# BayesCombo: A Quick Guide

*Bruno Contrino & Stanley E. Lazic*

*22 Jan 2017*

## Introduction

Scientists often evaluate theories or draw conclusions by informally combining results from several experiments. The experiments and the measured outcomes are usually diverse – making a meta-analyses inappropriate – and scientists therefore typically use the number of significant p-values to support their conclusion. P-values, however, are a poor way of integrating results, and since statistical power is often low, "conflicting results" are common. Informal methods of evaluating a series of experiments makes inefficient use of the data and can lead to incorrect conclusions and poor decisions. Here we show how to combine diverse evidence across experiments using Bayes factors [1,2], based on a method developed by Kuiper et al. [3].

The procedure is outlined in Figure 1 and consists of the following five steps:

1. Before seeing the data, specify the prior probability of three hypotheses: that the effect is less than zero, greater than zero, and exactly zero. An equal probability of 1/3 is usually appropriate for these exhaustive and mutually exclusive hypotheses. (It is also possible to have a range of values around zero instead of a point value of exactly zero, and the null hypothesis could be a value other than zero; we will ignore these details for now.) These probabilities will be updated after observing the data.

2. Specify a prior distribution for the effect size (ES). The ES could be a difference between means, the slope of a regression line, or an odds ratio. A sensible default prior is used if none is specified.

3. Calculate the effect size and standard error (SE) from an experiment, which can be obtained from output of a statistical analysis.

4. Calculate the Bayes factor (BF) for each hypothesis, which represents the evidence for a hypothesis after seeing the data, relative to the probability of a hypothesis before seeing the data. The BFs are calculated as a ratio of posterior to prior distributions over a range of parameter values.

5. Update the prior hypothesis probabilities with the Bayes factors to give the posterior probabilities for each hypothesis.

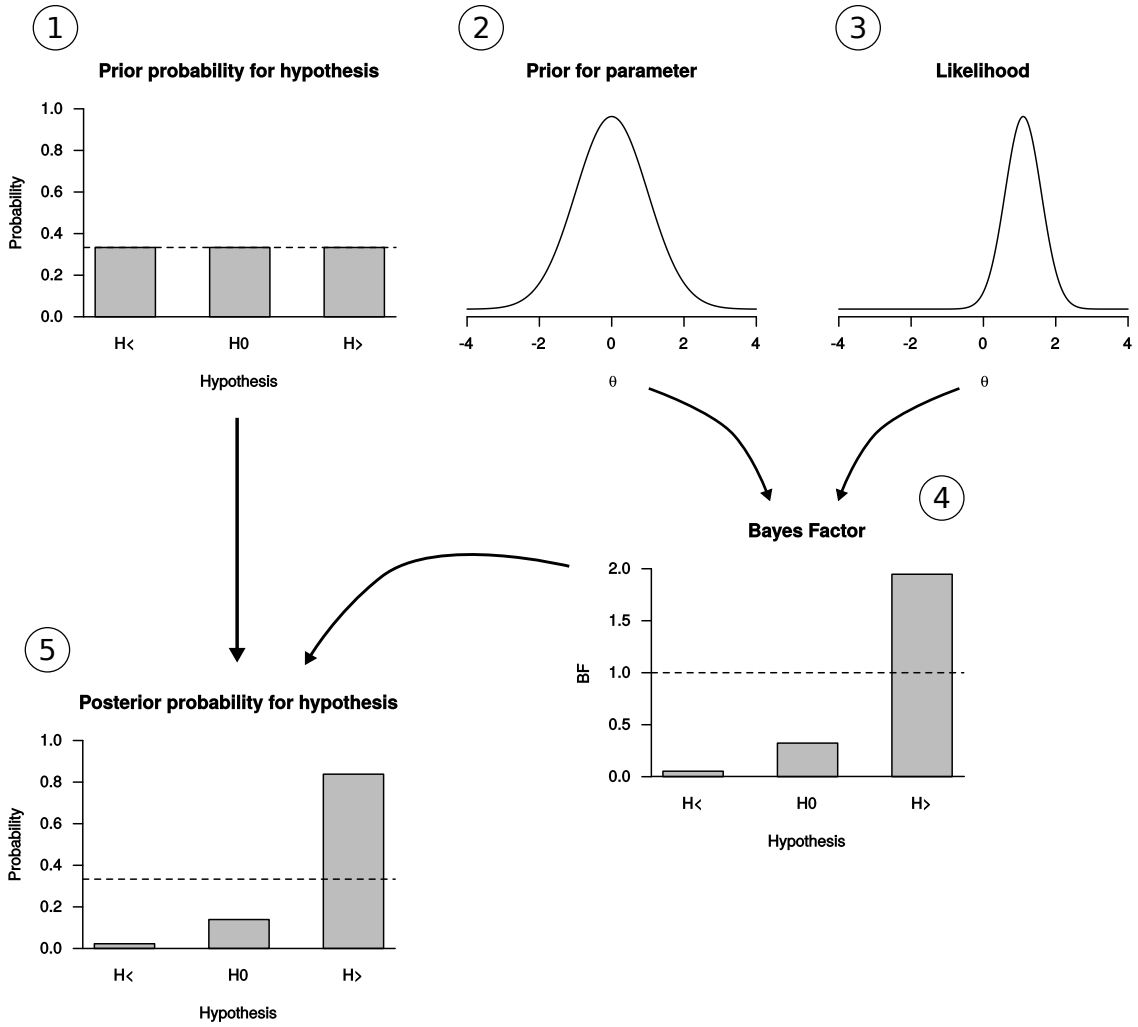


Figure 1: Five steps to get from a prior to a posterior probability of a hypothesis.

For the next experiment, use these updated probabilities to replace those defined in Step 1 and go through steps 2–5 again. Repeat for all experiments that you want to include. Let's work through an example.

# Posterior probability for a single experiment

## Step 1: Define priors for hypotheses

We have an experiment where 20 rats were randomised to one of four doses of the antidepressant fluoxetine given in the drinking water. The time that the rats spent immobile in the Forced Swim Test – a standard behaviour test of "depression" in rodents – was recorded.

We specify the prior probability of three hypotheses: that fluoxetine increases (`H>`), decreases (`H<`), or has no

effect (H0) on immobility time. These three hypotheses are exhaustive (include all possible outcomes) and mutually exclusive (only one can be true). Although fluoxetine is known to decrease immobility time, we will specify an equality probability of 1/3 for each hypothesis to illustrate the approach.

### Step 2: Define prior for effect size

Next, we need to specify a prior for the effect size (we define the effect size in the Step 3). For now we will use the default prior, which is calculated from the data. It is a normal prior, centred at zero, with the variance calculated such that the 99% confidence interval (CI) of the prior matches the 99% CI of the data distribution.

### Step 3: Calculate effect size and standard error

The data for this example are in the `labstats` package (available on CRAN) and plotted in Figure XXX. To calculate the effect size (and the prior in the previous step) we need to define the analysis. Here, dose is treated as a continuous variable (see reference [5]) and so a linear regression will quantify the relationship between fluoxetine and immobility time. No effect corresponds to a flat line (slope = 0) in Figure XXX.

```
library(labstats)
par(las=1)
plot(time.immob ~ dose, data=fluoxetine, col="royalblue",
    ylab="Time immobile (s)", ylim=c(0, 250), xlim=c(0, 250),
    xlab="Dose of fluoxetine (mg/L in drinking water)")
abline(lm(time.immob ~ dose, data=fluoxetine), col="royalblue")
```

The code below calculates the effect size and standard error. These are the required inputs and are returned from `lm()`, `glm()` and related functions (e.g. from ANOVAs, t-tests, or regressions with Gaussian, Poisson, or binomial outcomes).

```
summary(lm(time.immob ~ dose, data=fluoxetine))$coef
```

```
##             Estimate Std. Error  t value    Pr(>|t|)
## (Intercept)  170.440 14.8368131 11.48764 1.01541e-09
## dose          -0.252  0.0991326 -2.54205 2.04365e-02
```

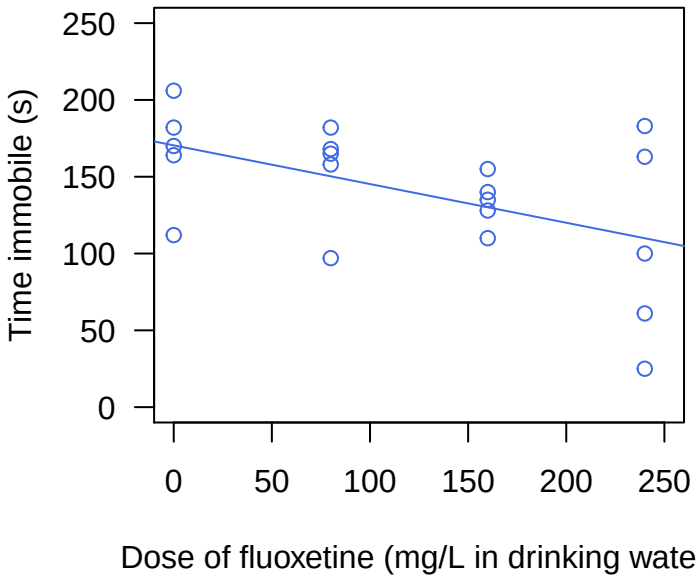


Figure 2: Effect of fluoxetine (Prozac) on rats in the Forced Swim Test. Data are from Lazic [4]

From the above output we see that the estimated slope is -0.252, with a standard error of 0.099, and a p-value of 0.020. We now have all the information.

### Step 4: Calculate Bayes factors

Bayes factors are used as an intermediate step in calculating the posterior probabilities of each hypothesis. The functions in the `BayesCombo` package calculate these automatically, but we illustrate how they are calculated to provide some insight into the method. The top panel of Figure XXX shows the data (likelihood) distribution, which has a mean of 0.75 and standard deviation of 1. The prior in the middle panel is normal with a mean of zero and standard deviation of 1.29. The bottom panel shows the posterior, which is shifted to positive values. The BFs for the three hypotheses are then calculated as the ratio of posterior to prior areas or heights on the curve:

$$BF_{H<0} = \frac{\text{area of d}}{\text{area of a}}$$

$$BF_{H=0} = \frac{\text{height of point e}}{\text{height of point b}}$$

$$BF_{H>0} = \frac{\text{area of f}}{\text{area of c}}$$

In Figure XXX 50% of the prior distribution is above 0 (region c), as is 72% of the posterior (region f). The interpretation is that the data have increases the plausibility of hypothesis H> from 50% to 72%. The ratio of these values is the Bayes factor and is equal to$0.72/0.5 = 1.4$. Thus we can say that H> is 1.4 times more likely. More generally, a BF $> 1$ means that the data supports a hypothesis, whereas a BF $< 1$ means that data do not support a hypothesis.

### Step 5: Use the BFs to update the prior hypothesis probabilities

The final step is to update the prior probability for each hypothesis with the BFs to get the posterior probabilities. The equation below shows the calculation for the hypothesis that the effect is greater than zero (H>), and other probabilities are calculated in the same way, just putting a different BF in the numerator. $Pr()$ are the prior probabilities for each hypothesis, and they cancel out from the equation when they are all equal.

$$P(H >) = \frac{Pr(H >)BF_{H>}}{Pr(H <)BF_{H<} + Pr(H0)BF_{H0} + Pr(H >)BF_{H>}}$$

All of the above steps can be conveniently calculated using the `pph()` function (Posterior Probability for Hypotheses). Returning to the fluoxetine example, we can calculate the probability that the slope is negative, positive, or zero. Below, we specify the slope (`beta = -0.252`) and its standard error (`se.beta = 0.099`) that we obtained previously from the output of the `lm()` function. The default settings are used for the other options and so there is nothing else to specify. The output below shows that the probability of a negative slope (`H<`) is 0.9120, a positive slope (`H>`) is 0.0106, and a zero slope (`H0`) is 0.0774.

```
x <- pph(beta = -0.252, se.beta = 0.099)
summary(x)
```

```
##     H<     H0     H>
## 0.9120 0.0774 0.0106
```

Plotting the output of the `pph()` function returns the data, prior, and posterior distributions (Fig. XXX). The mean of the data (dotted line) is centred on -0.252 and the standard error determines the width of this distribution. The posterior probabilities of the three hypotheses are calculated from these prior and posterior distributions.

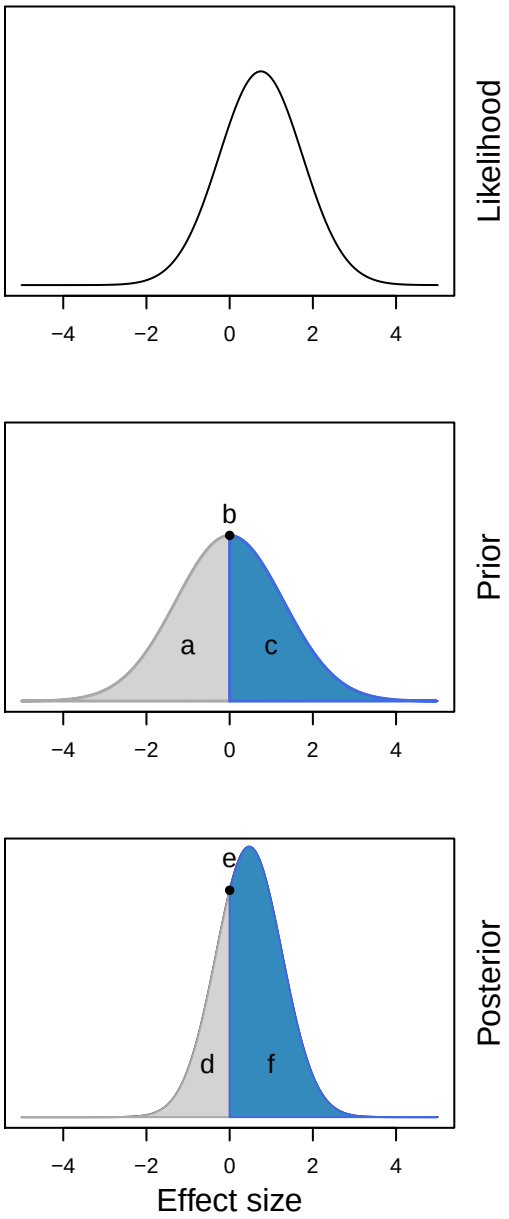


Figure 3: Likelihood, prior, and posterior for a experiment.

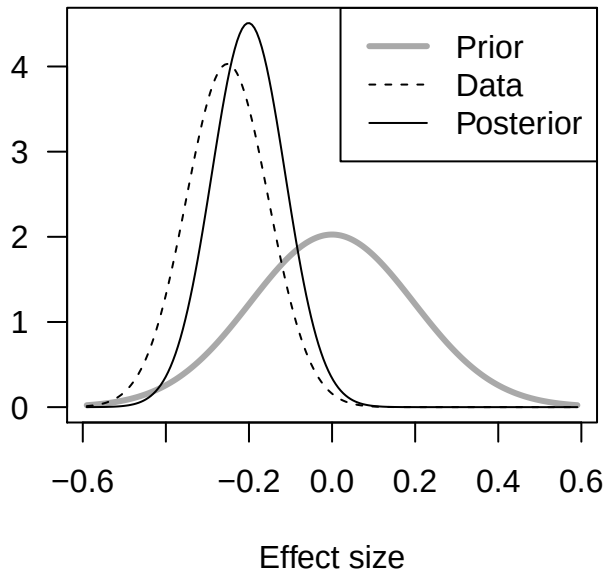


Figure 4: sfsd.

```
par(las=1)
plot(x, leg.loc = "topright")
```

The above default prior may be more informative than desired (note how the posterior is pulled towards the prior in Fig XXX) and the easiest way to decrease the influence of the prior is to make it wider by specifying a multiplier. The code below doubles the previous standard error (`se.mult = 2`), and the posterior is now much closer to the data distribution (Fig XXX). However, the posterior hypothesis probabilities have not changed much, rounded to two decimal places the probability that the slope is negative is still 0.91.

```
x2 <- pph(beta = -0.252, se.beta = 0.099, se.mult = 2)
summary(x2)
```

```
##     H<     H0     H>
## 0.9051 0.0887 0.0062
```

```
par(las=1)
plot(x2, leg.loc = "topright")
```

A final example illustrates other options. The prior on the slope is directly specified as having a mean (`beta0`) of 0 and standard error (`se0`) of 1.2. In the previous analyses H0 was always exactly equal to 0, but here we define H0 as a range of values close to zero with `H0 = c(-0.05, 0.05)`. Finally, the priors on the hypotheses

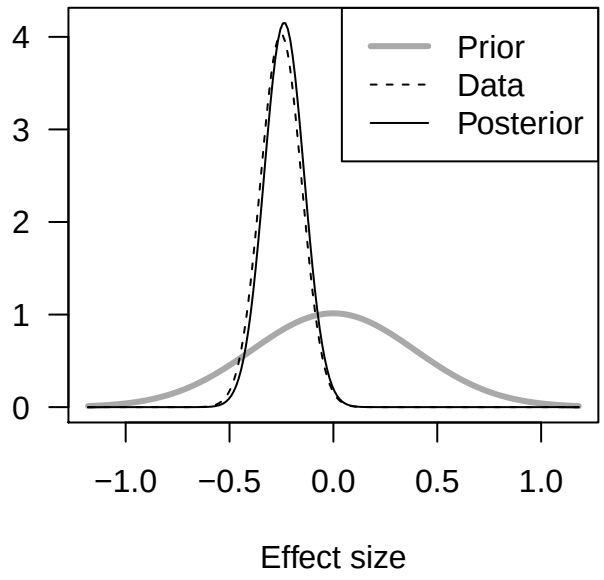


Figure 5: sfsd.

are also given as an argument to `H.priors`. The values indicate that the prior probability of the slope being negative, zero, and positive are 0.495, 0.495, 0.01, respectively. The interpretation is that we expect that fluoxetine either decreases immobility time or has no effect, but it is unlikely to increase immobility time.

```
x3 <- pph(beta = -0.252, se.beta = 0.099, beta0=0, se0=1.2,
          H0 = c(-0.05, 0.05), H.priors=c(0.495, 0.495, 0.01))
summary(x3)
```

```
##    H<    H0    H>
## 0.984 0.016 0.000
```

```
par(las=1)
plot(x3, leg.loc = "topright")
```

## Posterior probability for multiple experiments

We can use the above procedure to sequentially combine the results of experiments where the posterior probabilities for hypotheses from one experiment is used as the prior probability for the next experiment. It's up to the user, however, to ensure that the experiments are testing the same overall hypothesis or theory. In addition, the direction of the effects should be aligned; for example, if a positive effect size in one experiment

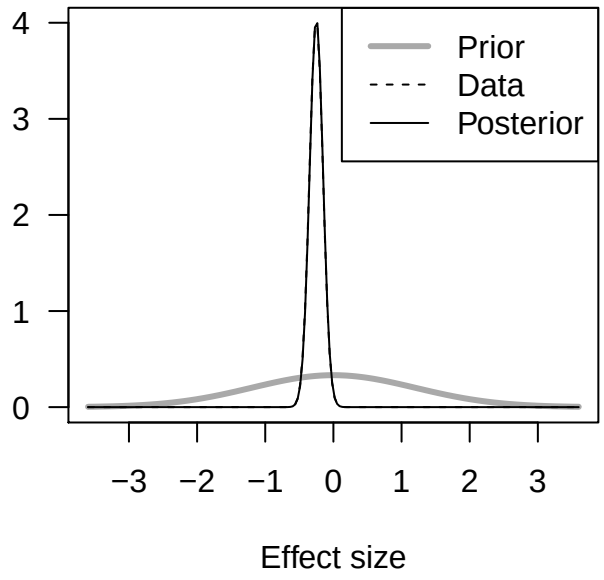


Figure 6: sfsd.

is interpreted as supporting a theory, but a negative effect size in another experiment also supports the theory, then the negative effect should be multiplied by -1 to change its direction.

`ev.combo()` is the key function and only requires the effect sizes (`beta`) and their standard errors (`se.beta`) as input, same as the `pph()` function. The default prior mean (`beta0 = 0`) is suitable for most analyses, as is the equal prior hypothesis probabilities (`H.priors = c(1/3, 1/3, 1/3)`) for each hypothesis (negative, zero, or positive effect). In the example below, assume we have four clinical trials where positive effect sizes indicate a beneficial effect of a treatment.

```
x4 <- ev.combo(beta = c(2.3, 1.2, 0.2, 0.44),
               se.beta = c(1.03, 0.75, 0.16, 0.28))
```

The `forestplot()` function makes a graph resembling a traditional forest plot, with the observed effect sizes and their 99% CI (black) and the prior effect sizes (grey). All four experiments have positive effect sizes but only the first experiment is significant at the usual 0.05 level.

```
par(las=1)
forestplot(x4)
abline(v=0, lty=2)
```

The results summary below shows how the support for the three hypothesis changes as each experiment is added. The first line in the output contains the prior hypothesis probability of 1/3 or 33%. When the first

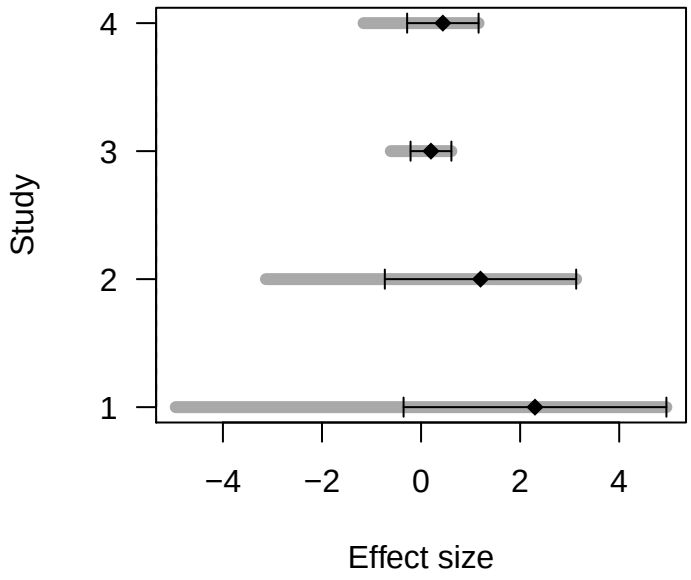


Figure 7: caption

experiment is included (second row) the probability that the effect size is greater than zero (`H>`) increases to 85%. As more experiments are included, the probability increases further to 98%, and the null has only 1.7% support.

```
summary(x4)
```

```
##          H<     H0     H>
## [1,] 0.3333 0.3333 0.3333
## [2,] 0.0213 0.1324 0.8464
## [3,] 0.0022 0.0605 0.9373
## [4,] 0.0004 0.0382 0.9614
## [5,] 0.0000 0.0167 0.9833
```

It is easier to see how these probabilities change as experiments are added with a graph.

```
par(las=1)
plot(x4, ylab="PPH", xlab="Experiment")
```

In addition to seeing how the probabilities of hypotheses change as experiments are added, we can also plot each experiment's individual contribution. This is equivalent to analysing each experiment with the `pph()` function and the results are shown in Fig XXX. The `H>` hypothesis is above 0.5 for all four experiments (left graph).

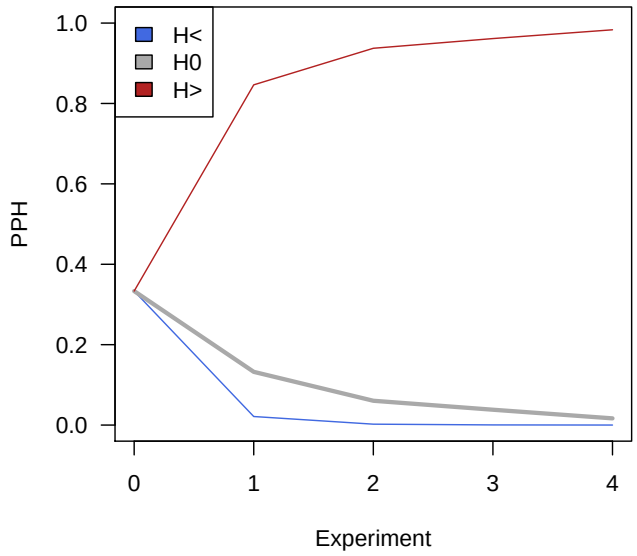


Figure 8: caption

```
par(mfrow=c(1,2))
dotchart(x4$pph.uniform, xlim=c(0,1), xlab="PPH", pch=21, bg="grey")
dotchart(t(x4$pph.uniform), xlim=c(0,1), xlab="PPH", pch=21, bg="grey")
```

## References


1. Wagenmakers E-J, Lodewyckx T, Kuriyal H, Grasman R (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method. *Cognitive Psychology* 60: 158-189.

2. Wetzels R, Grasman RP, Wagenmakers E-J (2010). An encompassing prior generalization of the Savage-Dickey density ratio. *Computational Statistics and Data Analysis* 54: 2094-2102.

3. Kuiper RM, Buskens V, Raub W, Hoijtink H (2012). Combining statistical evidence from several studies: A method using Bayesian updating and an example from research on trust problems in social and economic exchange. *Sociological Methods and Research* 42(1): 60-81.

4. Lazic SE (2008). Why we should use simpler models if the data allow this: relevance for ANOVA designs in experimental biology. *BMC Physiology* 8:16.

5. Lazic SE (2016). *Experimental Design for Laboratory Biologists: Maximising Information and Improving Reproducibility.* Cambridge University Press: Cambridge, UK

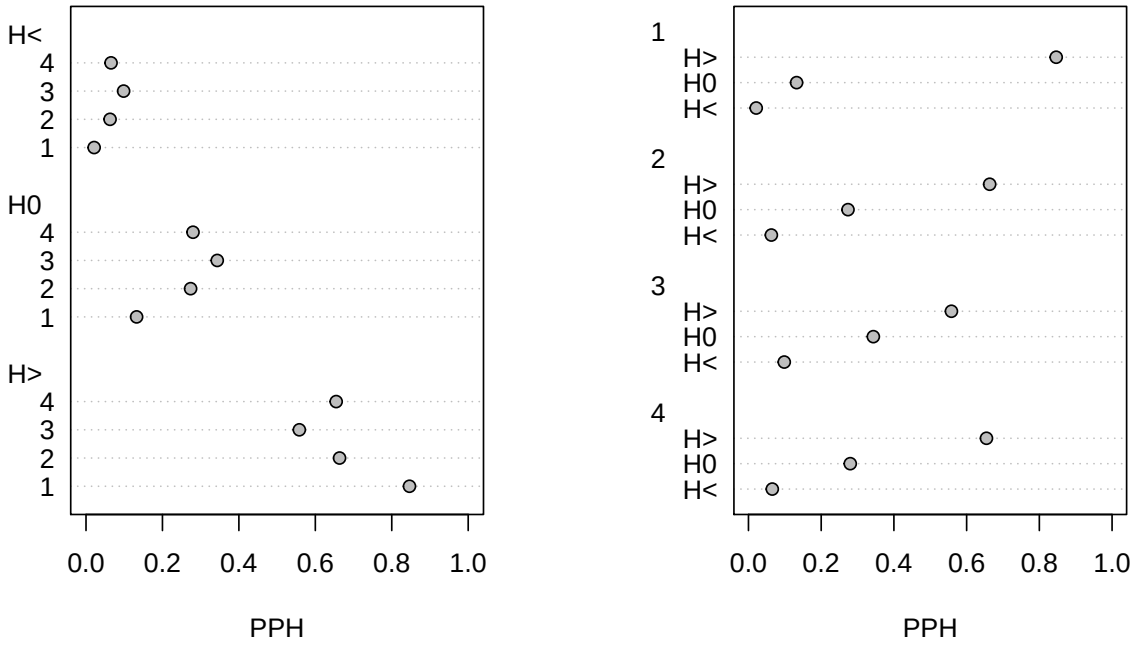


Figure 9: caption