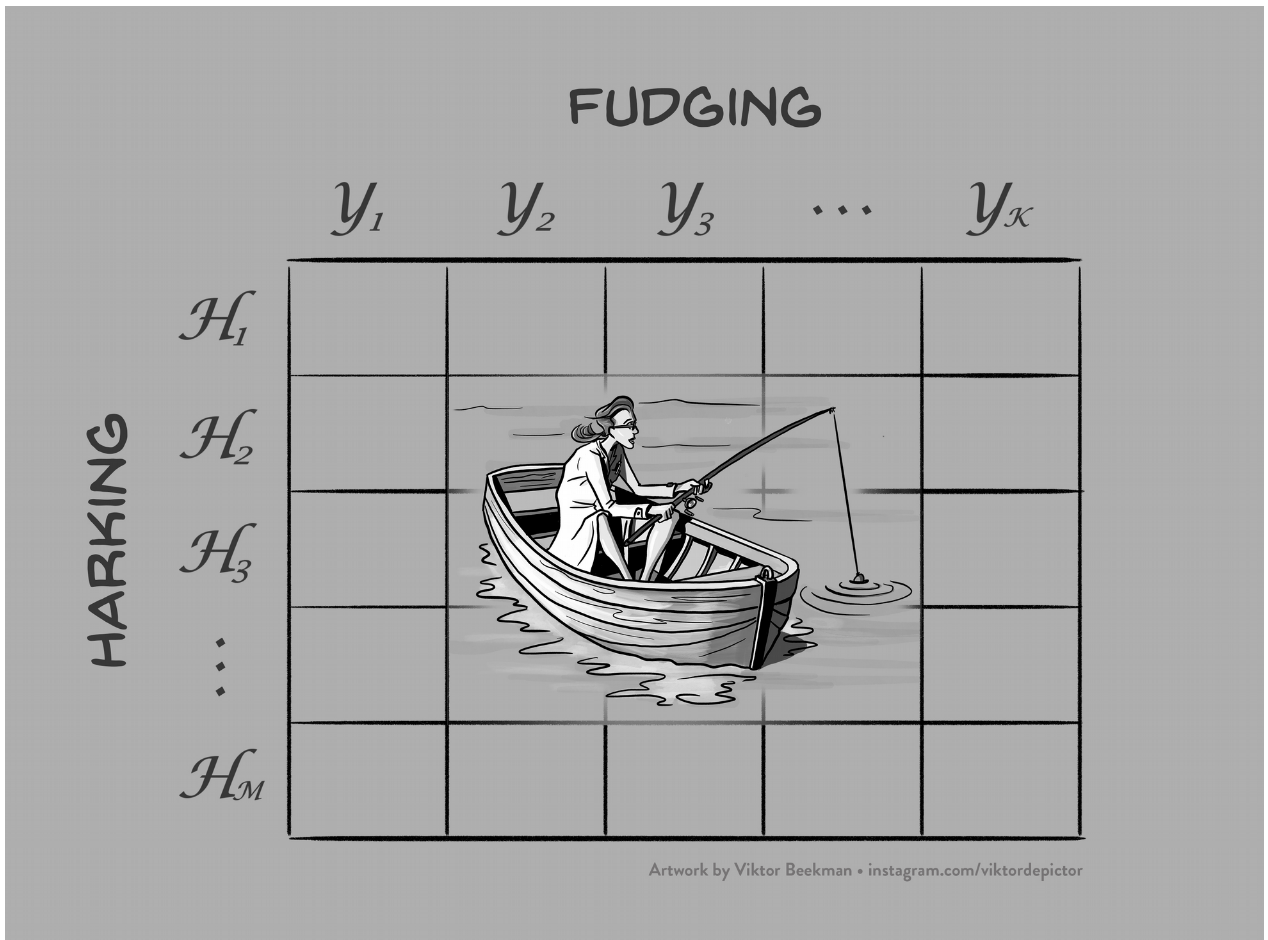
Eric-Jan Wagenmakers

The Case for Radical Transparency in Statistical Reporting



Overview

- What researchers want
- What the field gets
- How to uncover hidden uncertainty

The Main Dilemma

- Dr. X has a favorite theory that she has worked on and published about previously.
- Dr. X designs an experiment to test a prediction from her theory.
- Dr. X collects the data, a painstaking and costly process. Part of her career and those of her students ride on the outcome.

The Main Dilemma

- Now the data need to be analyzed.
- If $p < .05$, the experiment is deemed a *success*; if $p > .05$, it is deemed a *failure*.

Who is, without a shadow of a doubt, the most biased analyst in the entire galaxy, past, present, and future?

Who is, without a shadow of a doubt, the most biased analyst in the entire galaxy, past, present, and future?

$$x$$

Richard Feynman

"The first principle is that you must not fool yourself---and you are the easiest person to fool"

The Main Dilemma

- So the world's most biased analyst, Dr. X, the easiest person to fool, proceeds to analyze the data.
- Dr. X can do this alone, without any oversight whatsoever. In most cases, the data and analysis code never leave the lab.

A Perfect Storm

- Data are analyzed with no accountability, by the person who is easiest to fool, often with limited statistical training, who has every incentive imaginable to produce $p < .05$.
- When $p < .05$, the result is declared “significant”and any further doubt is frowned upon, as it violates an implicit social contract [at least in psychology].

John Tukey

[Statistical procedures should not be used] "...for sanctification, for the preservation of conclusions from all criticism, for the granting of an *imprimatur*."

What Researchers Want

- To discover the truth, but also:
  - To present compelling data that leave no room for doubt or dissent
  - To develop a coherent theoretical framework
  - To publish papers that make interesting claims

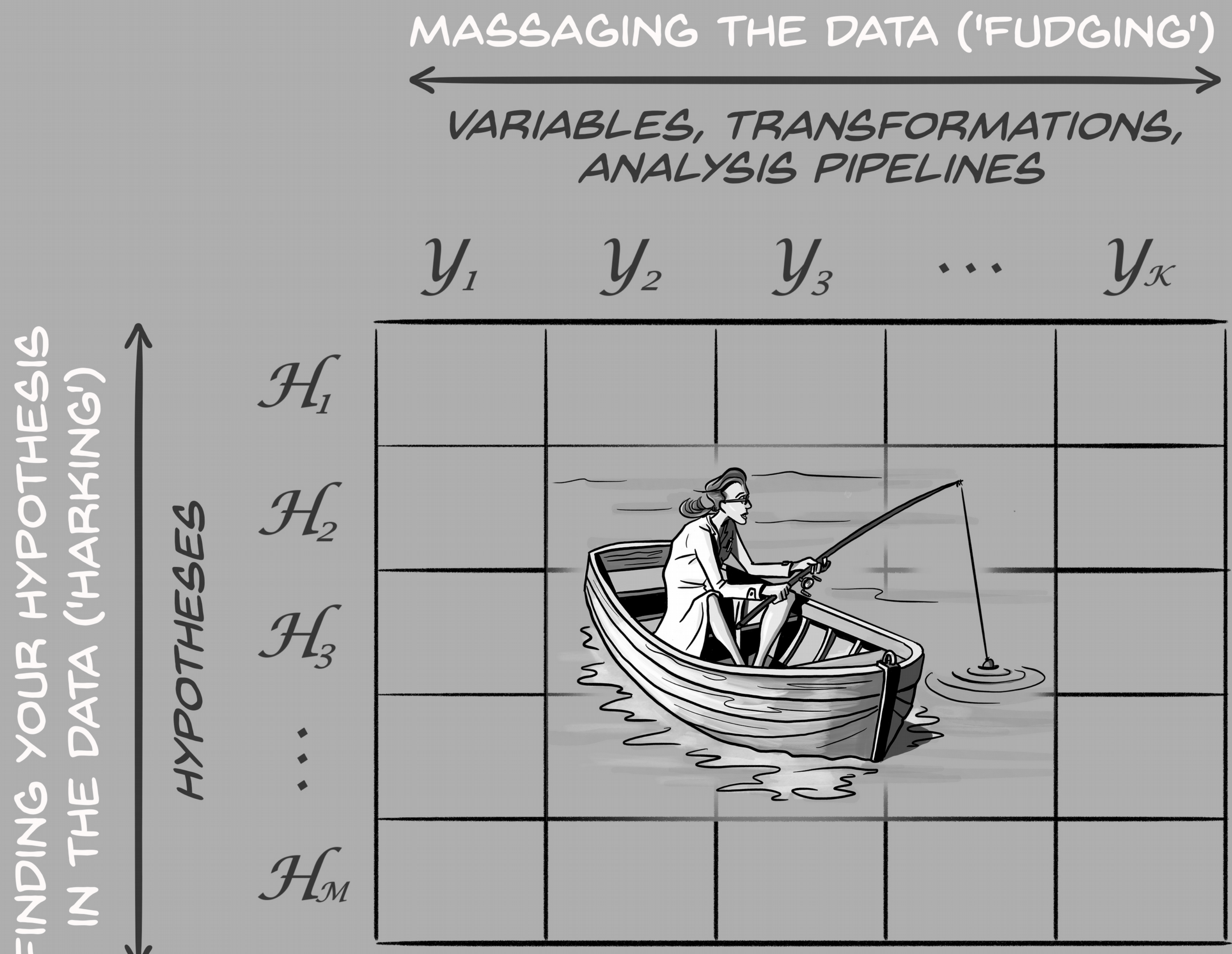
Researchers Abhor Uncertainty

Overview

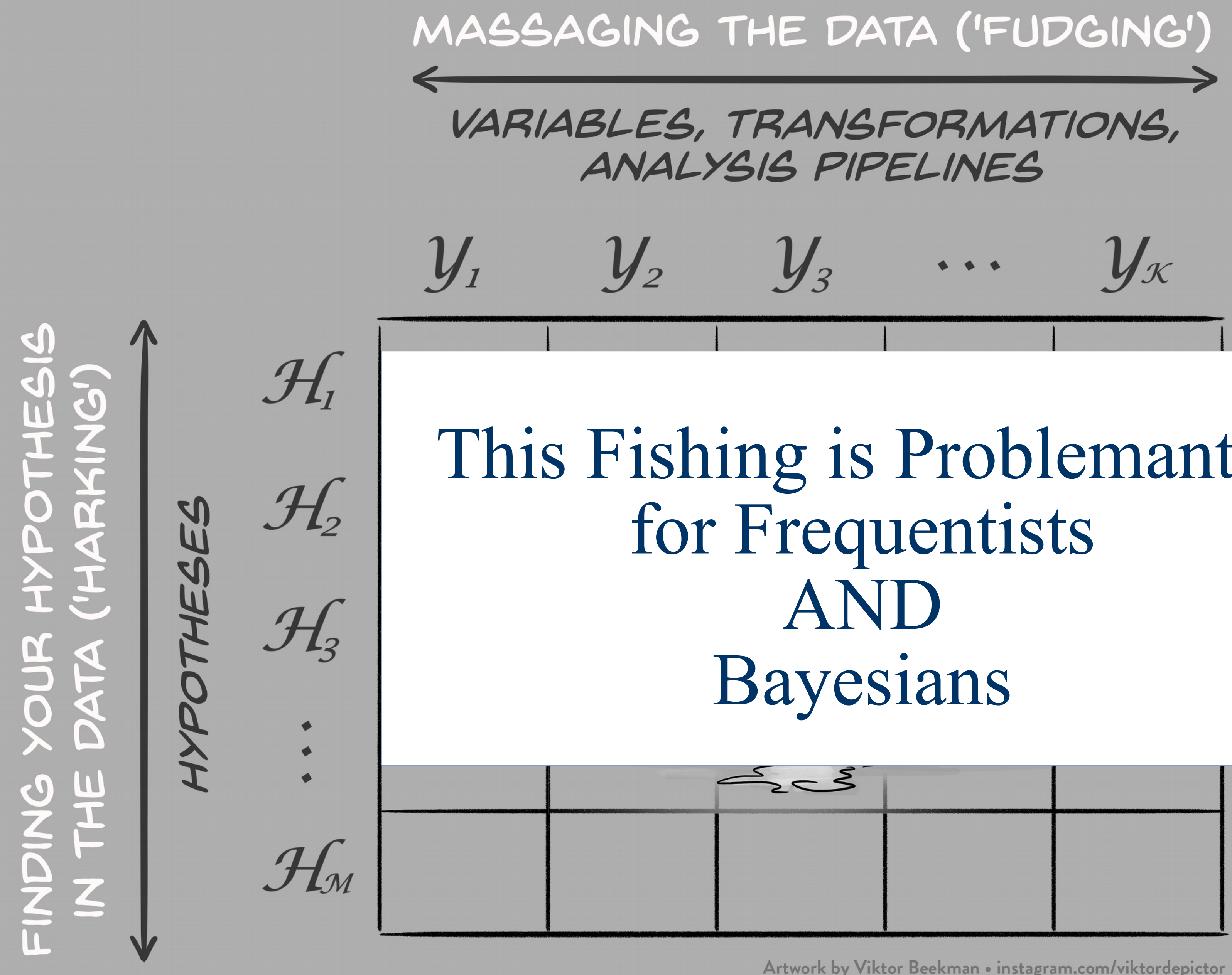
- What researchers want
- <u>What the field gets</u>
- How to uncover hidden uncertainty

Fruits of Perverse Incentives and Uncertainty-Allergy

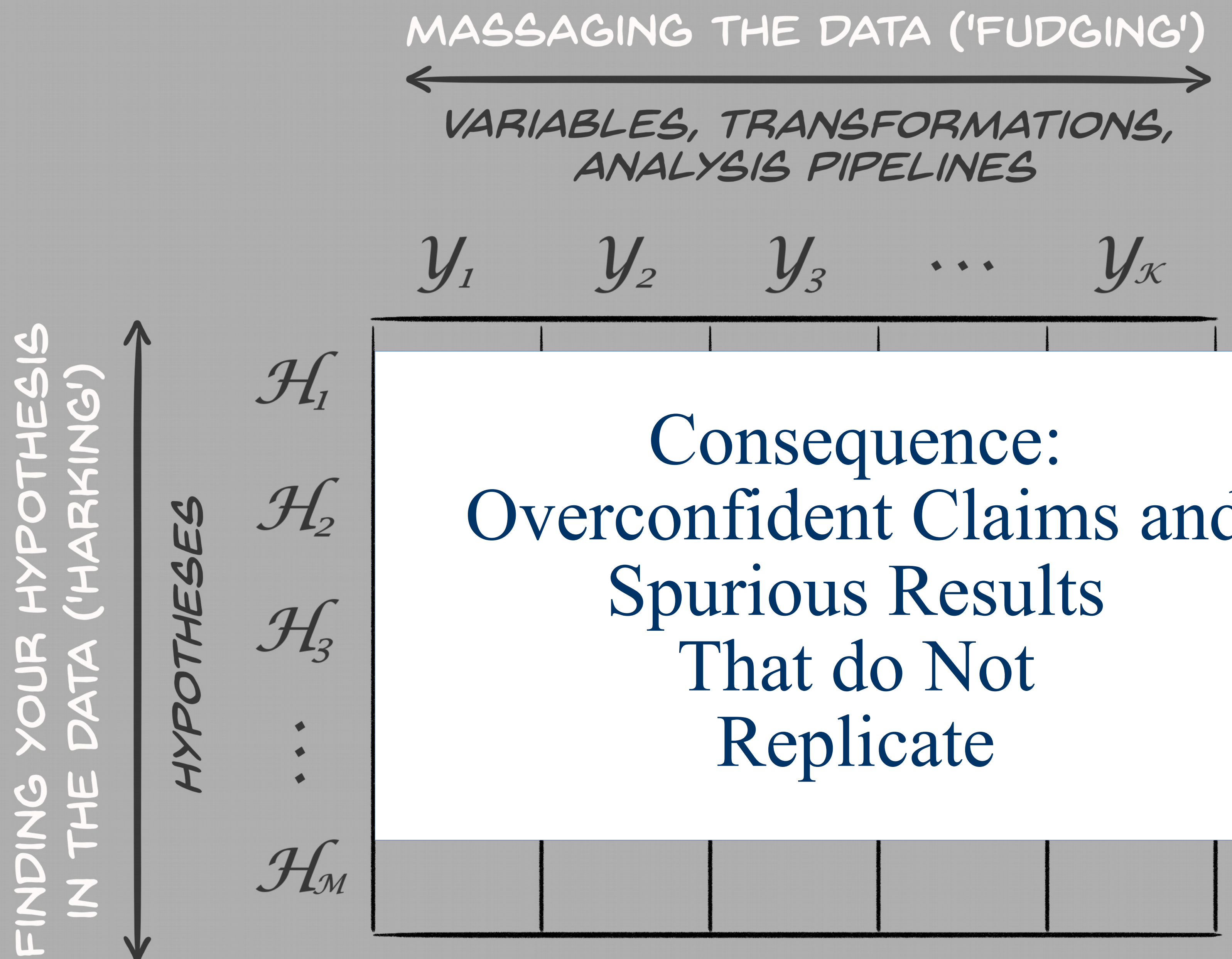
- Publication bias
- Fudging
- HARKing



Artwork by Viktor Beekman • instagram.com/viktordepictor



This Fishing is Problemantic for Frequentists AND Bayesians



Consequence: Overconfident Claims and Spurious Results That do Not Replicate

Artwork by Viktor Beekman • instagram.com/viktordepictor

Overview

- What researchers want
- What the field gets
- <u>How to uncover hidden uncertainty</u>

Method 1: Preregistration of Analysis Plans

- Separates what was post-hoc from what was pre-planned
- Prevents researchers from fooling themselves and others
- Does not rule out exploratory expeditions; just labels them as such

Method 2: Outcome-Independent Publication

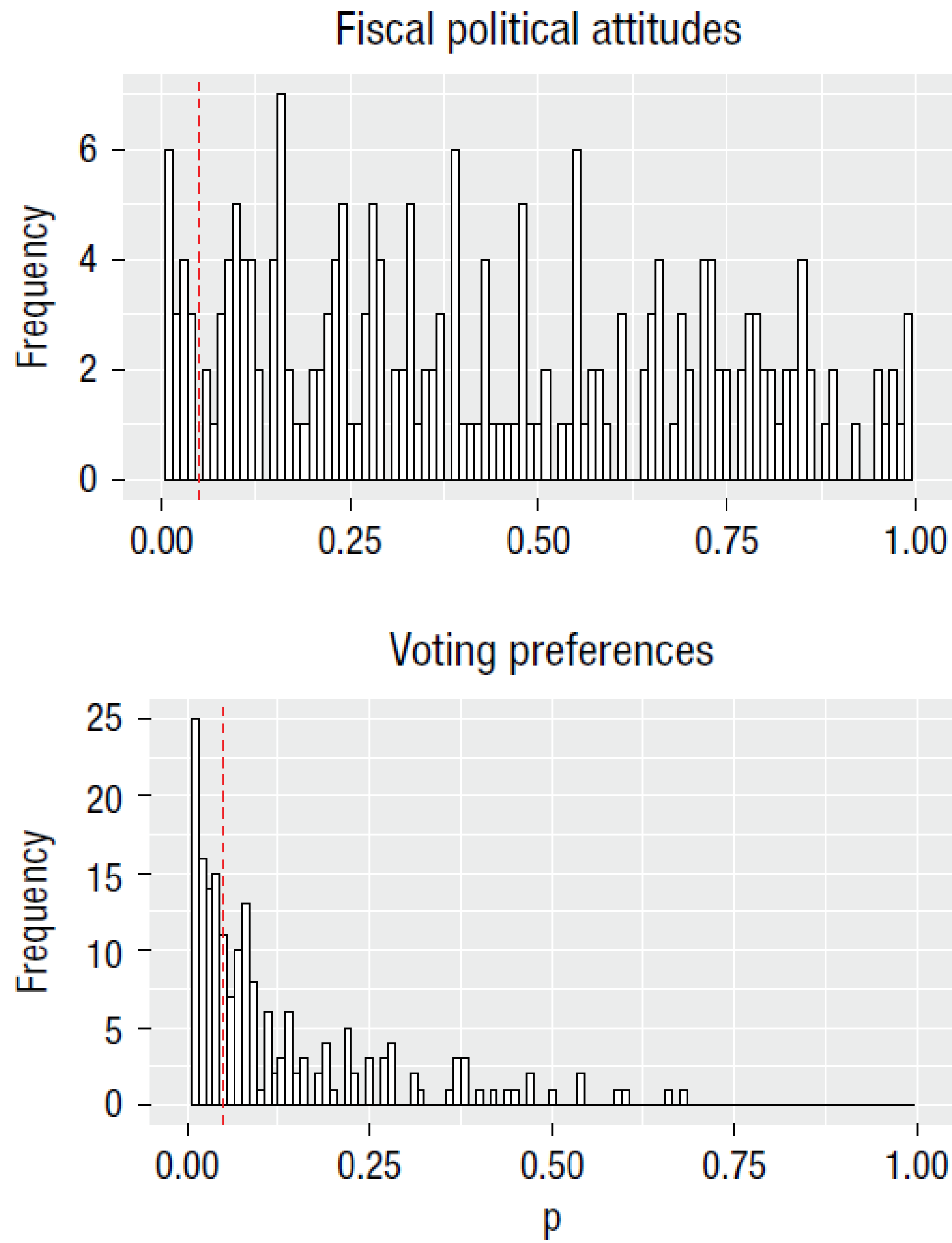
- Judge work based on quality of execution, not on whether $p <. 05$
- Best if combined with preregistration, as advocated by Chris Chambers

Method 3: Sensitivity Analyses

- Examine sensitivity to modeling choices: data, likelihood, and prior. For instance:
  - Multiverse analysis
  - Crowd sourcing
- Ideally, this is done by independent labs

Increasing Transparency Through a Multiverse Analysis

Sara Steegen[1], Francis Tuerlinckx[1], Andrew Gelman[2], and Wolf Vanpaemel[1]
[1]KU Leuven, University of Leuven and [2]Columbia University



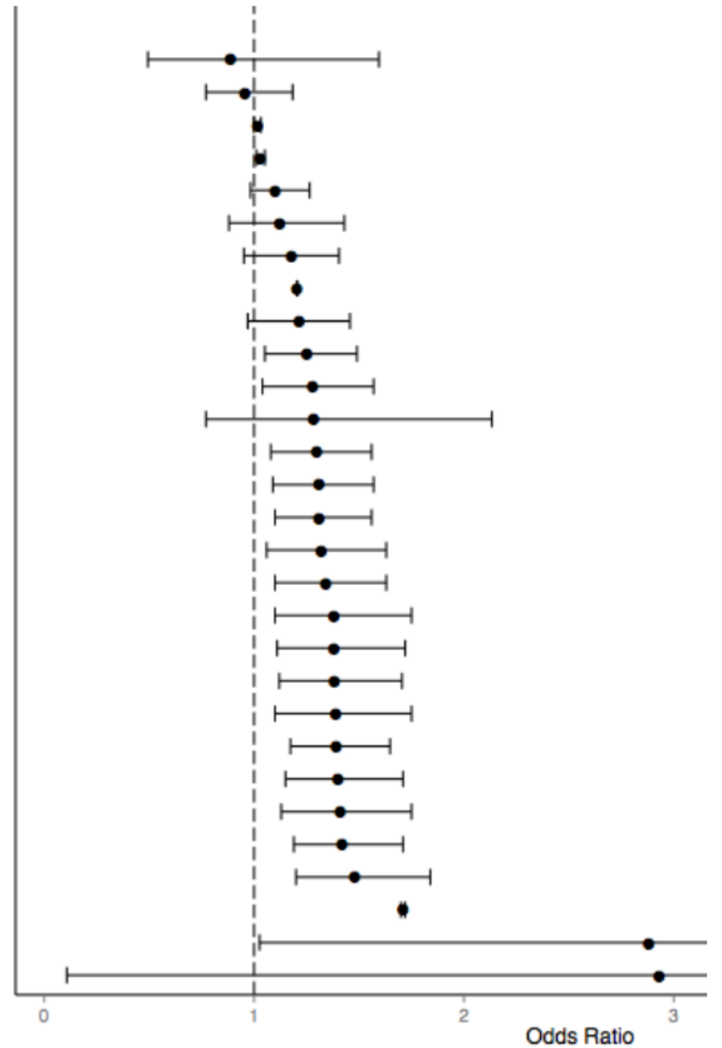
Fiscal political attitudes
Frequency
0
2
4
6
0.00
0.25
0.50
0.75
1.00
Voting preferences
Frequency
0
5
10
15
20
25
0.00
0.25
0.50
0.75
1.00
p

Many analysts, one dataset: Making transparent how variations in analytical choices affect results

Authors

Silberzahn R.[6], Uhlmann E. L.[8], Martin D. P.[35], Anselmi P.[32], Aust F.[26], Awtrey E.[37], Bahník Š.[39], Bai F.[25], Bannard C.[29], Bonnier E.[16], Carlsson R.[9], Cheung F.[13], Christensen G.[20], Clay R.[4], Craig M. A.[15], Dalla Rosa A.[32], Dam L.[28], Evans M. H.[30], Flores Cervantes I.[41], Fong N.[18], Gamez-Djokic M.[14], Glenz A.[40], Gordon-McKeon S.[7], Heaton T. J.[33], Hederos K.[17], Heene M.[11], Hofelich Mohr A. J.[31], Högden F.[26], Hui K.[12], Johannesson M.[16], Kalodimos J.[7], Kaszubowski E.[21], Kennedy D.M.[38], Lei R.[15], Lindsay T. A.[31], Liverani S.[3], Madan C. R.[22], Molden D.[14], Molleman E.[28], Morey R. D.[28], Mulder L. B.[28], Nijstad B. R.[28], Pope N. G.[19], Pope B.[2], Prenoveau J. M.[10], Rink F.[28], Robusto E.[32], Roderique H.[34], Sandberg A.[17], Schlüter E.[27], Schönbrodt F. D.[11], Sherman M. F.[10], Sommer S.A.[5], Sotak K.[1], Spain S.[1], Spörlein C.[24], Stafford T.[33], Stefanutti L.[32], Tauber S.[28], Ullrich J.[40], Vianello M.[32], Wagenmakers E.-J.[23], Witkowiak M.[42], Yoon S.[18], & Nosek B. A.[35, 36]

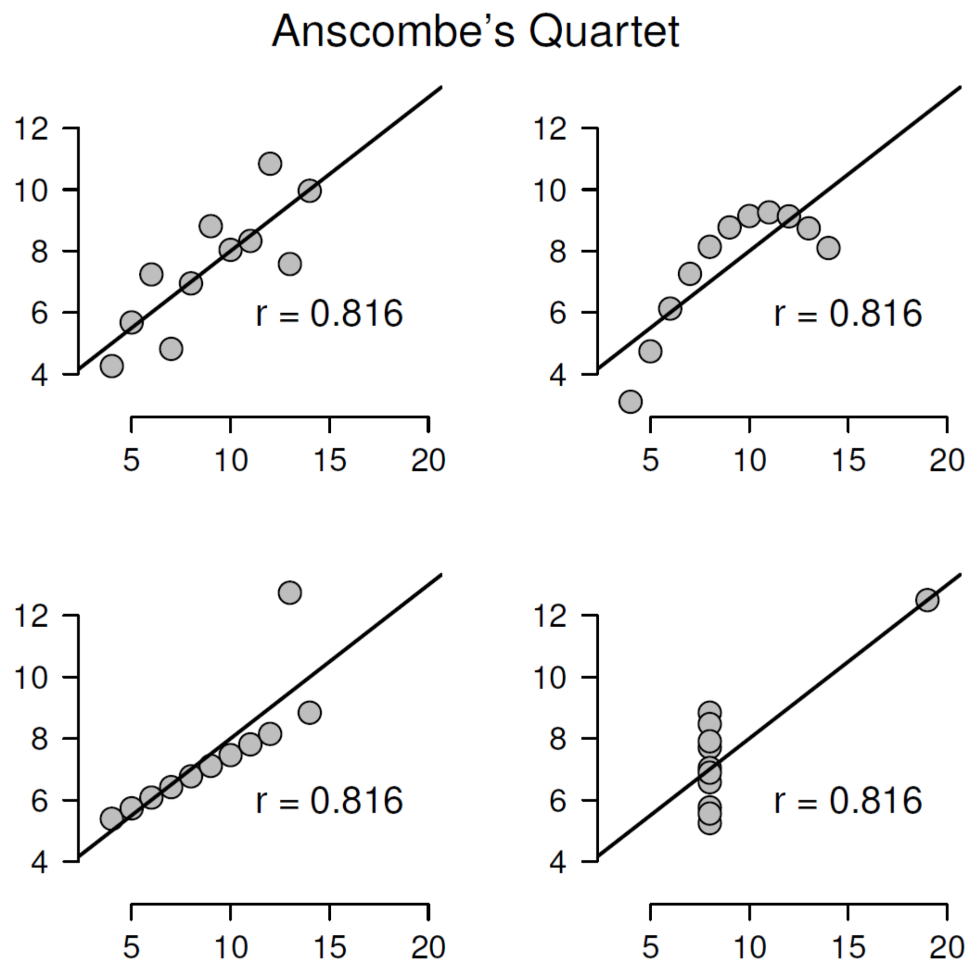
Team	Analytic Approach	OR
12	Zero–inflated Poisson regression	0.89
17	Bayesian logistic regression	0.96
15	Hierarchical log–linear modeling	1.02
10	Multilevel regression and logistic regression	1.03
18	Hierarchical Bayes model	1.10
31	Logistic regression	1.12
1	Ordinary least squares with robust standard errors, logistic regression	1.18
4	Spearman correlation	1.21
14	Weighted least squares regression with referee fixed–effects and clustered standard errors	1.21
11	Multiple linear regression	1.25
30	Clustered robust binomial logistic regression	1.28
6	Linear Probability Model	1.28
26	Three–level hierarchical generalized linear modeling with Poisson sampling	1.30
3	Multilevel Binomial Logistic Regression using bayesian inference	1.31
23	Mixed model logistic regression	1.31
16	Hierarchical Poisson Regression	1.32
2	Linear probability model, logistic regression	1.34
5	Generalized linear mixed models	1.38
24	Multilevel logistic regression	1.38
28	Mixed effects logistic regression	1.38
32	Generalized linear models for binary data	1.39
8	Negative binomial regression with a log link analysis	1.39
20	Cross–classified multilevel negative binomial model	1.40
13	Poisson Multi–level modeling	1.41
25	Multilevel logistic binomial regression	1.42
9	Generalized linear mixed effects models with a logit link function	1.48
7	Dirichlet process Bayesian clustering	1.71
21	Tobit regression	2.88
27	Poisson regression	2.93



Method 4: Share the Data

- Facilitates re-analysis
- In review process, allows reviewers to propose and carry out informative alternative analyses

Method 5: Plot the Data



Method 6: Adopt an Inclusive Inferential Approach

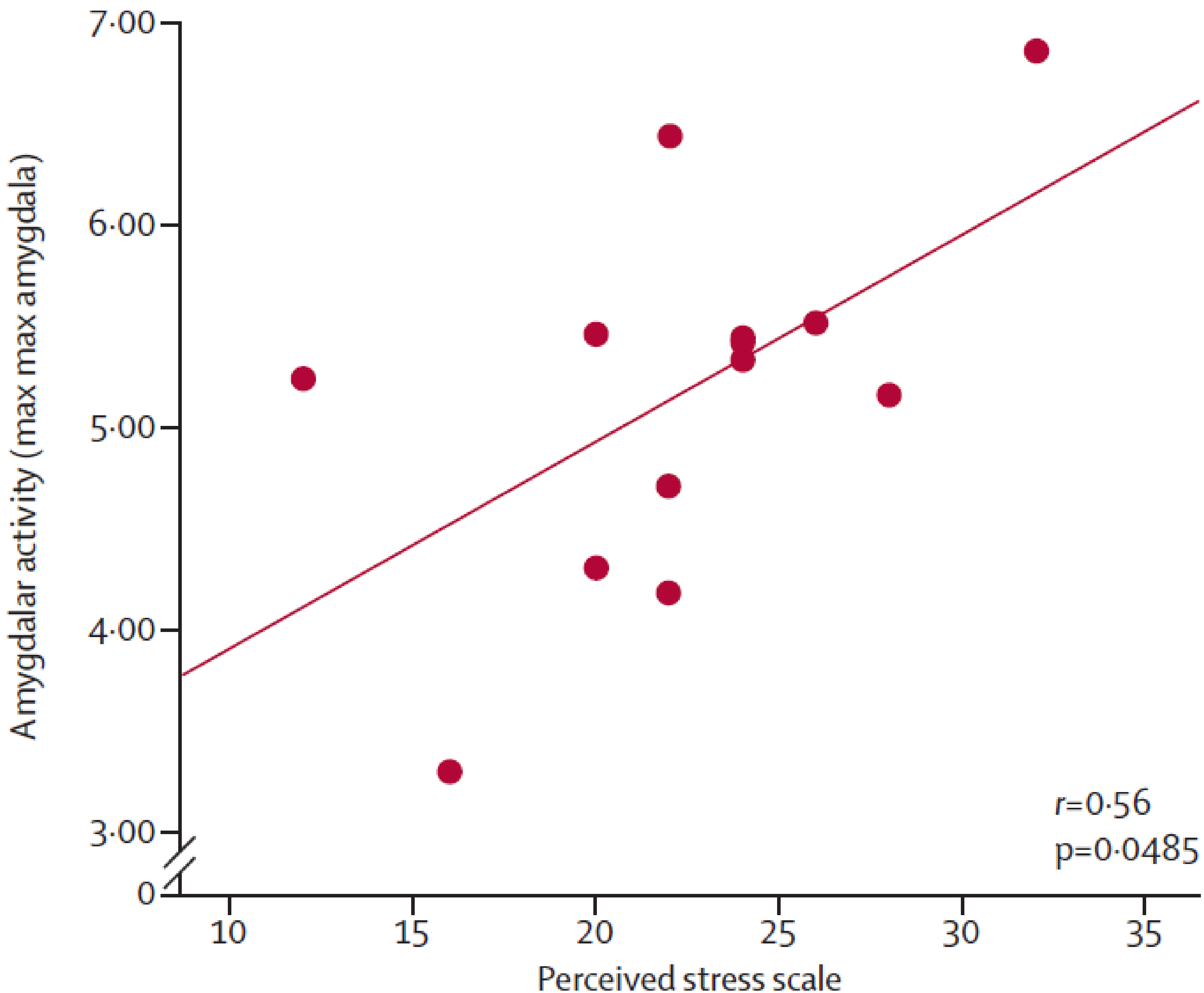
- A classical analysis may be reported as $r(13) = .58$, $p < .05$ (sometimes even without showing the data)
- This does not stimulate statistical curiosity; it is meant to suppress it

Method 6: Inclusive Analyses

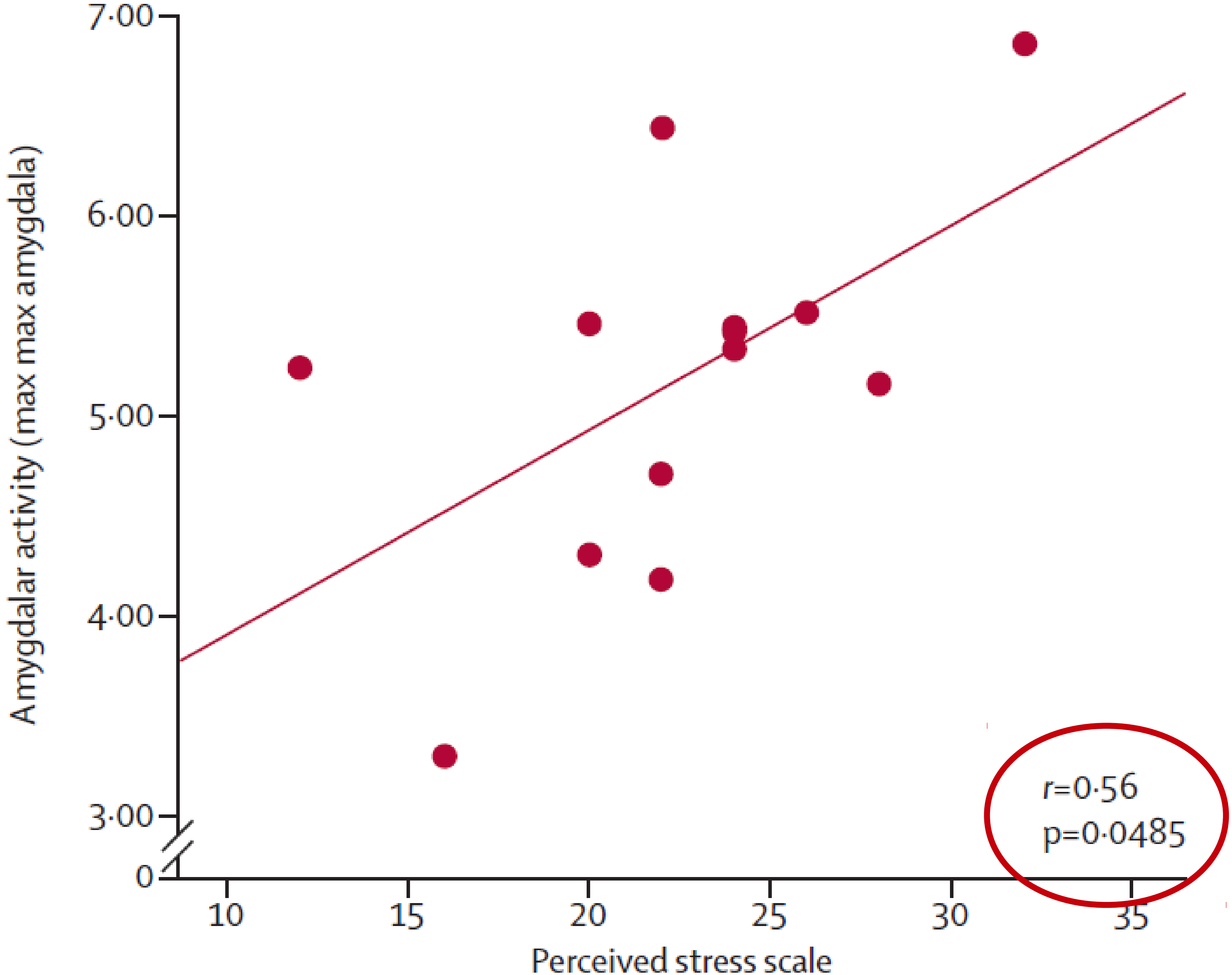
- Consider a paper published this year in the Lancet:

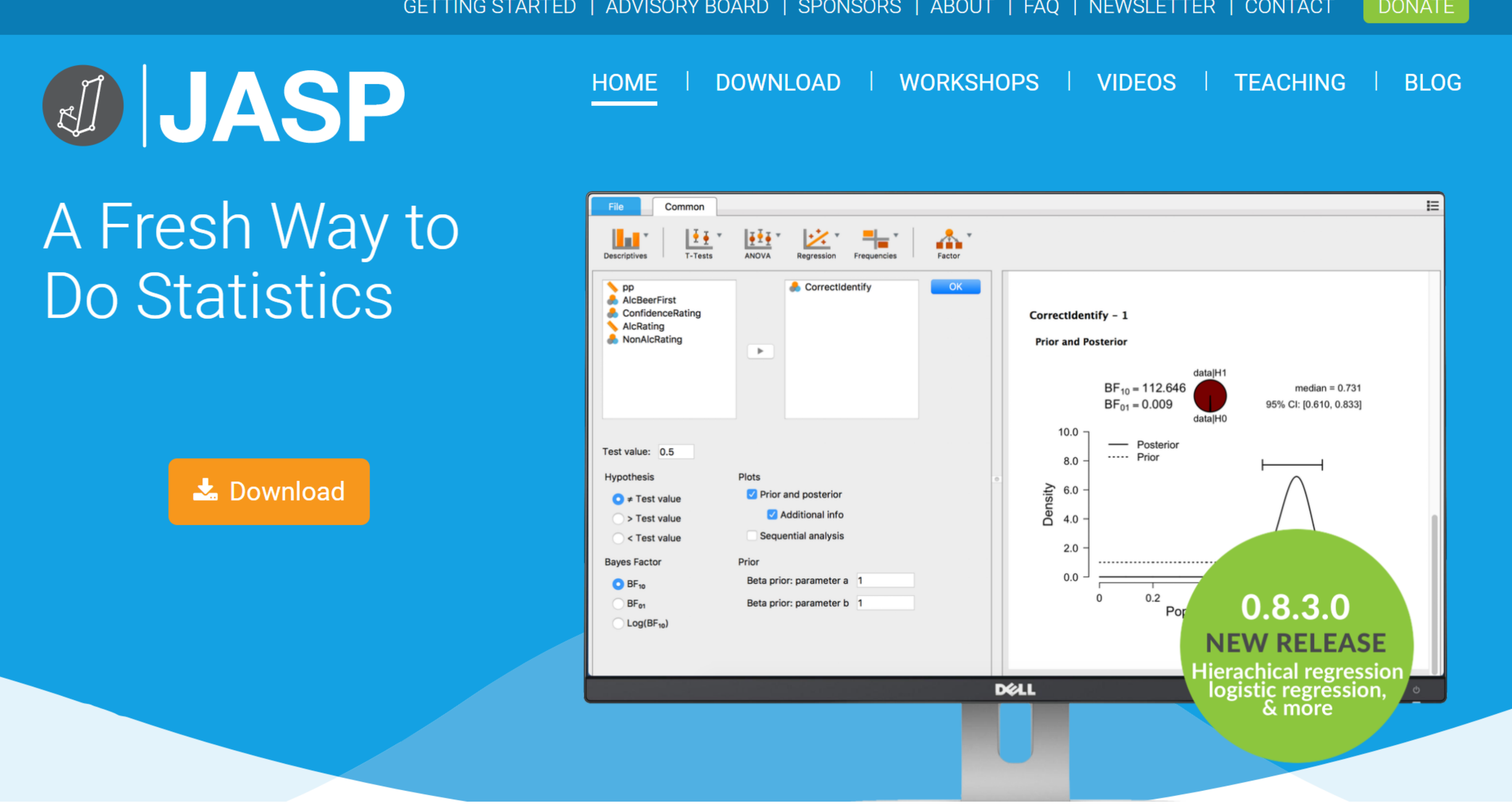
Relation between resting amygdalar activity and cardiovascular events: a longitudinal and cohort study

Ahmed Tawakol, Amorina Ishai*, Richard AP Takx, Amparo L Figueroa, Abdelrahman Ali, Yannick Kaiser, Quynh A Truong, Chloe JE Solomon, Claudia Calcagno, Venkatesh Mani, Cheuk Y Tang, Willem JM Mulder, James W Murrough, Udo Hoffmann, Matthias Nahrendorf, Lisa M Shin, Zahi A Fayad†, Roger K Pitman†*

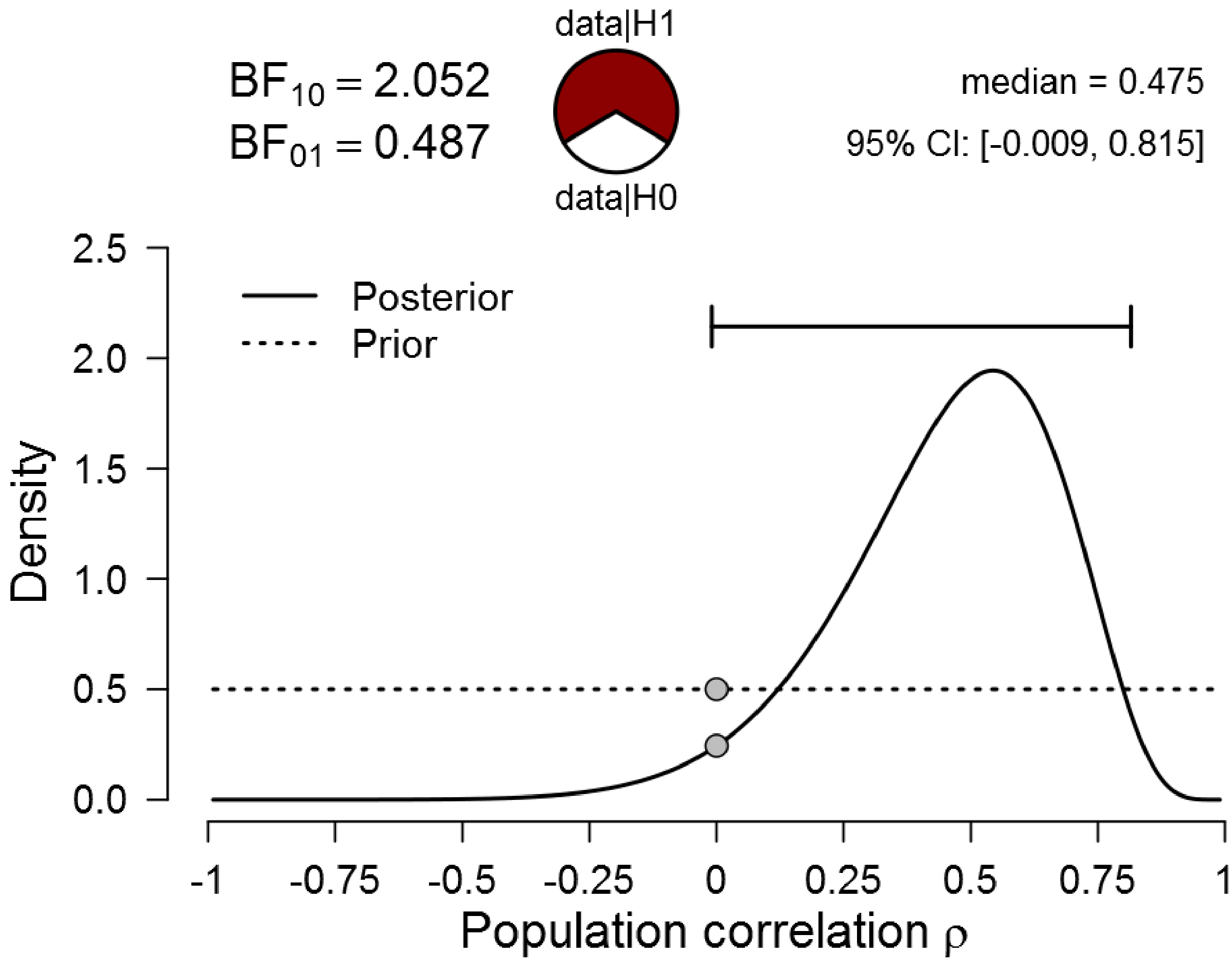


7·00
6·00
5·00
4·00
3·00
0
Amygdalar activity (max max amygdala)
10
15
20
25
30
35
Perceived stress scale
r=0·56
p=0·0485

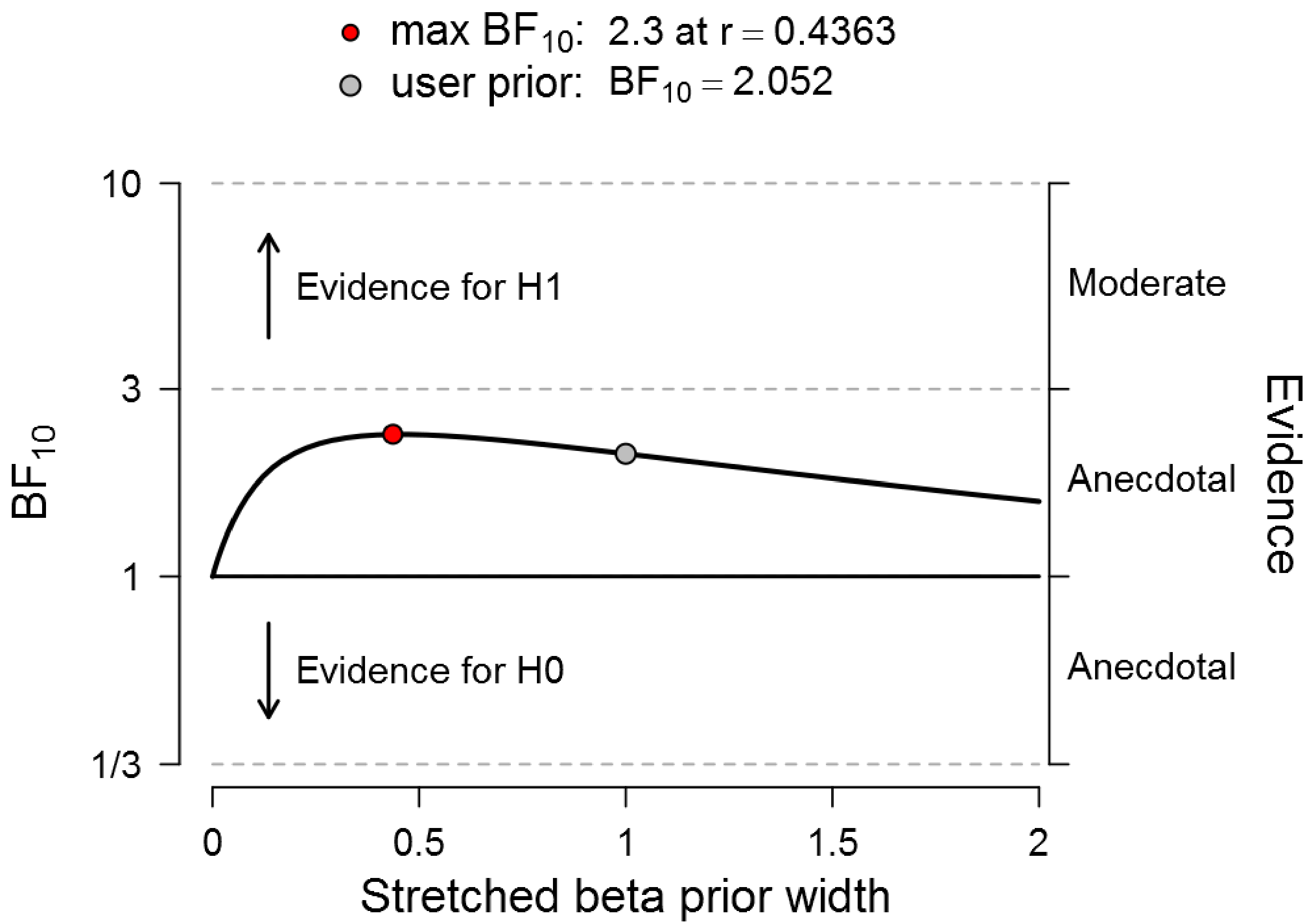

Amygdalar activity (max max amygdala)
7·00
6·00
5·00
4·00
3·00
0
10
15
20
25
30
35
Perceived stress scale
r=0·56
p=0·0485



jasp-stats.org



$BF_{10} = 2.052$
$BF_{01} = 0.487$
data|H1
data|H0
median = 0.475
95% CI: [-0.009, 0.815]
Posterior
Prior
Density
2.5
2.0
1.5
1.0
0.5
0.0
-1
-0.75
-0.5
-0.25
0
0.25
0.5
0.75
1
Population correlation ρ


max BF_{10}: 2.3 at r = 0.4363
user prior: BF_{10} = 2.052
10
3
1
1/3
BF_{10}
Evidence for H1
Evidence for H0
Moderate
Anecdotal
Anecdotal
Evidence
0
0.5
1
1.5
2
Stretched beta prior width

Concluding Comments I

- More transparency is *sorely* needed
- Transparency means mental hygiene: the scientific equivalent of brushing your teeth, or washing your hands after visiting the restroom
- This requires a change in culture

Concluding Comments II

- Journals and funders can start demanding mental hygiene
- Mental hygiene can also be rewarded. For instance, journals could prefer papers that conduct multiverse analyses etc.

Concluding Comments III

- Journals could publish reviewers' reports when these contain useful re-analyses, promoting a crowd sourcing approach and rewarding reviewers for their efforts

Thanks for Your Attention