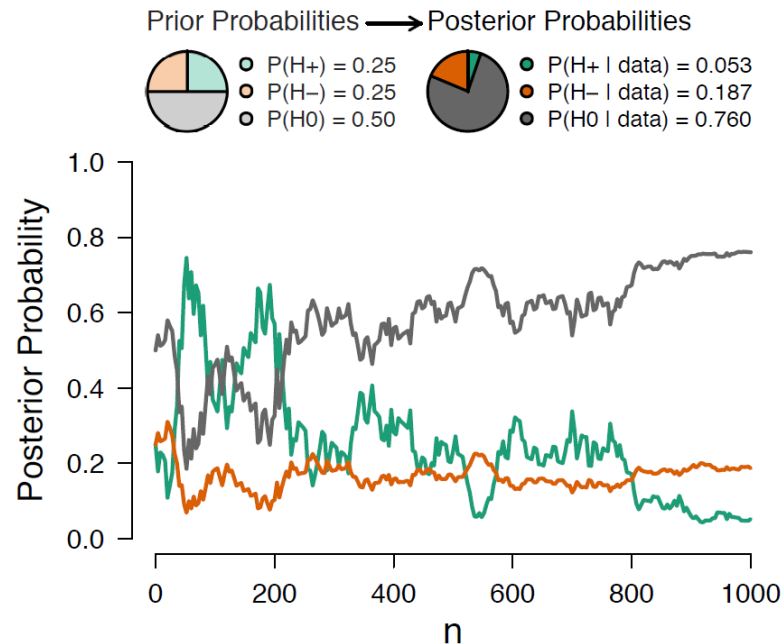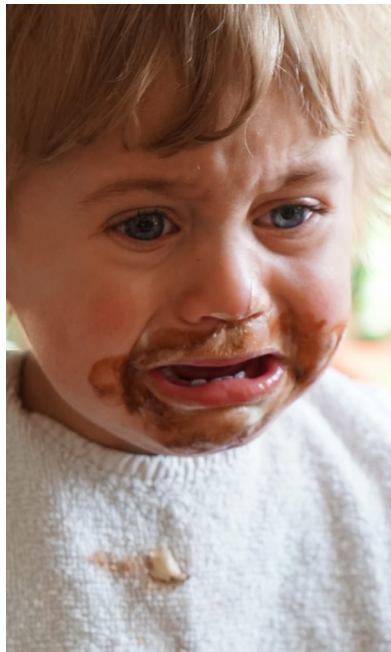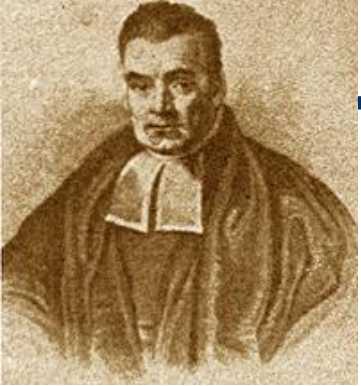# Bayesian Benefits for the Pragmatic Biostatistician

E.-J. Wagenmakers

Main Claims

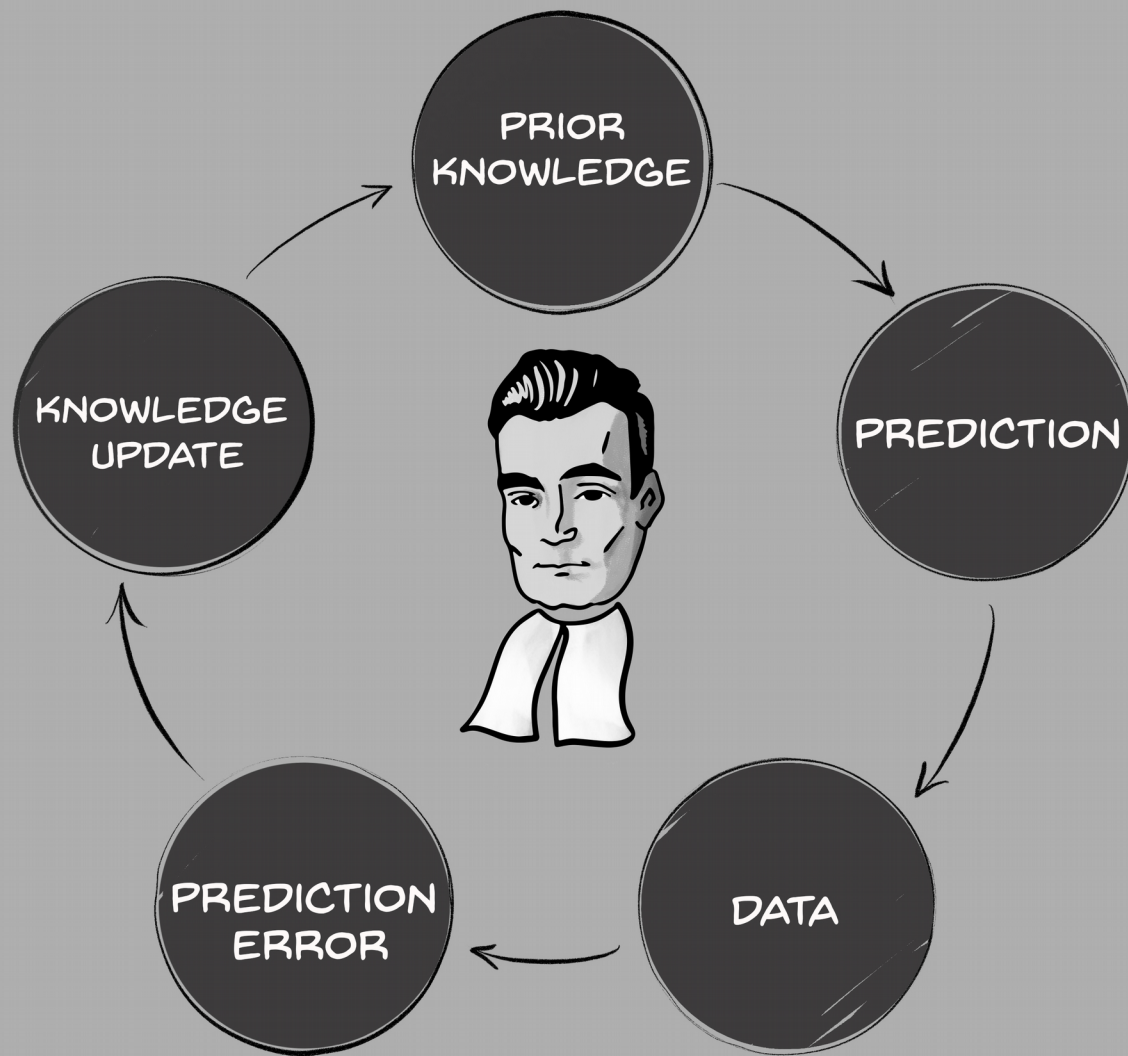- Frequentist inference is superficial at best, and misleading at worst. Yet it continues to dominate the field.

- Bayesian inference provides a series of benefits that are beyond the scope of frequentist methods.

- These Bayesian benefits are obtained in JASP (jasp-stats.org) with minimal effort.

Bayes' Rule

$$\underbrace{p(\theta \mid \text{data})}_{\substack{\text{Posterior beliefs} \\ \text{about parameters}}} = \underbrace{p(\theta)}_{\substack{\text{Prior beliefs} \\ \text{about parameters}}} \times \underbrace{\frac{p(\text{data} \mid \theta)}{p(\text{data})}}_{\substack{\text{Predictive} \\ \text{updating factor}}}$$

BAYESIAN INDUCTIVE CYCLE

PRIOR KNOWLEDGE
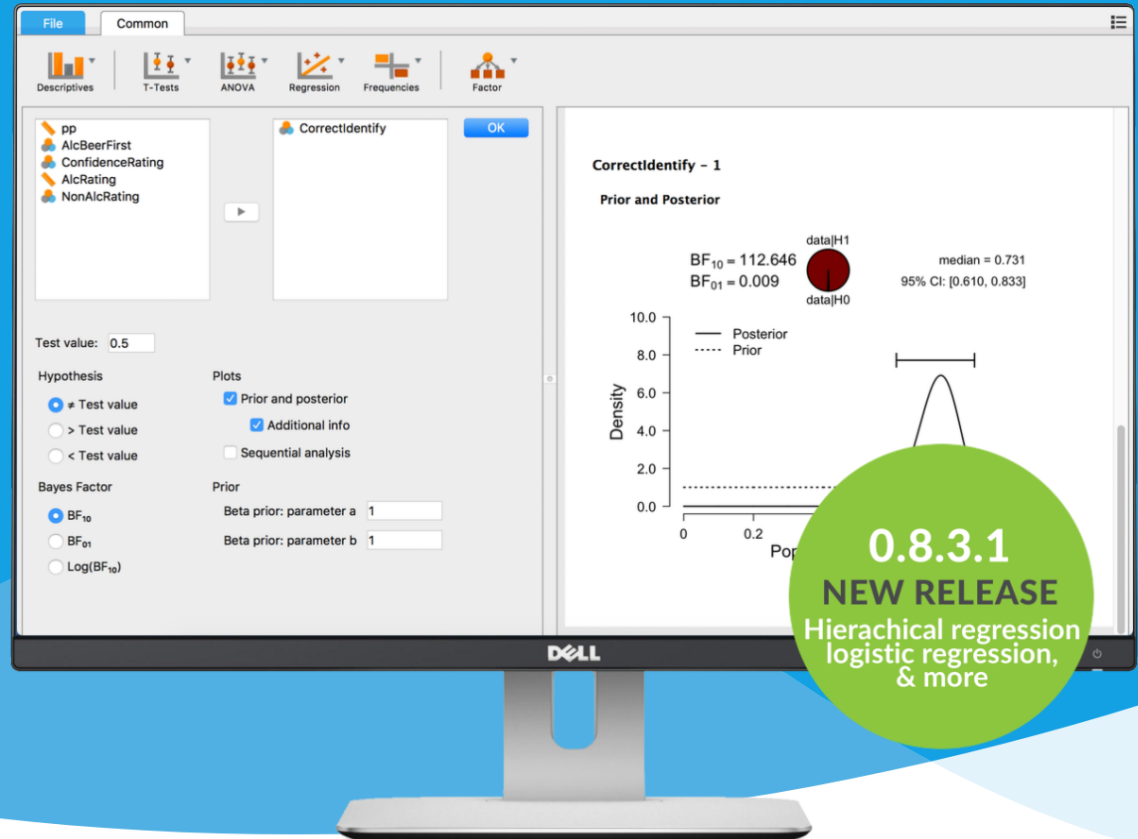
PREDICTION

DATA

PREDICTION ERROR

KNOWLEDGE UPDATE

Examples

- Example 1: Does progesterone prevent miscarriages? (NEJM, this month)

- Example 2: Are movies with Adam Sandler profitable regardless of their quality?

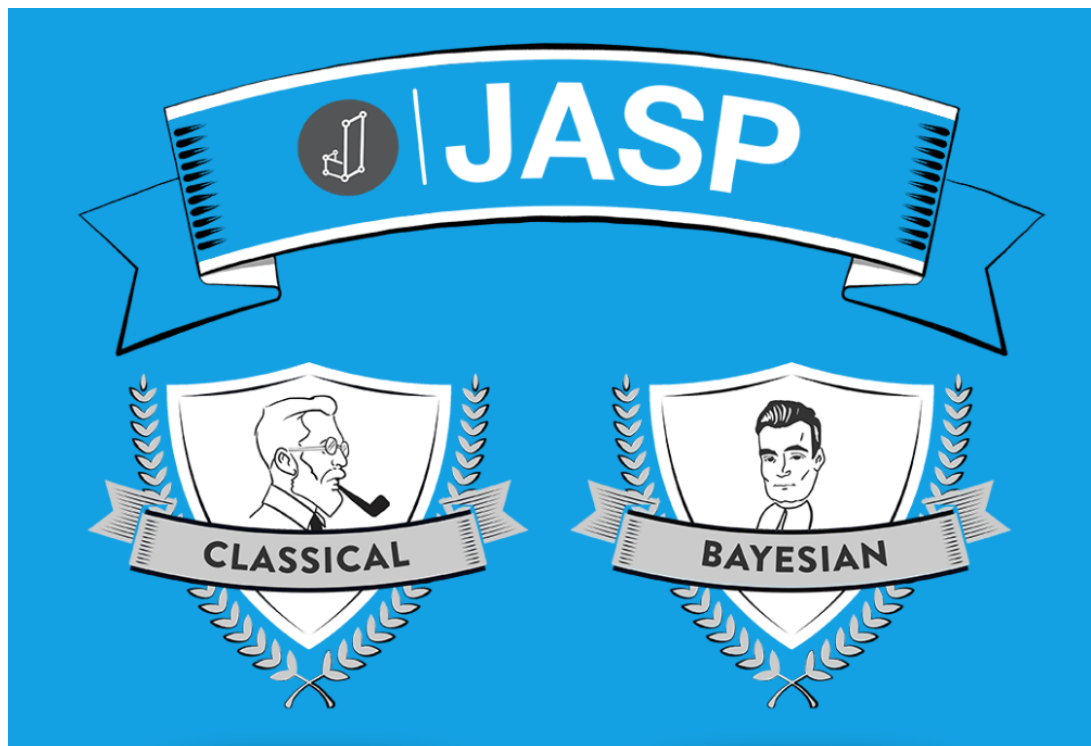- Example 3 [if time permits]: Do men with bigger balls neglect their children more? (PNAS, 2013)
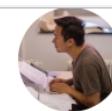
JASP

A Fresh Way to Do Statistics

⬇ Download

File | Common

Descriptives | T-Tests | ANOVA | Regression | Frequencies | Factor

pp
AlcBeerFirst
ConfidenceRating
AlcRating
NonAlcRating

▶

CorrectIdentify   OK

CorrectIdentify – 1

Prior and Posterior

data|H1

$BF_{10} = 112.646$
$BF_{01} = 0.009$

median = 0.731
95% CI: [0.610, 0.833]

data|H0

Test value: 0.5

Hypothesis
○ ≠ Test value
○ > Test value
○ < Test value

Plots
☑ Prior and posterior
   ☑ Additional info
☐ Sequential analysis

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

Prior
Beta prior: parameter a   1
Beta prior: parameter b   1

Density — Posterior ---- Prior

10.0
8.0
6.0
4.0
2.0
0.0

0   0.2   Pop

DELL

0.8.3.1
NEW RELEASE
Hierachical regression
logistic regression,
& more

More information at
jasp-stats.org

Eric-Jan Wagenmakers
CEO / Founder. Guides the development of JASP.

Alexander Ly
CTO. Responsible for guiding JASP's scientific and technological strategy and developer of some Bayesian tests.

Bruno Boutin
Lead Software Developer. Responsible for the core development of JASP.

Frans Meerhoff
Software Developer. Responsible for the core development of JASP.

Akash Raj
Software Developer. Responsible for the implementation of UI elements. Implemented the Summary Stats module.

Quentin Gronau
Analyst. Responsible for the t-tests and the binomial test. Implemented the figures for the Bayesian analyses.

Alexandra Sarafoglou
Analyst. Contributing to the multinomial analysis, the video tutorials, and the JASP workshop.

Jan G. Voelkel
Software Developer. Responsible for improving the R analyses.

Maarten Marsman
Analyst. Responsible for the Bayesian linear models (e.g., ANOVA and regression).

Don van den Bergh
Analyst. Responsible for the frequentist and Bayesian reliability analysis, the machine learning module, and the network module. Also part of the workshop organisation team.

Johnny van Doorn
Analyst. Responsible for Bayesian nonparametric analyses and part of the workshop organization team.

Dora Matzke
Analyst. Responsible for developing and maintaining the help functionality and the JASP documentation.

Sacha Epskamp
Analyst. Responsible for factor analysis and the SEM module.

Alexander Etz
The voice of many JASP video tutorials and other videos on our Youtube channel.

Erik-Jan van Kesteren
Software developer. Responsible for adding plots, functions, and UI elements, and interfacing R and C++.

Raoul Grasman
Data scientist and code contributor. Responsible for improving code and developing new modules.

Herbert Hoijtink
Contributing to the Informative Hypotheses module.

Joris Mulder
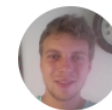Contributing to the Informative Hypotheses module.

Xin Gu
Contributing to the Informative Hypotheses module.

Richard Morey
Author and maintainer of the BayesFactor package.

Tim Draws
Marketing and Communication Manager. Responsible for marketing strategy, website, blog, and the YouTube channel.

Koen Derks
Contributing to the Machine Learning module, and the Bayesian Informative Hypothesis Testing module.

Joris Goosen
Software developer. Responsible for the core development of JASP.

Lotte Kehrer
Contributing to the blog, YouTube channel and manual of JASP.

ORIGINAL ARTICLE

A Randomized Trial of Progesterone in Women with Bleeding in Early Pregnancy

A. Coomarasamy, A.J. Devall, V. Cheed, H. Harb, L.J. Middleton, I.D. Gallos,
H. Williams, A.K. Eapen, T. Roberts, C.C. Ogwulu, I. Goranitis, J.P. Daniels,
A. Ahmed, R. Bender-Atik, K. Bhatia, C. Bottomley, J. Brewin, M. Choudhary,
F. Crosfill, S. Deb, W.C. Duncan, A. Ewer, K. Hinshaw, T. Holland, F. Izzat,
J. Johns, K. Kriedt, M.-A. Lumsden, P. Manda, J.E. Norman, N. Nunes,
C.E. Overton, S. Quenby, S. Rao, J. Ross, A. Shahid, M. Underwood,
N. Vaithilingam, L. Watkins, C. Wykes, A. Horne, and D. Jurkovic

BACKGROUND

Bleeding in early pregnancy is strongly associated with pregnancy loss. Progesterone is essential for the maintenance of pregnancy. Several small trials have suggested that progesterone therapy may improve pregnancy outcomes in women who have bleeding in early pregnancy.

METHODS

We conducted a multicenter, randomized, double-blind, placebo-controlled trial to evaluate progesterone, as compared with placebo, in women with vaginal bleeding in early pregnancy. Women were randomly assigned to receive vaginal suppositories containing either 400 mg of progesterone or matching placebo twice daily, from the time at which they presented with bleeding through 16 weeks of gestation. The primary outcome was the birth of a live-born baby after at least 34 weeks of gestation. The primary analysis was performed in all participants for whom data on the primary outcome were available. A sensitivity analysis of the primary outcome that included all the participants was performed with the use of multiple imputation to account for missing data.

RESULTS

A total of 4153 women, recruited at 48 hospitals in the United Kingdom, were randomly assigned to receive progesterone (2079 women) or placebo (2074 women). The percentage of women with available data for the primary outcome was 97% (4038 of 4153 women). The incidence of live births after at least 34 weeks of gestation was 75% (1513 of 2025 women) in the progesterone group and 72% (1459 of 2013 women) in the placebo group (relative rate, 1.03; 95% confidence interval [CI], 1.00 to 1.07; P=0.08). The sensitivity analysis, in which missing primary outcome data were imputed, resulted in a similar finding (relative rate, 1.03; 95% CI, 1.00 to 1.07; P=0.08). The incidence of adverse events did not differ significantly between the groups.
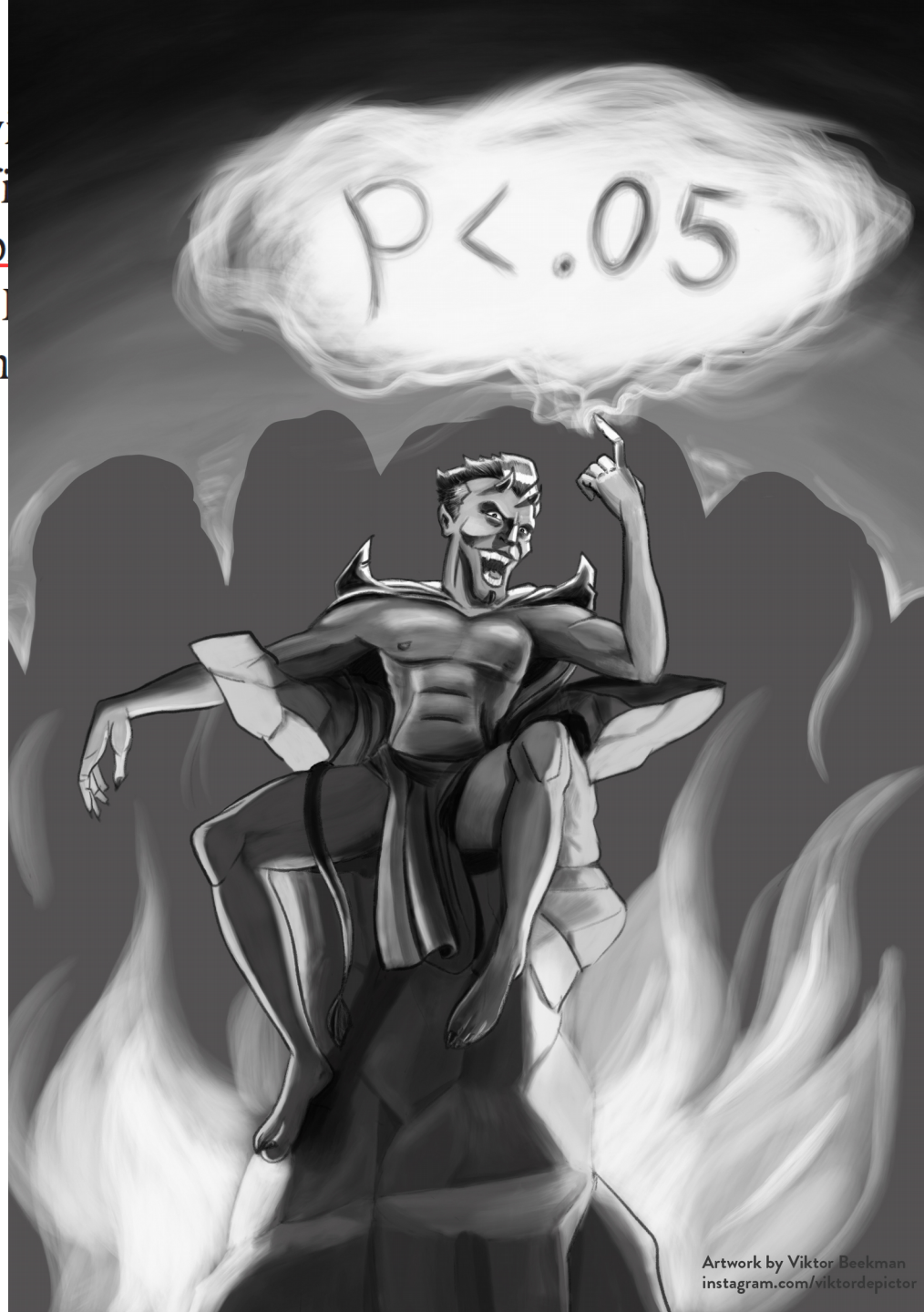
CONCLUSIONS

Among women with bleeding in early pregnancy, progesterone therapy administered during the first trimester did not result in a significantly higher incidence of live births than placebo. (Funded by the United Kingdom National Institute for Health Research Health Technology Assessment program; PRISM Current Controlled Trials number, ISRCTN14163439.)

CONCLUSIONS

Among women w... ...therapy administered during the fi... ...igher incidence of live births than p... ...onal Institute for Health Research ... ...SM Current Controlled Trials num...

P<.05

Artwork by Viktor Beekman
instagram.com/viktordepictor

Professor Arri Coomarasamy, MBChB, MD, FRCOG @arricoom... · May 11 ∨

10/ The P-value (0.08) didn't cross the magical bright-line (P-value = 0.05), but it wasn't far from it, for those who seek reassurance in a P-value. For those who prefer to look at 95% confidence interval, well, it is almost there, going from 1.00 to 1.07.

💬 1   ⟲   ♡ 1   ✉

Professor Arri Coomarasamy, MBChB, MD, FRCOG @arricoom... · May 11 ∨

11/ So, we are left with some (only some) statistical uncertainty, but scientific inference is much more than statistical inference. What is the context? What is biological rationale? Is there any supportive subgroup behaviour? What is the external evidence?

💬 1   ⟲   ♡ 1   ✉

Professor Arri Coomarasamy, MBChB, MD, FRCOG @arricoom... · May 11 ∨

12/ CONTEXT: In 2012, NICE considered the evidence and in its draft guideline, suggested: "Consider progesterone for women with threatened miscarriage"; "Inform women that although there is some evidence that progesterone can prevent a miscarriage, this evidence is not strong".

💬 1   ⟲ 1   ♡ 1   ✉

Professor Arri Coomarasamy, MBChB, MD, FRCOG @arricoom... · May 11 ∨

14/ BIOLOGICAL RATIONALE: Progesterone is vital to support a pregnancy. Removal of corpus luteum (which produces progesterone) always results in a miscarriage. Anti-progestogens cause termination. It is therefore plausible progesterone could be helpful for some women.

💬 2   ⟲ 1   ♡ 2   ✉

Professor Arri Coomarasamy, MBChB, MD, FRCOG @arricoom... · May 11

18/ BIOLOGICAL GRADIENT: If the risk of euploid miscarriage increases with increasing number of previous losses, then we would expect a 'biological gradient' i.e., increasing benefit with progesterone with increasing number of previous miscarriages. We found a powerful gradient:

	Progesterone (n/N)	Placebo (n/N)		Risk Ratio [95% CI]	P-value for interaction
First Subgroup Analysis					
Number of previous miscarriages					
0	824 / 1111	840 / 1127		0.99 (0.95-1.04)	0.01
≥ 1	689 / 914	619 / 886		1.09 (1.03-1.15)	
Second Subgroup Analysis					
Number of previous miscarriages					
0	824 / 1111	840 / 1127		0.99 (0.95-1.04)	
1	413 / 547	367 / 502		1.04 (0.97-1.12)	0.02
2	178 / 230	167 / 236		1.08 (0.97-1.19)	
≥ 3	98 / 137	85 / 148		1.28 (1.08-1.51)	
All Participants	1513 / 2025	1459 / 2013		1.03 (1.00-1.07)	

0.75 1 1.25

← Favors Placebo   Favors Progesterone →

IMSR
INSTITUTE OF METABOLISM AND SYSTEMS RESEARCH

FUNDED BY
NIHR | National Institute for Health Research

UNIVERSITY OF BIRMINGHAM
BCTU Birmingham Clinical Trials Unit

Tommy's
@Birmingham
National Centre for Miscarriage Research

💬 2   🔁   ♡ 3   ✉

19/ EXTERNAL EVIDENCE: PRISM is a large and high quality study. But there are other studies on the subject of progesterone use in the first trimester to prevent miscarriages. Here is a summary of meta-analysed evidence (limited to vaginal progesterone):

Meta-analysis
Vaginal progesterone: live birth or ongoing pregnancy

Study or Subgroup	Progesterone Events	Total	Placebo or no treatment Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI	Year
2.1.1 Threatened miscarriage							
Palagiano 2004	21	25	17	25	0.8%	1.24 [0.90, 1.70]	2004
Alimohamadia 2013	59	71	61	71	3.0%	0.97 [0.84, 1.11]	2013
Yassaee 2014	24	30	20	30	1.0%	1.20 [0.88, 1.64]	2014
PRISM 2019	1513	2025	1459	2013	72.3%	1.03 [0.99, 1.07]	2019
Subtotal (95% CI)		2151		2139	77.2%	1.03 [1.00, 1.07]	
Total events	1617		1557				

Heterogeneity: Chi² = 2.95, df = 3 (P = 0.40); I² = 0%
Test for overall effect: Z = 1.78 (P = 0.07)

Study or Subgroup	Progesterone Events	Total	Placebo or no treatment Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI	Year
2.1.2 Recurrent miscarriage							
PROMISE 2015	262	398	271	428	12.9%	1.04 [0.94, 1.15]	2015
Ismail 2017	273	340	199	335	9.9%	1.35 [1.22, 1.50]	2017
Subtotal (95% CI)		738		763	22.8%	1.18 [1.09, 1.26]	
Total events	535		470				

Heterogeneity: Chi² = 12.74, df = 1 (P = 0.0004); I² = 92%
Test for overall effect: Z = 4.39 (P < 0.0001)

Study or Subgroup	Progesterone Events	Total	Placebo or no treatment Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI	Year
Total (95% CI)		2889		2902	100.0%	1.07 [1.03, 1.10]	
Total events	2152		2027				

Heterogeneity: Chi² = 26.98, df = 5 (P < 0.0001); I² = 81%
Test for overall effect: Z = 3.87 (P = 0.0001)
Test for subgroup differences: Chi² = 9.93, df = 1 (P = 0.002), I² = 89.9%

0.7 0.85 1 1.2 1.5
Favours Placebo or no treatment   Favours Progesterone

IMSR
INSTITUTE OF METABOLISM AND SYSTEMS RESEARCH

FUNDED BY
NIHR | National Institute for Health Research

UNIVERSITY OF BIRMINGHAM
BCTU Birmingham Clinical Trials Unit

Tommy's
@Birmingham
National Centre for Miscarriage Research

💬 2   ⟲ 1   ♡ 2   ✉

Professor Arri Coomarasamy, MBChB, MD, FRCOG @arricoom... · May 11 ⌄

20/ PRISM trial interpretation: Given a) the biological rationale, b) subgroup effects, c) biological gradient and d) supportive external evidence, we judge progesterone to have a probable effect in women with early pregnancy bleeding and a history of any number of miscarriages.

💬 1   ⟲   ♡ 1   ✉

Professor Arri Coomarasamy, MBChB, MD, FRCOG @arricoom... · May 11 ⌄

21/ Given progesterone is safe (our and external evidence), affordable (health economic evaluation out soon), acceptable to patients, and given the low NNTs for live birth, we believe the evidence has policy implications. We will liaise with NICE to provide all the evidence.

Vaginal bleeding at ≤ 12 weeks of pregnancy

No previous losses → Standard care

One or more previous pregnancy loss(es) → UTROGESTAN 200mg, 400mg, twice daily

Outcome: Live birth
RR = 1.09 (1.03 – 1.15)
NNT (1 or more miscarriages) = 20
NNT (3 or more miscarriages) = 8

IMSR INSTITUTE OF METABOLISM AND SYSTEMS RESEARCH

FUNDED BY **NIHR** | National Institute for Health Research

UNIVERSITY OF BIRMINGHAM
BCTU Birmingham Clinical Trials Unit

Tommy's @Birmingham National Centre for Miscarriage Research

💬   ⟲ 4   ♡ 8   ✉

Progesterone Summary

- With progesterone: 1513/2025 (75%) live births; with placebo, 1459/2013 (72%) live births.

- P=.08

- How much *evidence* do the data provide for the hypothesis that progesterone is helpful?

Evidence is what causes beliefs to change and so evidence is measured by change in belief

Evans

Measuring Statistical Evidence Using Relative Belief

Michael Evans

Bayesian Reanalysis

◆ Logistic regression with treatment vs. placebo coded as +0.5 / –0.5:

$$\log\left(\frac{p_1}{1-p_1}\right) = \beta - \frac{\psi}{2}$$

$$\log\left(\frac{p_2}{1-p_2}\right) = \beta + \frac{\psi}{2}$$

$$y_1 \sim \text{Binomial}(n_1, p_1)$$

$$y_2 \sim \text{Binomial}(n_2, p_2).$$

Bayesian Reanalysis

- An uninformative prior on the grand mean, a test-irrelevant nuisance parameter.

- Under H1, a weakly informative prior on the log odds ratio, the test-relevant prior that quantifies the (anticipated) treatment effect.

Approximate Bayes Factors and Orthogonal Parameters, with Application to Testing Equality of Two Binomial Proportions

By ROBERT E. KASS† and SURESH K. VAIDYANATHAN

Approximate Bayes Factors and Orthogonal Parameters, with Application to Testing Equality of Two Binomial Proportions

By ROBERT E. KASS† and SURESH K. VAIDYANATHAN

Implemented (with some bells and whistles) in the CRAN R package "abtest" and also in JASP

BayesianSpectacles

Powered by JASP

Informed Bayesian Inference for the A/B Test

POSTED ON MAY 16TH, 2019

This post is an extended synopsis of a preprint that is available on arXiv: http://arxiv.org/abs/1905.02068

Abstract

Booming in business and a staple analysis in medical trials, the A/B test assesses the effect of an intervention or treatment by comparing its success rate

Informed Bayesian Inference for the A/B Test

Quentin F. Gronau
University of Amsterdam
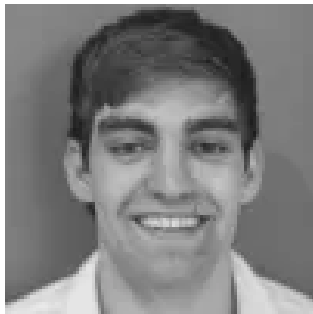
Akash Raj K. N.
University of Amsterdam

Eric-Jan Wagenmakers
University of Amsterdam

Bayesian Reanalysis: Summary I

- Ignoring H0, there is good evidence that the effect is helpful rather than harmful.

- Under the default one-sided prior, H0 outpredicts H+ by a factor of 2. This is very weak evidence ("not worth more than a bare mention") that the effect is absent rather than helpful.
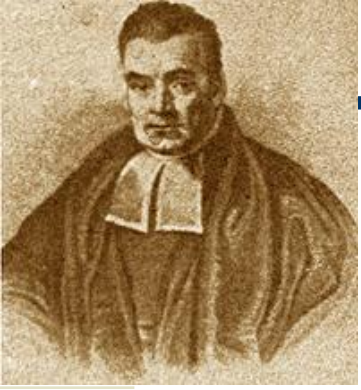
Bayesian Reanalysis: Summary II

- Priors more consistent with the observed data do not change the qualitative pattern: there is *absence of evidence*, not evidence of absence or evidence of presence.

- More data can be added and learning can continue.

A Fresh Way to Learn Bayesian Statistics

August 22 & 23, 2019
University of Amsterdam

Eighth Annual JAGS and WinBUGS Workshop
Bayesian Modeling for Cognitive Science
http://bayescourse.socsci.uva.nl/

artwork: instagram.com/viktordepictor
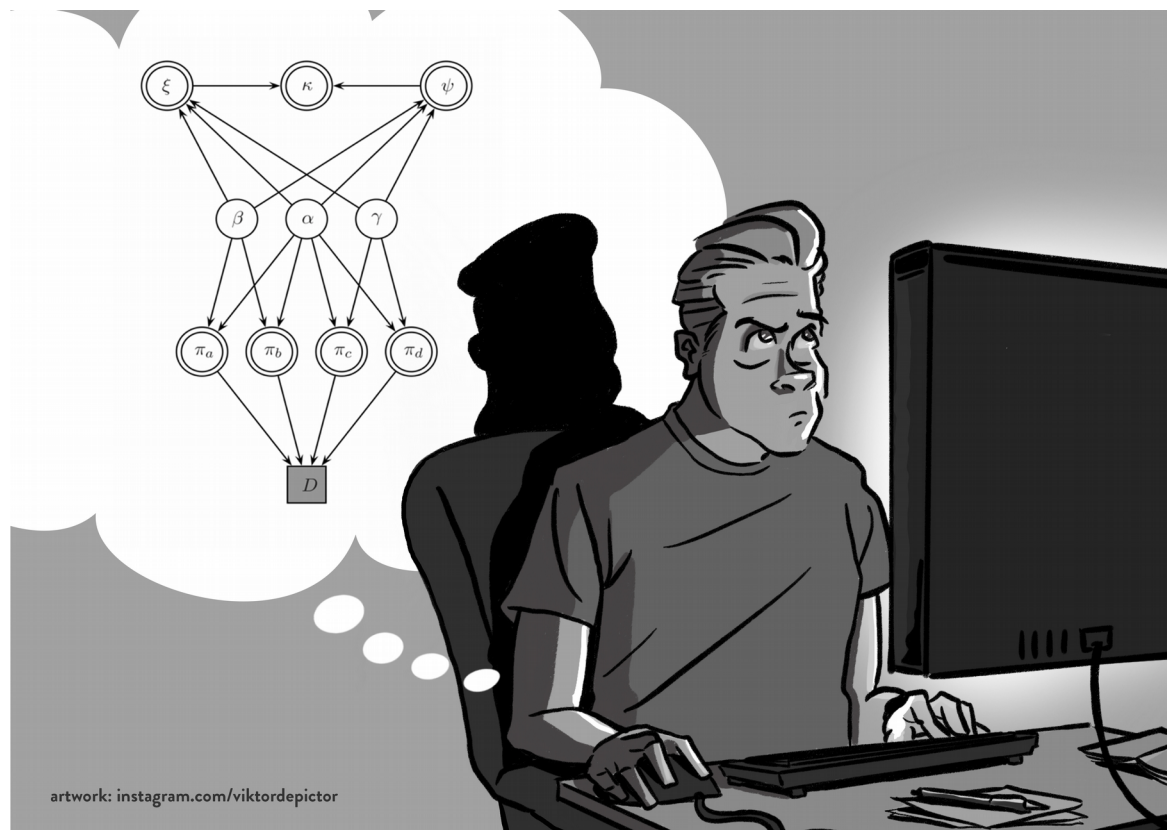
August 26-30, 2019
University of Amsterdam
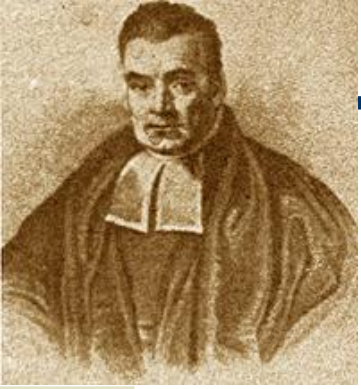
Examples

- Example 1: Does progesterone prevent miscarriages? (NEJM, this month)

- Example 2: Are movies with Adam Sandler profitable regardless of their quality?

- Example 3 [if time permits]: Do men with bigger balls neglect their children more? (PNAS, 2013)

AWESOME-O

- South park episode 166.
- Eric Cartman pretends to be a robot, the A.W.E.S.O.M.-O 4000.
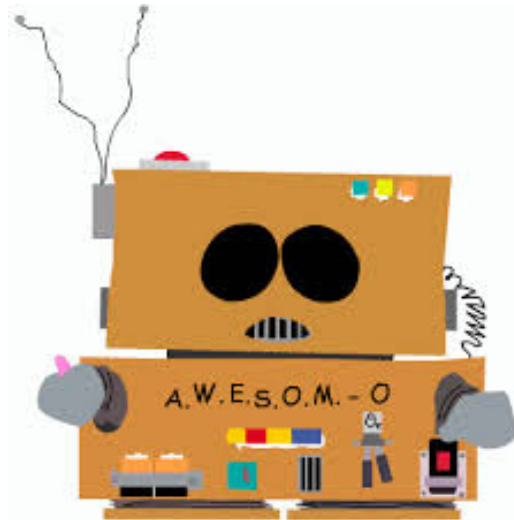
AWESOME-O

- Hollywood movie-producers kidnap the robot and force it to generate profitable movie ideas.

- The A.W.E.S.O.M.-O 4000 generates more than 2,000 silly movie ideas, 800 of which star Adam Sandler.
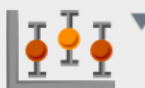
South Park Hypothesis (Implied)

- General: "Adam Sandler movies are profitable regardless of their quality"
- Specific: "For Adam Sandler movies, box office success does not correlate with freshness ratings on Rotten Tomatoes"

Descriptives | **T-Tests** | **ANOVA** | **Regression** | **Frequencies** | **Factor**

	Year	Freshness	Box Office ($M)	Movie Title
1	2000	0.22	38.5	Little Nicky
2	2001	0.3	54.4	The Animal
3	2002	0.22	126.2	Mr. Deeds
4	2002	0.01	40.3	The Master of Disguise
5	2002	0.21	34.9	The Hot Chick
6	2002	0.79	17.8	Punch-Drunk Love
7	2002	0.12	23.3	Adam Sandler's Eight Crazy Nights
8	2003	0.43	133.8	Anger Management
9	2003	0.23	22.7	Dickie Roberts: Former Child Star
10	2004	0.53	42	Spanglish
11	2004	0.44	120.8	50 First Dates
12	2005	0.09	22.3	Deuce Bigalow: European Gigolo
13	2005	0.31	158	The Longest Yard
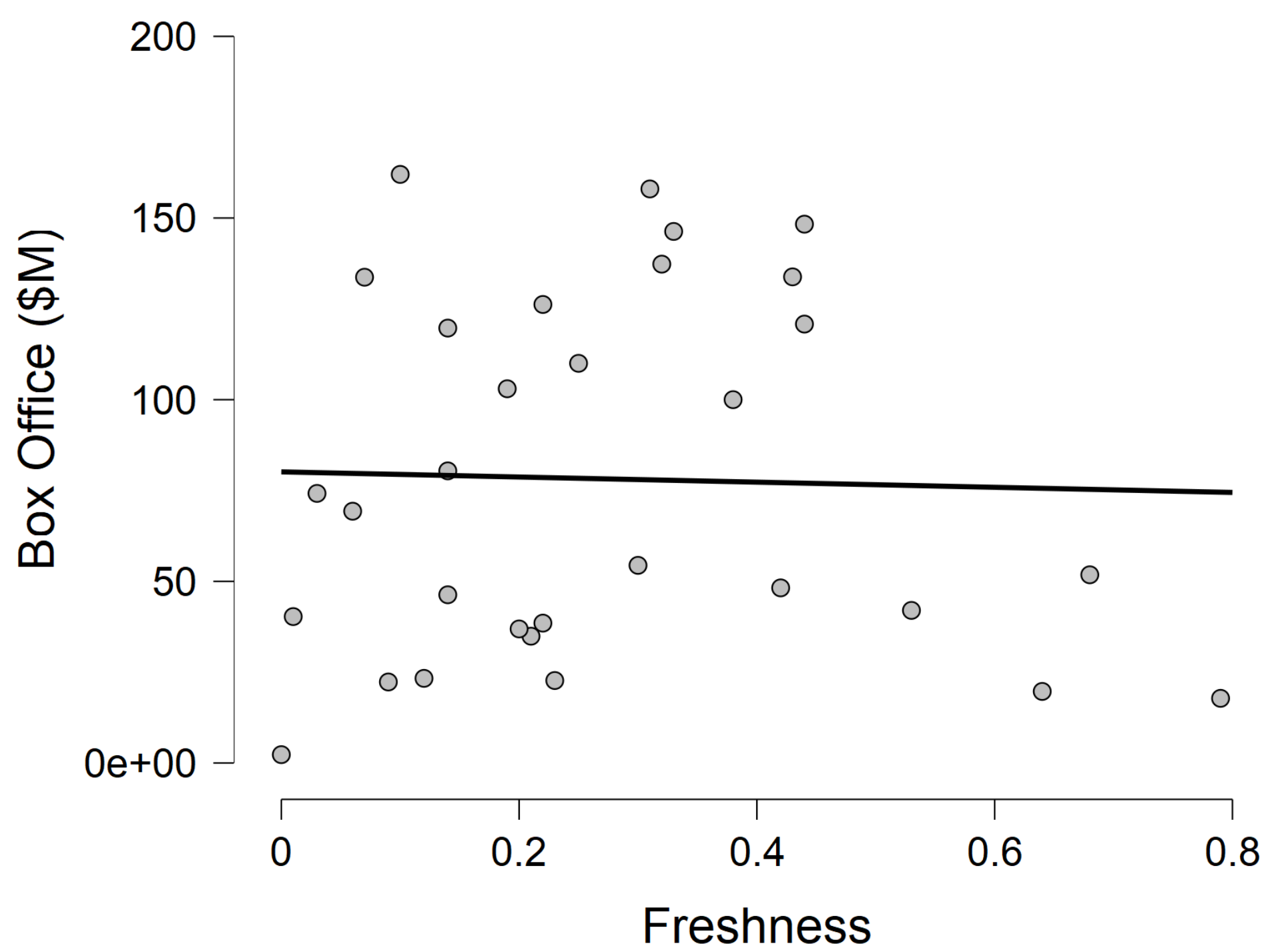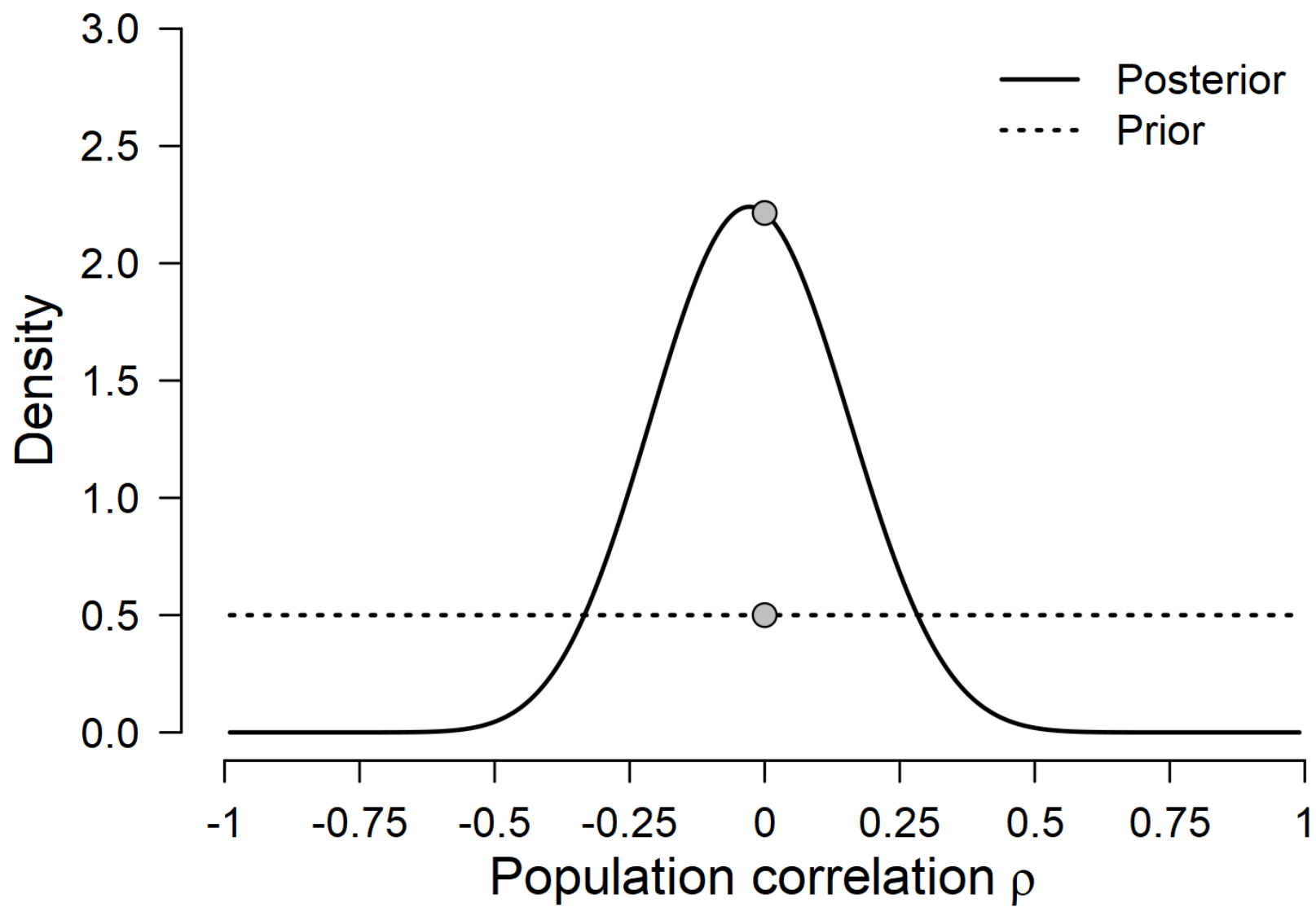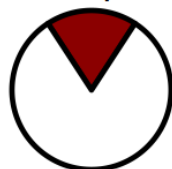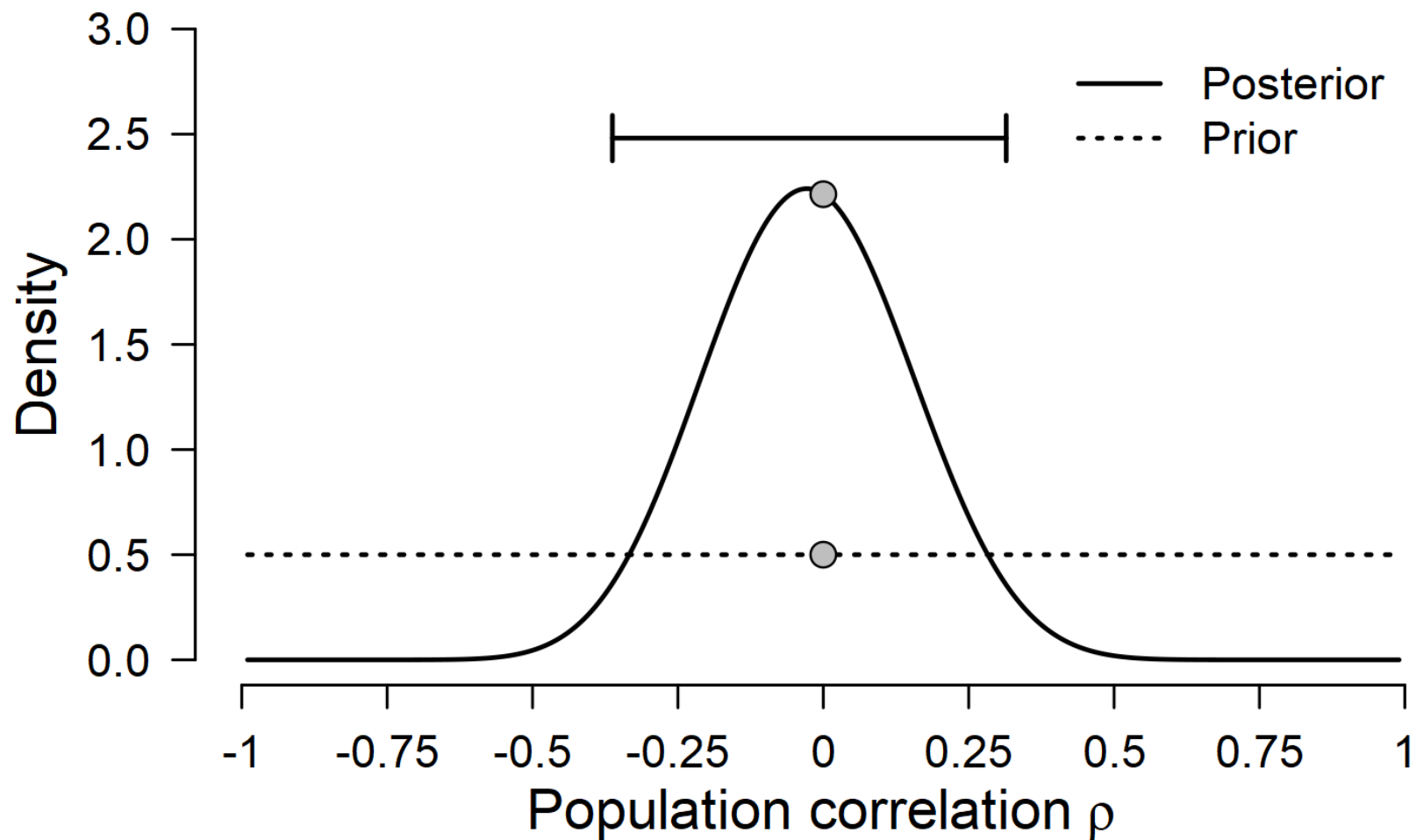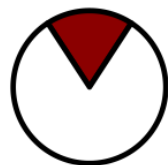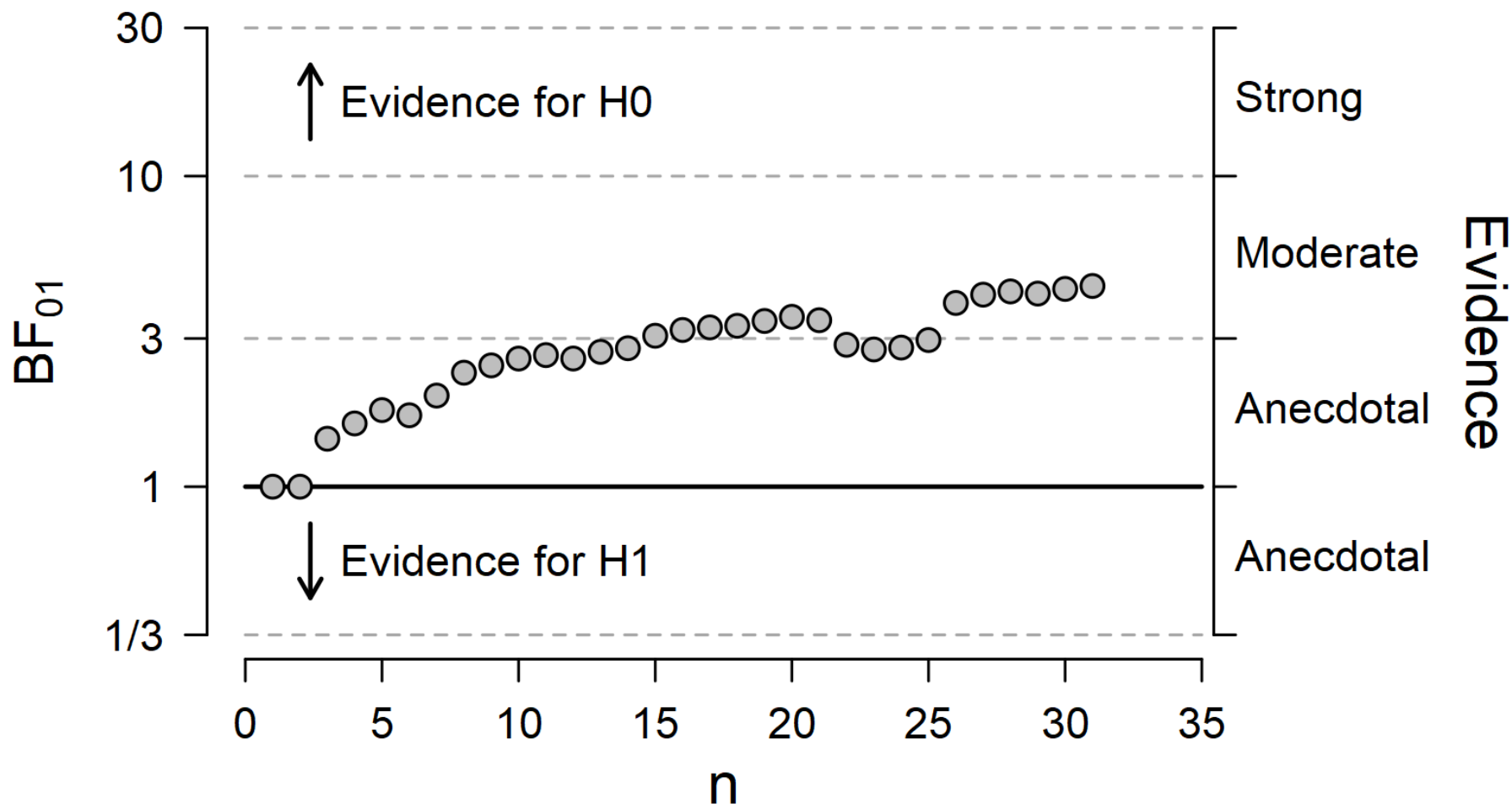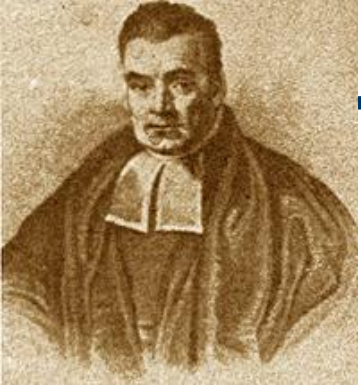
Bayesian Analysis of the South Park Hypothesis

- The south park hypothesis receives support from the data, albeit only modestly.

- Evidence can be quantified in support of H0.

- Evidence may be monitored as the data accumulate (aka "learning").

Examples

- Example 1: Does progesterone prevent miscarriages? (NEJM, this month)
- Example 2: Are movies with Adam Sandler profitable regardless of their quality?
- Example 3 [if time permits]: Do men with bigger balls neglect their children more? (PNAS, 2013)

Testicular volume is inversely correlated with nurturing-related brain activity in human fathers

Jennifer S. Mascaro[a,b,c], Patrick D. Hackett[a], and James K. Rilling[a,b,c,d,1]

Testicular volume is inversely correlated with nurturing-related brain activity in human fathers

Jennifer S. Mascaro[a,b,c], Patrick D. Hackett[a], and James K. Rilling[a,b,c,d,1]

Hypothesis [paraphrased]:

Men with Big Balls are so busy spreading their semen that they lack the time to raise their children.

Results

Reproductive Biology and Parenting Behavior. Although testes volume was not related to body mass, there was a significant linear correlation between testes volume and height [$r(53) = 0.27, P < 0.05$]. Therefore, residual testes volume, controlling for height, was used in subsequent analyses. Residual testes volume was negatively related to paternal caregiving [$r(52) = -0.29, P < 0.05$]

Discussion

Collectively, these data provide the most direct support to date that the biology of human males reflects a trade-off between mating and parenting effort. Fathers' testicular volume and testosterone levels were inversely related to parental investment
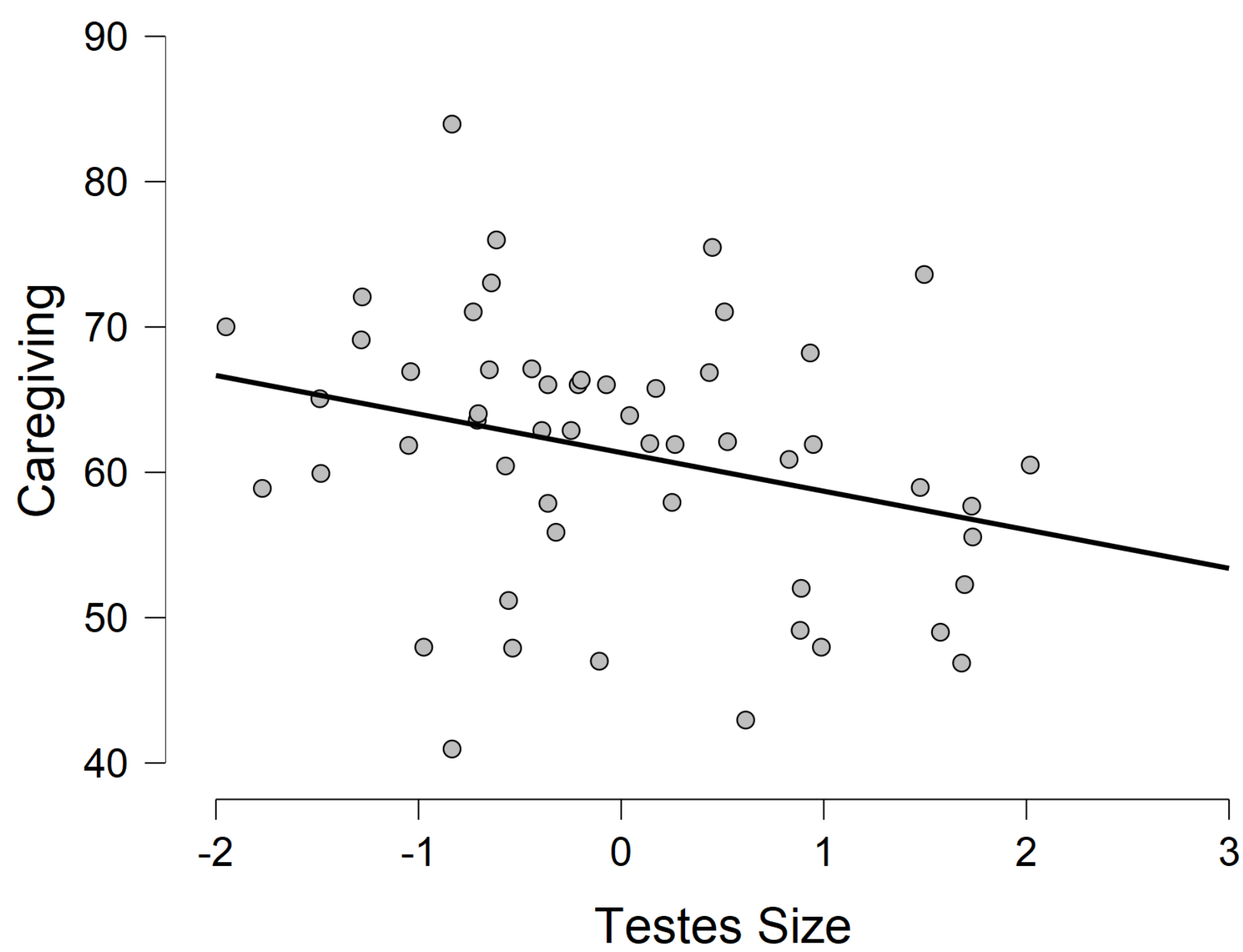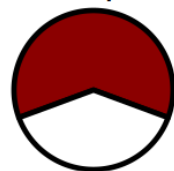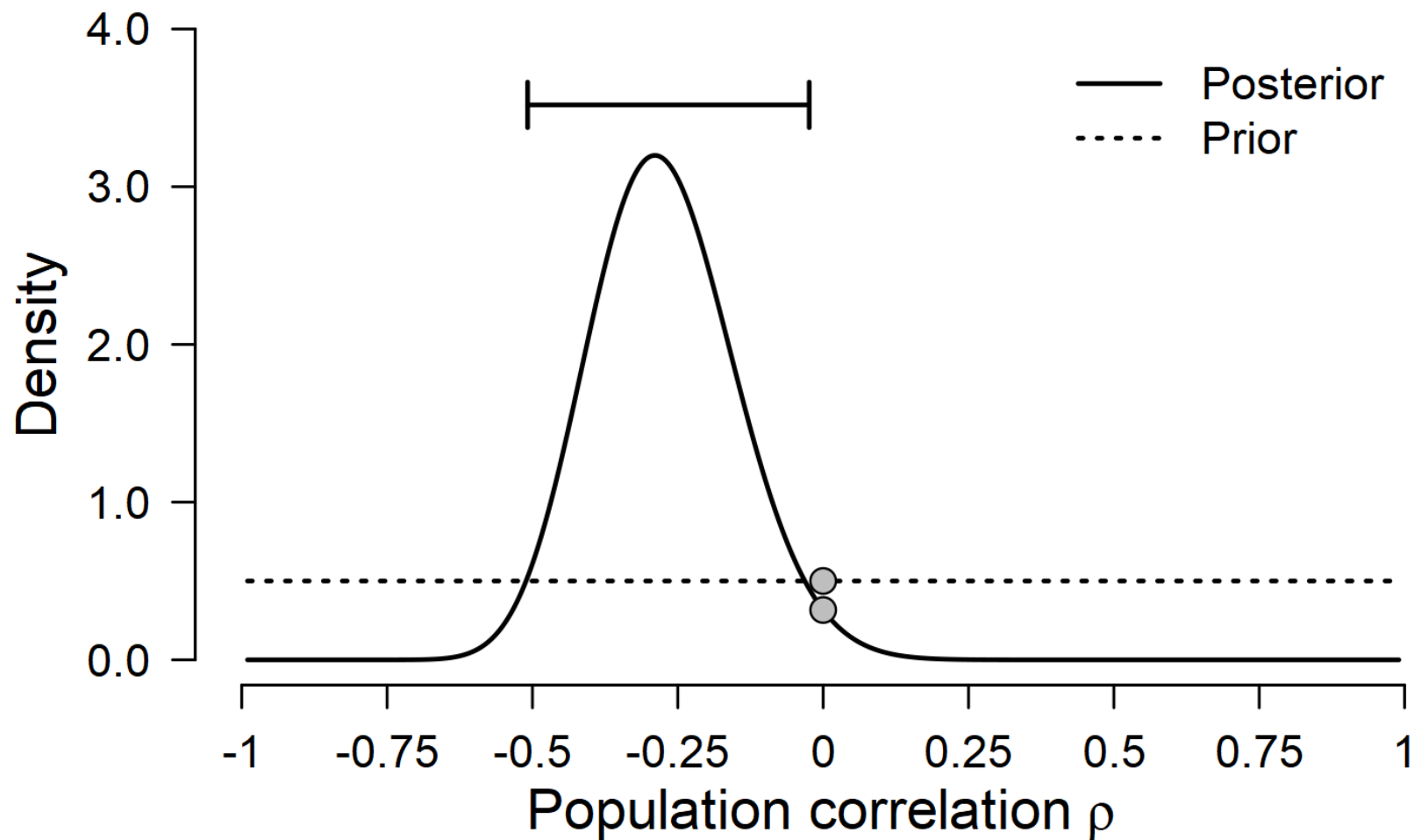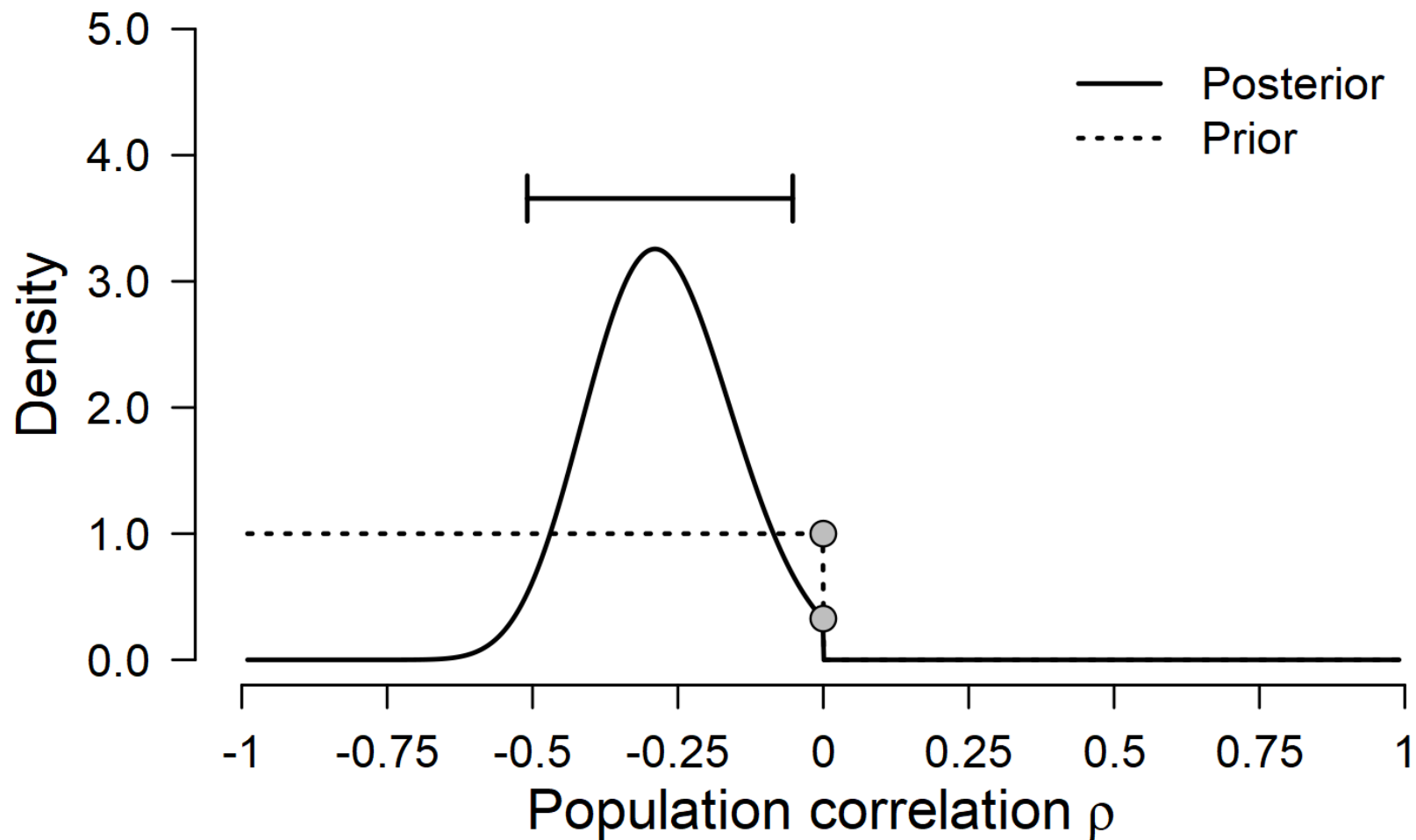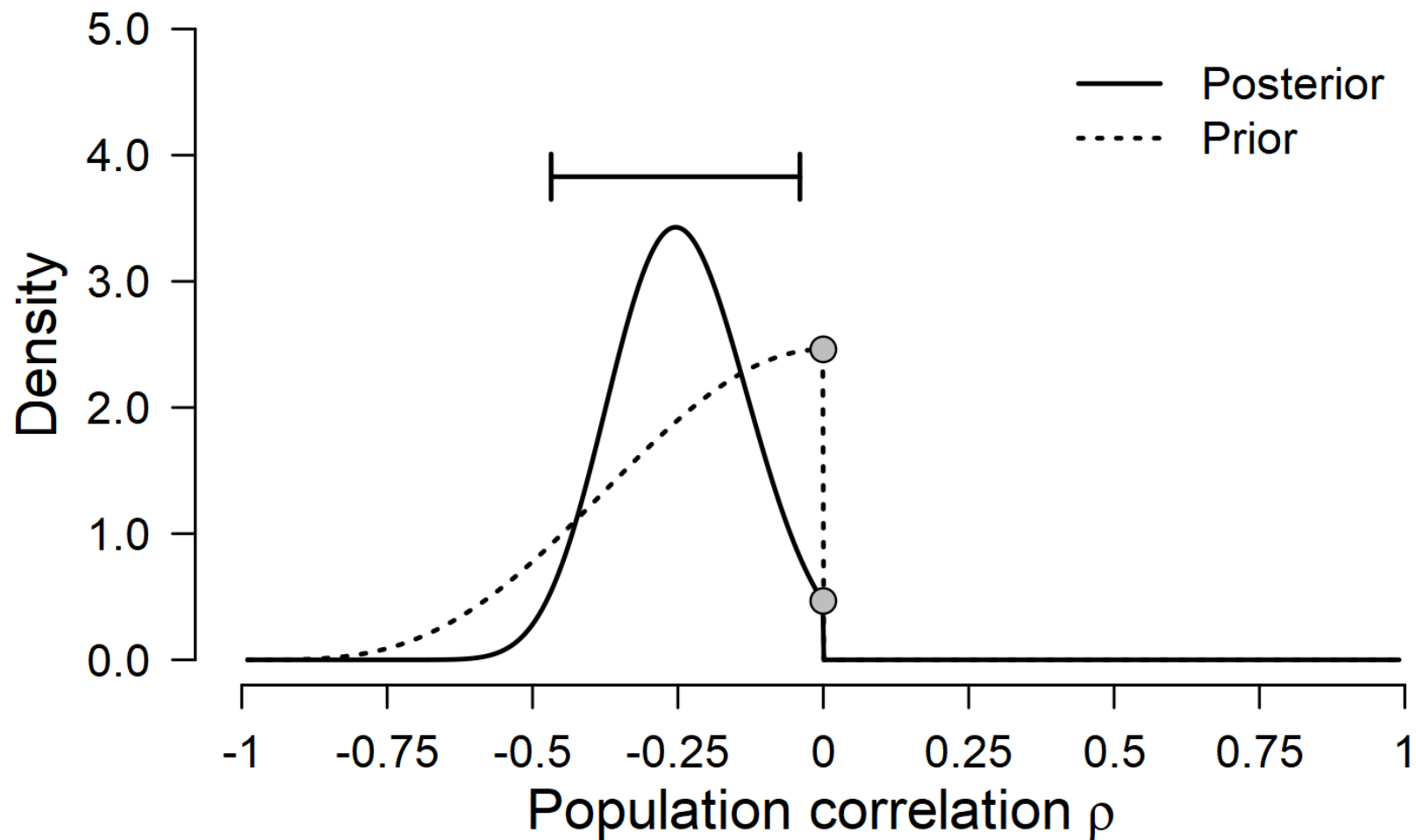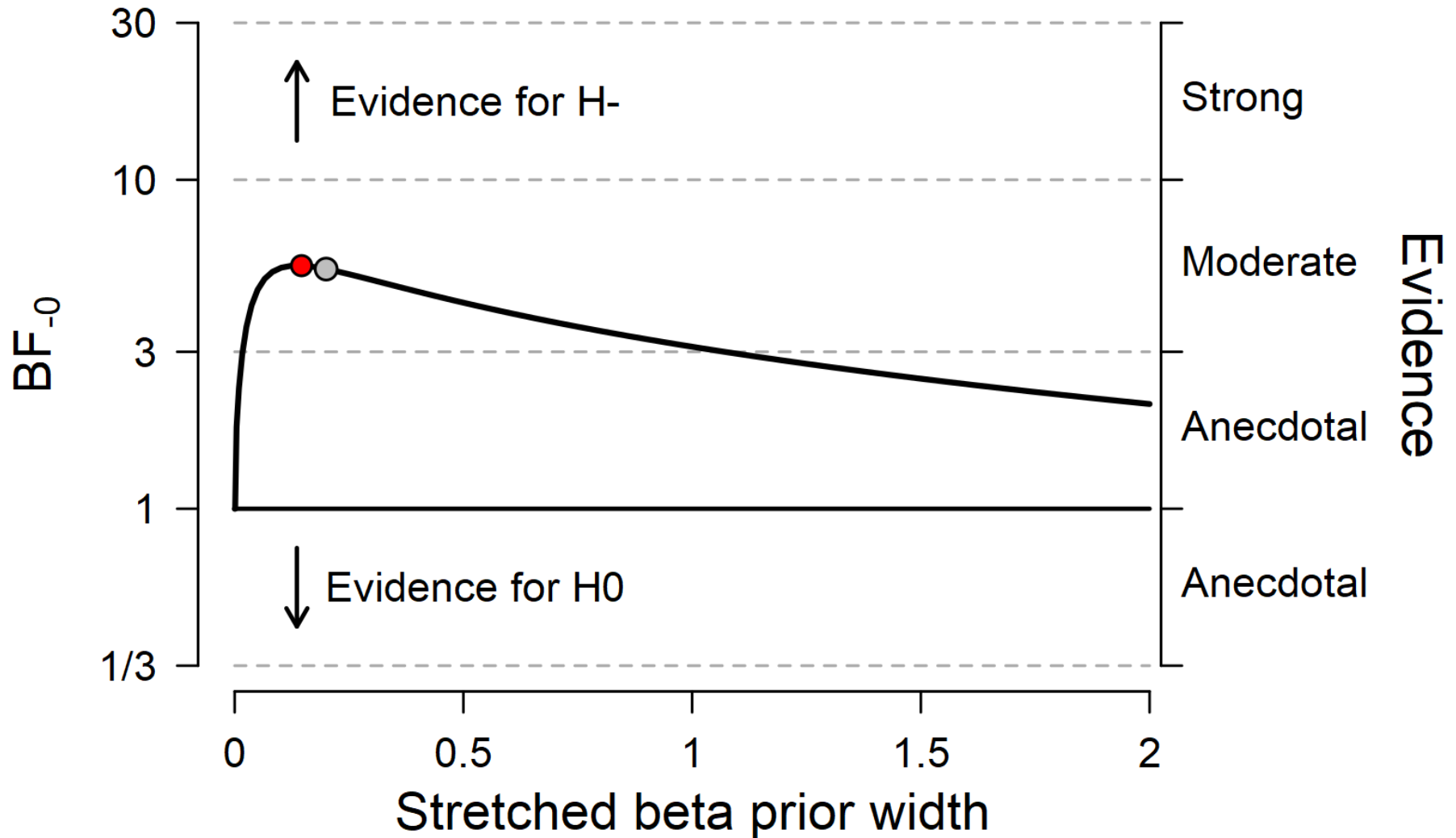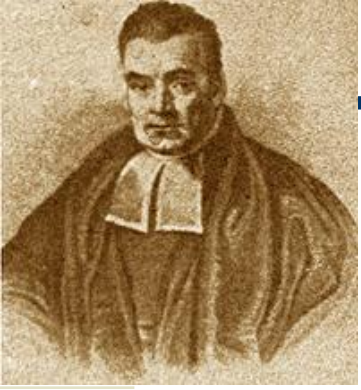
Pragmatic Bayesian Advantages

- ◆ Attach probabilities to parameters and hypotheses;

- ◆ Quantify evidence, for any hypothesis you care to specify;

- ◆ Monitor evidential flow as the data come in;

- ◆ Assess sensitivity of the conclusion to alternative prior specification.

Thanks for your Attention!