

# BayesCombo: A Quick Guide

*Bruno Contrino & Stanley E. Lazic*

*15 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-analysis inappropriate – and scientists therefore typically use the number of significant p-values to support a theory. 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 diverse experiments based on a method developed by Kuiper et al. [1].

The basic procedure is the following:

1. Define three hypotheses and their probability before seeing the data. The hypotheses are that the effect is either less than zero, greater than zero, or exactly zero (it is also possible to have a range of values around zero, we will ignore this for now). Usually an equal probability of 1/3 is appropriate.
2. Obtain an effect size and standard error from an experiment. The effect size could be a difference between means, or the slope of a regression line for example.
3. Calculate the Bayes factor (BF) for each hypothesis, which is calculated as a ratio of posterior to prior distribution over different parameter values.
4. Convert the Bayes factors for each hypothesis into the probability for each hypothesis.
5. Go back to Step 1 and set the probability of the hypotheses (previously 1/3) to the values calculated in the previous step. Then, go through the cycle again with data from a new experiment.

Figure 1 shows an example where the data (likelihood) has a mean of 0.75 and standard deviation of one. The prior is normal with a mean of zero and standard deviation of 1.29. The BFs for the three hypotheses are then calculated as the ratio of posterior to prior areas or heights on the curve:

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

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

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

Converting the BFs into probabilities gives 0.16 0.41 0.43

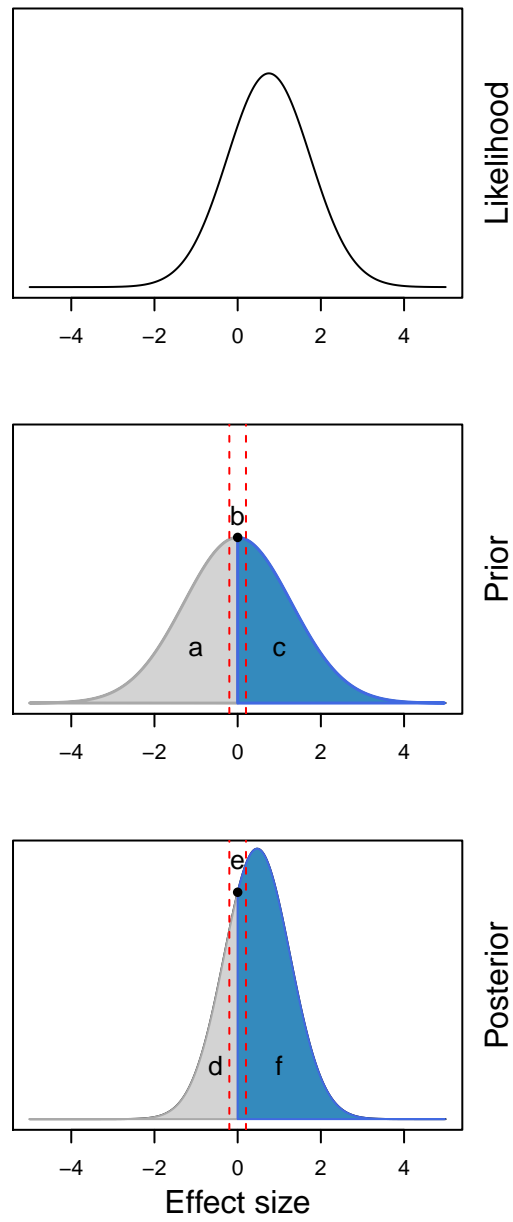


Figure 1: Likelihood, prior, and posterior for a experiment.

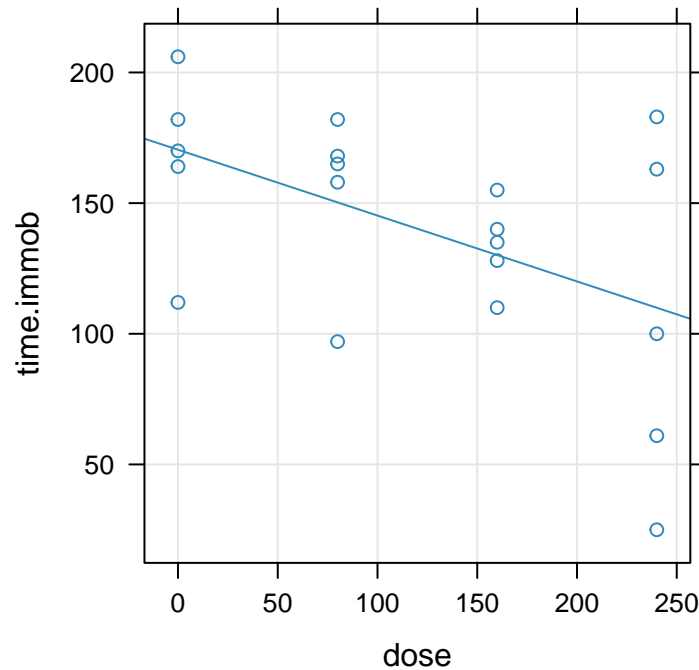


Figure 2: Effect of fluoxetine (Prozac) on rats in the Forced Swim Test. Data are from Lazic [2]

## Usage

The required inputs are coefficients and standard errors obtained from the output of `lm()` and `glm()` functions (e.g. from ANOVAs, t-tests, or regressions with Gaussian, Poisson, or binomial outcomes). It's up to the user to ensure that the data being combined 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 study is interpreted as supporting a theory, but a negative effect size in another study also supports the theory, then the negative effect should be multiplied by -1 to change its direction.

## Posterior probability for a single study

The figure below plots the data from an experiment where 20 rats were randomised to one of four treatment groups (dose of the antidepressant fluoxetine given in the drinking water) and the time (s) that the rats spent immobile in the Forced Swim Test was recorded. These data are in the `labstats` package.

```
library(labstats)
library(lattice)
xyplot(time.immob ~ dose, data=fluoxetine, type=c("g","p","r"), col="#348ABD")
```

We run a linear regression treating dose as a continuous variable (see reference [3]) and are interested in testing if fluoxetine has an effect. No effect corresponds to a flat line (slope = 0) in the above Figure.

```
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
```

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. Using the `pmp()` function, we can calculate the probability that the slope is negative, positive, or zero. The probability of each of these outcomes are called the posterior model probabilities (PMPs), and they sum to one. But we need to specify two types of prior probability. First, we need to define the prior probability of each of these models, and the default is an equality probability of 1/3 for each. Second, we need to define a prior for the effect size. The default is normal prior, centred at zero and variance such that the 99% confidence interval (CI) of the prior matches the 99% CI of the data distribution.

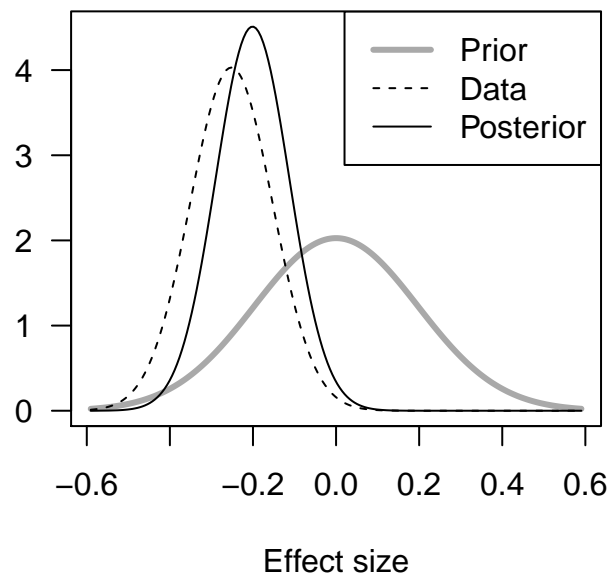
Below, we specify the slope (`beta = -0.252`) and its standard error (`se.beta = 0.099`). The default settings are used and the output shows that the probability of a negative slope ( $H<$ ) is 0.9120, that it is positive ( $H>$ ) is 0.0106, and that it is exactly zero ( $H0$ ) is 0.0774.

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

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

The plot below shows the data, the default prior, and the posterior distributions. The mean of the data (dotted line) is centred on -0.252 and the standard error determines the width of this distribution. The PMPs are then calculated from these prior and posterior distributions.

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

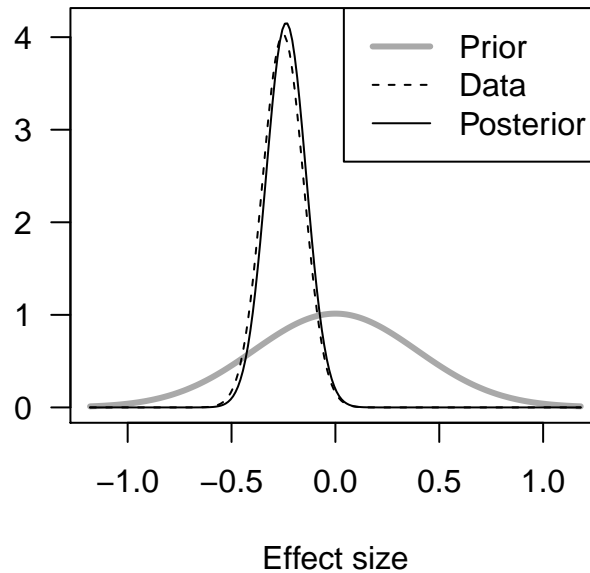


The above default prior may be more informative than desired (note how the posterior is pulled towards the prior) and the easiest way to decrease the influence of the prior is to make it wider by specifying a variance multiplier. The code below doubles the previous variance (`var.mult = 2`), and the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

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

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

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

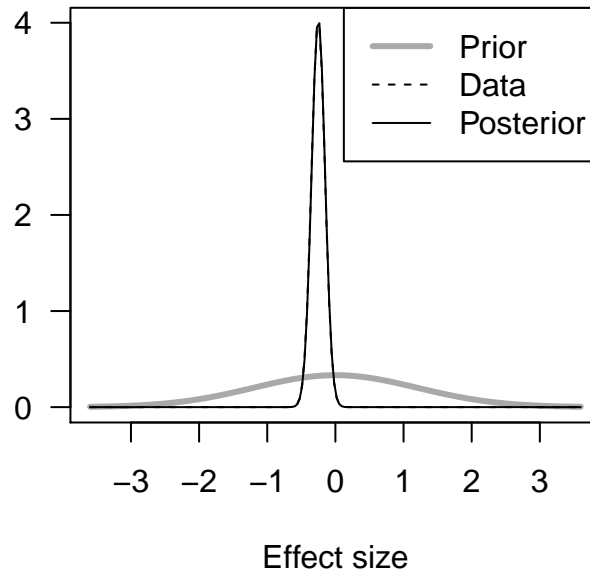


A final example is given below to illustrate other options. The prior on the parameter is directly specified as having a mean (`beta0`) of 0 and standard deviation (`se0`) of 1.2. In the previous analyses `H0` was always exactly equal to 0, but here we define `H0` as a range of values close to zero with `H0 = c(-0.05, 0.05)`. Finally, the priors on the models are also given as an argument to `mod.priors`. The values indicate that the prior probability of the slope being negative, zero, or positive is 0.495, 0.495, 0.01, respectively.

```
x3 <- pmp(beta = -0.252, se.beta = 0.099, beta0=0, se0=1.2,
          H0 = c(-0.05, 0.05), mod.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")
```



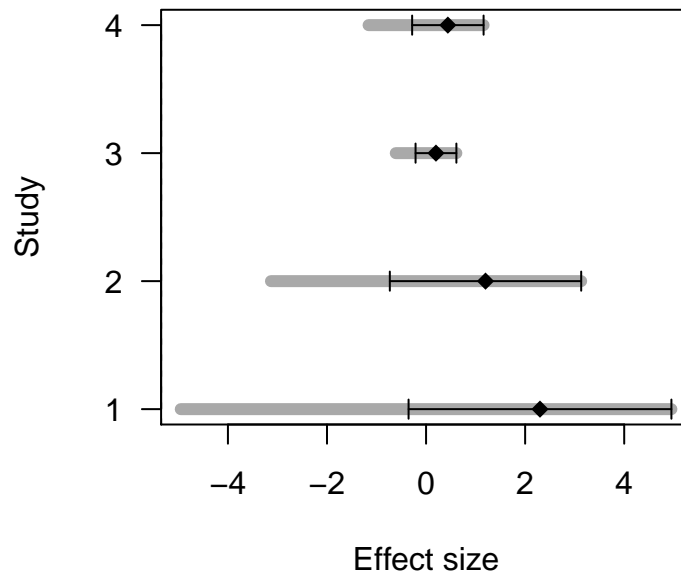
## Posterior probability for multiple studies

PMPs from several studies can be combined to accumulate evidence for a hypothesis or theory. `pmp.combo()` is the main analysis function and only requires the effect sizes (`beta`) and their standard errors (`se.beta`) as input, same as the `pmp()` function. The default prior mean (`beta0 = 0`) is suitable for most analyses, as is the equal prior model probabilities (`mod.priors = c(1/3, 1/3, 1/3)`) for each model (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 <- pmp.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 studies have positive effect sizes but only the first study is significant at the usual 0.05 level.

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



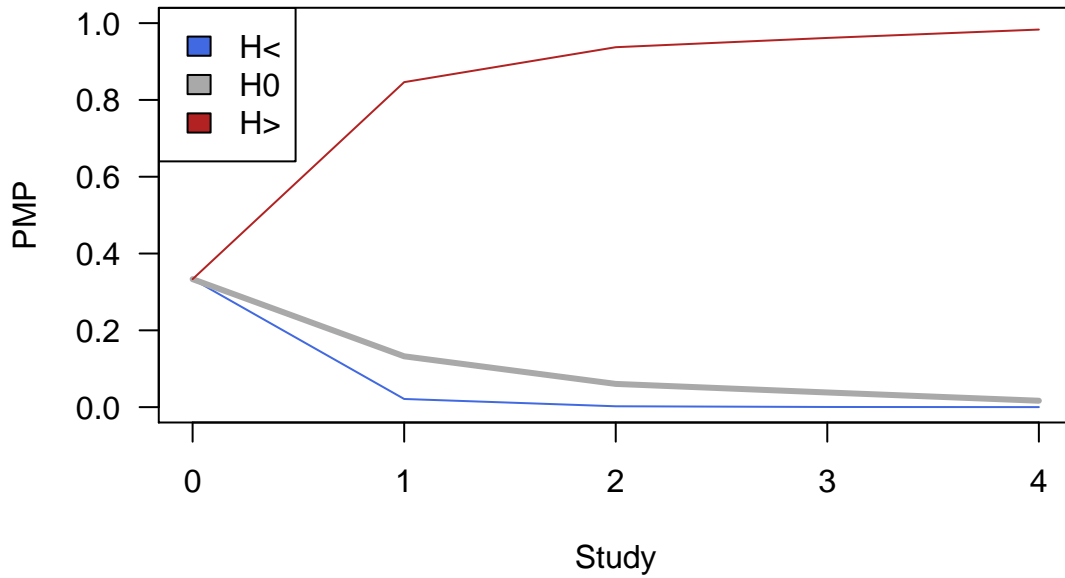
A summary of the results shows how the support for the three hypothesis changes as each study is added. For example, in the output below we see that when only the first study is included (second row) the probability that the effect size is greater than zero ( $H>$ ) is about 85%. As more studies are included, the probability increases to 98%, and the null has only 1.8% 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
```

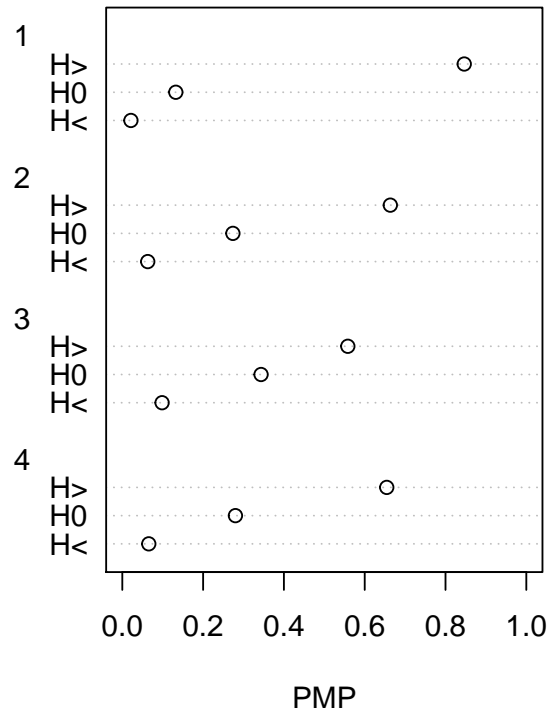
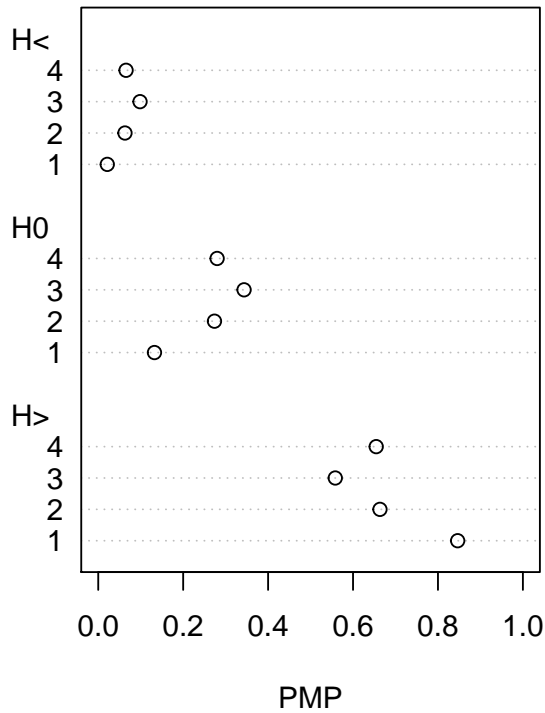
It is easier to see how these probabilities change as studies are added with a graph.

```
par(las=1)
plot(x4, ylab="PMP", xlab="Study")
```



In addition to seeing how the PMPs change as studies are added, we can also extract and plot each study analysed on its own. This is equivalent to analysing each study with the `pmp()` function and the results are shown below with two related graphs. It is clear to see, for example that the  $H>$  hypothesis was above 0.5 for all four experiments (left graph).

```
par(mfrow=c(1,2))
dotchart(x4$pmp.uniform, xlim=c(0,1), xlab="PMP")
dotchart(t(x4$pmp.uniform), xlim=c(0,1), xlab="PMP")
```





## References

1. 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.
2. 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.
3. Lazic SE (2016). *Experimental Design for Laboratory Biologists: Maximising Information and Improving Reproducibility*. Cambridge University Press: Cambridge, UK