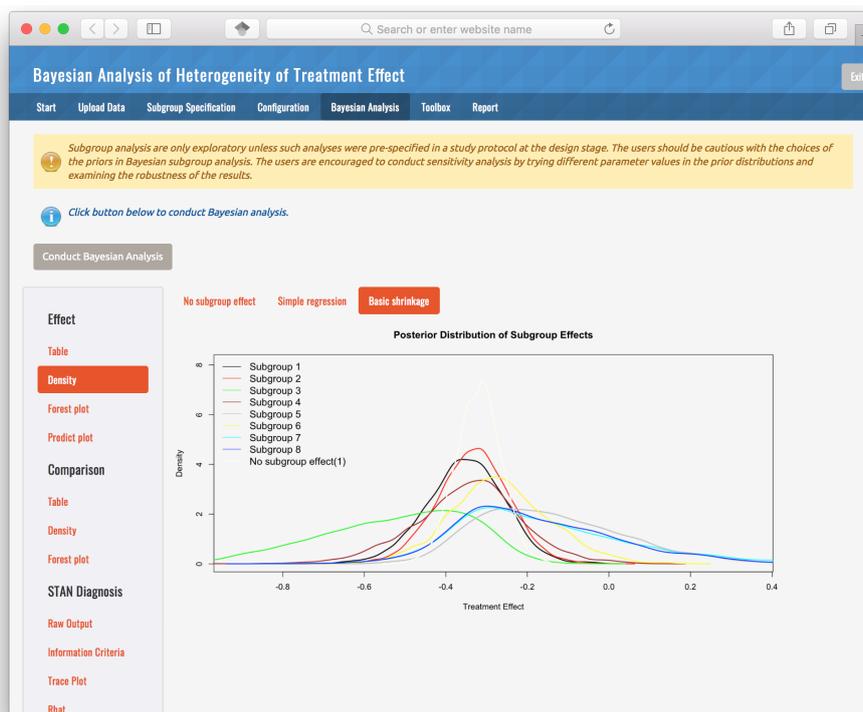


Figure 9: Bayesian HTE analysis: density plot of subgroup treatment effects from the basic shrinkage model.

No subgroup effect model as a reference (Figure 7).

Step 4. Next, we move to the Bayesian Analysis panel. A progress bar will appear indicating the progress of the Stan analysis after clicking the `Conduct Bayesian Analysis` button. When the analysis is completed, the results will be presented according to the configuration options from the Configuration panel.

Figures 8–10 present the subgroup treatment effect table, density plot and forest plot for the **Basic shrinkage** model with the given priors. We observe that the overall Enalapril effect is significant if we assume that there is no subgroup effect. With the **Basic shrinkage** model, the subgroups with ejection fraction $\leq 29\%$ (subgroups 1–4) are all showing strong evidence for the Enalapril effect. Nonetheless, for all the subgroups with ejection fraction $> 29\%$ (subgroups 5–8), there is not sufficient evidence to show that the Enalapril is superior to the placebo.

Figure 11 presents the results of the posterior predictive check by comparing the posterior distributions of median, standard deviation, minimum and maximum of the subgroup treatment effect with the observed statistics. The check shows that there are no major systematic discrepancies between the posterior predictive distributions and the observed data.

As a comparison, Figure 12 presents the subgroup treatment effect forest plot for the **Simple regression** model. The results from the **Simple regression** model are similar

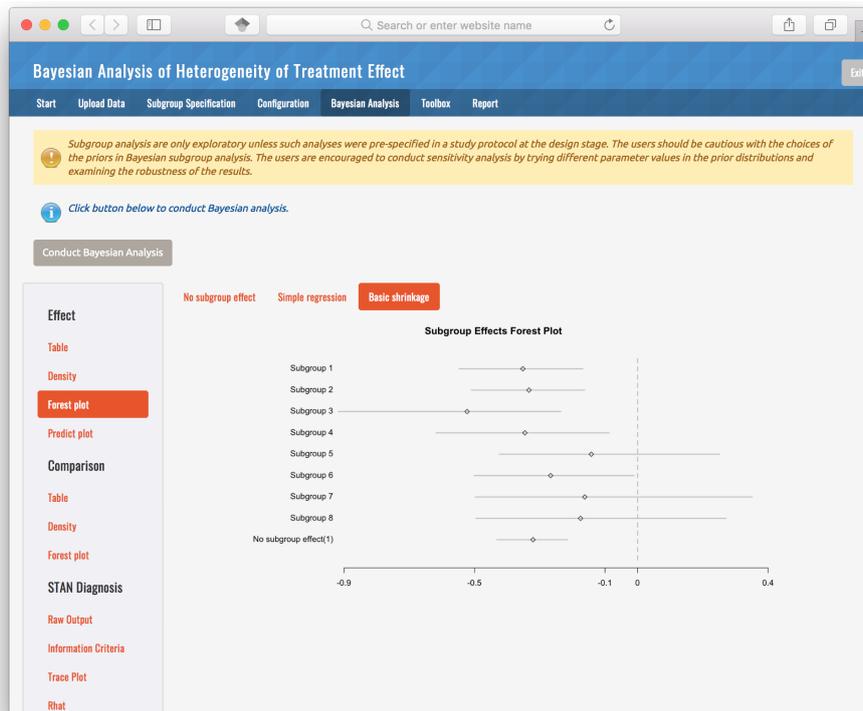


Figure 10: Bayesian HTE analysis: forest plot of subgroup treatment effects from the basic shrinkage model.

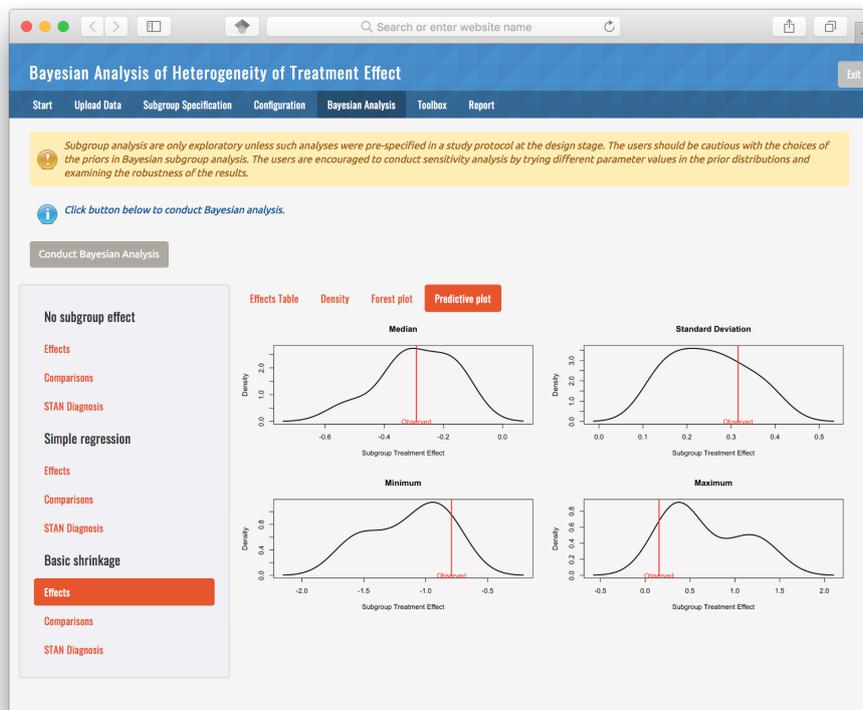


Figure 11: Bayesian HTE analysis: posterior predictive checks for the basic shrinkage model.

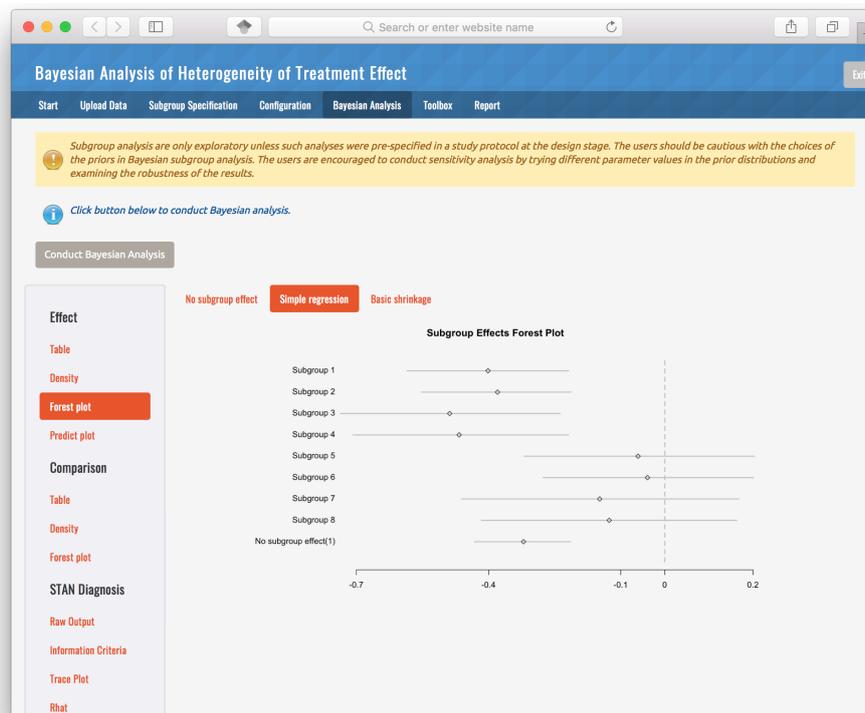


Figure 12: Bayesian HTE analysis: forest plot of subgroup treatment effects from the simple regression model.

to the **Basic shrinkage** model: significant Enalapril effects are only observed in subgroups 1–4.

Next, for patients with sodium ≤ 141 and no vasodilator, we evaluate the treatment effect difference between the subgroups with ejection fraction $\leq 29\%$ (subgroup 1) and $> 29\%$ (subgroup 5). Figure 13 shows how to select the subgroups of interest in the **Display parameter** section on the **Configuration** panel. Figure 14 presents the comparison forest plot based on the **Basic shrinkage** model. We observe evidence that suggests there may exist treatment by ejection fraction interaction: the Enalapril drug seems to be more effective for the patients with ejection fraction $\leq 29\%$, sodium ≤ 141 and no vasodilator, compared to the patients with ejection fraction $> 29\%$, sodium ≤ 141 and no vasodilator, but the difference is not significant.

Step 5. At the end, on the **Report** panel, a summary result table is provided based on the model with the smallest LOOIC. The analysis report can be downloaded as a PDF, HTML, or Word document (Figure 15).

The report contains sections **Data Summary**, **Subgroup Definition**, **Analysis Results** for each selected model, and **Summary**. Figure 16 shows the contents page of an example report.

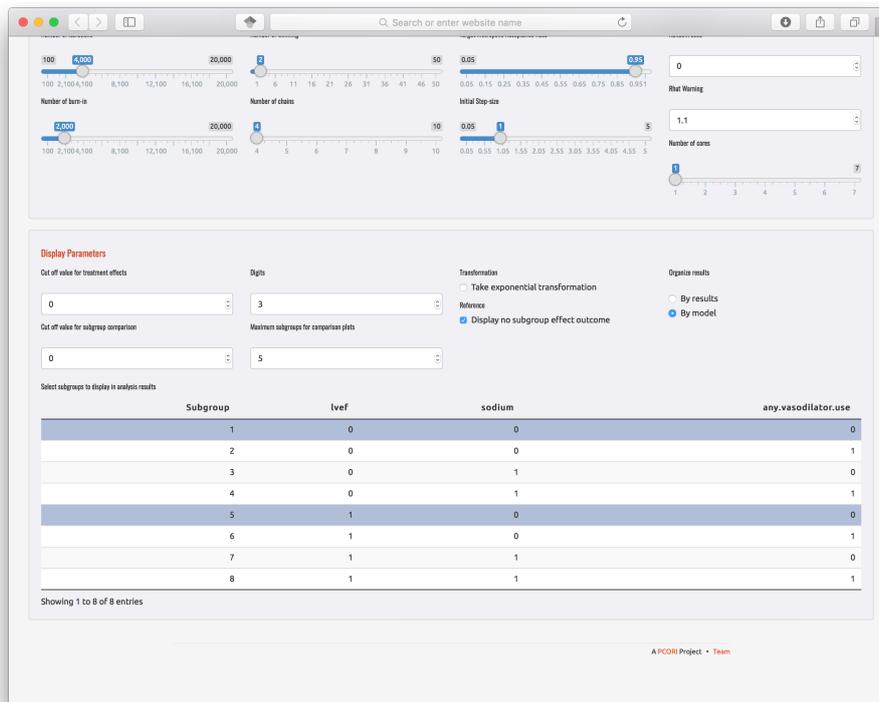


Figure 13: Configuration: select a subset of subgroups for analysis result presentation.

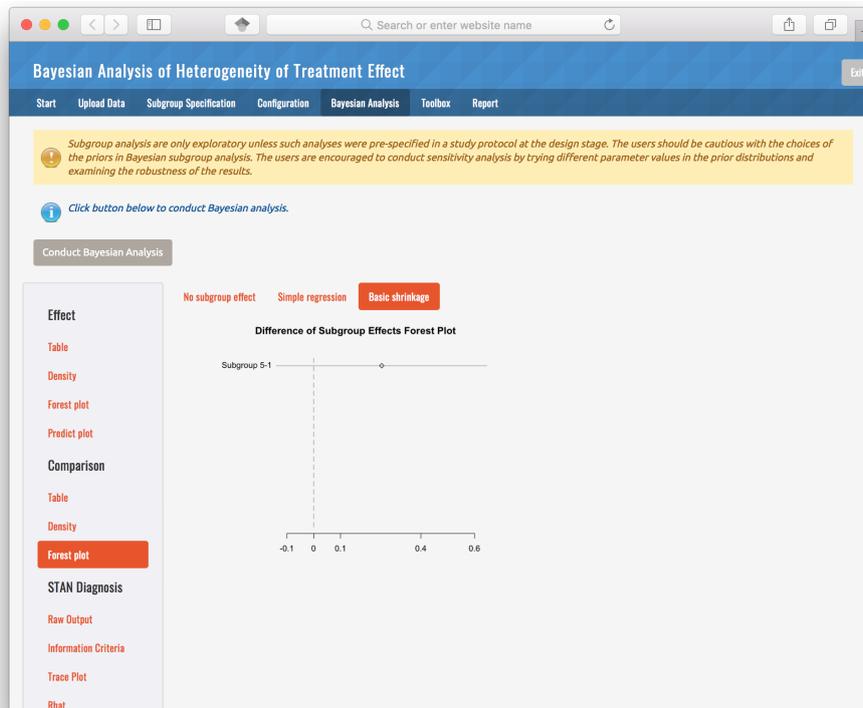


Figure 14: Bayesian HTE analysis: difference between treatment effects of subgroups 1 and 5 from the basic shrinkage model.

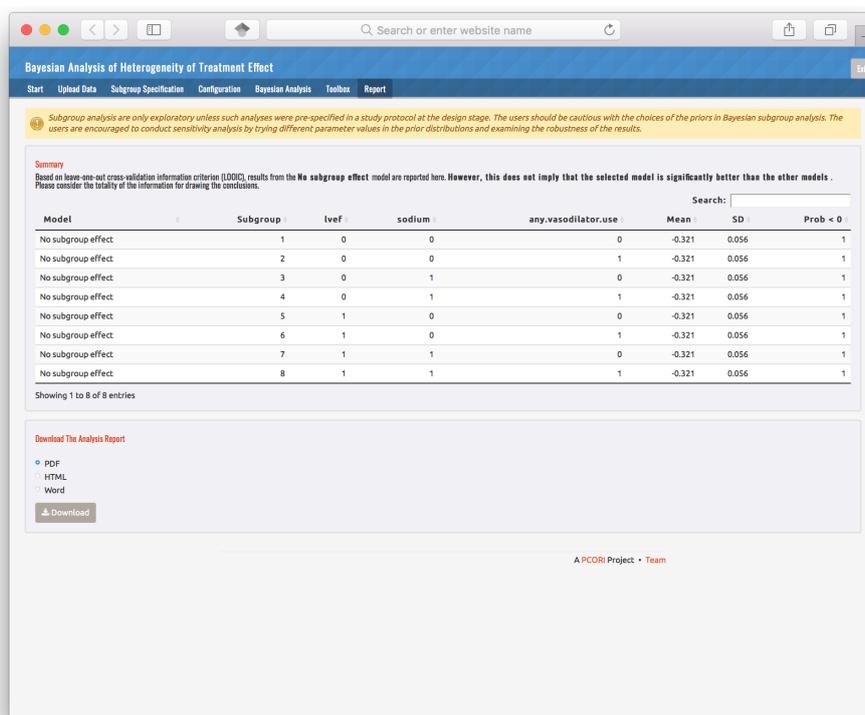


Figure 15: Report panel.

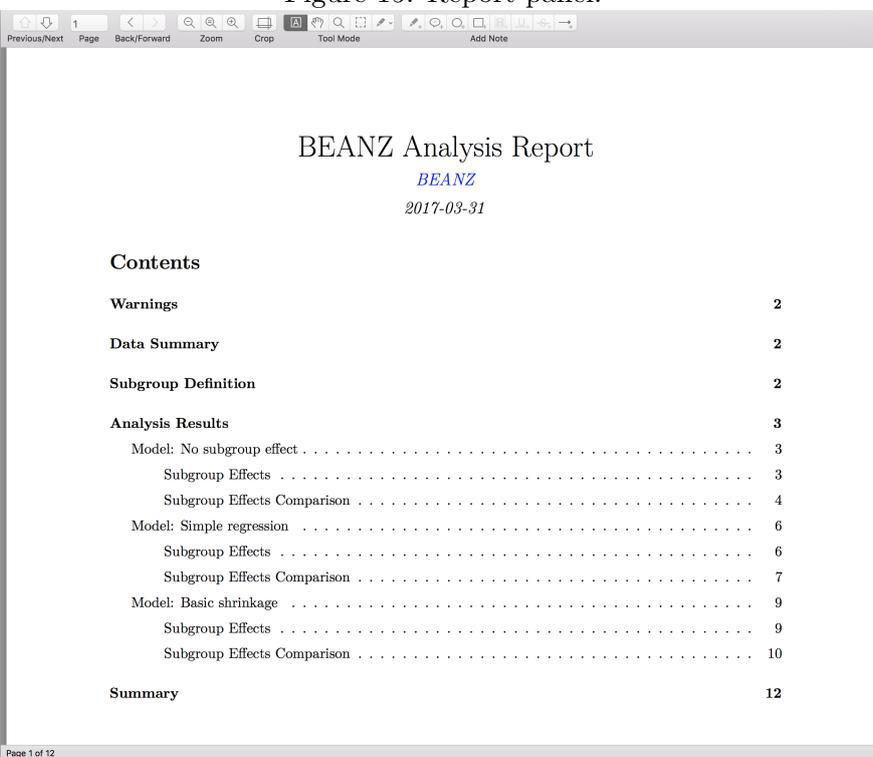


Figure 16: Report contents page example.

6. Discussion

In this era of precision medicine, it is not sufficient to evaluate the efficacy of a medical intervention based only on its average treatment effect. It is critical to be able to better understand the nature and sources of heterogeneity in treatment effects in order to help patients and stakeholders make informed health care decisions.

The Bayesian approach offers a principled and flexible way for evaluating the differences in treatment effects across patient characteristics. It is a formal approach to combine evidence from the study at hand with external information or subjective beliefs to answer decision-driven questions from the varied perspectives of different stakeholders.

A major impediment to the conduct of Bayesian subgroup analysis is the lack of user-friendly software. We attempt to solve this problem by providing, for the first time, an R package software tool with web-based GUI, which makes it easy to carry out Bayesian subgroup analysis. Our software readily facilitates integral aspects of a Bayesian analysis including prior specification, model specification (choosing from a suite of models), control of MCMC simulation settings, visualization of posterior densities, tabular presentation of numerical summaries, and report generation.

We also provide a toolbox for conducting frequentist analysis via fitting regression models with interactions between treatment and baseline covariates, by linking to the R package **anoimt**, which provides an extensive array of subgroup analysis techniques and interaction models. The toolbox also provides a likelihood ratio test for detecting qualitative interactions (Gail and Simon 1985).

For Bayesian analysis, specifying a prior distribution is challenging. There is no default method that will handle all situations. In the absence of prior (empirical) information, we recommend low-information priors, ones that provide support for a broad range of values. Even here care is needed, especially with respect to the scale of a parameter. For example, a $N(0, 10000)$ can be either low information or high information depending on the scale of the regressor and consequently the slope. Similarly, a $\text{Uniform}(0, 10)$ prior for a sampling standard deviation can be either low or high information depending on the measurement scale. Specifically, Neuenschwander, Capkun-Niggli, Branson, and Spiegelhalter (2010) introduced a prior maximum sample size concept that can be used to quantify the information level in such priors. We recommend careful attention to these issues. If empirical information is available, we recommend summarizing it by a probability distribution, for example the normalized likelihood, or if the estimate and standard error (SE) are given, by $N(\text{estimate}, c^2 \times \text{SE}^2)$, with $c > 1$, a “power prior” penalizing to reflect that the external data are not completely exchangeable with the current data. If the analyst also determined that the external data were biased relative to the current study, a $N(\text{estimate} + \text{offset}, c^2 \times \text{SE}^2)$ prior can be used. Potential bias can also be captured by selecting an appropriate value for c . For all important analyses, we recommend sensitivity analyses relative to these very explicit analytic choices.

Given the exploratory nature of the subgroup analysis, it is important to evaluate whether a particular Bayesian model provides a good fit to the data (Gelman 2003; Rubin *et al.* 1984). The software currently provides LOOIC and DIC criteria that can be the basis for comparing the goodness-of-fit. The software also provides posterior predictive checks for checking if the modeled summaries of subgroup-specific treatment effects such as the median, minimum, maximum, and standard deviation are consistent with the corresponding observed data summaries. Other techniques such as cross-validation may also be considered for this

task. We plan to include this in a future version of the software.

A potential limitation of the models represented in our software is that the subgroup treatment effects are assumed to follow a normal distribution. For certain scenarios such as analysis of rare events, the normal approximation may not be appropriate (Jones *et al.* 2011). Nevertheless, maximum likelihood estimates with large information will be approximately normal at least in some scale for most cases. Moreover, the normal approximation allows for relatively straightforward interpretation of the models considered for the mean of $\hat{\theta}_g$. In practice, one can transform for improved normality although such transformation may move the analysis away from the clinically relevant scale. This is the first release of the software; we plan to include other distributions in subsequent versions of the software.

Acknowledgments

This work was partially supported through a Patient-Centered Outcomes Research Institute (PCORI) Award (ME-1303-5896) and the Cervical Cancer SPORE (NCI 1 P50 CA098252).

References

- Assmann SF, Pocock SJ, Enos LE, Kasten LE (2000). “Subgroup Analysis and Other (Mis)uses of Baseline Data in Clinical Trials.” *The Lancet*, **355**(9209), 1064–1069. doi: [10.1016/s0140-6736\(00\)02039-0](https://doi.org/10.1016/s0140-6736(00)02039-0).
- Carpenter B, Gelman A, Hoffman M, 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](https://doi.org/10.18637/jss.v076.i01).
- Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J (2017). *shiny: Web Application Framework for R*. R package version 1.0.5, URL <https://CRAN.R-project.org/package=shiny>.
- Dixon DO, Simon R (1991). “Bayesian Subset Analysis.” *Biometrics*, **47**(3), 871–881. doi: [10.2307/2532645](https://doi.org/10.2307/2532645).
- Dusseldorp E, Doove L, van Mechelen I (2015). “quint: An R Package for the Identification of Subgroups of Clients Who Differ in Which Treatment Alternative Is Best for Them.” *Behavior Research Methods*, **48**(2), 650–663. doi: [10.3758/s13428-015-0594-z](https://doi.org/10.3758/s13428-015-0594-z).
- Egami N, Ratkovic M, Imai K (2018). *FindIt: Finding Heterogeneous Treatment Effects*. R package version 1.1.3, URL <https://CRAN.R-project.org/package=FindIt>.
- Gail M, Simon R (1985). “Testing for Qualitative Interactions between Treatment Effects and Patient Subsets.” *Biometrics*, **41**(2), 361–372. doi: [10.2307/2530862](https://doi.org/10.2307/2530862).
- Gelman A (2003). “A Bayesian Formulation of Exploratory Data Analysis and Goodness-of-Fit Testing.” *International Statistical Review*, **71**(2), 369–382. doi: [10.1111/j.1751-5823.2003.tb00203.x](https://doi.org/10.1111/j.1751-5823.2003.tb00203.x).
- Gelman A, Rubin DB (1992). “Inference from Iterative Simulation Using Multiple Sequences.” *Statistical Science*, **7**(4), 457–472. doi: [10.1214/ss/1177011136](https://doi.org/10.1214/ss/1177011136).

- Henderson NC, Louis TA, Wang C, Varadhan R (2016). “Bayesian Analysis of Heterogeneous Treatment Effects for Patient-Centered Outcomes Research.” *Health Services and Outcomes Research Methodology*, **16**(4), 213–233. doi:10.1007/s10742-016-0159-3.
- Holland PW (1986). “Statistics and Causal Inference.” *Journal of the American Statistical Association*, **81**(396), 945–960. doi:10.2307/2289064.
- Jones HE, Ohlssen DI, Neuenschwander B, Racine A, Branson M (2011). “Bayesian Models for Subgroup Analysis in Clinical Trials.” *Clinical Trials*, **8**(2), 129–143. doi:10.1177/1740774510396933.
- Lewis S, Clarke M (2001). “Forest Plots: Trying to See the Wood and the Trees.” *BMJ: British Medical Journal*, **322**(7300), 1479–1480. doi:10.1136/bmj.322.7300.1479.
- 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.
- Pocock SJ, Assmann SE, Enos LE, Kasten LE (2002). “Subgroup Analysis, Covariate Adjustment and Baseline Comparisons in Clinical Trial Reporting: Current Practice and Problems.” *Statistics in Medicine*, **21**(19), 2917–2930. doi:10.1002/sim.1296.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rubin DB, *et al.* (1984). “Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician.” *The Annals of Statistics*, **12**(4), 1151–1172. doi:10.1214/aos/1176346785.
- Schou IM, Marschner IC (2014). “Methods for Exploring Treatment Effect Heterogeneity in Subgroup Analysis: An Application to Global Clinical Trials.” *Pharmaceutical Statistics*, **14**(1), 44–55. doi:10.1002/pst.1656.
- SOLVD Investigators, *et al.* (1991). “Effect of Enalapril on Survival in Patients with Reduced Left Ventricular Ejection Fraction and Congestive Heart Failure.” *New England Journal of Medicine*, **325**(5), 293–302. doi:10.1056/nejm199108013250501.
- Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A (2002). “Bayesian Measures of Model Complexity and Fit.” *Journal of the Royal Statistical Society B*, **64**(4), 583–639. doi:10.1111/1467-9868.00353.
- Therneau TM (2017). *survival: Survival Analysis*. R package version 2.41-3, URL <https://CRAN.R-project.org/package=survival>.
- Varadhan R, Kovalchik S (2015). *anoaint: Analysis of Interactions*. R package version 1.4, URL <https://CRAN.R-project.org/package=anoaint>.
- Varadhan R, Segal JB, Boyd CM, Wu AW, Weiss CO (2013). “A Framework for the Analysis of Heterogeneity of Treatment Effect in Patient-Centered Outcomes Research.” *Journal of Clinical Epidemiology*, **66**(8), 818–825. doi:10.1016/j.jclinepi.2013.02.009.

- Varadhan R, Yao W (2014). **DSBayes**: *Bayesian Subgroup Analysis in Clinical Trials*. R package version 1.1, URL <https://CRAN.R-project.org/package=DSBayes>.
- Vehtari A, Gelman A, Gabry J (2017). “Practical Bayesian Model Evaluation Using Leave-One-Out Cross-Validation and WAIC.” *Statistics and Computing*, **27**(5), 1413–1432. doi: [10.1007/s11222-016-9709-3](https://doi.org/10.1007/s11222-016-9709-3).
- Wang C, Varadhan R, Trustees of Columbia University (2018). **beanz**: *Bayesian Analysis of Heterogeneous Treatment Effect*. R package version 2.3, URL <https://CRAN.R-project.org/package=beanz>.

Affiliation:

Chenguang Wang, Nicholas C. Henderson, Ravi Varadhan
Division of Biostatistics and Bioinformatics
Sidney Kimmel Comprehensive Cancer Center
Johns Hopkins University
550 N. Broadway Suite 1103
Baltimore MD, 21205, United States of America
Email: cwang68@jhmi.edu, nhender5@jhmi.edu, ravi.varadhan@jhu.edu

Thomas A. Louis
Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health
615 North Wolfe Street, E3545
Baltimore MD, 21205, United States of America
Email: tlouis@jhsphe.edu

Carlos O. Weiss
Mercy Health Physician Partners – Geriatrics
200 Jefferson SE
Grand Rapids, MI 49503, United States of America
Email: weissco@mercyhealth.com