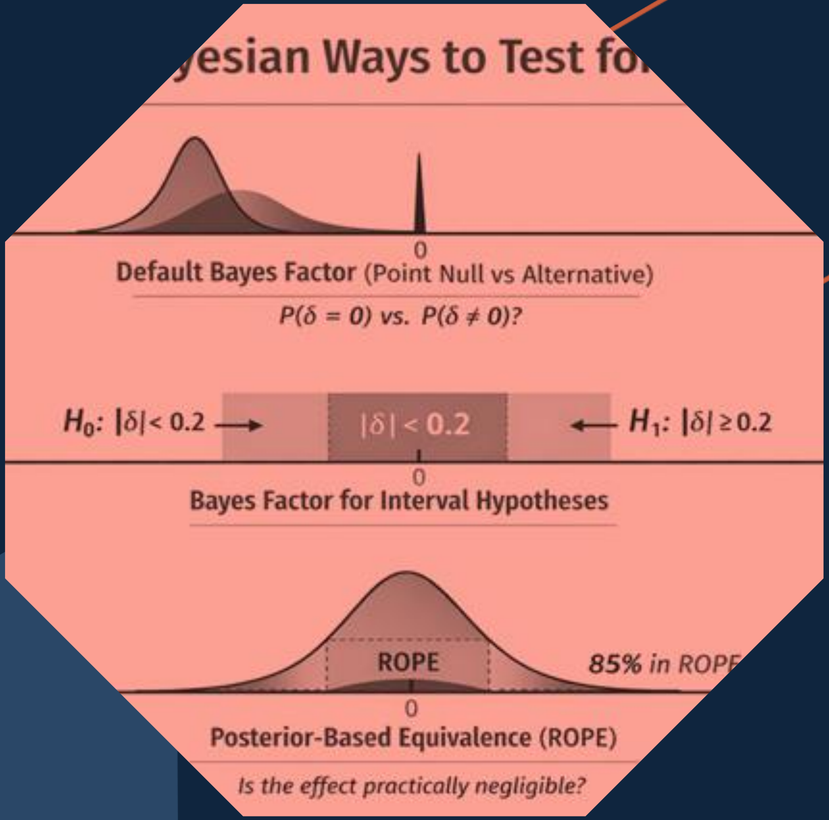
When nothing happens:
Bayesian approaches to testing for 'no effect'

Mircea Zloteanu

King's College London


yesian Ways to Test fo
0
Default Bayes Factor (Point Null vs Alternative)
P(δ = 0) vs. P(δ ≠ 0)?
H₀: |δ| < 0.2
|δ| < 0.2
H₁: |δ| ≥ 0.2
0
Bayes Factor for Interval Hypotheses
ROPE
85% in ROPE
0
Posterior-Based Equivalence (ROPE)
Is the effect practically negligible?

Agenda

- What is “no effect”
- Bayes Factors and Posteriors
- Model comparison (Nill)
- Interval comparison (Null)
- Posterior probability (ROPE)

What is so special about “no effect”?

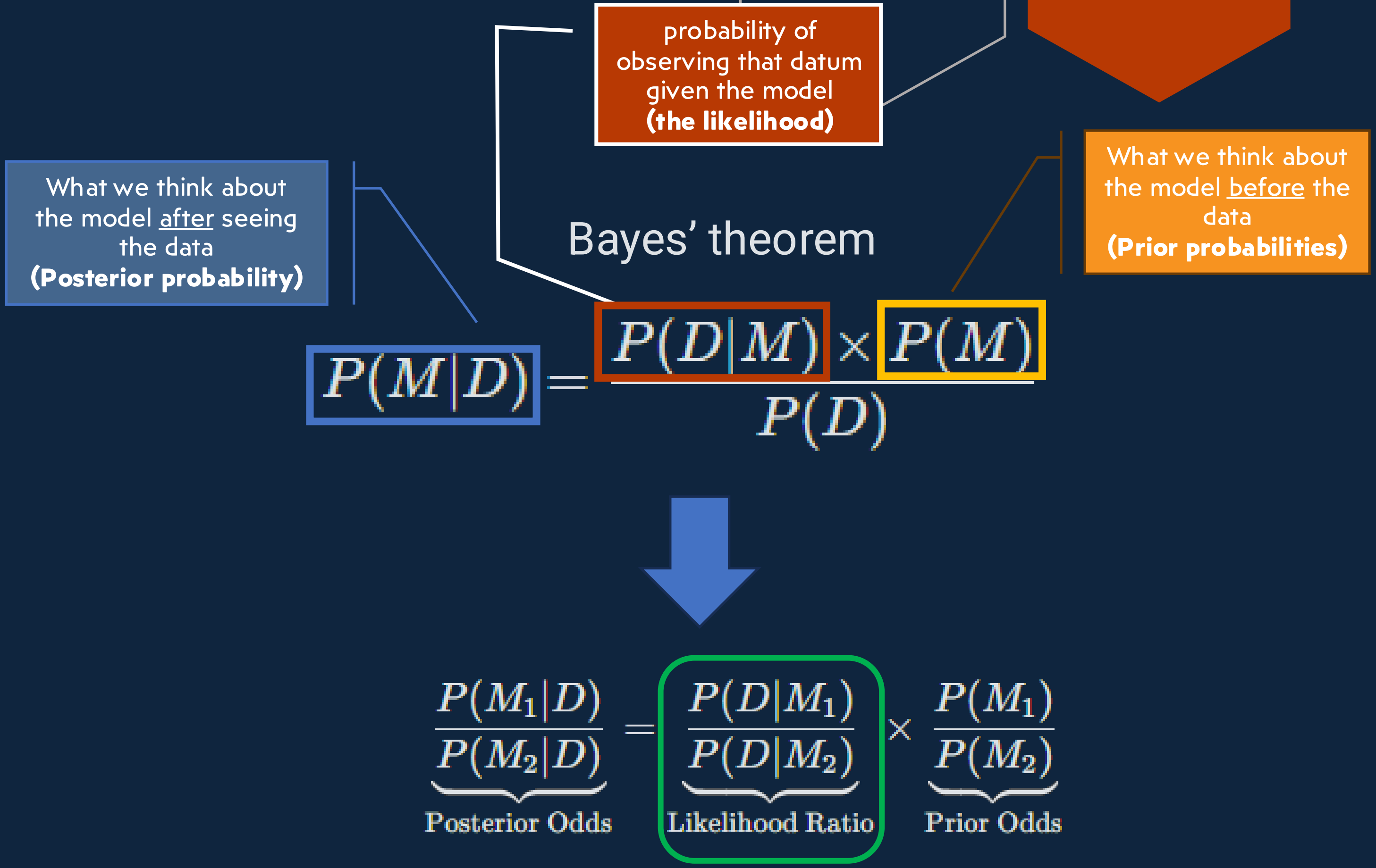
- In the frequentist framework, NHST, you often determine the presence of a non-zero effect using a p-value.
- However, the p-value alone cannot be used to determine “no effect” (for that you need additional testing)
- Planning for such tests is very different than planning to just reject non-zero.
- Often, if a researcher has not planned out things very well, they write “there was no effect, p>.05”, and a Reviewer (me) comes along and says “you can’t say that”. So, they ask a friend, who tells them “have you tried Bayes Factors?”
- Today, I will show that while the Bayesian framework can be used to make such claims, it isn’t that simple.

Bayes Factors & Posteriors

Quick overview of how we do things in the Bayesian framework

Bayes Factors

- Bayes Factors (BFs) are indices of *relative* evidence of one “model” over another
- The “relative” part is important
- All they do is compare models
- If your models are wrong, so is your conclusion
- If you have two bad models, then this just says “one of these nonsense models is better than the other nonsense model”



What we think about the model after seeing the data **(Posterior probability)**

probability of observing that datum given the model **(the likelihood)**

What we think about the model before the data **(Prior probabilities)**

Bayes’ theorem

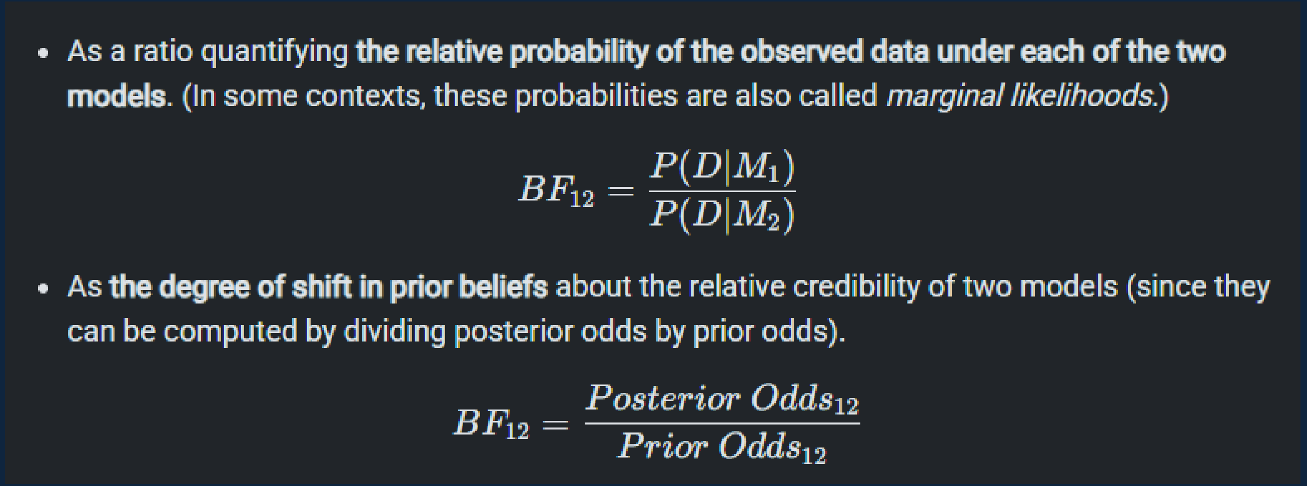
$$\boxed{P(M|D)} = \frac{\boxed{P(D|M)} \times \boxed{P(M)}}{P(D)}$$

$$\underbrace{\frac{P(M_1|D)}{P(M_2|D)}}_{\text{Posterior Odds}} = \boxed{\underbrace{\frac{P(D|M_1)}{P(D|M_2)}}_{\text{Likelihood Ratio}}} \times \underbrace{\frac{P(M_1)}{P(M_2)}}_{\text{Prior Odds}}$$

We can use the above equation to compare the probability-odds of two models

Bayes Factors

- **The BF is just a Likelihood Ratio**
- It is the ***factor*** by which some **prior odds** have been updated after observing the data to **posterior odds**.
- It is purely **objective!**
- (what isn't objective is the Prior Probabilities you assign models)



- As a ratio quantifying **the relative probability of the observed data under each of the two models**. (In some contexts, these probabilities are also called *marginal likelihoods*.)

$$BF_{12} = \frac{P(D|M_1)}{P(D|M_2)}$$

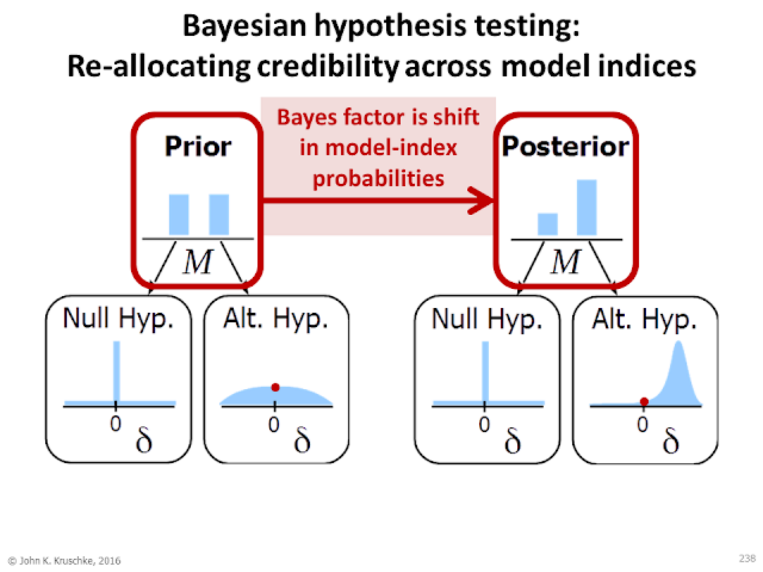
- As **the degree of shift in prior beliefs** about the relative credibility of two models (since they can be computed by dividing posterior odds by prior odds).

$$BF_{12} = \frac{Posterior\ Odds_{12}}{Prior\ Odds_{12}}$$

> "I would also bet that most people aren't even aware BFs test only **priors** and not posteriors"
> – Mattan Beh-Shachar

BF tells you: *Given the prior assumptions of each model, which model predicted the observed data better?*

BFs can be used in two ways



Bayesian hypothesis testing:
Re-allocating credibility across model indices

© John K. Kruschke, 2016

238

- **Testing single parameters (coefficients) within a model**
- **Comparing statistical models themselves**
- **Often, these approaches are referred to as:** Null hypothesis Bayesian testing (NHBT)

Note:

Many researchers think the BF answers:

  - “How plausible is the hypothesis after seeing the data?”

But what it actually answers is:

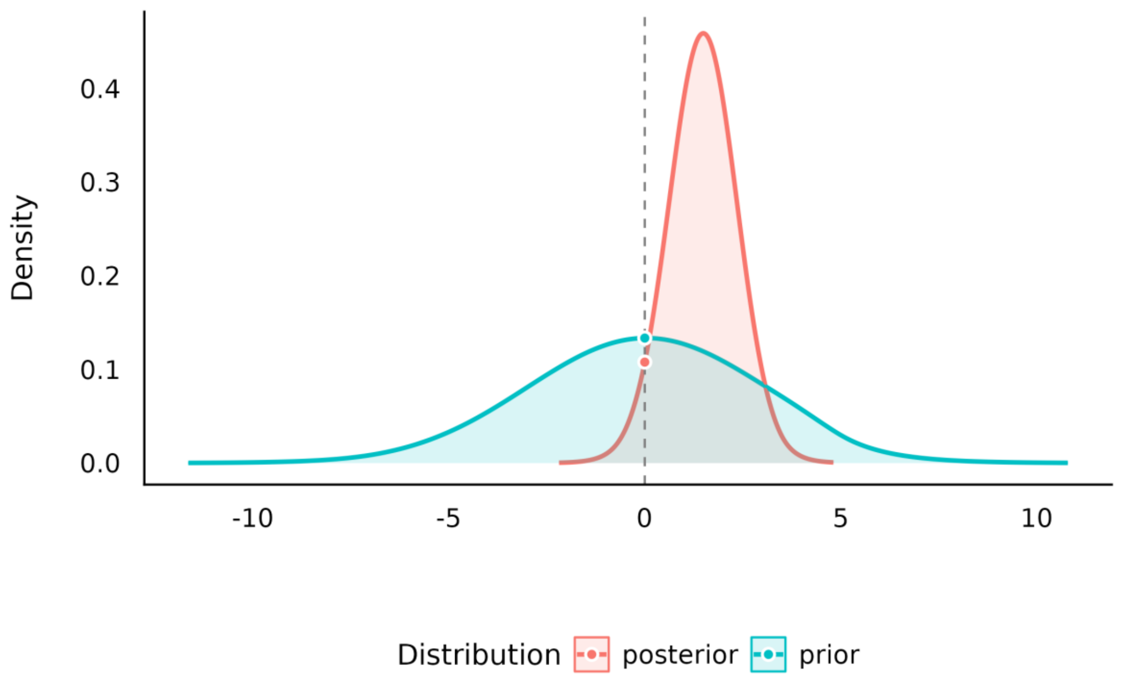
- “Which model predicted the observed data better *before seeing it*?”

No effect = parameter no better than 0

BF for parameter testing 2 ways

NHBT

- For Bayesian parameter estimation, interest centers on the posterior distribution of the model parameters.
- The posterior distribution reflects the relative plausibility of the parameter values after prior knowledge has been updated by means of the data.
- Testing refers to the relative evidence for M_0 or M_1 given the data. Not about the parameter itself.

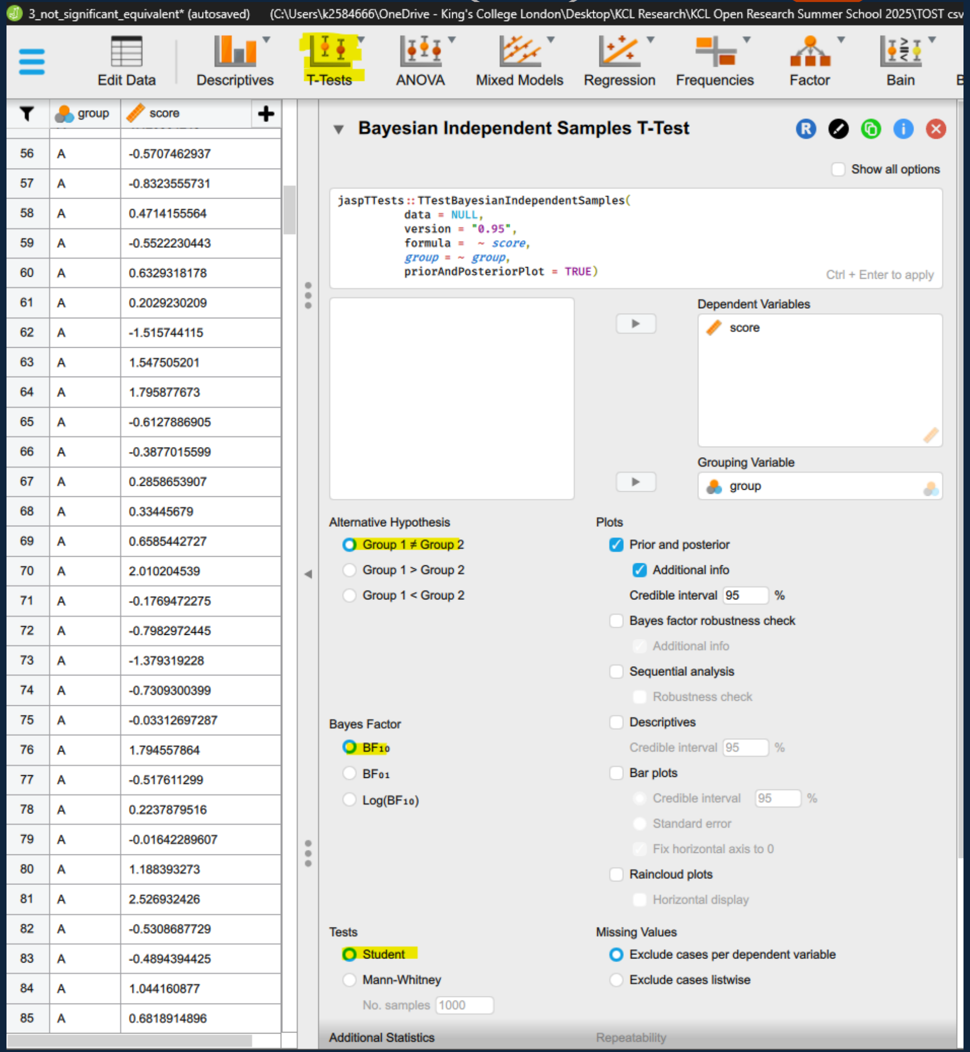


Parameter Estimation =! Hypothesis Testing

Berger (2006, p. 383): “[...] Bayesians cannot test precise hypotheses using confidence intervals. In classical statistics one frequently sees testing done by forming a confidence region for the parameter, and then rejecting a null value of the parameter if it does not lie in the confidence region. This is simply wrong if done in a Bayesian formulation (and if the null value of the parameter is believable as a hypothesis).”

Example 1: (true) No effect

- First, I show what we can obtain with the different methods.
- Using JASP can be an easy start



3_not_significant_equivalent* (autosaved)   (C:\Users\k2584666\OneDrive - King's College London\Desktop\KCL Research\KCL Open Research Summer School 2025\TOST csv

	group	score
56	A	-0.5707462937
57	A	-0.8323555731
58	A	0.4714155564
59	A	-0.5522230443
60	A	0.6329318178
61	A	0.2029230209
62	A	-1.515744115
63	A	1.547505201
64	A	1.795877673
65	A	-0.6127886905
66	A	-0.3877015599
67	A	0.2858653907
68	A	0.33445679
69	A	0.6585442727
70	A	2.010204539
71	A	-0.1769472275
72	A	-0.7982972445
73	A	-1.379319228
74	A	-0.7309300399
75	A	-0.03312697287
76	A	1.794557864
77	A	-0.517611299
78	A	0.2237879516
79	A	-0.01642289607
80	A	1.188393273
81	A	2.526932426
82	A	-0.5308687729
83	A	-0.4894394425
84	A	1.044160877
85	A	0.6818914896

Bayesian Independent Samples T-Test

□ Show all options

```
jaspTTests::TTestBayesianIndependentSamples(
          data = NULL,
          version = "0.95",
          formula =  ~ score,
          group = ~ group,
          priorAndPosteriorPlot = TRUE)
```

Ctrl + Enter to apply

Dependent Variables: score

Grouping Variable: group

Alternative Hypothesis
- ● Group 1 ≠ Group 2
- ○ Group 1 > Group 2
- ○ Group 1 < Group 2

Plots
- ☑ Prior and posterior
  - ☑ Additional info
  - Credible interval 95 %
- □ Bayes factor robustness check
  - ☑ Additional info
- □ Sequential analysis
  - □ Robustness check
- □ Descriptives
  - Credible interval 95 %
- □ Bar plots
  - ○ Credible interval 95 %
  - ○ Standard error
  - ☑ Fix horizontal axis to 0
- □ Raincloud plots
  - □ Horizontal display

Bayes Factor
- ● BF_{10}
- ○ BF_{01}
- ○ $Log(BF_{10})$

Tests
- ● Student
- ○ Mann-Whitney
  - No. samples 1000

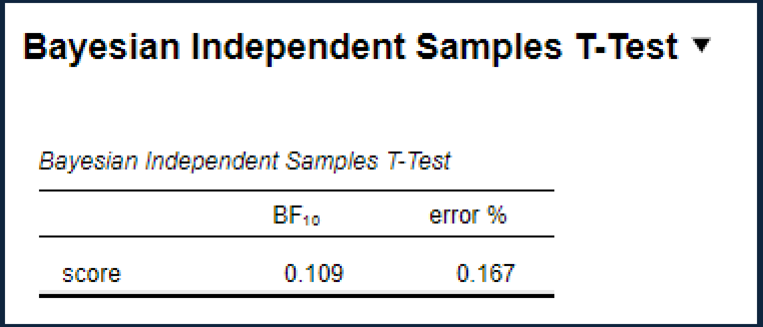
Missing Values
- ● Exclude cases per dependent variable
- ○ Exclude cases listwise

Additional Statistics

Repeatability

Example 1: (true) No effect

- Running the test shows the BF for M1 vs M0
- This Table is all that is needed to make a judgements.
- Here, we can say that if we assume:
  - H_0 is $\delta = 0$
  - H_1 is $\delta \sim$ prior (Cauchy, 0, 0.707)
- Given this, we see the $BF_{10} = 0.11$, or if we flip it, $BF_{01} = 1/BF_{10} = 9.2$



Bayesian Independent Samples T-Test ▾

Bayesian Independent Samples T-Test

	BF_{10}	error %
score	0.109	0.167

The observed data are 9x more likely under H_0 than under H_1

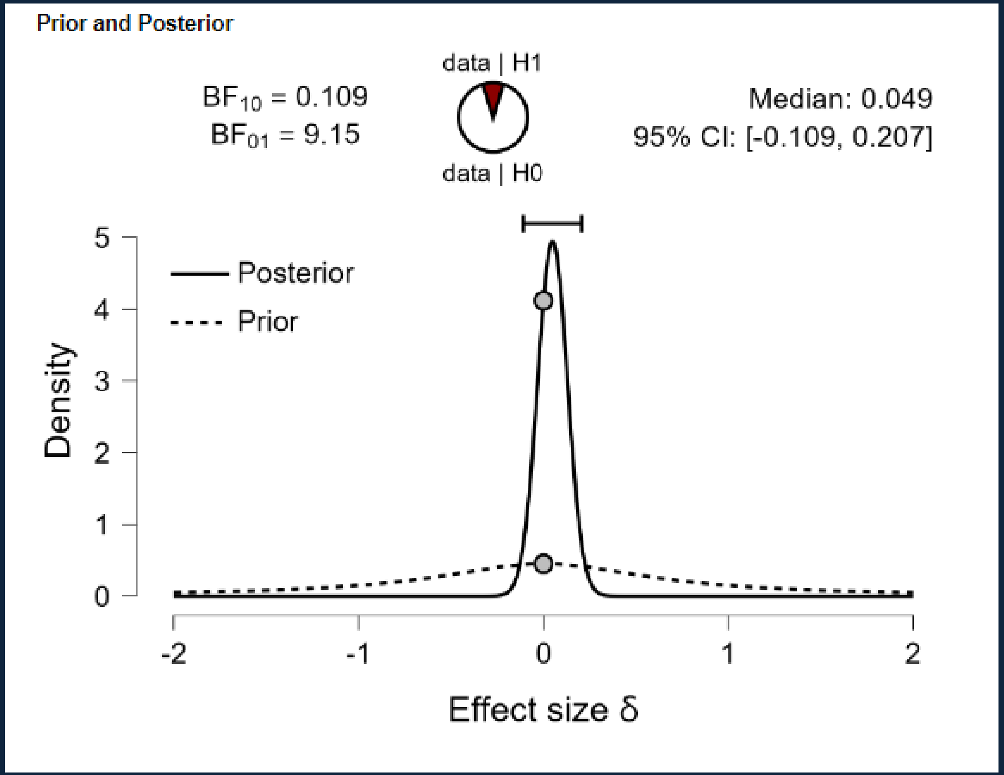
The observed data tell us that we should revise our relative initial belief by a factor of 9 to 1 in favour of H_0 .

Someone with no prior preference for either hypothesis (i.e., prior odds = 1) should now believe that the null model is 9 times more probable than this alternative model (i.e., posterior odds = 9 x 1 = 9).

"[…] the Bayes factor for H_0 versus H_1 could be obtained by analytically integrating out the model parameter θ. However, the Bayes factor may likewise be obtained by only considering H_1, and dividing the height of the posterior for θ by the height of the prior for θ, at the point of interest." (Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010)

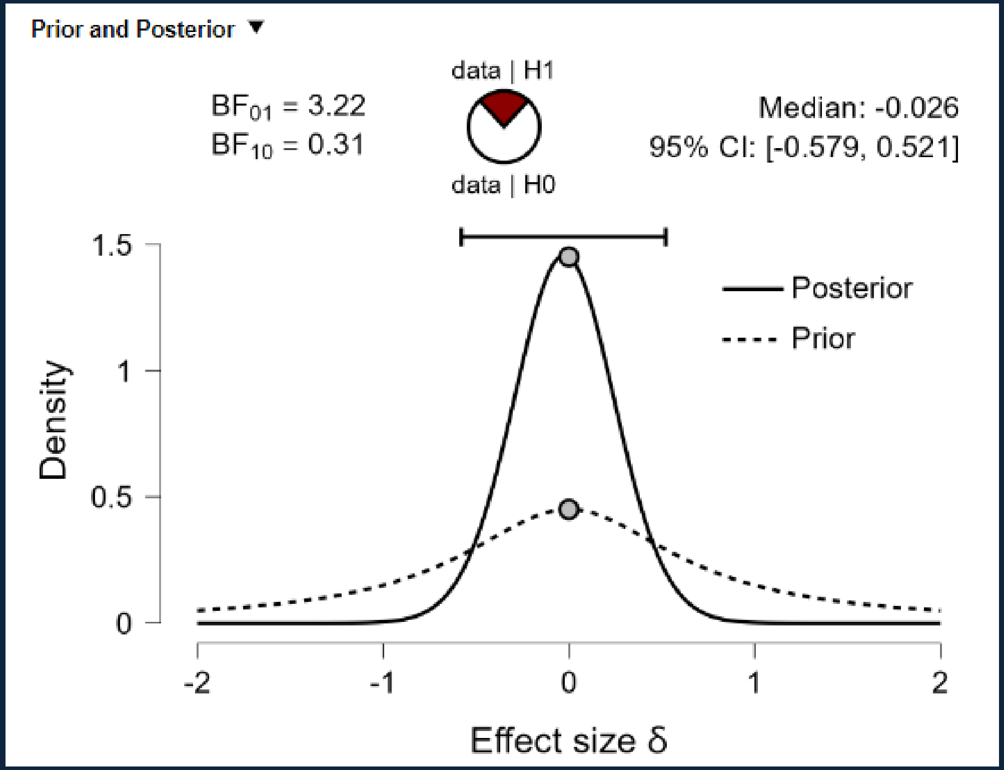
Example 1: (true) No effect

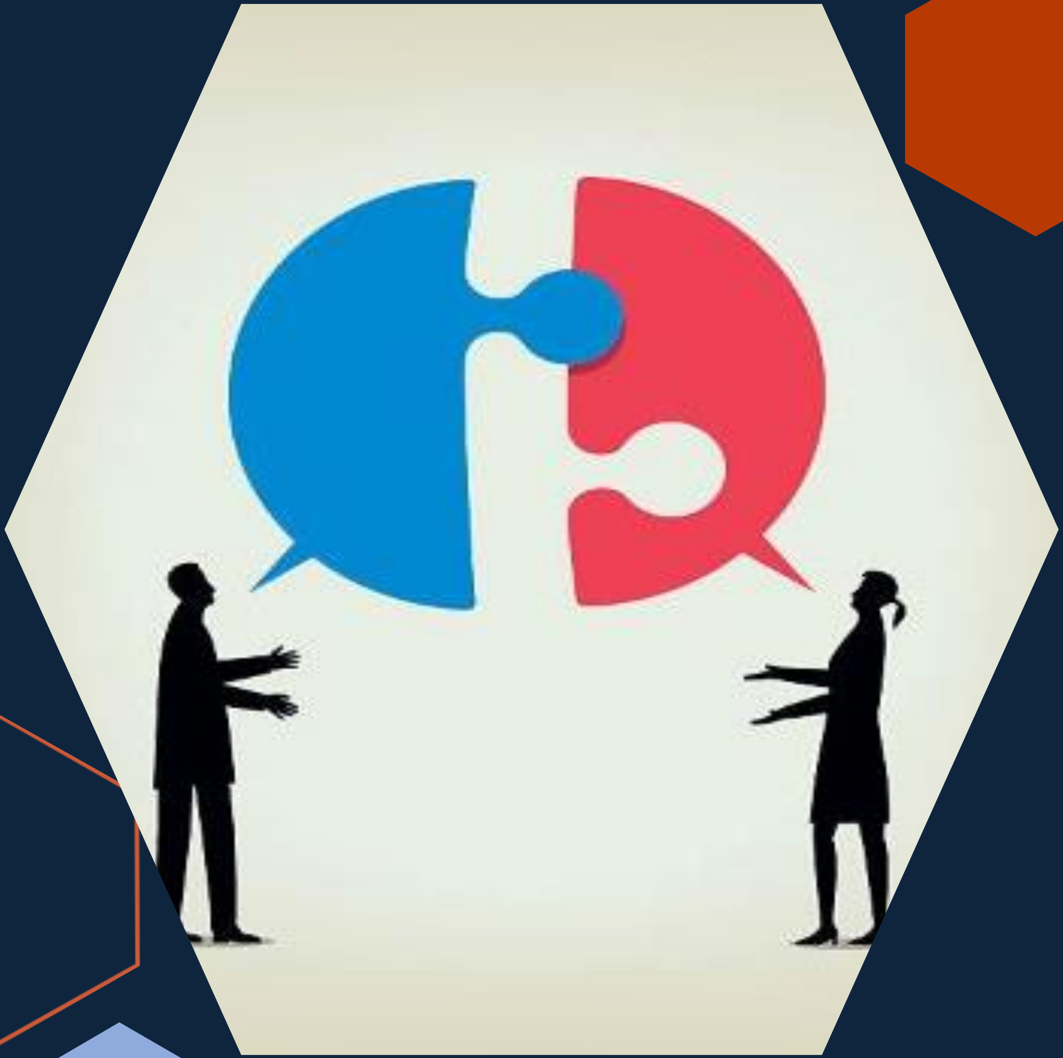
- We can also investigate the posteriors and credible intervals, δ_{Med} = 0.05 [-0.11, 0.21]
- These seem consistent with our conclusion.
  - Because my SESOI = $\delta \pm 0.40$;
  - Values in this range are too small to care about (equivalent to 0)
- But, that may not always be the case!



Example 2: (false) No effect

- Running the test shows the BF for M1 vs M0
- Again, we assume:
  - H_0 is $\delta = 0$
  - H_1 is $\delta \sim$ prior (Cauchy, 0, 0.707)
- Given this, we see the $BF_{01} = 3.22$
- We can say the H_0 is 3x more likely., but:
  - $\delta_{Med} = -0.03$ [-0.58, 0.52]
  - Given our SESOI, this include important values





This is not a true conflict.

Hypothesis testing =!
Parameter estimation

We can even have different priors for each approach.

But it highlights an issue on what we mean by “no effect”?

Example 1: (true) No effect

Setting δ = 0 is the same as comparing two models

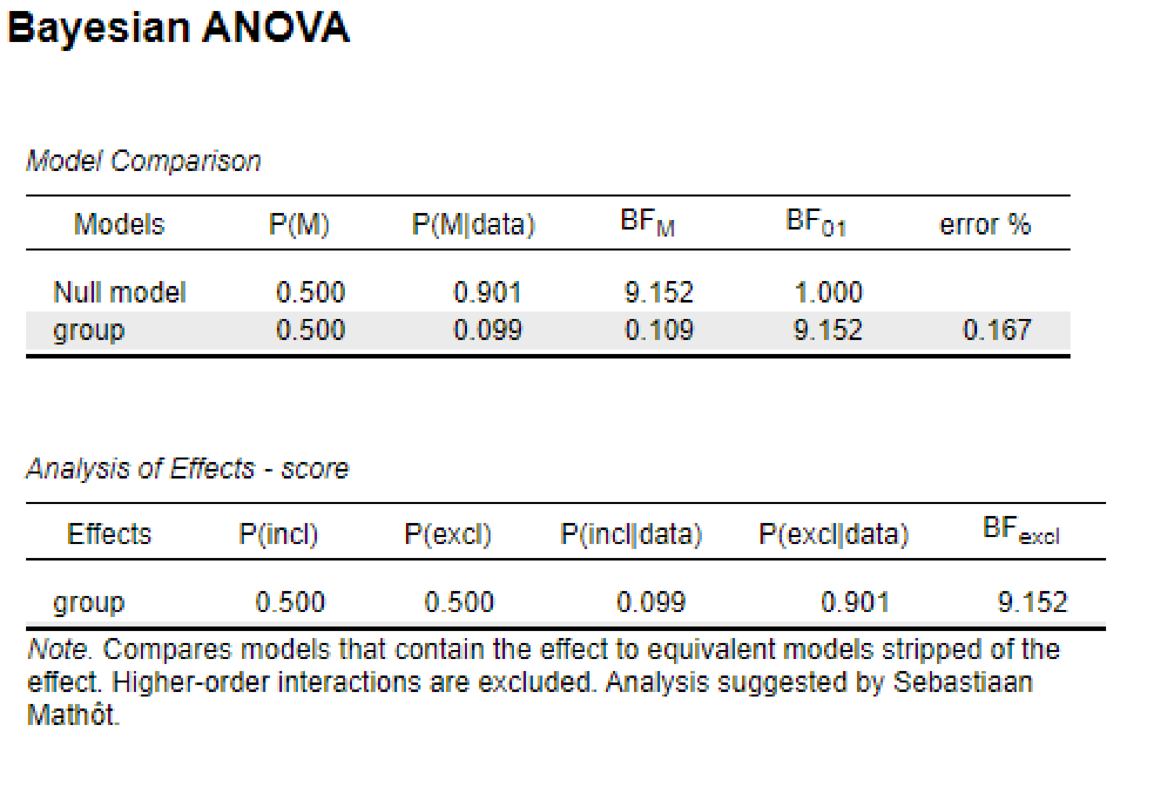
M_0

m0 = rating ~ 1 + 0*group (no effect)
m0 = rating ~ 1

M_1

m1 = rating ~ 1 + group

1/ $\mathrm{BF}_{\mathrm{incl}}$ = Bfexcl = 9.2;
Evidence you should consider the model without "group" as 9x more probable.



Bayesian ANOVA

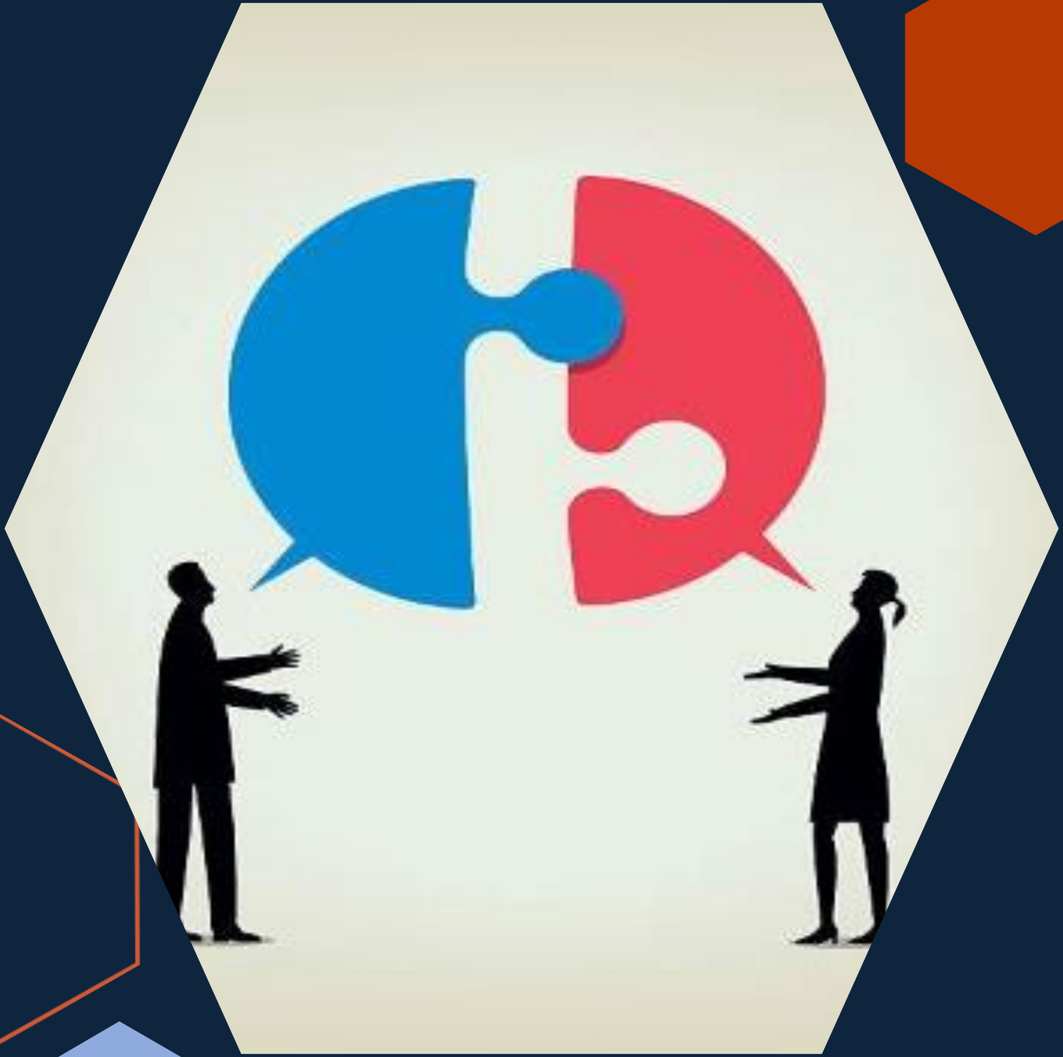
Model Comparison

Models	P(M)	P(M|data)	BF_{M}	BF_{01}	error %
Null model	0.500	0.901	9.152	1.000	
group	0.500	0.099	0.109	9.152	0.167

Analysis of Effects - score

Effects	P(incl)	P(excl)	P(incl|data)	P(excl|data)	$\mathrm{BF}_{\mathrm{excl}}$
group	0.500	0.500	0.099	0.901	9.152

Note. Compares models that contain the effect to equivalent models stripped of the effect. Higher-order interactions are excluded. Analysis suggested by Sebastiaan Mathôt.



This is not a true conflict.

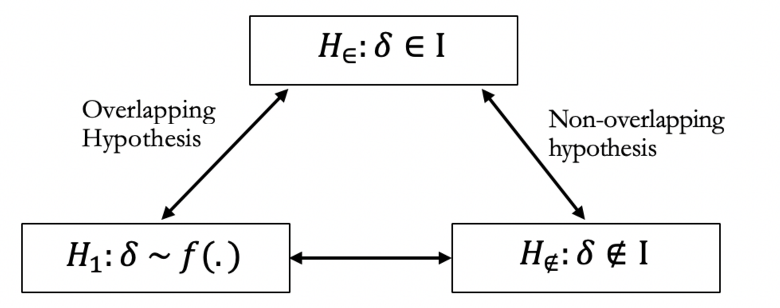
Hypothesis testing =!
Parameter estimation

We can even have different priors for each approach.

But it highlights an issue on what we mean by “no effect”?

Bayes Equivalence Tests (BEQ)

- Testing against 0 may not always be sensible.
- We can stay within the BF framework and do better.
- What if we explicitly tell the model to compare the marginal likelihoods against a region?
- Region of Practical Equivalence (ROPE)



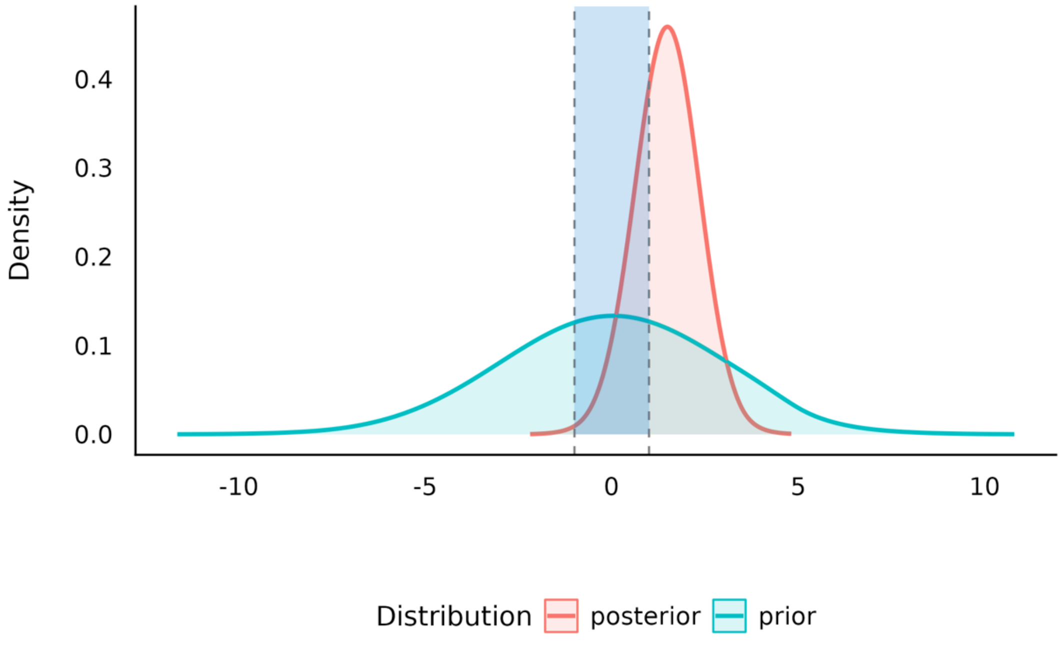
$H_{\in}: \delta \in \mathrm{I}$

Overlapping Hypothesis

Non-overlapping hypothesis

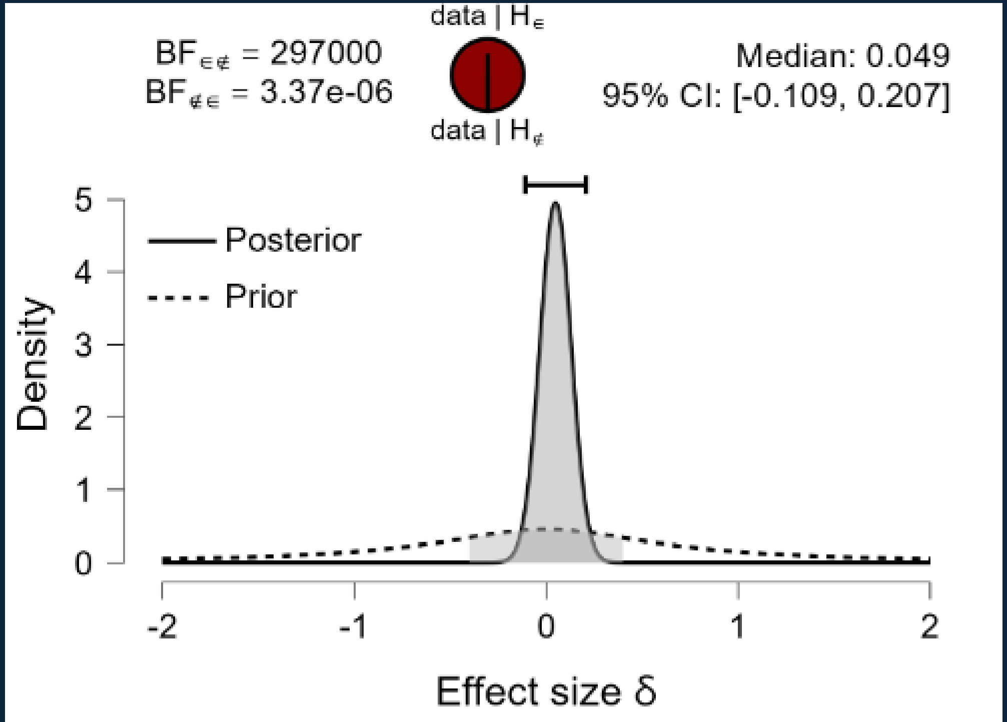
$H_1: \delta \sim f(.)$

$H_{\notin}: \delta \notin \mathrm{I}$



Example 1: (true) No effect

- We use Example 1 again, but now use a different testing ensemble
- We test against a ROPE = δ ± 0.40
- The “null” has now changed into a region.



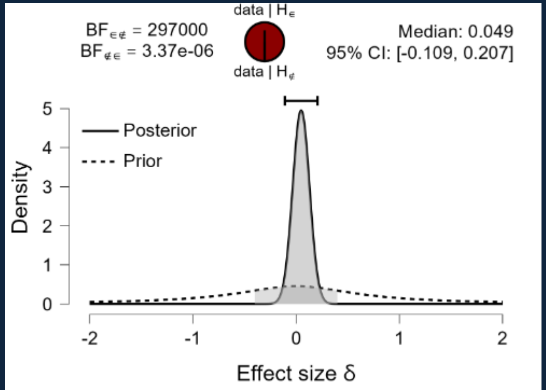
Example 1: (true) No effect

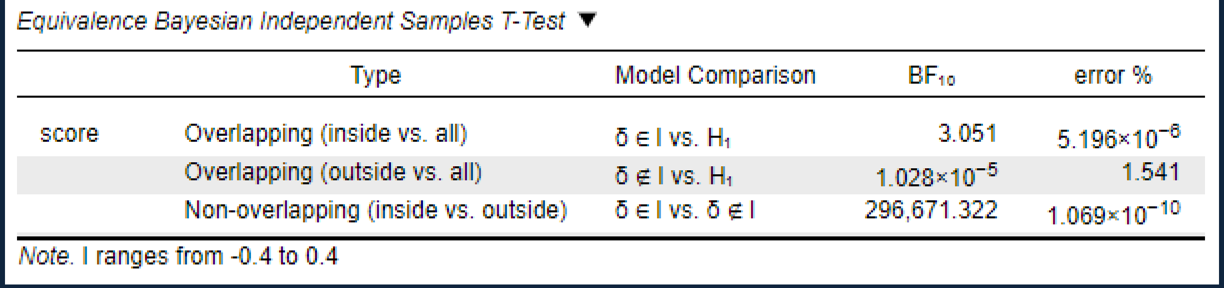
We use Example 1 again, but now use a different testing ensemble.

We test against a ROPE = $\delta \pm 0.40 = I$ (in JASP)

This allows us to test:

M_E, Equivalent model	$\delta \in I$
M_O, Outside model	$\delta \in/ I$
M_u, Unrestricted model	δ~Cauchy prior





Equivalence Bayesian Independent Samples T-Test ▼

	Type	Model Comparison	BF_{10}	error %
score	Overlapping (inside vs. all)	$\delta \in I$ vs. H_1	3.051	5.196×10^{-6}
	Overlapping (outside vs. all)	$\delta \notin I$ vs. H_1	1.028×10^{-5}	1.541
	Non-overlapping (inside vs. outside)	$\delta \in I$ vs. $\delta \notin I$	296,671.322	1.069×10^{-10}

Note. I ranges from -0.4 to 0.4

Example 1: (true) No effect

This region can be tested against! 3 tests:

Row 1: Does restricting my effect to this region make sense? [Inside ROPE vs Unrestricted]

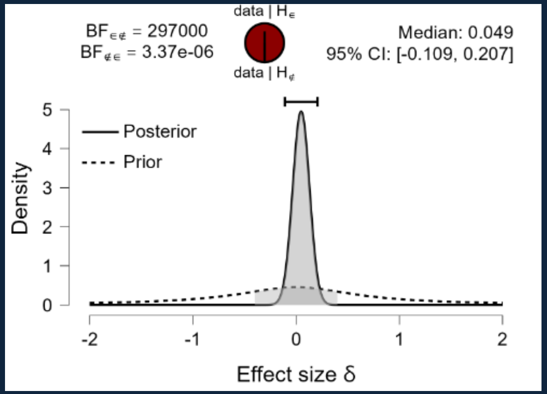
- A: The data are **about 3× more likely under the equivalence model** than under the unrestricted model.

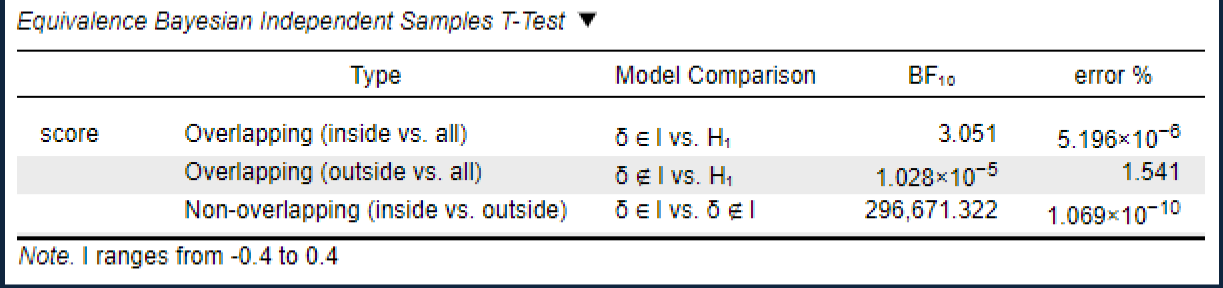
Row 2: Does restricting my effect to this region make sense? [Outside ROPE vs Unrestricted]

- A: The data are **~97,000x less likely** under the "meaningful effect" (SESOI) model than under the unrestricted model. (1/ **1.028 × 10^{-5}**)

Row 3: Does my effect fall within this region? [Inside ROPE vs Outside ROPE]

- The data are **~297,000x more likely** if the true effect lies **within the equivalence interval** than outside it.
- **Extremely strong evidence for <u>equivalence</u>.**

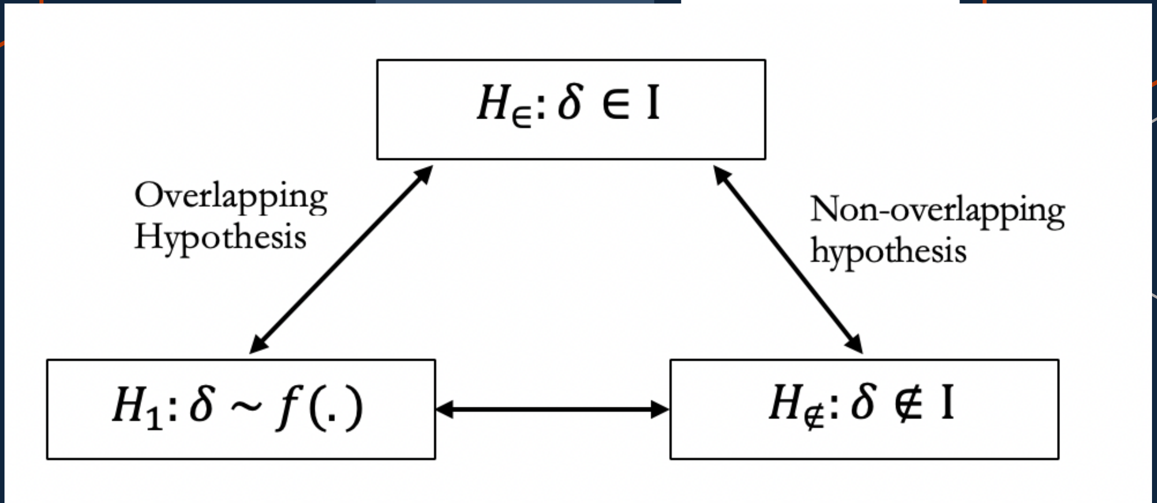




Equivalence Bayesian Independent Samples T-Test ▼

	Type	Model Comparison	BF_{10}	error %
score	Overlapping (inside vs. all)	$\delta \in I$ vs. H_1	3.051	5.196×10^{-6}
	Overlapping (outside vs. all)	$\delta \notin I$ vs. H_1	1.028×10^{-5}	1.541
	Non-overlapping (inside vs. outside)	$\delta \in I$ vs. $\delta \notin I$	296,671.322	1.069×10^{-10}

Note. I ranges from -0.4 to 0.4



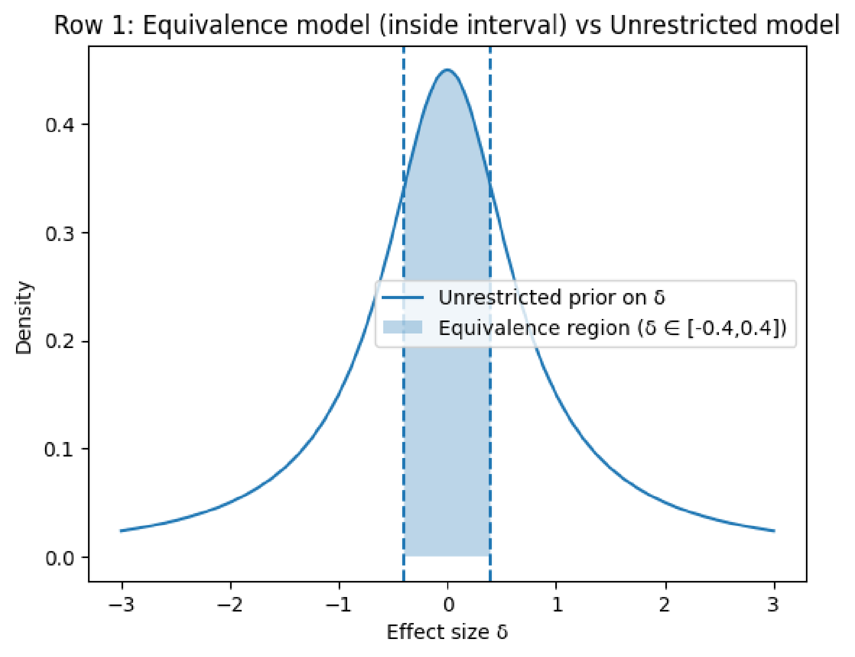
$H_{\in}: \delta \in \mathrm{I}$

Overlapping Hypothesis

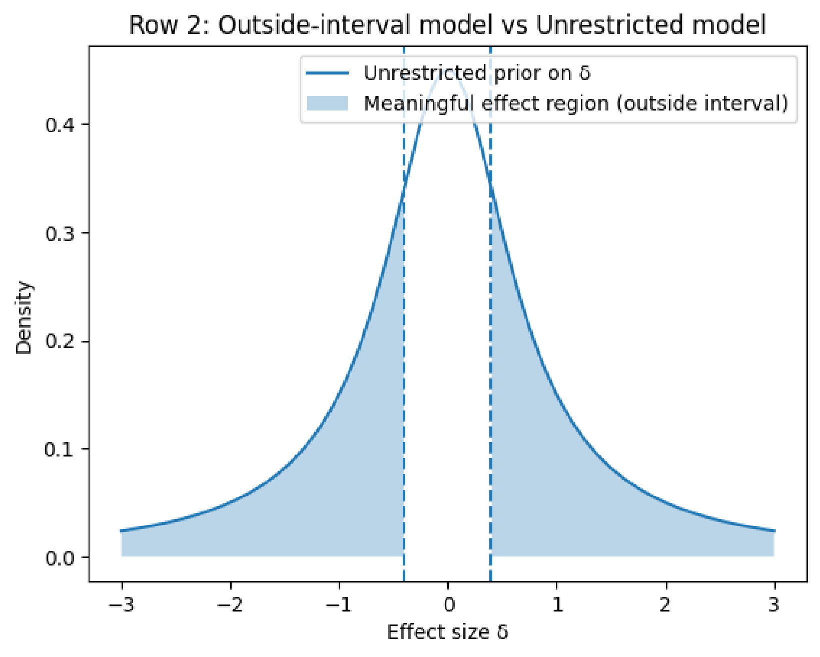
Non-overlapping hypothesis

$H_1: \delta \sim f(.)$

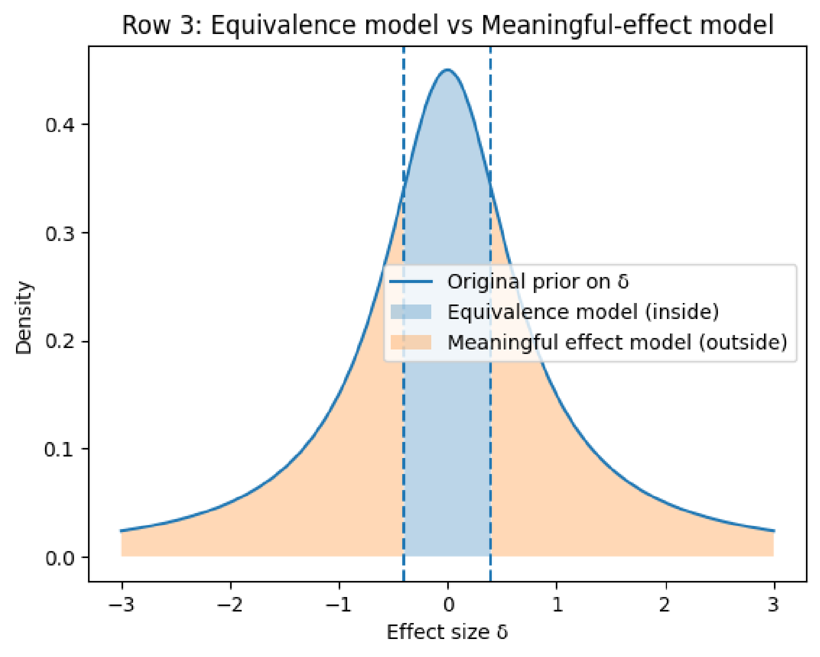
$H_{\notin}: \delta \notin \mathrm{I}$



Row 1: Equivalence model (inside interval) vs Unrestricted model



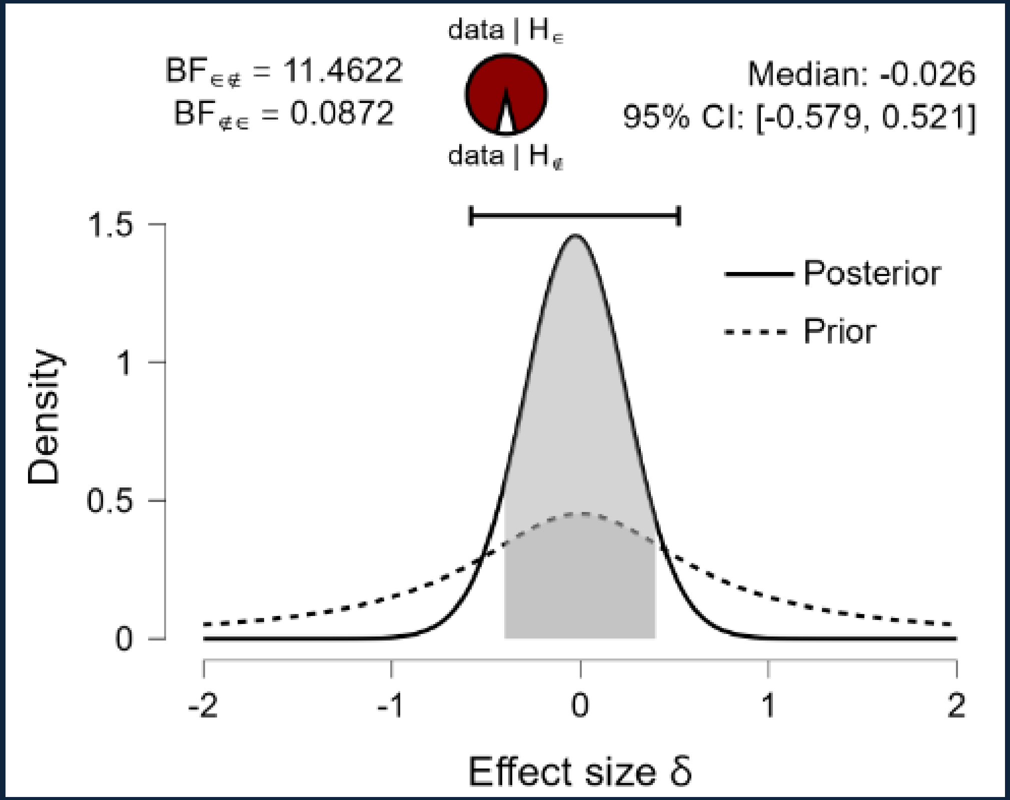
Row 2: Outside-interval model vs Unrestricted model



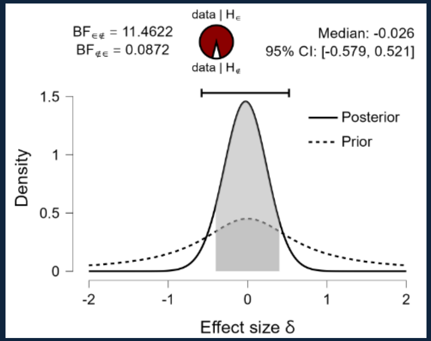
Row 3: Equivalence model vs Meaningful-effect model

Example 2: (false) No effect

- Now lets see what happens to our Example 2 when we do a BEQ
- Just looking at the plot, we see how things will go.
- But also provides relative insights that we would not get under a frequentist framework.



Example 2: (false) No effect



Row 1: Inside ROPE vs Unrestricted

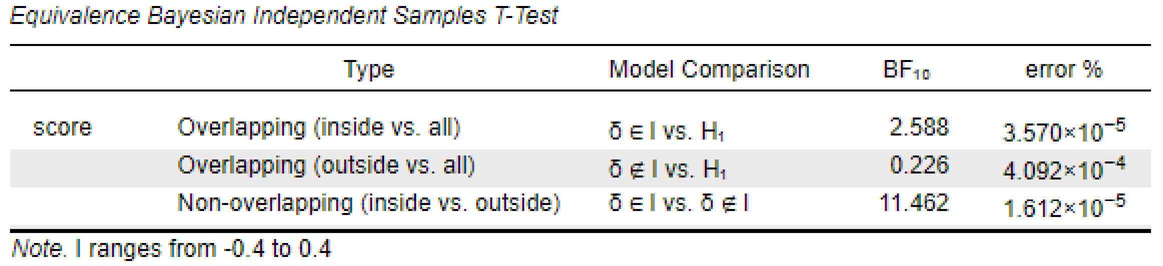
- A: 2.6× evidence for **equivalence**

Row 2: Outside ROPE vs Unrestricted

- A: 4.4× *less* likely under "δ ∉ I“
- i.e., moderate evidence *against* a non-trivial effect. Disfavors non-equivalence.

Row 3: Equivalence [Inside ROPE vs Outside ROPE]

- **11.5× more likely** if the true effect is within [−0.4, 0.4] than outside it.
- **Strong evidence for equivalence**



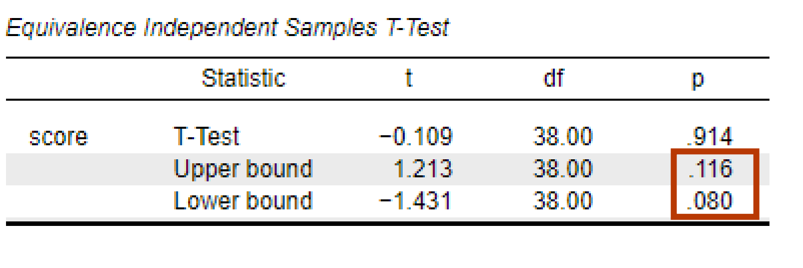
Equivalence Bayesian Independent Samples T-Test

	Type	Model Comparison	BF_{10}	error %
score	Overlapping (inside vs. all)	δ ∈ I vs. H_1	2.588	3.570×10^{-5}
	Overlapping (outside vs. all)	δ ∉ I vs. H_1	0.226	4.092×10^{-4}
	Non-overlapping (inside vs. outside)	δ ∈ I vs. δ ∉ I	11.462	1.612×10^{-5}

Note. I ranges from -0.4 to 0.4

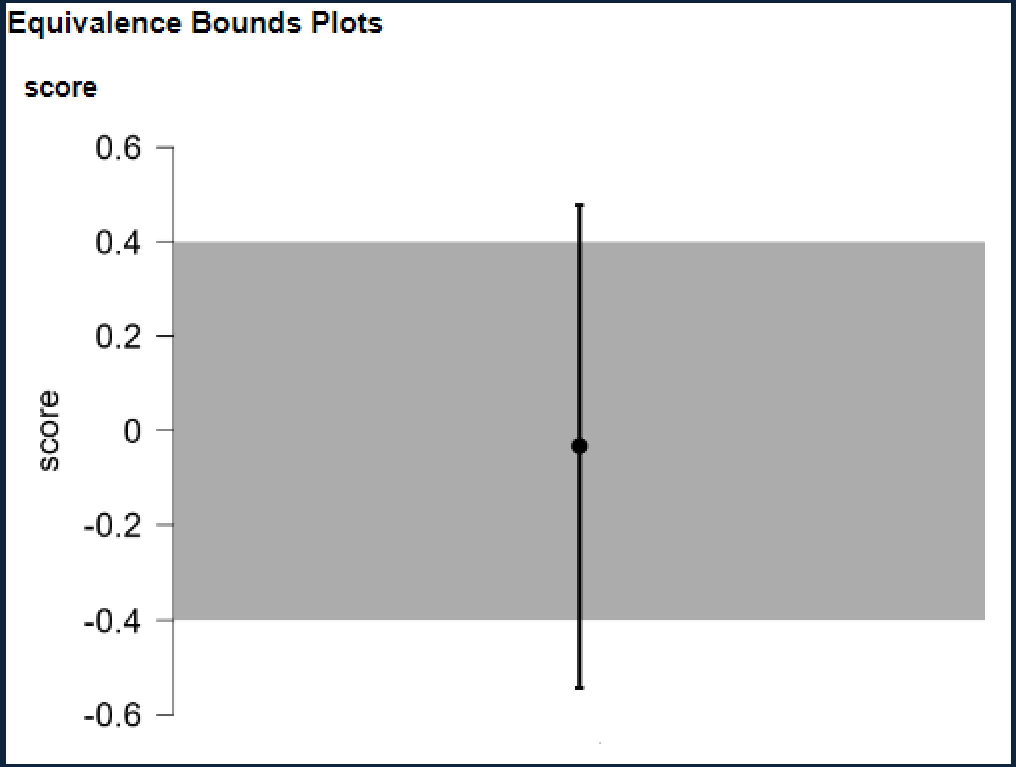
Why did I call it “(false) no effect”?

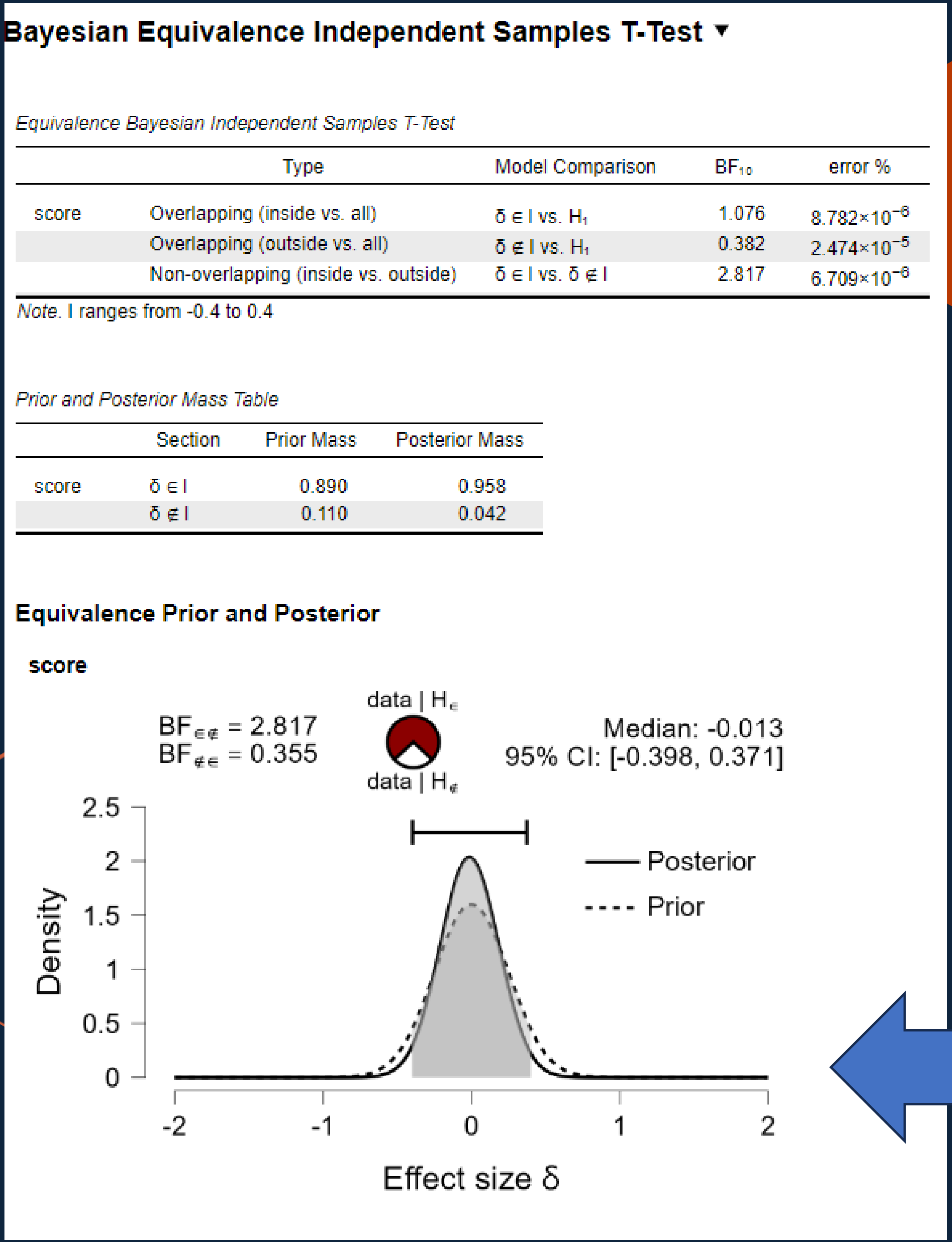
- Under a frequentist framework, the results are “non-significant and not equivalent”



Equivalence Independent Samples T-Test

	Statistic	t	df	p
score	T-Test	−0.109	38.00	.914
	Upper bound	1.213	38.00	.116
	Lower bound	−1.431	38.00	.080





Bayesian Equivalence Independent Samples T-Test ▼

Equivalence Bayesian Independent Samples T-Test

	Type	Model Comparison	BF_{10}	error %
score	Overlapping (inside vs. all)	$\delta \in I$ vs. H_1	1.076	8.782×10^{-6}
	Overlapping (outside vs. all)	$\delta \notin I$ vs. H_1	0.382	2.474×10^{-5}
	Non-overlapping (inside vs. outside)	$\delta \in I$ vs. $\delta \notin I$	2.817	6.709×10^{-6}

Note. I ranges from -0.4 to 0.4

Prior and Posterior Mass Table

	Section	Prior Mass	Posterior Mass
score	$\delta \in I$	0.890	0.958
	$\delta \notin I$	0.110	0.042

Equivalence Prior and Posterior

score

Benefits: BF interval null procedure is better at discriminating between equivalence and nonequivalence, particularly for relatively small sample sizes and narrow equivalence intervals.

Findings can differ based on your:

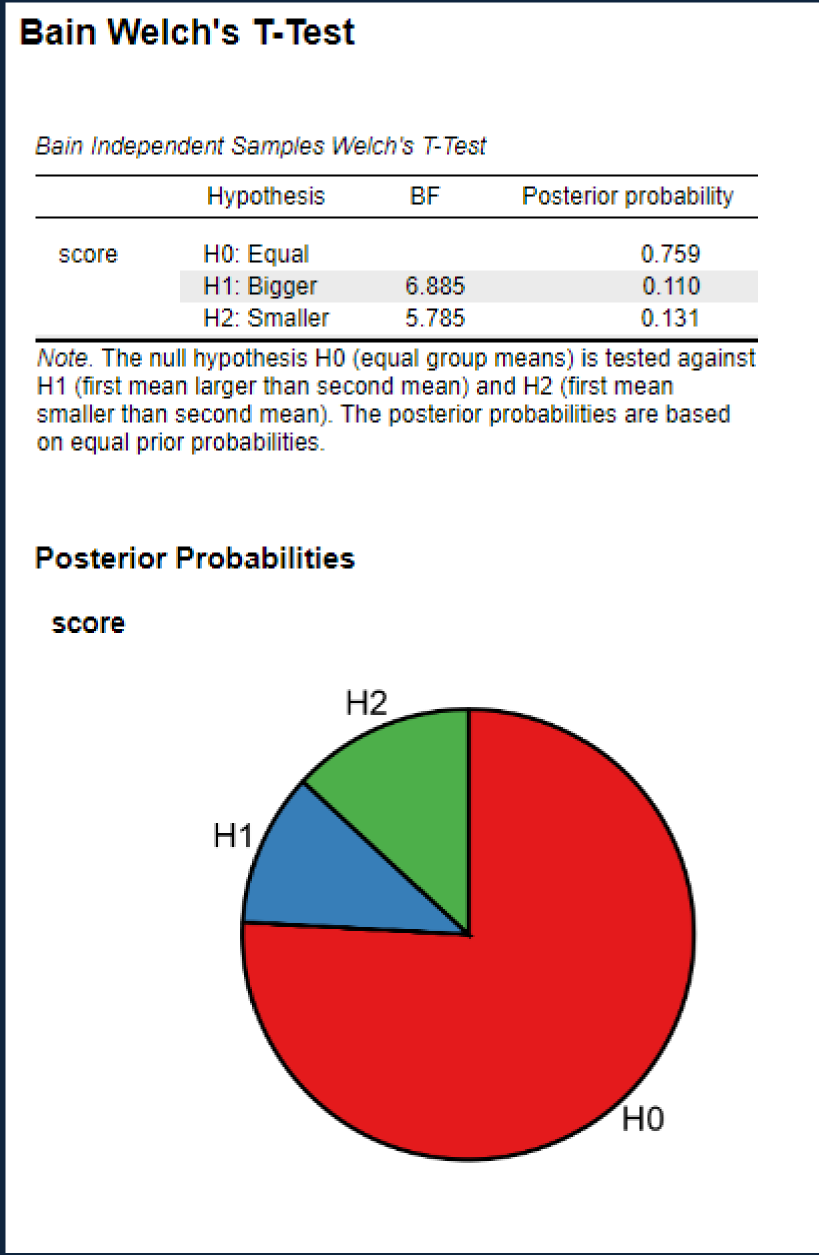
- framework (freq vs Bayes)
- approach
- priors

If I tighten my priors for "no effect" as N(0, 0.25), my current data is too uncertain to claim equivalence.

(this is similar to a spike-and-slab prior)

Bonus

- BAIN package in JASP
- Test multiple hypotheses at once:
  - Unequal vs equal
  - Bigger
  - Smaller
  - All



Bain Welch's T-Test

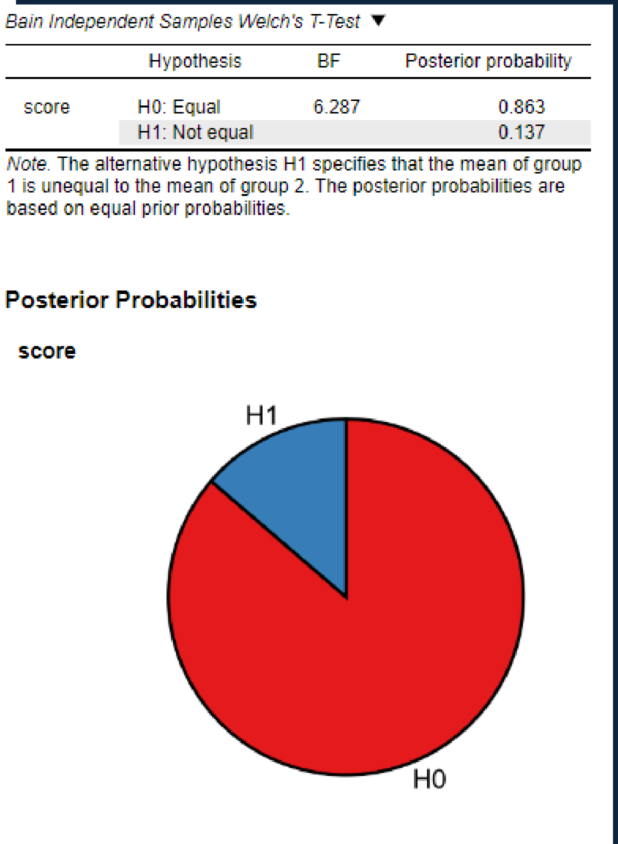
Bain Independent Samples Welch's T-Test

	Hypothesis	BF	Posterior probability
score	H0: Equal		0.759
	H1: Bigger	6.885	0.110
	H2: Smaller	5.785	0.131

Note. The null hypothesis H0 (equal group means) is tested against H1 (first mean larger than second mean) and H2 (first mean smaller than second mean). The posterior probabilities are based on equal prior probabilities.

Posterior Probabilities

score



Bain Independent Samples Welch's T-Test ▼

	Hypothesis	BF	Posterior probability
score	H0: Equal	6.287	0.863
	H1: Not equal		0.137

Note. The alternative hypothesis H1 specifies that the mean of group 1 is unequal to the mean of group 2. The posterior probabilities are based on equal prior probabilities.

Posterior Probabilities

score

A note on Interval BFs

“The NOH Bayes factor gives a ratio of support for H_1: $\delta \in I$ versus H_0: $\delta \notin I$, which may seem to be appealing, but as we show, it does not give a direct expression of the probability that the population effect size is in interval *I*.”

- Kiers et al. (2025)

The NOH BF considers the posterior odds to back compute the BF.

What we may want to know is the Probability that the posterior is in the ROPE ->

HDI + ROPE procedure

HDI + ROPE procedure

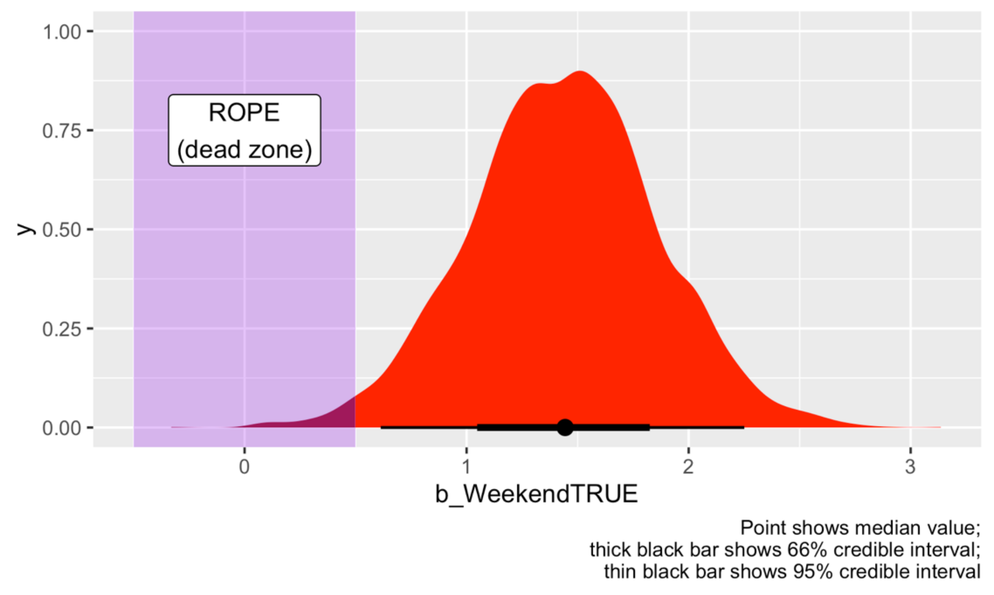
Bayes factor is not a valid measure of the effect size! If you increase N the BF will also increase/decrease, even if the effect stays the same.

Instead of testing, we can use estimation to obtain a similar goal

- We set a ROPE
- We compute the proportion of the HDI of a posterior distribution that lies within a region of practical equivalence.
- We determine if the effect is substantial or not

Strengths: Provides information related to the practical relevance of the effects.

Limitations: A ROPE range needs to be arbitrarily defined. Sensitive to the scale (the unit) of the predictors. Not sensitive to highly significant effects. Invalid under multicollinearity.



Point shows median value;
thick black bar shows 66% credible interval;
thin black bar shows 95% credible interval

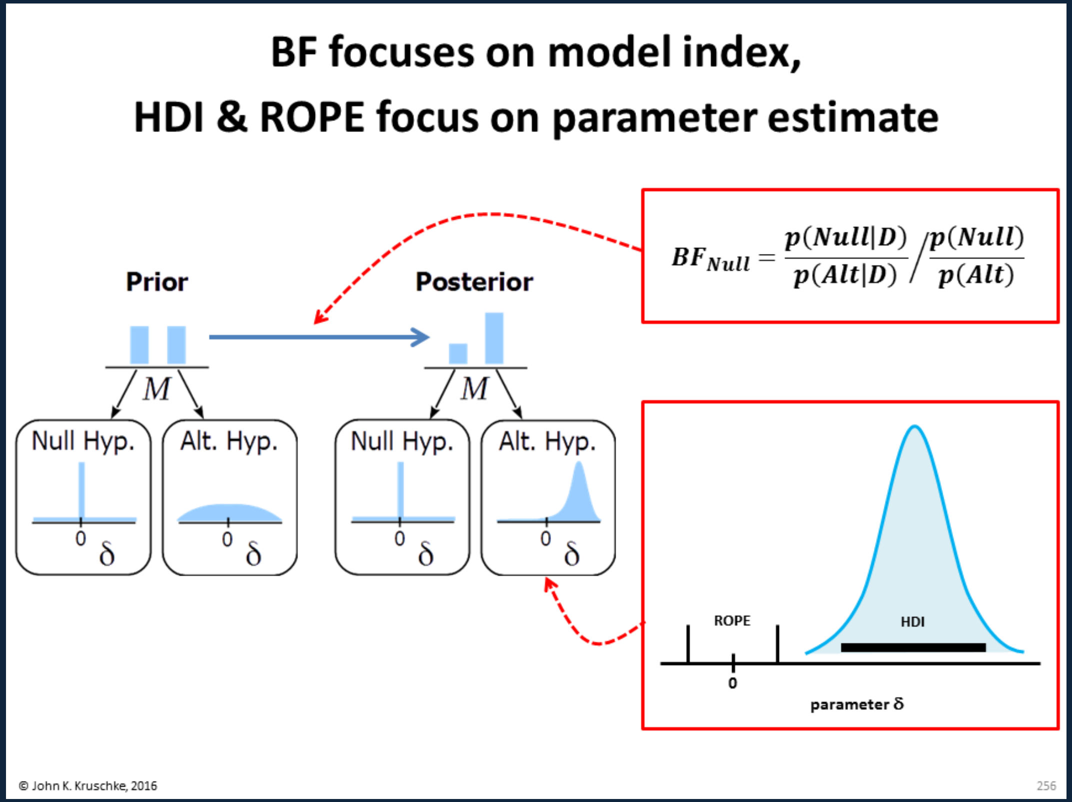
The BF focuses on the model index, whereas the HDI and ROPE focus on the parameter estimate

Difference between ROPE and BF

ROPE is not hypothesis testing!

The ROPE is *part of the decision rule,* not part of the null hypothesis. The ROPE does not constitute an interval null hypothesis; the null hypothesis here is a point value.

The ROPE is part of the decision rule for two main purposes: First, it allows decisions to accept the null. Second, it makes the decision rule asymptotically correct: As data sample size increases, the rule will come to the correct decision, either practically equivalent to the null value (within the ROPE) or not (outside the ROPE).



BF focuses on model index,

HDI & ROPE focus on parameter estimate

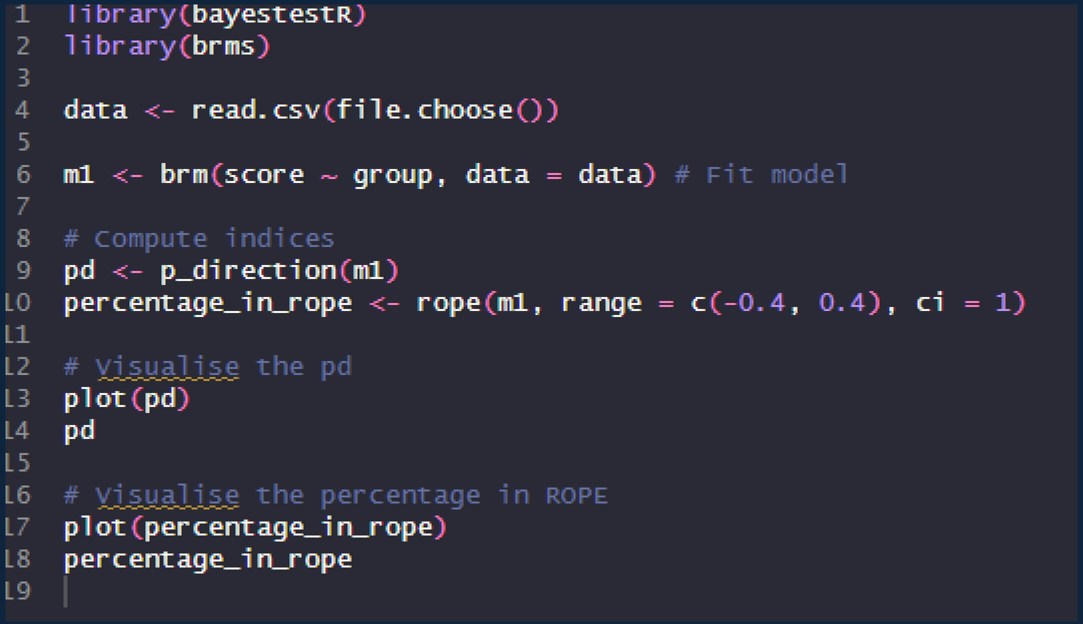
$$BF_{Null} = \frac{p(Null|D)}{p(Alt|D)} \Big/ \frac{p(Null)}{p(Alt)}$$

© John K. Kruschke, 2016

256

Example 1: (true) No effect

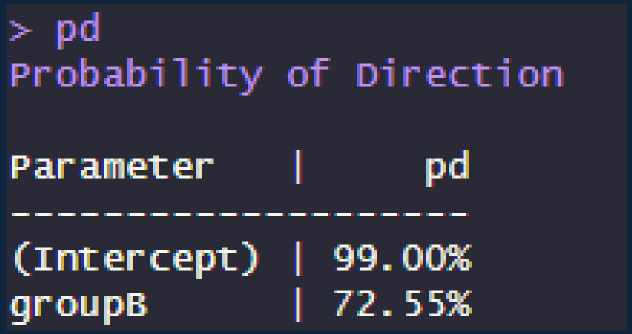
- We use Example 1 again
- Now we can compute any quantities we want from the posterior
- Inside ROPE
- Probability of direction (pd)



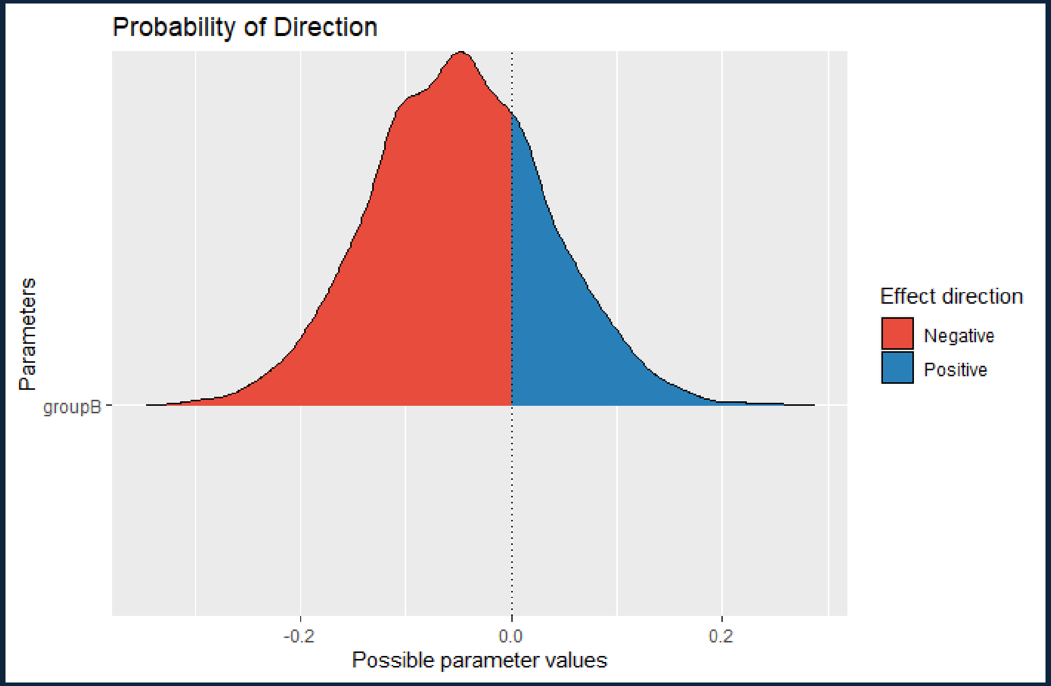
```r
library(bayestestR)
library(brms)

data <- read.csv(file.choose())

m1 <- brm(score ~ group, data = data) # Fit model

# Compute indices
pd <- p_direction(m1)
percentage_in_rope <- rope(m1, range = c(-0.4, 0.4), ci = 1)

# Visualise the pd
plot(pd)
pd

# Visualise the percentage in ROPE
plot(percentage_in_rope)
percentage_in_rope

```

Bayesian Posteriors

- Interrogate any quantity you wish
- What is the probability that my effect is positive/negative? (ranges from 50% to 100%)



```
> pd
Probability of Direction

Parameter   |     pd
----------------------
(Intercept) | 99.00%
groupB      | 72.55%
```



Effect existence with PD

For convenience, we suggest the following reference values as an interpretation helpers:

https://easystats.github.io/bayestestR/articles/guidelines.html

pd <= **95%** ~ *p* > .1: uncertain

pd > **95%** ~ *p* < .1: possibly existing

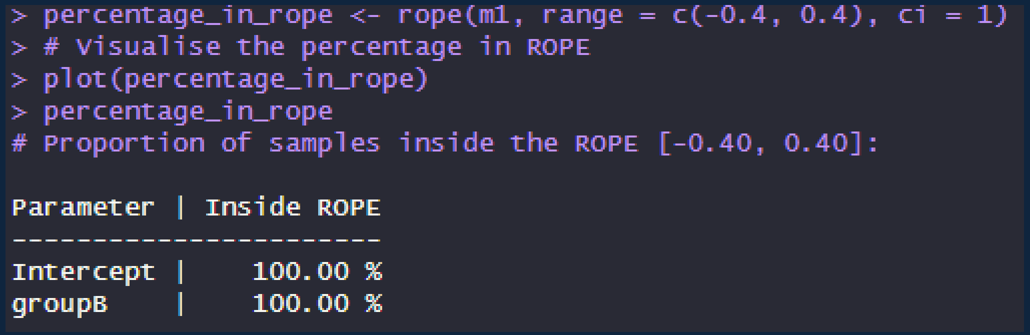
pd > **97%**: likely existing

pd > **99%**: probably existing

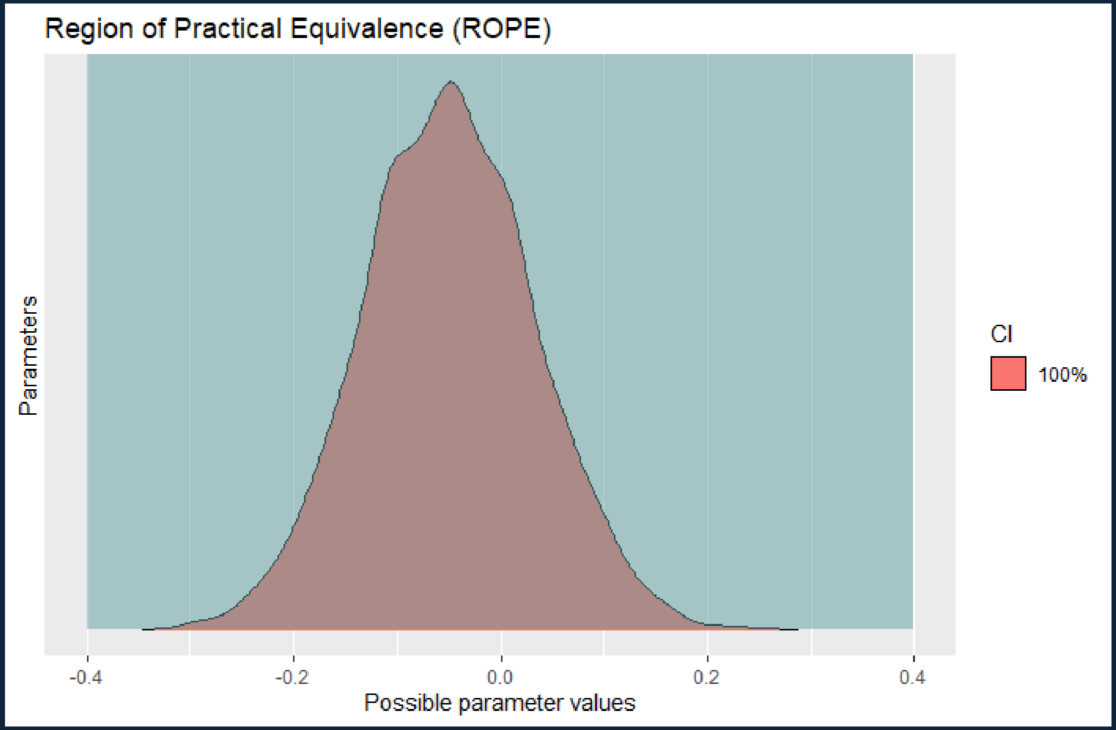
pd > **99.9%**: certainly existing

Bayesian Posteriors

- Interrogate any quantity you wish
- What proportion of my entire posteriors (ROPE 100% rule) falls within the ROPE?



```
> percentage_in_rope <- rope(m1, range = c(-0.4, 0.4), ci = 1)
> # Visualise the percentage in ROPE
> plot(percentage_in_rope)
> percentage_in_rope
# Proportion of samples inside the ROPE [-0.40, 0.40]:

Parameter | Inside ROPE
-----------------------
Intercept |    100.00 %
groupB    |    100.00 %
```



Region of Practical Equivalence (ROPE)

“95% HDI+ROPE decision rule” (Kruschke, 2014) – don’t use!

What percentage in ROPE to accept or to reject?

95%HDI rule

- If the HDI is **completely outside** the ROPE, the “null hypothesis” for this parameter is “rejected”.
- If the ROPE **completely covers** the HDI, *i.e.*, all most credible values of a parameter are inside the region of practical equivalence, the null hypothesis is accepted.
- Else, it’s unclear whether the null hypothesis should be accepted or rejected.

ROPE 100% rule

- If the **full ROPE** is used (*i.e.*, 100% of the HDI), then the null hypothesis is rejected or accepted if the percentage of the posterior within the ROPE is smaller than to 2.5% or greater than 97.5%.
- Desirable results are low proportions inside the ROPE (the closer to zero the better).

Effect “significance”

The percentage in **ROPE** is a index of **significance** (in its primary meaning), informing us whether a parameter is related or not to a non-negligible change (in terms of magnitude) in the outcome. Rather than using it as a binary, all-or-nothing decision criterion, such as suggested by the original equivalence test, we recommend using the percentage as a *continuous* index of significance. However, based on simulation data, we suggest the following reference values as an interpretation helpers:

> **99%** in ROPE: negligible (we can accept the null hypothesis)

> **97.5%** in ROPE: probably negligible

<= **97.5%** & >= **2.5%** in ROPE: undecided significance

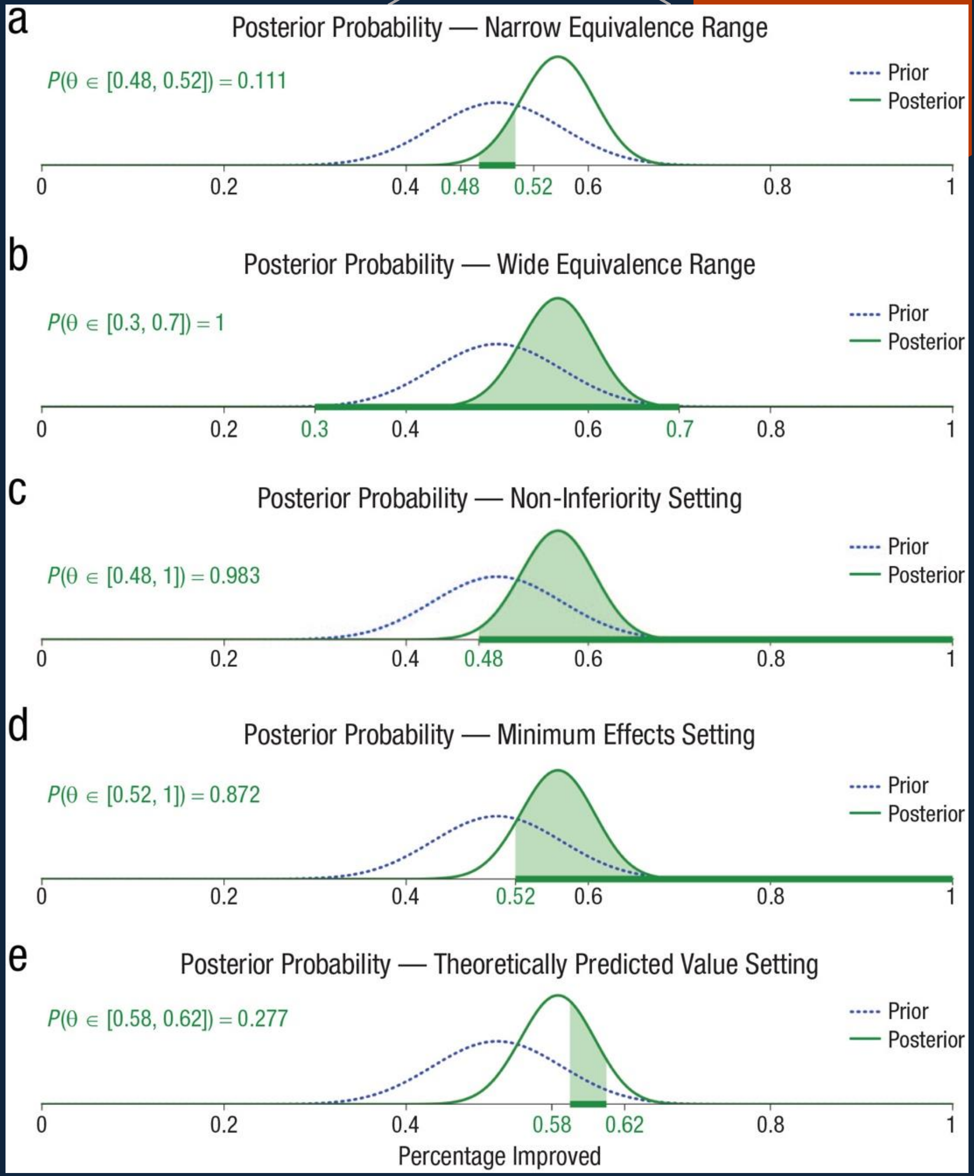
< **2.5%** in ROPE: probably significant

< **1%** in ROPE: significant (we can reject the null hypothesis)

Note that extra caution is required as its interpretation highly depends on other parameters such as sample size and ROPE range (see here).

More options for ROPE

- The beauty of Bayesian stats is that we have an entire posterior distribution to use for our inferences.



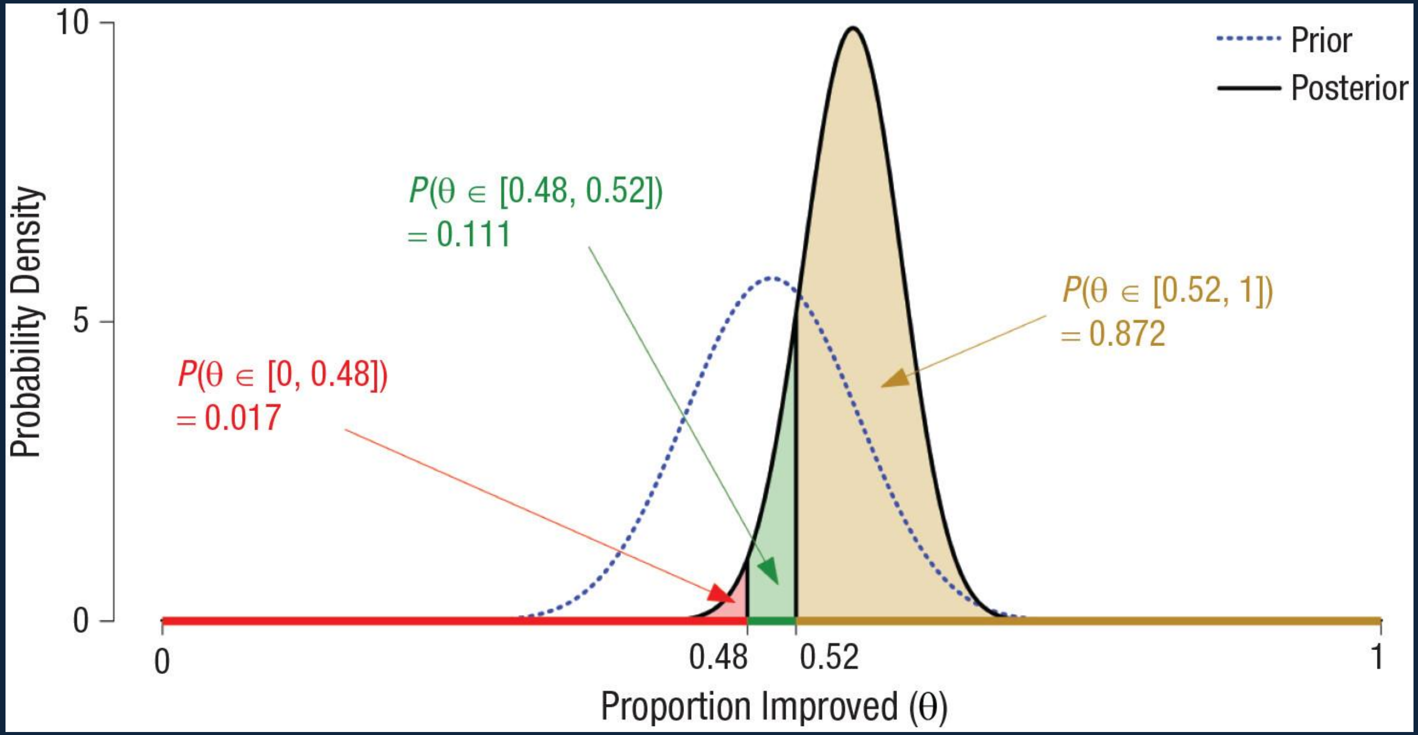
More options for ROPE

We can section regions to mean:

“got worse”

“stayed the same”

“improved”

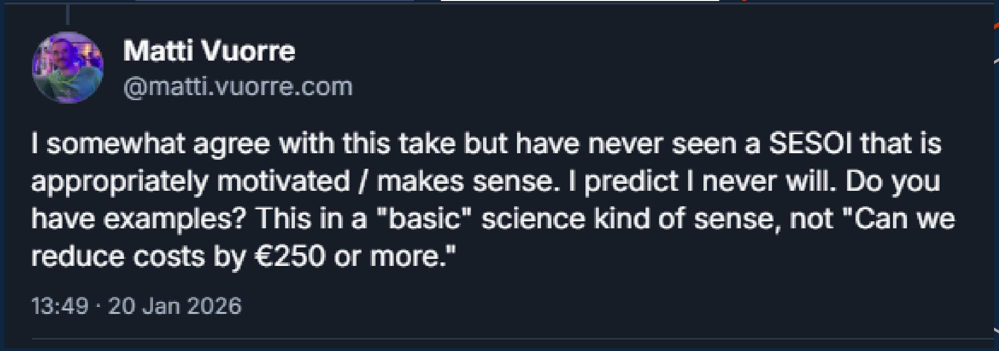


Complexities

Are BFs or BEQs or ROPE a better way to make inferences?

IDK

There are plenty of ongoing debates regarding both.

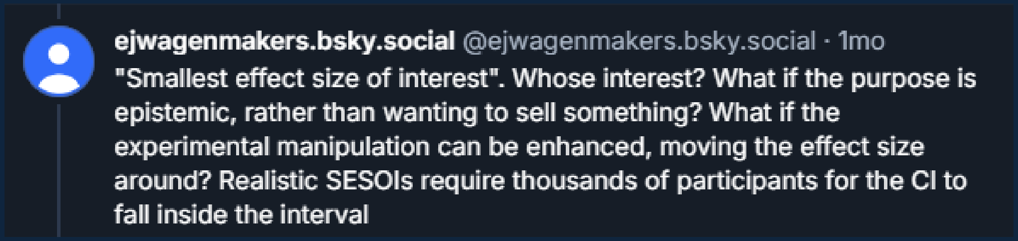


Matti Vuorre
@matti.vuorre.com

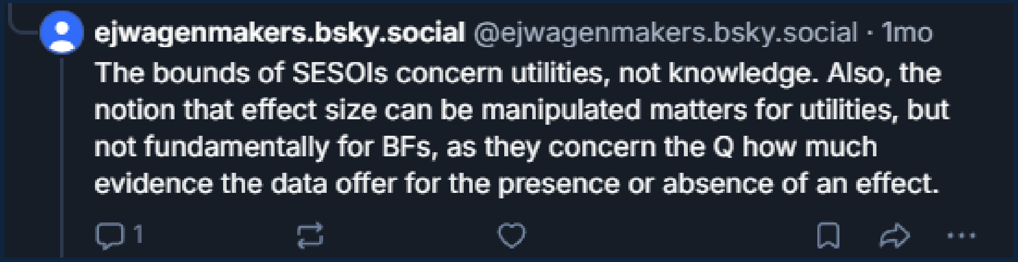
I somewhat agree with this take but have never seen a SESOI that is appropriately motivated / makes sense. I predict I never will. Do you have examples? This in a "basic" science kind of sense, not "Can we reduce costs by €250 or more."

13:49 · 20 Jan 2026

Against ROPE, but still anti BFs

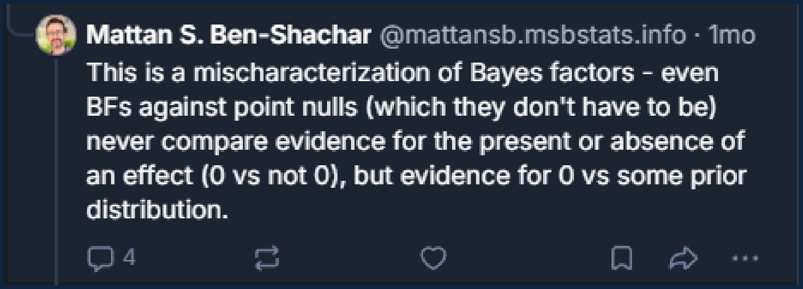


ejwagenmakers.bsky.social @ejwagenmakers.bsky.social · 1mo
"Smallest effect size of interest". Whose interest? What if the purpose is epistemic, rather than wanting to sell something? What if the experimental manipulation can be enhanced, moving the effect size around? Realistic SESOIs require thousands of participants for the CI to fall inside the interval



ejwagenmakers.bsky.social @ejwagenmakers.bsky.social · 1mo
The bounds of SESOIs concern utilities, not knowledge. Also, the notion that effect size can be manipulated matters for utilities, but not fundamentally for BFs, as they concern the Q how much evidence the data offer for the presence or absence of an effect.

Against ROPE, but pro BFs



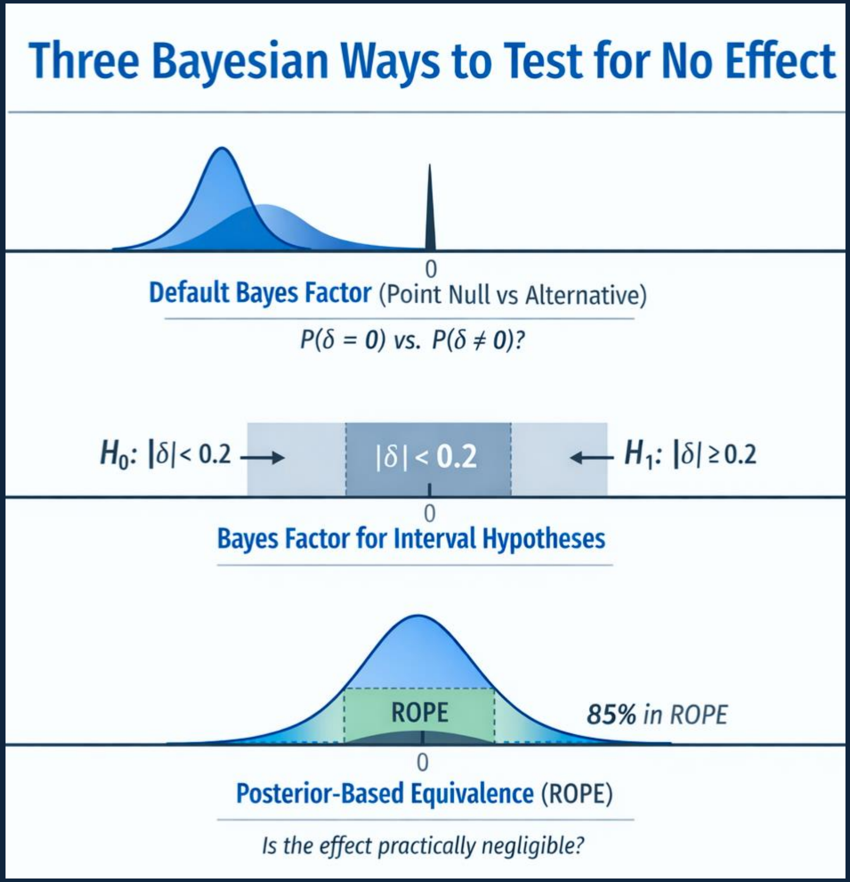
Mattan S. Ben-Shachar @mattansb.msbstats.info · 1mo
This is a mischaracterization of Bayes factors - even BFs against point nulls (which they don't have to be) never compare evidence for the present or absence of an effect (0 vs not 0), but evidence for 0 vs some prior distribution.

Pro ROPE, against BFs
(wrote the bayestestR package)

NHBT | BEQ | ROPE + HDI | PD

“No effect”

1. Can be determined in several ways
2. The question answered will determine which method is most appropriate
3. Be mindful of overinterpreting a single result, especially with weak priors
4. Determine your SESOI (if possible)
5. Your priors matter for BF a lot
6. Your scale and model matter for ROPE a lot


Three Bayesian Ways to Test for No Effect
0
Default Bayes Factor (Point Null vs Alternative)
P(δ = 0) vs. P(δ ≠ 0)?
H₀: |δ|< 0.2
|δ| < 0.2
H₁: |δ| ≥ 0.2
0
Bayes Factor for Interval Hypotheses
ROPE
85% in ROPE
0
Posterior-Based Equivalence (ROPE)
Is the effect practically negligible?

Thank you

Mircea Zloteanu

mircea.zloteanu@kcl.ac.uk

Figuring Stuff Out https://mzloteanu.substack.com/

@mzloteanu.bsky.social

linkedin.com/in/mirceaz/

KING'S College LONDON

References

Tendeiro, J., Kiers, H., Hoekstra, R., Wong, T. K., & Morey, R. D. (2022). Diagnosing the Misuse of the Bayes factor in Applied Research. https://doi.org/10.31234/osf.io/du3fc

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. Journal of Open Source Software, 4(40), 1541. https://doi.org/10.21105/joss.01541

Kiers, H. A. L., & Tendeiro, J. N. (2025). Bridging Null Hypothesis Testing and Estimation: A Practical Guide to Statistical Conclusion Drawing From Research in Psychology. Advances in Methods and Practices in Psychological Science, 8(3). https://doi.org/10.1177/25152459251365960

Kruschke, J. K., & Liddell, T. M. (2017). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. Psychonomic Bulletin & Review, 25(1), 178–206. https://doi.org/10.3758/s13423-016-1221-4

Dienes, Z. (2021). Obtaining Evidence for No Effect. Collabra: Psychology, 7(1). https://doi.org/10.1525/collabra.28202