Likelihood and its Discontents

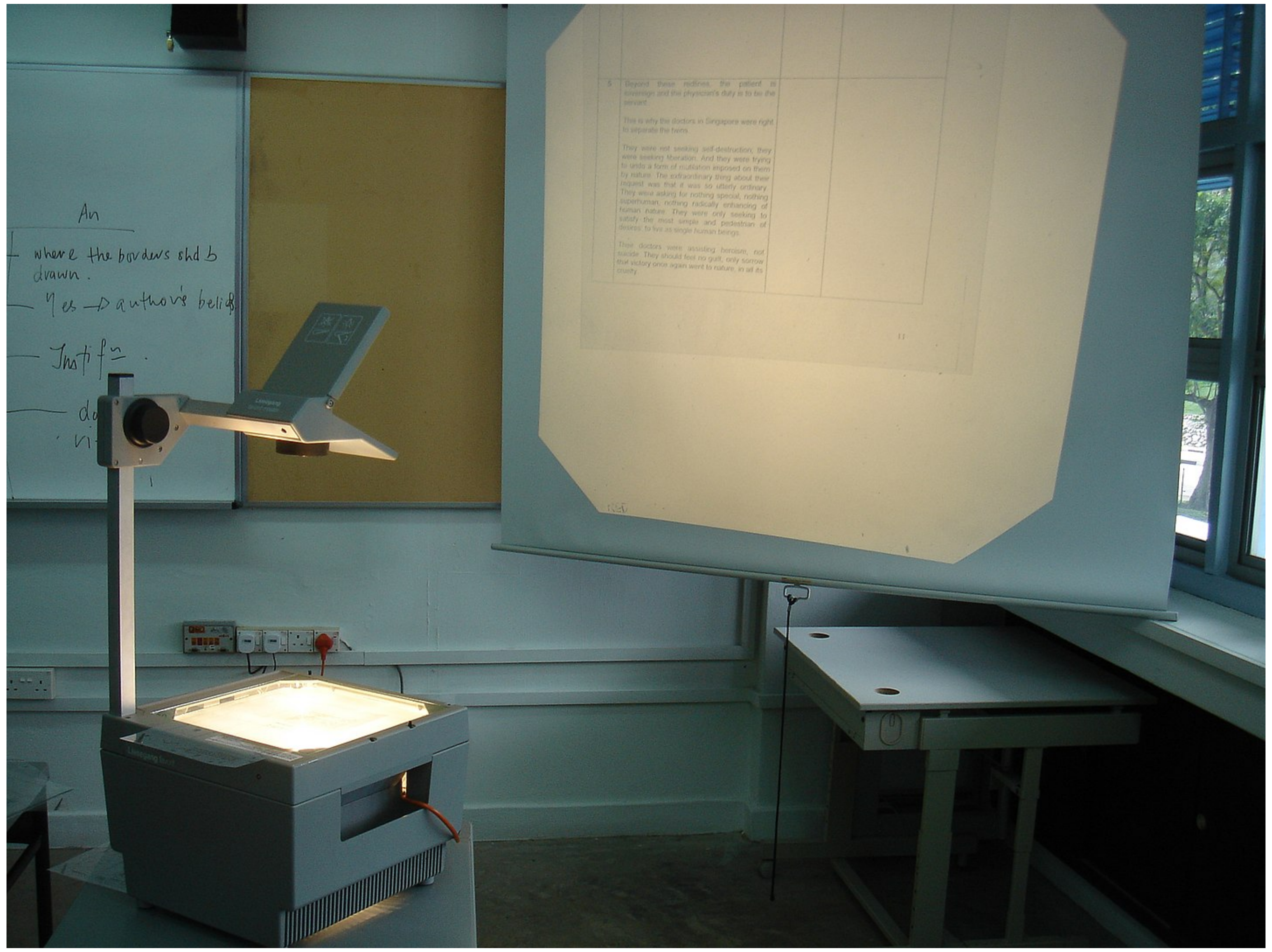
Nancy Reid
University of Toronto

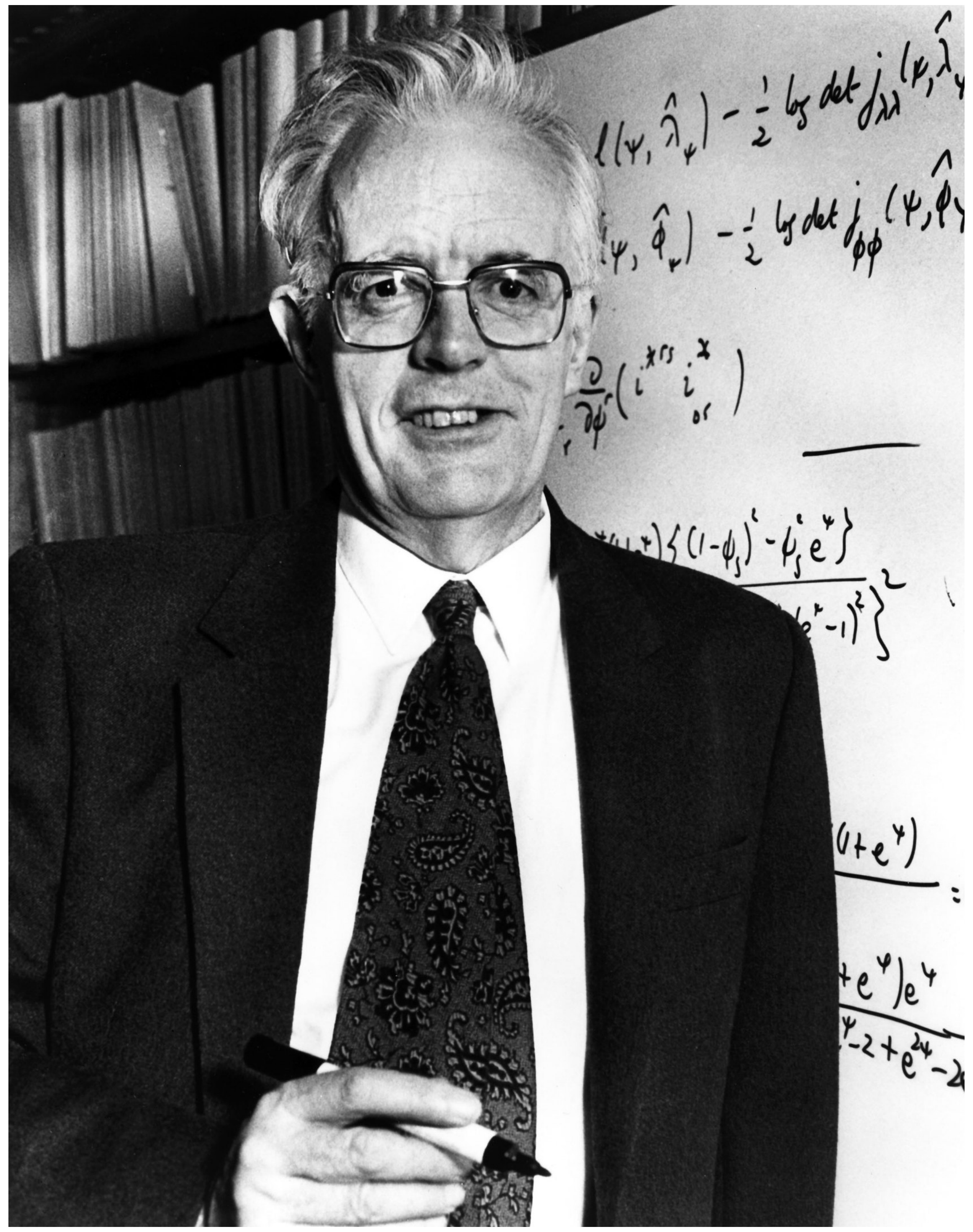
August 10, 2022

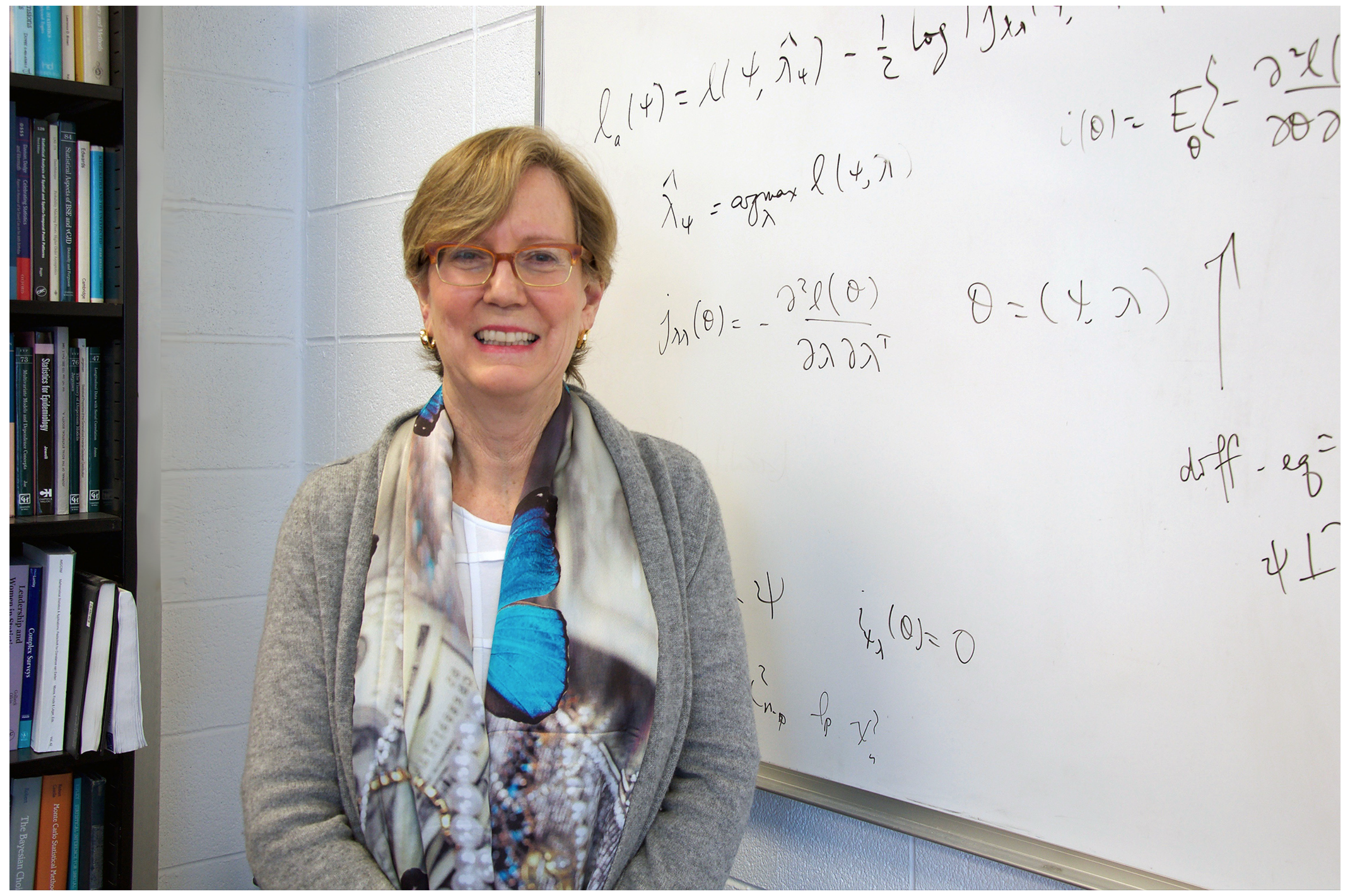
AUGUST 6–11
2022
WASHINGTON, DC
JSM
STATISTICS: A FOUNDATION FOR INNOVATION

JSM 1989

Probability models:
their role in statistical analysis

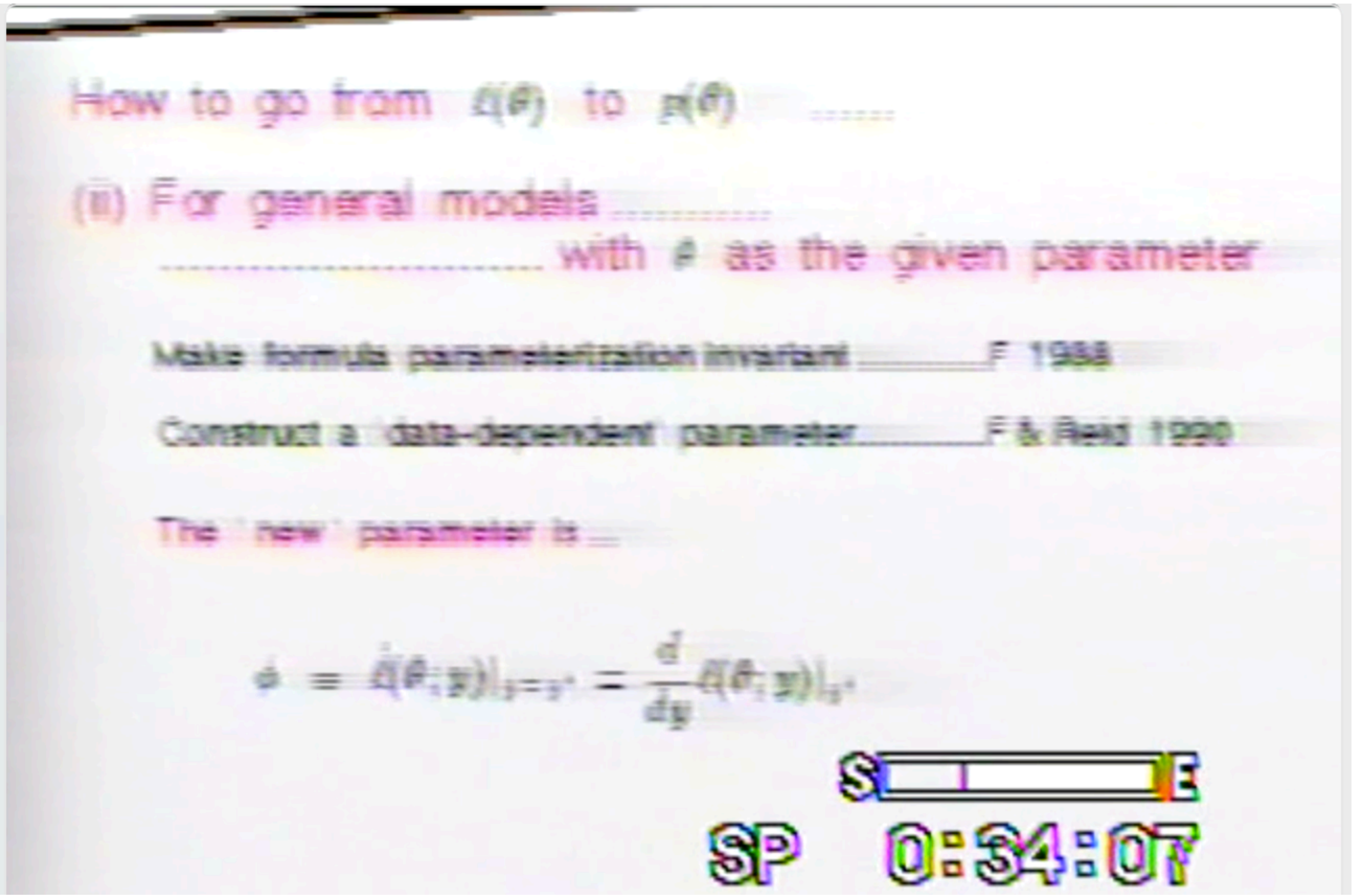


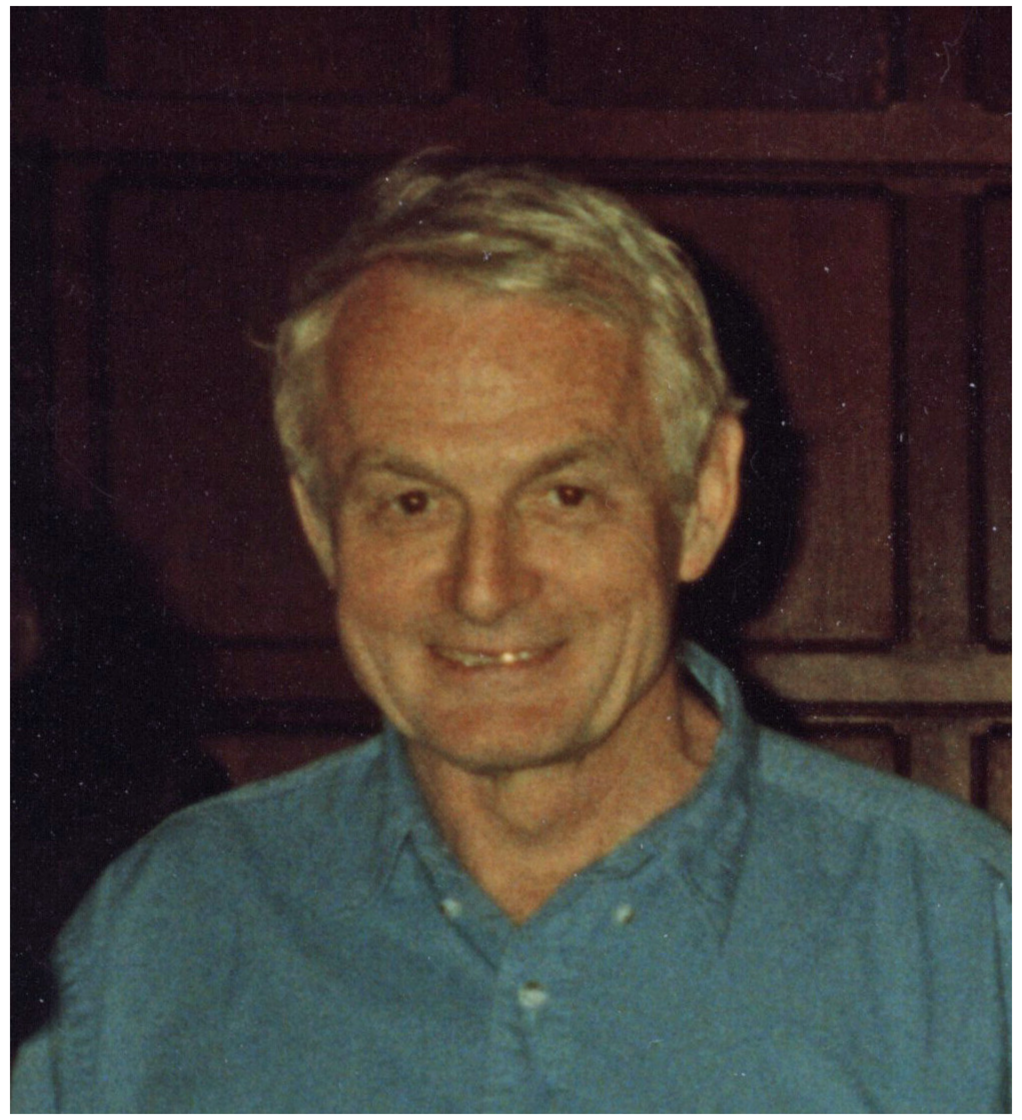




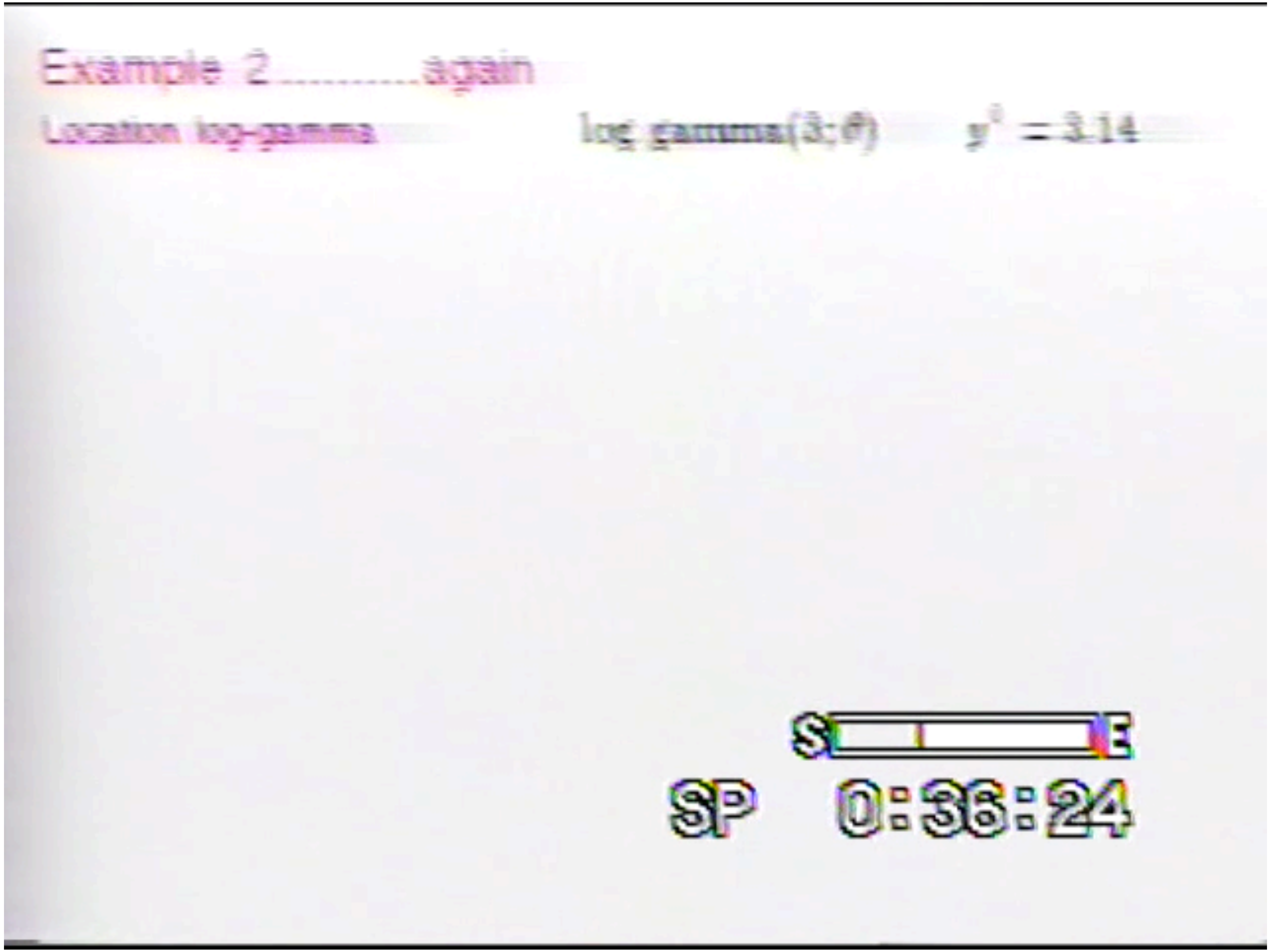
JSM 1990

From likelihood to significance

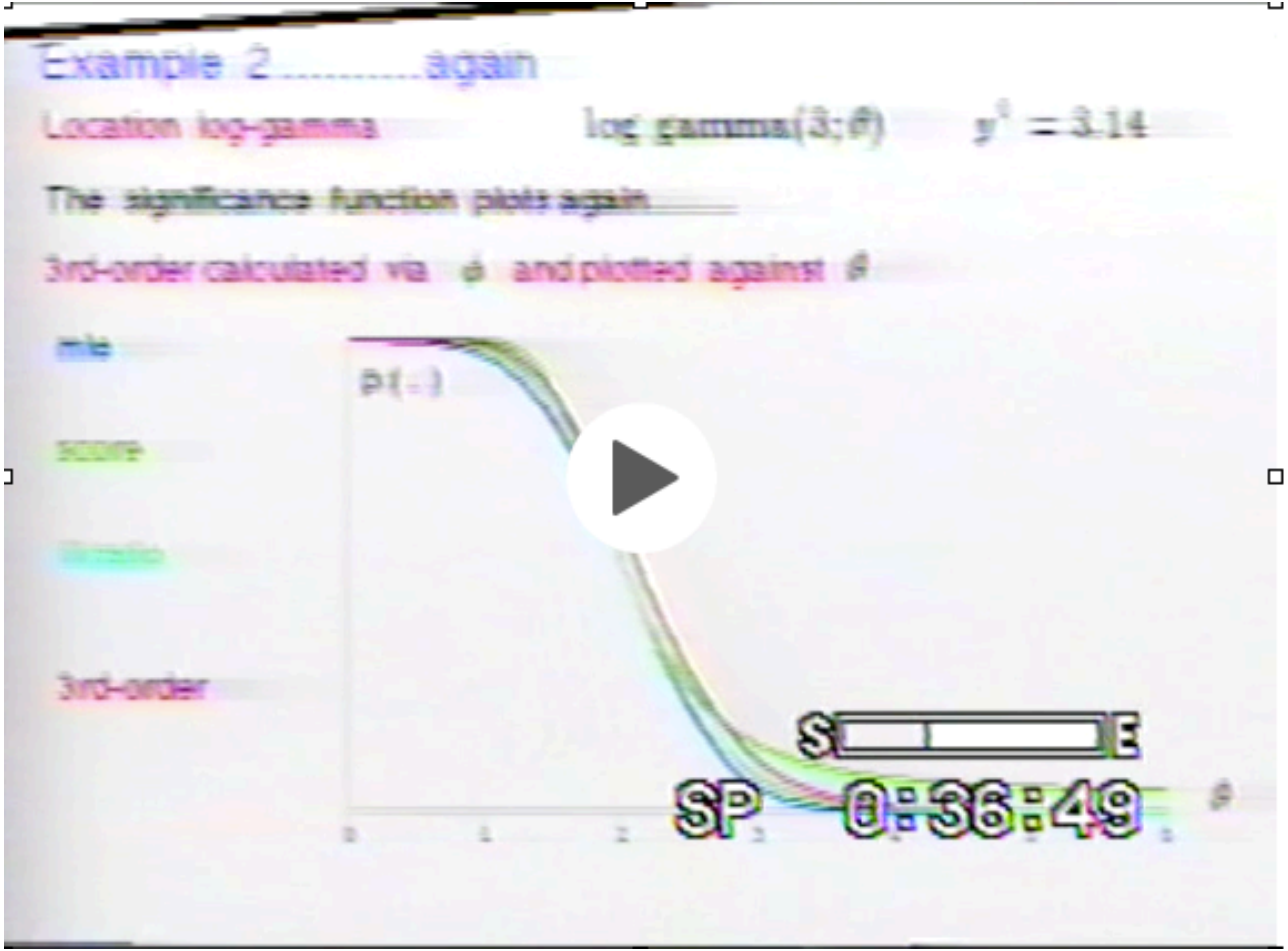




JSM 1990



JSM 1990


Example 2 ....again
Location log-gamma
The significance function plots again
mle
score
3rd-order
SP 0:36:49

Outline

1. Likelihood inference
2. The discontents
3. Theory and applications

1. Likelihood inference

Some notation

Model $y \sim f(\,\cdot\,;\theta), \quad \theta \in \Theta$

Data $y_1, \ldots, y_n$ independent

Likelihood function $L(\theta; \underline{y}) \propto f(\underline{y}; \theta) = \Pi_{i=1}^{n} f(y_i; \theta)$

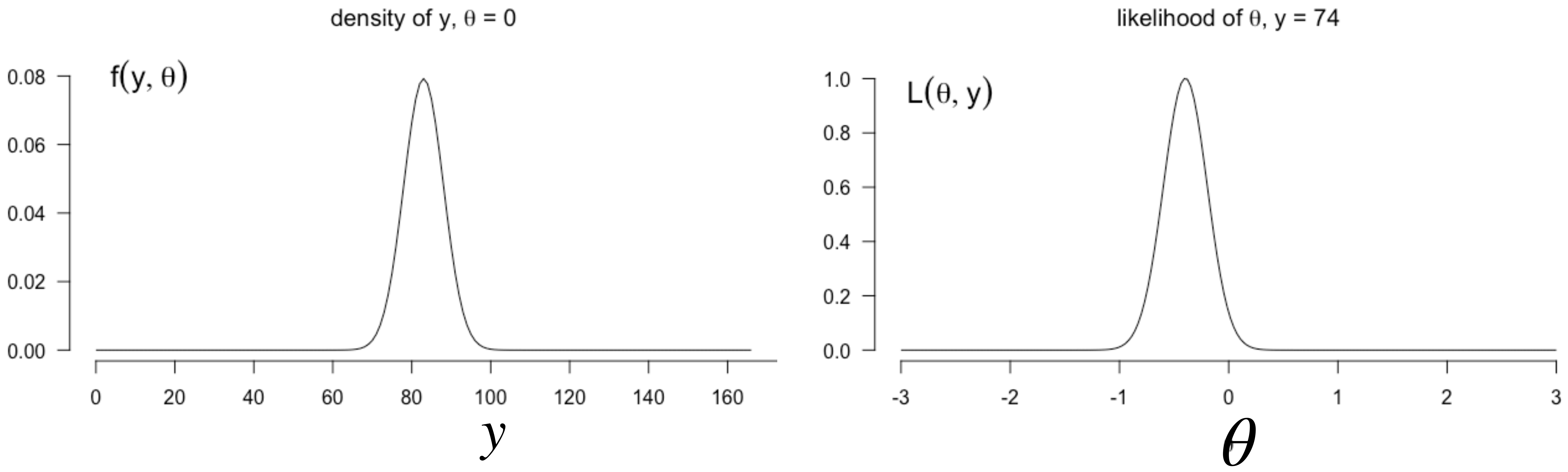
Log-likelihood function $\ell(\theta; y) = \log L(\theta; y) = \Sigma_{i=1}^{n} \log f(y_i; \theta)$

Maximum Likelihood Estimator $\hat{\theta} = \arg\sup \ell(\theta; y)$

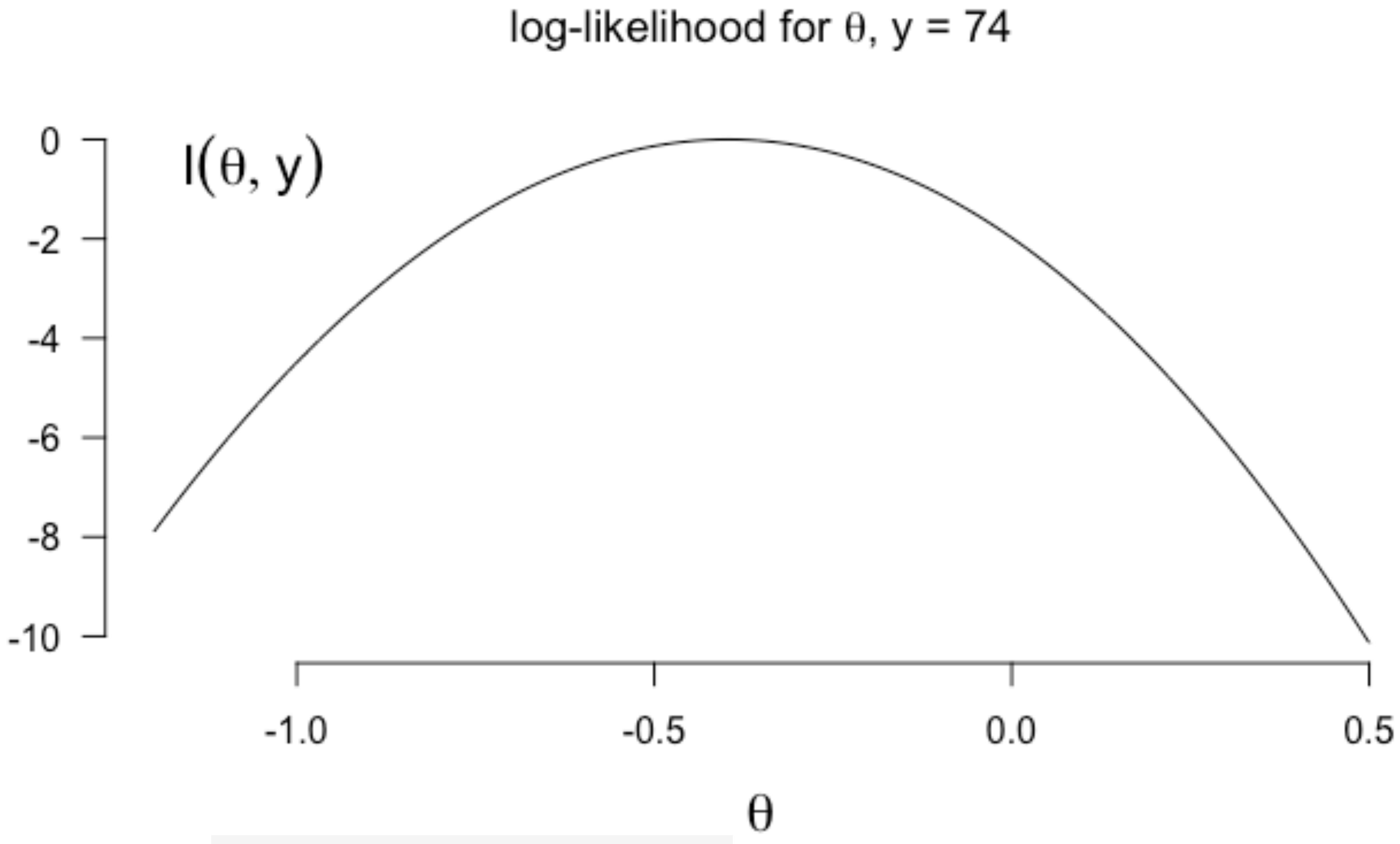
Profile likelihood function $L_{prof}(\psi) = L(\psi, \widehat{\lambda}_{\psi}) \quad \theta = (\psi, \lambda)$ Constrained max. lik. est.

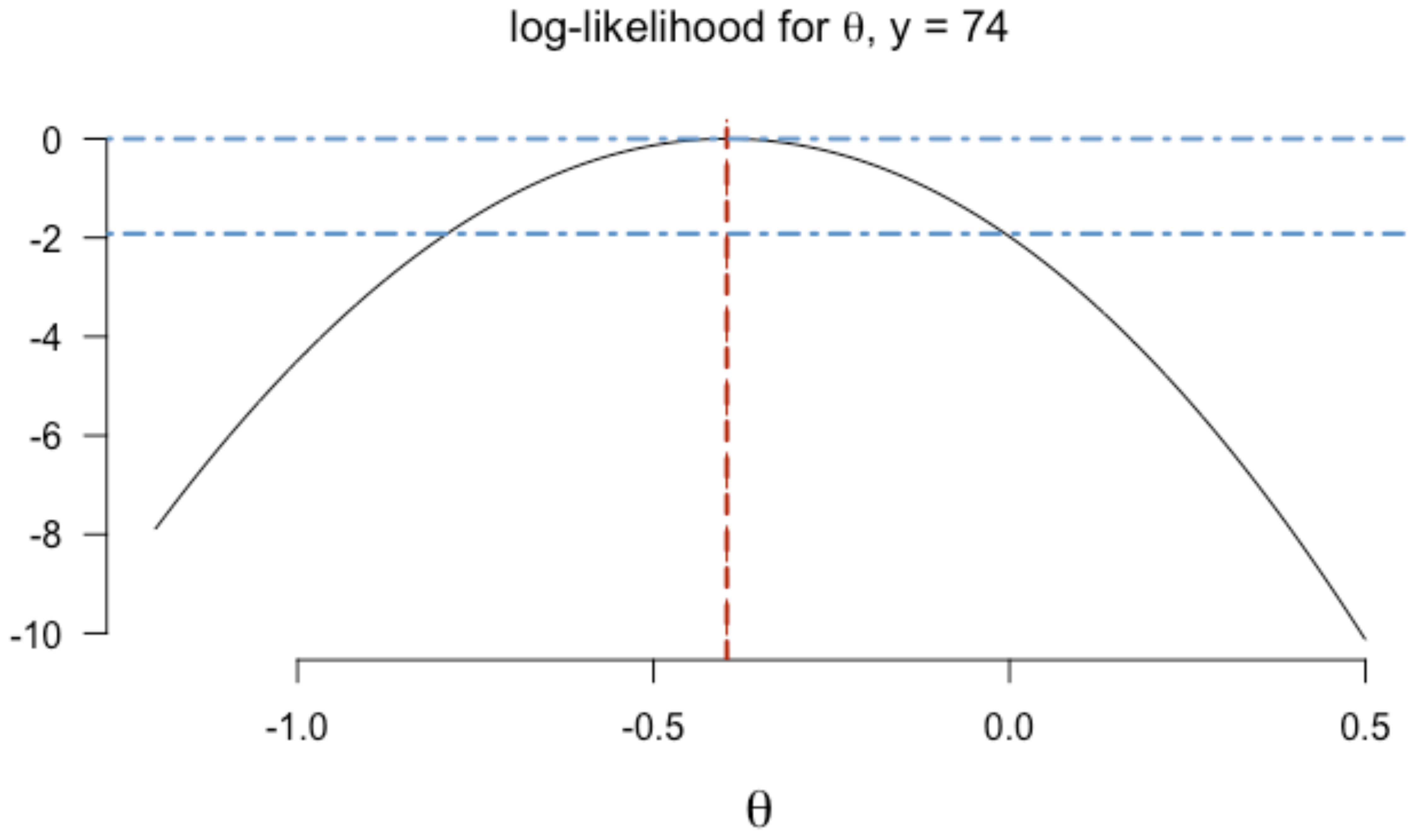
The likelihood function

Fisher 1922



Likelihood inference





	Died	Lived	
New	74	138	212
Old	92	120	212
Total	166	258	424

2-sided *p*-value = 0.07

likelihood ratio test
no adjustment for covariates

$$\hat{\theta} \pm 1.96\, \hat{\sigma}_\theta$$

$$(-0.75, +0.03)$$

$$2\{\ell(\hat{\theta}) - \ell(\theta)\} = 3.84$$

$$(-0.72, +0.03)$$

Profile log-likelihood: $2\{\ell_{prof}(\hat{\psi}) - \ell_{prof}(\psi)\} = 3.84$

Haphazard selection

Recent literature

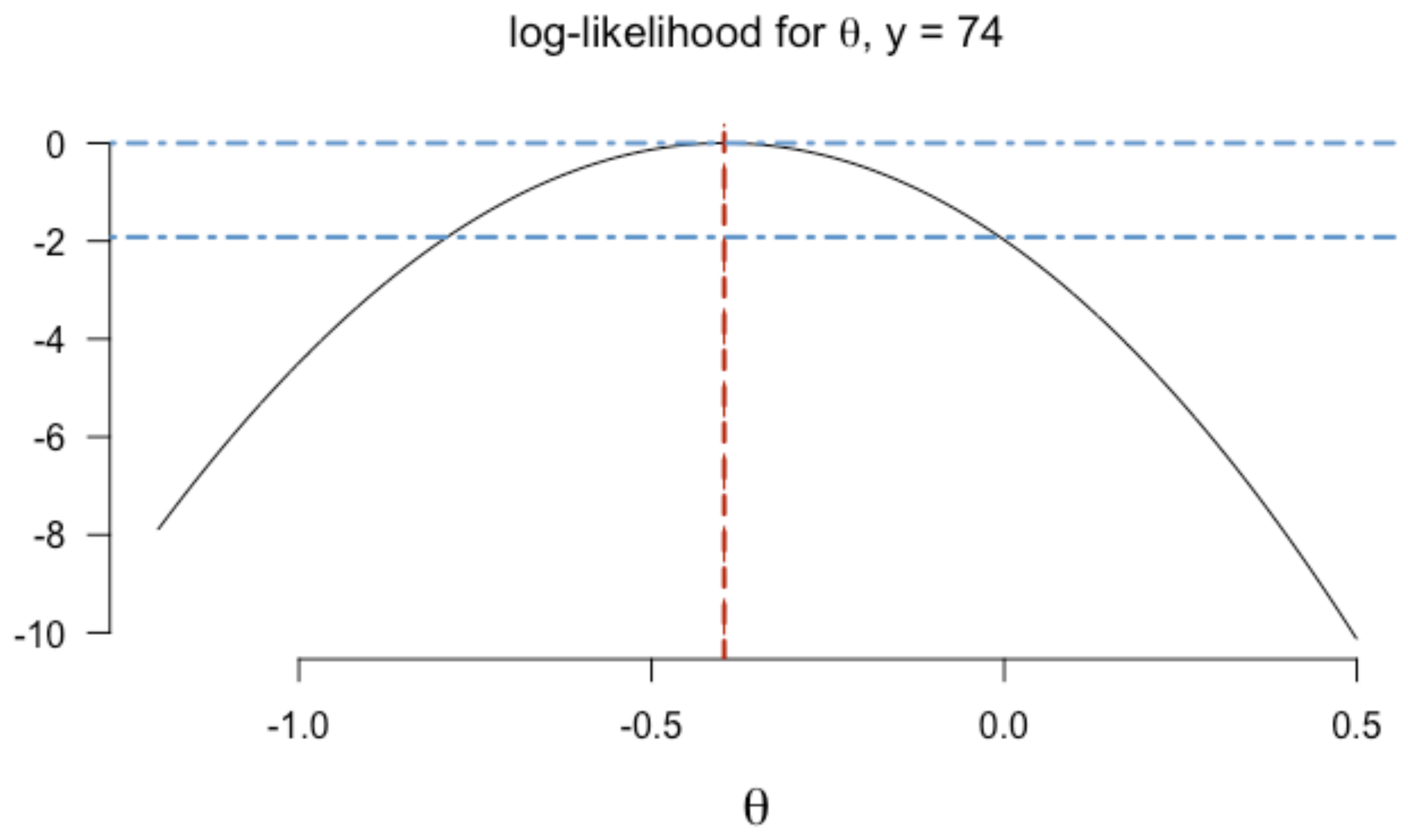
- "Likelihood-based inference for partially observed epidemics …" Bu et al *JASA* 2022
- "Likelihood-based bacterial identification approach …" Ryu *AoAS* 2022
- "Likelihood-based model selection for stochastic block models" Wang & Bickel *AoS* 2022
- "Graphical models for extremes" Engelke & Hitz *JRSS B* 2022
- ``General maximum likelihood empirical Bayes …" Jiang & Zhang *AoS* 2010

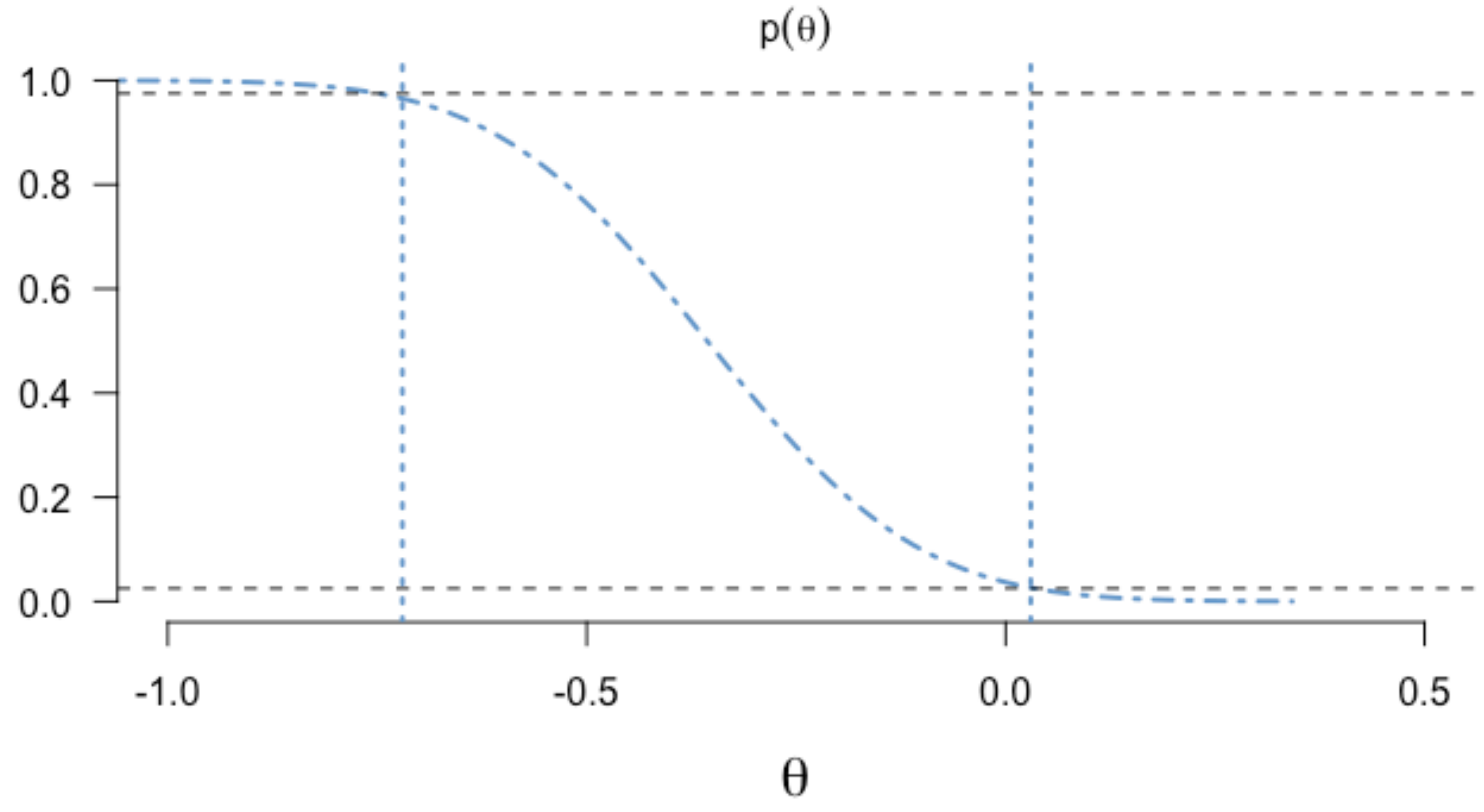
Why so useful?

- **Puts modelling first** $L(\theta; y) \propto f(y; \theta), \quad \theta \in \Theta$
- Provides **reliable** summary measures
  - maximum likelihood estimate, likelihood ratio test
- Can be converted to a probability, using Bayesian arguments
- Can be penalized to encourage variable selection or avoid over-fitting With a prior
- Can be converted to a significance function, using asymptotic theory Lasso +

Converting likelihood to significance

- Limit theory $q(\theta) = (\hat{\theta} - \theta) j^{1/2}(\hat{\theta}) \to N(0,1)$ $r(\theta) = \pm [2\{\ell(\hat{\theta}) - \ell(\theta)\}]^{1/2} \to N(0,1)$
- Significance function $p(\theta) \approx \Phi\{q(\theta)\}$ $p(\theta) \approx \Phi\{r(\theta)\}$ normal cdf



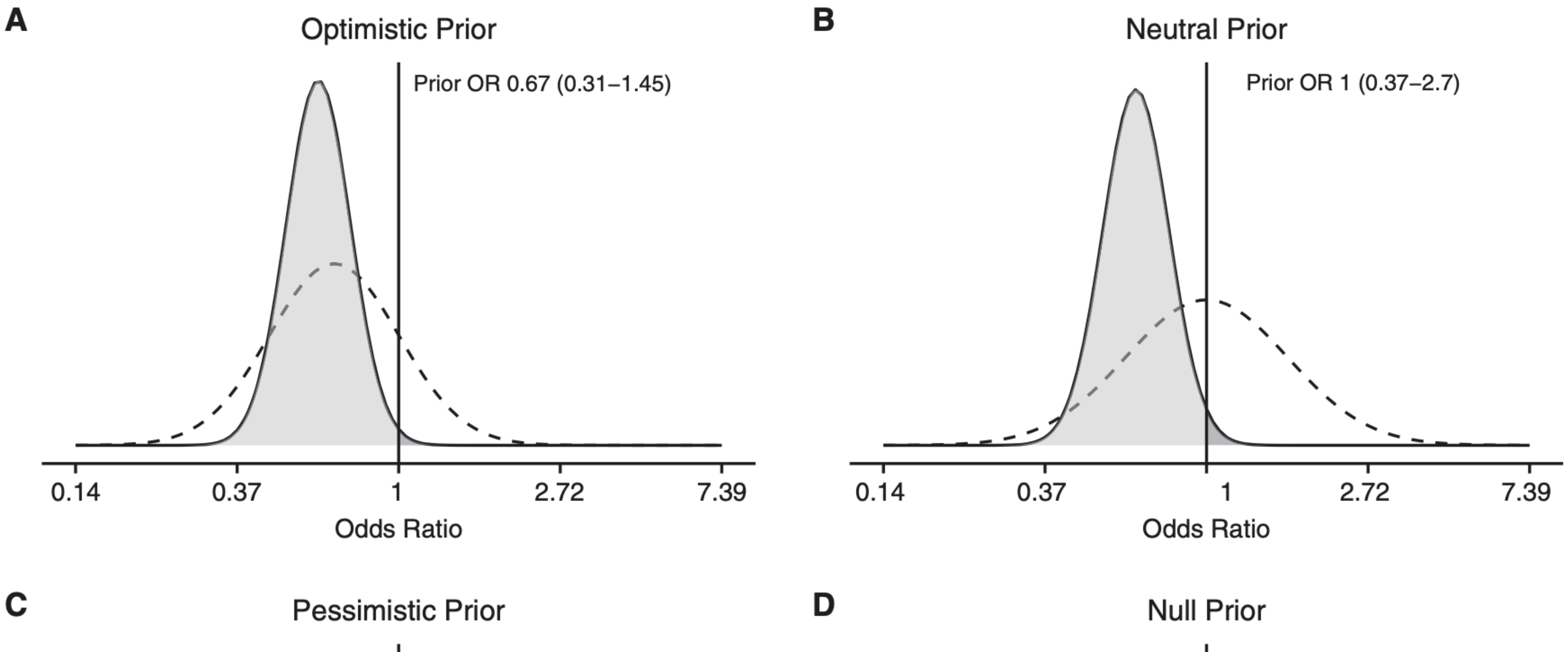


Converting likelihood to probabilities

Zampieri et al 2021 Fig 2

$$\pi(\theta \mid y) = \frac{L(\theta; y)\pi(\theta)}{\int L(\theta; y)\pi(\theta)d\theta}$$

$$\pi_m(\psi \mid y) = \int \pi(\theta \mid y)d\lambda \qquad \theta = (\psi, \lambda)$$



2. The discontents

Some challenges

- "The usual regularity conditions"
- High-dimensional parameter space
- Computational intractability
- The model is wrong
- Likelihood is not a probability

Some strategies

- New asymptotic theory
- Other forms of likelihood
- "Likelihood-like" functions
- Semi- and non-parametric approaches
- "Objective" Bayes

Some challenges

- "The usual regularity conditions"
- High-dimensional parameter space
- Computational intractability
- The model is wrong
- Likelihood is not a probability

Some strategies

- New asymptotic theory
- Other forms of likelihood
- "Likelihood-like" functions
- Semi- and non-parametric approaches
- "Objective" Bayes

Regularity conditions

Brazzale & Mamelli 2022

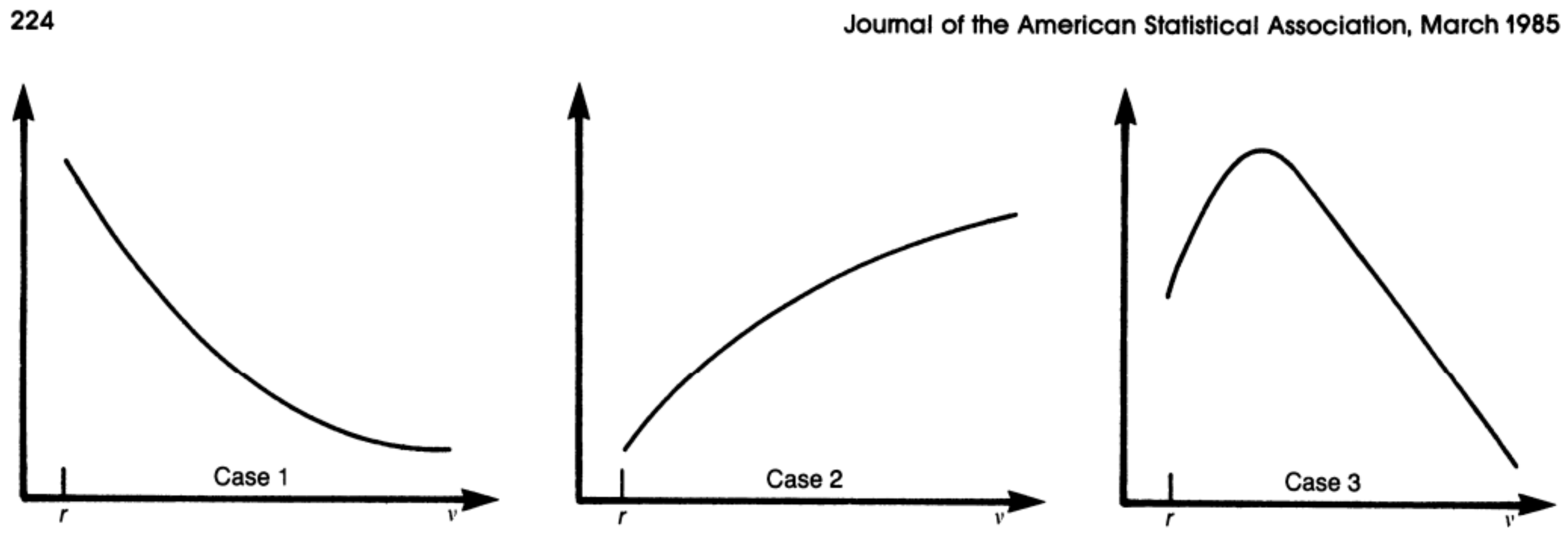


Figure 1. Plots of $g(v \mid x, r)$.

Joe & R 1985

- Variance component models $\sigma^2_{between} \geq 0$
- Logistic regression complete separability: $\hat{\beta} \approx \infty$
- Background + signal $Y \sim \text{Poisson}(b + \mu), \quad \mu > 0$

Regularity: some strategies

- Variance components $\sigma^2_{between} \geq 0$

  New asymptotics — Weighted sum of χ^2 Chernoff 54; Self & Liang 87

- Logistic regression; complete separability; $\hat{\beta} \approx \infty$

  Exact conditional likelihood Cox 58; Mehta & Patel 95

  Adjust maximum likelihood equation (de-bias) Firth 93; Kosmidis & Firth 20

- Background + signal

  Significance function seems satisfactory Fraser Reid Wong 2004

Identifiability

- Mixture models $\pi f(y;\theta_1) + (1-\pi) f(y;\theta_2)$ Chen & Chen 03, McLachlan et al 2019

  If $\pi = 1$, θ not identifiable; if $\theta_1 = \theta_2$, then π not identifiable

- Change-point problems

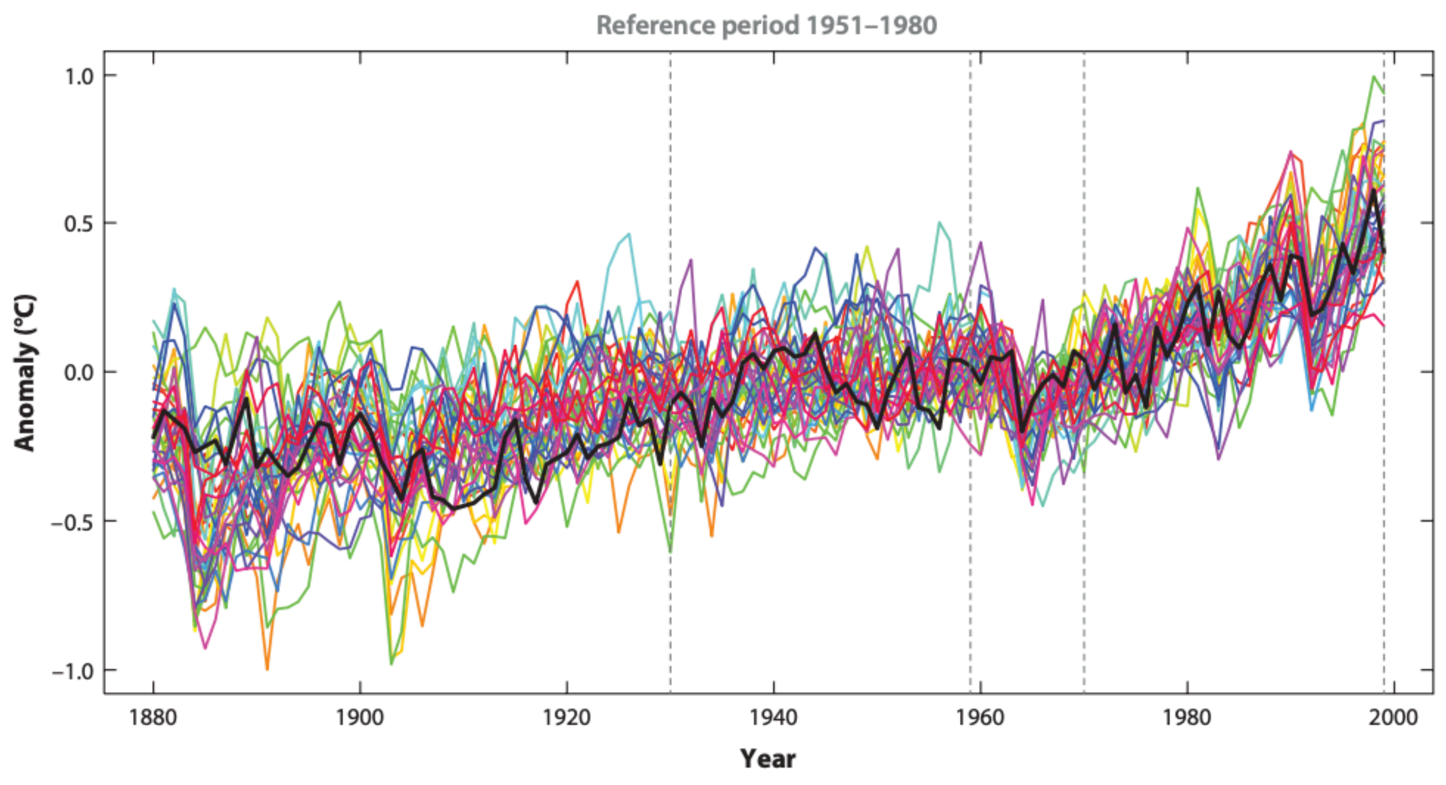


Figure 5

Global average temperature anomalies (*colored paths*) for 1880–1999 relative to the reference period 1951–1980 from 38 models in the Coupled Model Intercomparison Project Phase 5 (CMIP5). The thick black line is the Goddard Institute for Space Studies estimate of global mean temperature. The dashed vertical lines are time periods used in **Figure 6** to compare distributions of models and data.

Guttorp 2014

- New asymptotic arguments; involve maxima of Gaussian processes Cox 2006

Some challenges

- "The usual regularity conditions"
- High-dimensional parameter space
- Computational intractability
- The model is wrong
- Likelihood is not a probability

Some strategies

- New asymptotic theory
- Other forms of likelihood
- "Likelihood-like" functions
- Semi- and non-parametric approaches
- "Objective" Bayes

Many parameters

Neyman-Scott 1948

- example: several 2×2 tables with common odds ratio ψ
- $y_{ij} \sim Binom(m_{ij}; p_i), \quad j = 1, 2; \quad i = 1, \dots, n$
- maximum likelihood estimator not consistent as $n \to \infty$ — if $m_{ij} \equiv 1$, $\widehat{\psi} \to 2\psi$
- too many nuisance parameters — parameter space dimension growing with n

Many parameters: strategies

Example:

$$y_{ij} \sim Binom(m_{ij}, p_{ij}), \qquad \text{logit}(p_{ij}) = \psi$$

- conditional likelihood

$f(y_{i1} \mid y_{i1} + y_{i2}; \psi)$ is free of λ

$$\boldsymbol{L_c(\psi; \underline{y})} = \prod_{i=1}^{n} f(y_{i1} \mid y_{i1} + y_{i2}; \psi) = L_c(\psi; y_{+1})$$

- Y_{+1} is conditionally sufficient

Fisher's exact test

Many parameters: strategies

Cox 1958

Example:

$$y_{ij} \sim Binom(m_{ij}, p_{ij}), \qquad \text{logit}(p_{ij}) = \psi$$

- conditional likelihood

$f(y_{i1} \mid y_{i1} + y_{i2}; \psi)$ is free of λ

$$L_c(\psi; \underline{y}) = \prod_{i=1}^{n} f(y_{i1} \mid y_{i1} + y_{i2}; \psi) = L_c(\psi; y_{+1})$$

- Y_{+1} is conditionally sufficient

Fisher's exact test

- Y_{+1} carries all the information about ψ

"measures ψ"

- many similar models admit conditional or marginal likelihoods

Battey & Cox, 2020, 22

Many parameters: Likelihood

- we can let number of parameters $\rightarrow \infty$, if we can eliminate nuisance parameters to create a well-behaved likelihood function

matched pairs

Many parameters: Likelihood

- we can let number of parameters $\to \infty$, if we can eliminate nuisance parameters to create a well-behaved likelihood function

matched pairs

- conditional and marginal likelihood functions can be approximated by adjusting the profile likelihood function

$\ell_{prof}(\psi) = \ell(\psi, \hat{\lambda}_\psi)$

$$\ell_{adj}(\psi) = \ell_{prof}(\psi) - \frac{1}{2}\log|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)| + A(\psi) \qquad j_{\lambda\lambda} = -\partial^2\ell(\psi, \lambda)/\partial\lambda\partial\lambda^T$$

Cox & R 1987

Many parameters: Likelihood

- we can let number of parameters $\to \infty$, if we can eliminate nuisance parameters to create a well-behaved likelihood function

matched pairs

- conditional and marginal likelihood functions can be approximated by adjusting the profile likelihood function

$\ell_{prof}(\psi) = \ell(\psi, \hat{\lambda}_\psi)$

$$\ell_{adj}(\psi) = \ell_{prof}(\psi) - \frac{1}{2}\log|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi) + A(\psi) \qquad j_{\lambda\lambda} = -\partial^2\ell(\psi, \lambda)/\partial\lambda\partial\lambda^T$$

Cox & R 1987

- $\ell_{adj}(\psi)$ can be converted to a significance function, using higher-order approximations

beyond CLT

- but these approximations rely on regularity conditions!, especially fixed dimension

Very many parameters p_n

Sur et al 2019

- can let number of parameters increase with n, if nuisance parameters can be eliminated to create a well-behaved likelihood function

matched pairs

- but, if $p_n/n \to$ constant, a new asymptotic theory is needed

$$2\{\ell_{\text{prof}}(\hat\psi) - \ell_{\text{prof}}(\psi)\} \xrightarrow{d} \frac{\sigma_*^2}{\lambda_*}\chi_1^2$$

Sur, Chen, Candès 2019

logistic regression, $\psi = \beta_j$

also depends on $\lim_{n\to\infty} p_n/n$

(σ_*, λ_*) characterized

by studying optimization path

490 P. Sur et al.

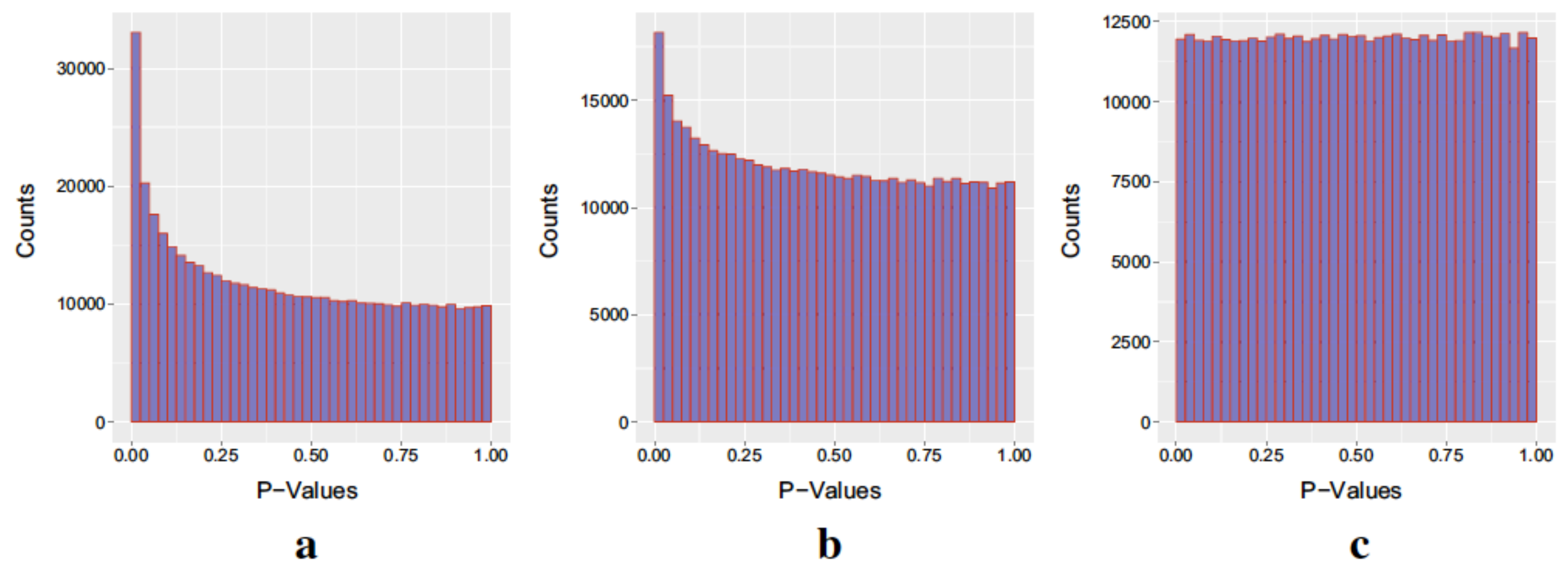


Fig. 1 Histogram of p-values for logistic regression under i.i.d. Gaussian design, when $\boldsymbol{\beta} = \mathbf{0}$, $n = 4000$, $p = 1200$, and $\kappa = 0.3$; **a** classically computed p-values; **b** Bartlett corrected p-values; **c** adjusted p-values

Very many parameters p_n

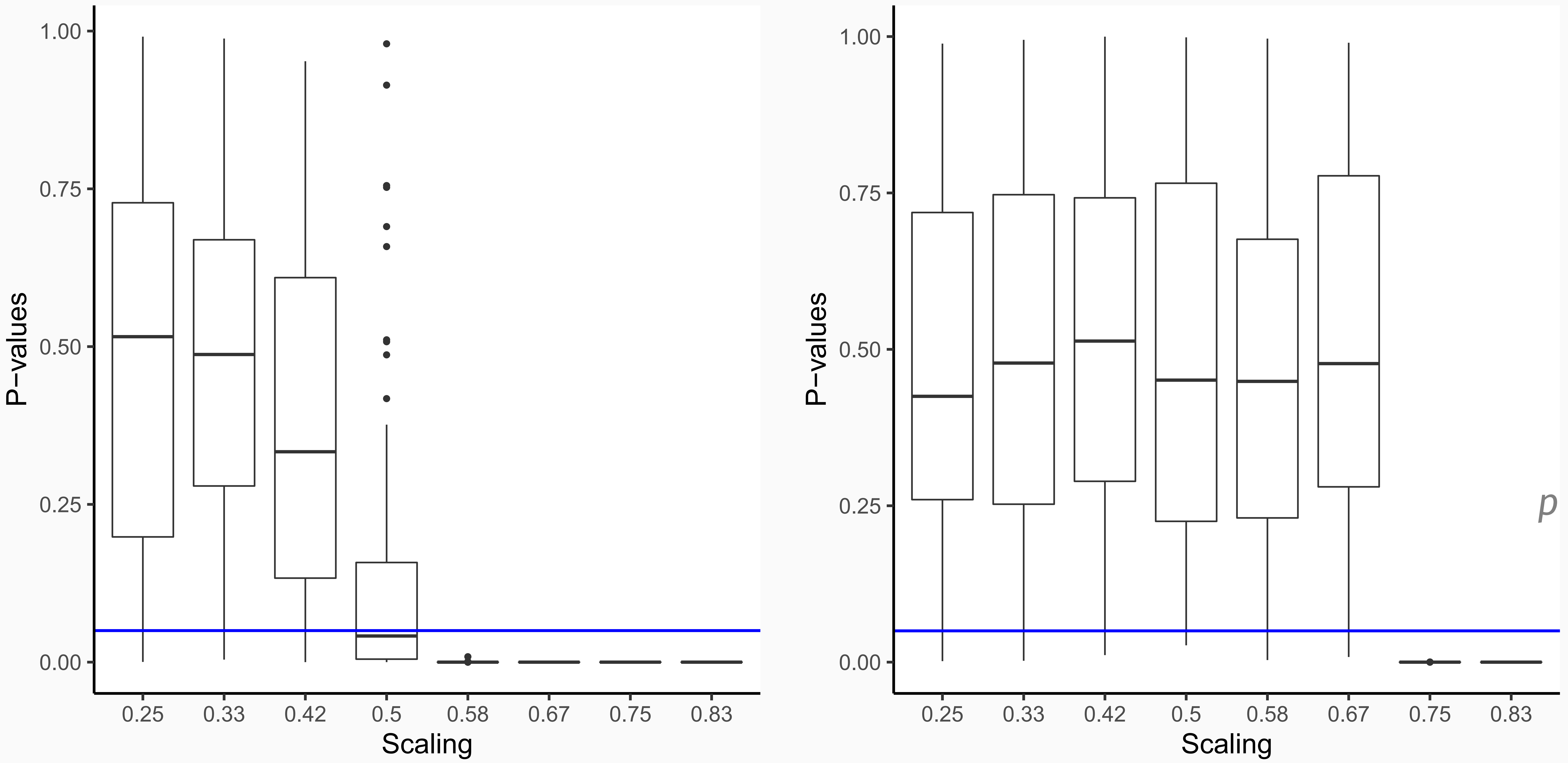
- in generalized linear models, maximum likelihood estimate not asymptotically normal unless p increases slowly with n
- under $H_0 : \beta = 0$, want p-values based on $(\hat{\beta} - \beta)/\hat{\sigma}_\beta$ to be $U(0,1)$
- this fails unless $p \sim n^{1/3}$ — Fan et al 2019

- challenge: theoretical analysis is somewhat specialized — random X; global null
- challenge: estimating σ_β, or σ_*^2/λ_*

- can methods of conditional or marginal likelihood be used for "fairly large" p?
- shows early promise — Tang & R 2020; Lunardon 2019; Battey & R 2022
- but much work remains

HoA for large p

First-order (normal) approximation | Higher-order approximation



$p \sim n^{1/3}$

Fig. 3. Plots for logistic regression illustrating the difference in the breakdown point of uniformity of the p-value distribution based on the standard normal approximation to the distribution of (a) r and of (b) r^*: we see that p-values based on the r^*-approximation appear to be uniformly distributed up to about $p = O(n^{2/3})$, whereas those based on the normal approximation to the distribution of r begin to exhibit non-uniformity at about $p = O(n^{1/2})$.

Some challenges

- “The usual regularity conditions”
- High-dimensional parameter space
- Computational intractability
- The model is wrong
- Likelihood is not a probability

Some strategies

- New asymptotic theory
- Other forms of likelihood
- “Likelihood-like” functions
- Semi- and non-parametric approaches
- “Objective” Bayes

The model is wrong

Some strategies:

- develop new theory for likelihood inference — Cox 1961,2;
- recent example: “Assumption lean inference for generalized linear models” — Vansteelandt & Dukes 2022
- limiting normal distribution for maximum likelihood estimator, but asymptotic variance larger than for correct model — related results for likelihood ratio tests

The model is wrong

Some strategies:

- develop new theory for likelihood inference — Cox 1961,2;
- recent example: "Assumption lean inference for generalized linear models" — Vansteelandt & Dukes 2022
- limiting normal distribution for maximum likelihood estimator, but asymptotic variance larger than for correct model — related results for likelihood ratio tests

- make the model more flexible
- Example: proportional hazards model $\lambda(y; x, \beta) = \lambda_0(y) \exp(x^T \beta)$ — Cox 1972
- justified as a partial likelihood — Cox 1975
- can also be interpreted as profile likelihood — Murphy & Van der Vaart 2000

The model is wrong

Some strategies:

- develop new theory for likelihood inference — Cox 1961,2;
- recent example: “Assumption lean inference for generalized linear models” — Vansteelandt & Dukes 2022
- limiting normal distribution for maximum likelihood estimator, but asymptotic variance larger than for correct model — related results for likelihood ratio tests

- make the model more flexible
- Example: proportional hazards model $\lambda(y; x, \beta) = \lambda_0(y)\exp(x^T\beta)$ — Cox 1972
- justified as a partial likelihood — Cox 1975
- can also be interpreted as profile likelihood — Murphy & Van der Vaart 2000

- go fully nonparametric: empirical likelihood, constrained density estimation, … — Balabdaoui et al. 2009; Robeva et al. 2021

Some challenges

- "The usual regularity conditions"
- High-dimensional parameter space
- Computational intractability
- The model is wrong
- Likelihood is not a probability

Some strategies

- New asymptotic theory
- Other forms of likelihood
- "Likelihood-like" functions
- Semi- and non-parametric approaches
- "Objective" Bayes

The likelihood function is intractable

- example: latent Gaussian model

Rue et al. 2017

$$f(y \mid z; \theta) = \prod_{i \in \mathcal{I}} f(y_i \mid z_i; \theta), \quad z \sim N\{\mu(\tau), \Sigma(\tau)\}$$

$$L(\theta; y) = \int f(y \mid z; \theta) f(z; \tau) \pi(\tau) dz d\tau$$

- spatial processes, network models, multivariate extremes, agent-based models

- one strategy: composite likelihood

pseudo-likelihood

- replace $L(\theta; \underline{y}) \propto f(\underline{y}; \theta)$ by, e.g.,

$$cL(\theta; \underline{y}) = \prod_{j<k} f_2(y_j, y_k; \theta)$$

- a type of wrong model, with some nice properties
- asymptotic theory has been developed as on previous slide

Molenberghs& Verbeke, 2005; Lindsay, 1988

Some challenges

- "The usual regularity conditions"
- High-dimensional parameter space
- Computational intractability
- The model is wrong
- Likelihood is not a probability

Some strategies

- New asymptotic theory
- Other forms of likelihood
- "Likelihood-like" functions
- Semi- and non-parametric approaches
- "Objective" Bayes

The likelihood function is not a probability

- can’t be integrated, without introducing some prior for the parameters
- in complex models priors can be difficult to specify

The likelihood function is not a probability

- can't be integrated, without introducing some prior for the parameters
- in complex models priors can be difficult to specify
- strategy: find a prior "we can all agree on" — objective, weakly informative
- posterior may give inference methods with good performance under the model — calibrated inference
- this needs to be checked in each application

The likelihood function is not a probability

- can't be integrated, without introducing some prior for the parameters
- in complex models priors can be difficult to specify
- strategy: find a prior "we can all agree on" — objective, weakly informative
- posterior may give inference methods with good performance under the model — calibrated inference
- this needs to be checked in each application
- satellite conjunction analysis — Elkantassi & Davison 2022
- inference for length of a normal vector — Stein 1959

Example

- $y_i \sim N(\theta_i, 1/n), \quad i = 1, \dots, k; \quad \pi(\underline{\theta}) \propto 1$
- posterior distribution of $a^{\mathrm{T}}\theta$ is well-calibrated
- marginal posterior distribution of $\psi = \Sigma\theta_i^2$ is not

Example

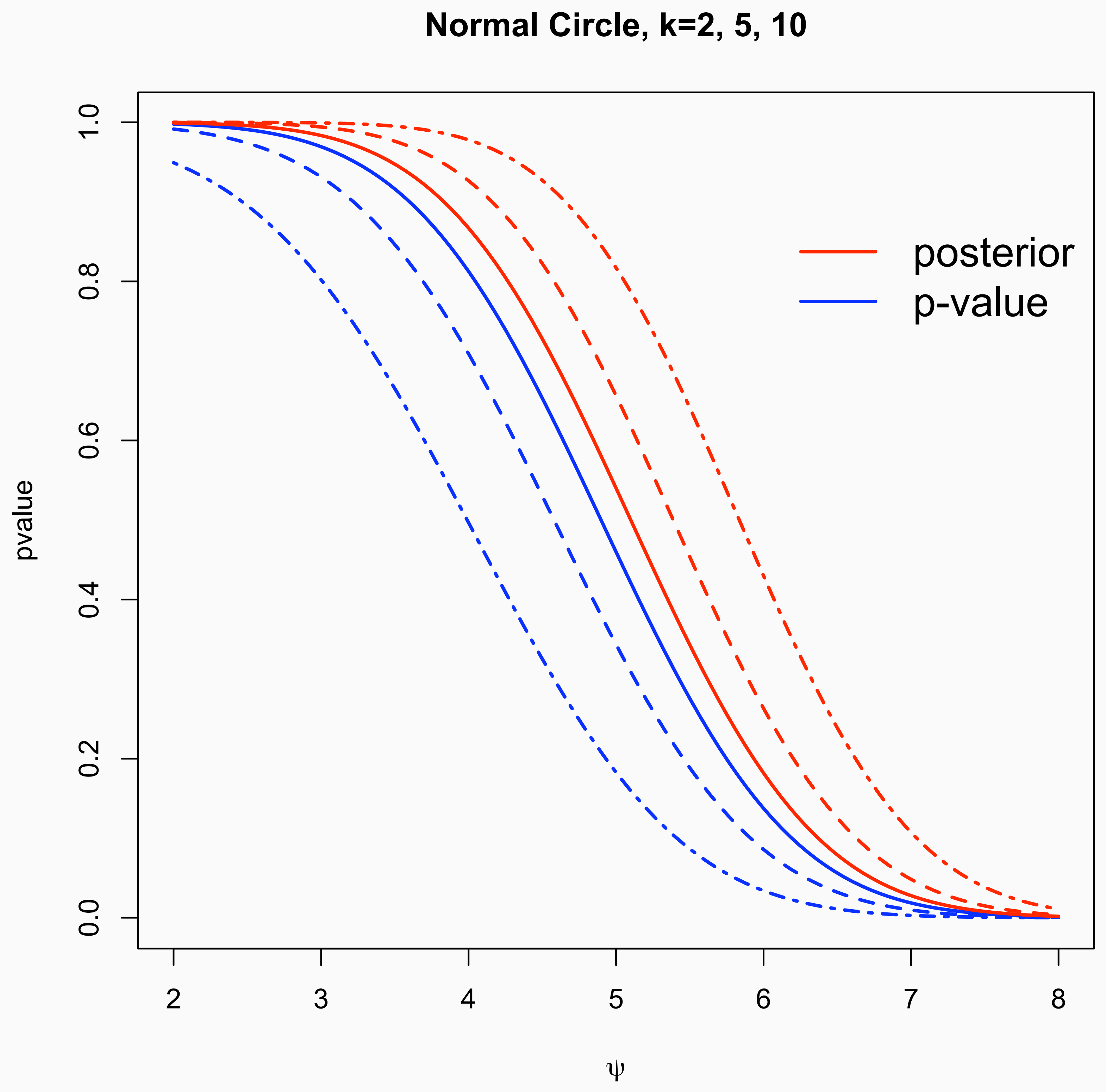
- $y_i \sim N(\theta_i, 1/n), \quad i = 1, \ldots, k; \quad \pi(\underline{\theta}) \propto 1$
- posterior distribution of $a^{\mathrm{T}}\theta$ is well-calibrated
- marginal posterior distribution of $\psi = \Sigma\theta_i^2$ is not
- discrepancy is a function of $\dfrac{k-1}{\psi\sqrt{n}}$

Example

- $y_i \sim N(\theta_i, 1/n), \quad i = 1, \dots, k; \quad \pi(\underline{\theta}) \propto 1$
- posterior distribution of $a^{\mathrm{T}}\theta$ is well-calibrated
- marginal posterior distribution of $\psi = \Sigma\theta_i^2$ is not
- discrepancy is a function of $\dfrac{k-1}{\psi\sqrt{n}}$
- we could construct a prior targetted on $\psi = \Sigma\theta_i^2$
- a toy example? Yes, but has recently re-emerged

Example

- $y_i \sim N(\theta_i, 1/n), \quad i = 1, \ldots, k; \quad \pi(\underline{\theta}) \propto 1$
- posterior distribution of $a^{\mathrm{T}}\theta$ is well-calibrated
- marginal posterior distribution of $\psi = \Sigma\theta_i^2$ is not
- discrepancy is a function of $\dfrac{k-1}{\psi\sqrt{n}}$
- we could construct a prior targetted on $\psi = \Sigma\theta_i^2$
- a toy example? Yes, but has recently re-emerged



Satellite conjunction analysis

- estimating probability of satellites passing too close to each other — collision risk
- simplified version requires estimation of length of a normal vector
- Bayesian version highlighted in Balch et al.


PROCEEDINGS A

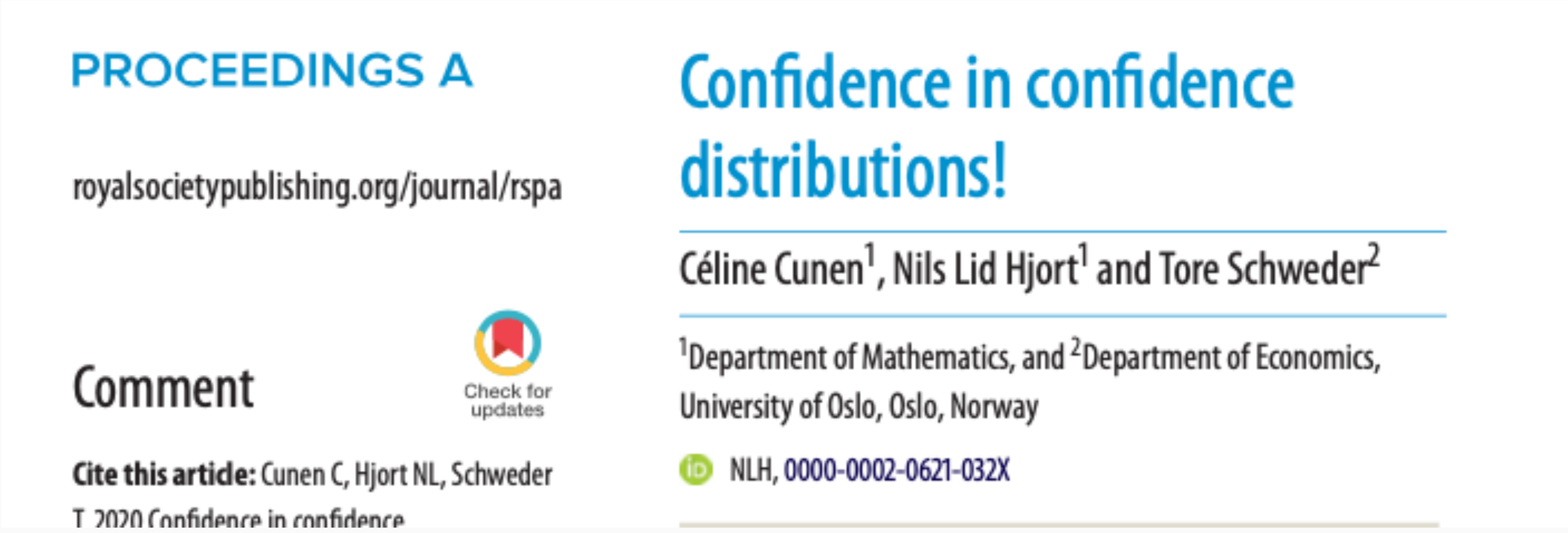
royalsocietypublishing.org/journal/rspa

Research

Check for updates

Satellite conjunction analysis and the false confidence theorem

Michael Scott Balch[1], Ryan Martin[2] and Scott Ferson[3]


PROCEEDINGS A

royalsocietypublishing.org/journal/rspa

Comment

Check for updates

Cite this article: Cunen C, Hjort NL, Schweder T. 2020 Confidence in confidence

Confidence in confidence distributions!

Céline Cunen[1], Nils Lid Hjort[1] and Tore Schweder[2]

[1]Department of Mathematics, and [2]Department of Economics, University of Oslo, Oslo, Norway

NLH, 0000-0002-0621-032X

- response by Cunen et al.
- detailed treatment in Elkantassi & Davison 2022 — J Guidance Control and Dynamics

further response by Balch et al. to Cunen on interpretation of confidence distributions

So much detail!

- Many adjective-likelihood functions

  Marginal, conditional, partial, composite, pseudo

  Quasi, empirical, bootstrap, simulated, sieve, penalized

- Common theme: provide inference strategies with well-understood properties
- Enables us to move away from the specific problem at hand
- Theory provides guidance for a range of similar applications

3. Theory and Applications

Statistics in the news

Economist, July 29

The Economist

Menu | Weekly edition | Search

Science & technology | Nudge factor

Evidence for behavioural interventions looks increasingly shaky

The academic literature is plagued by publication bias

"plagued by publication bias"

The source

Mertens et al. 2021

PNAS

RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES

The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains

Stephanie Mertens, Mario Herberz, Ulf J. J. Hahnel, and Tobias Brosch Authors Info & Affiliations

Edited by Susan Fiske, Psychology Department, Princeton University, Princeton, NJ; received April 27, 2021; accepted November 24, 2021

December 30, 2021 | 119 (1) e2107346118 | https://doi.org/10.1073/pnas.2107346118

THIS ARTICLE HAS BEEN UPDATED

The response

LETTER | JULY 19, 2022

No reason to expect large and consistent effects of nudge interventions

Barnabas Szaszi, Anthony Higney, [...] Elizabeth Tipton

VIEW THE ORIGINAL ARTICLE:

The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains

THIS ARTICLE HAS A REPLY:

Reply to Maier et al., Szaszi et al., and Bakdash and Marusich: The present and future of choice architecture research

LETTER | JULY 19, 2022

Left-truncated effects and overestimated meta-analytic means

Jonathan Z. Bakdash and Laura R. Marusich

VIEW THE ORIGINAL ARTICLE:

The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains

THIS ARTICLE HAS A REPLY:

Reply to Maier et al., Szaszi et al., and Bakdash and Marusich: The present and future of choice architecture research

LETTER | JULY 19, 2022

No evidence for nudging after adjusting for publication bias

Maximilian Maier, František Bartoš, [...] Eric-Jan Wagenmakers

VIEW THE ORIGINAL ARTICLE:

The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains

THIS ARTICLE HAS A REPLY:

Reply to Maier et al., Szaszi et al., and Bakdash and Marusich: The present and future of choice architecture research

REPLY | JULY 19, 2022

Reply to Maier et al., Szaszi et al., and Bakdash and Marusich: The present and future of choice architecture research

Stephanie Mertens, Mario Herberz, [...] Tobias Brosch

THIS ARTICLE REPLIES TO:

No evidence for nudging after adjusting for publication bias

Left-truncated effects and overestimated meta-analytic means

No reason to expect large and consistent effects of nudge interventions

In fairness

Mertens et al. 2021

OSFHOME

Search Support Donate Sign Up Sign In

The effectiveness of nudging: A meta-a... | Files | Wiki | Analytics | Registrations

0.0B Public 0

The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains

Contributors: Stephanie Mertens, Mario Herberz, Ulf J.J. Hahnel, Tobias Brosch

Date created: 2020-09-01 05:46 PM | Last Updated: 2021-12-15 06:52 PM

Identifier: DOI 10.17605/OSF.IO/FYWAE

Category: Project

Description: *This project is funded by the Swiss National Science Foundation and the Swiss Federal Office of Energy. It investigates the effectiveness of choice architecture interventions across behavioral domains.*

Files

Filter

Name	Modified

Citation

Components

Some details

Mertens et al. 2021

"Materials and methods"

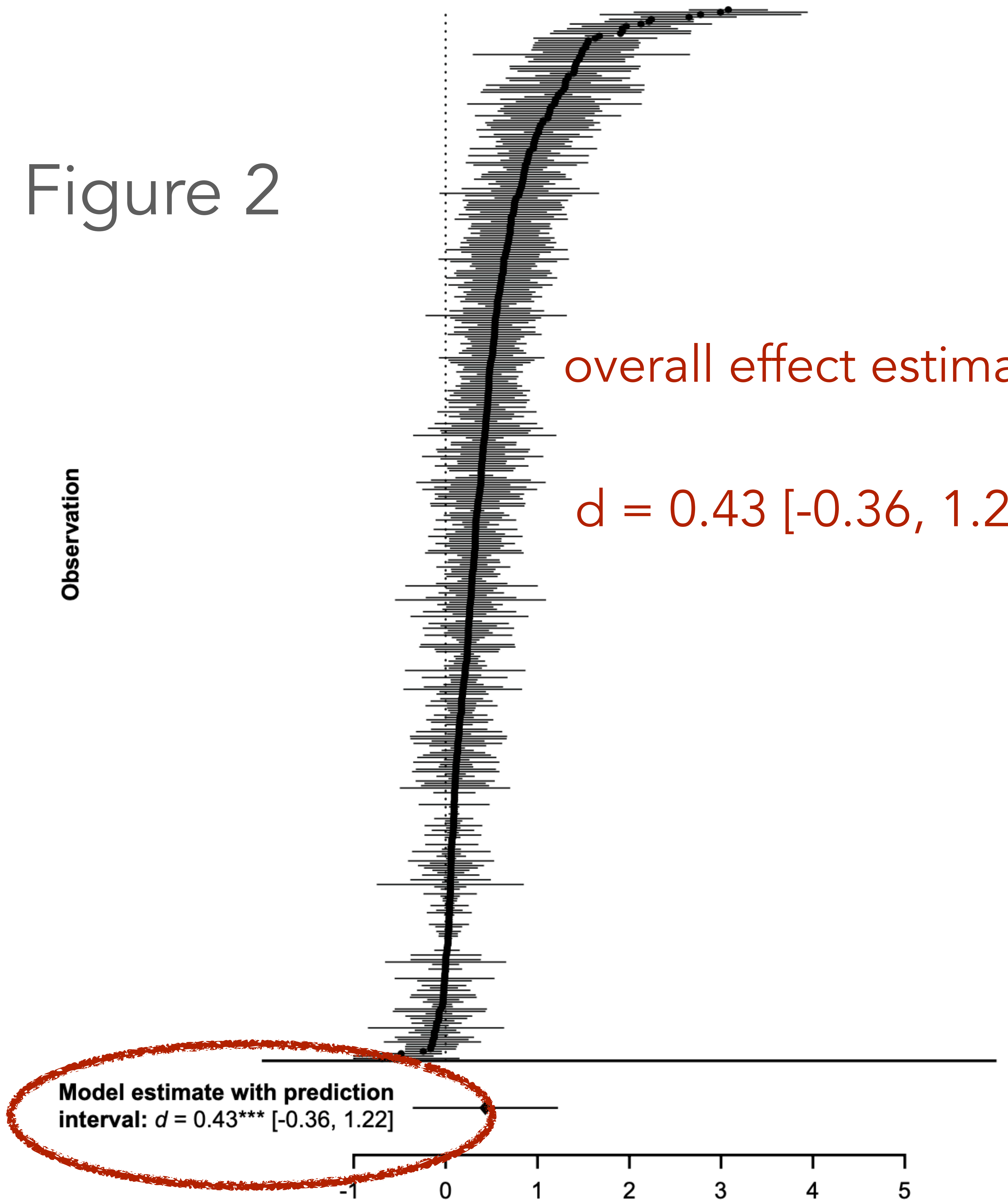
- 440 estimates of effect size: (treatment — control mean)/(estimated std error)
- 212 unique publications; sometimes several tmts with the same control
- Random effects to accommodate this
- Additional fixed effects (moderators) for secondary analysis —

  types of interventions; behavioural domain; study characteristics
- Publication bias assessed by plotting standard error vs effect size   Egger's test

Some results

Mertens et al. 2021

Figure 2



overall effect estimate

d = 0.43 [-0.36, 1.22]

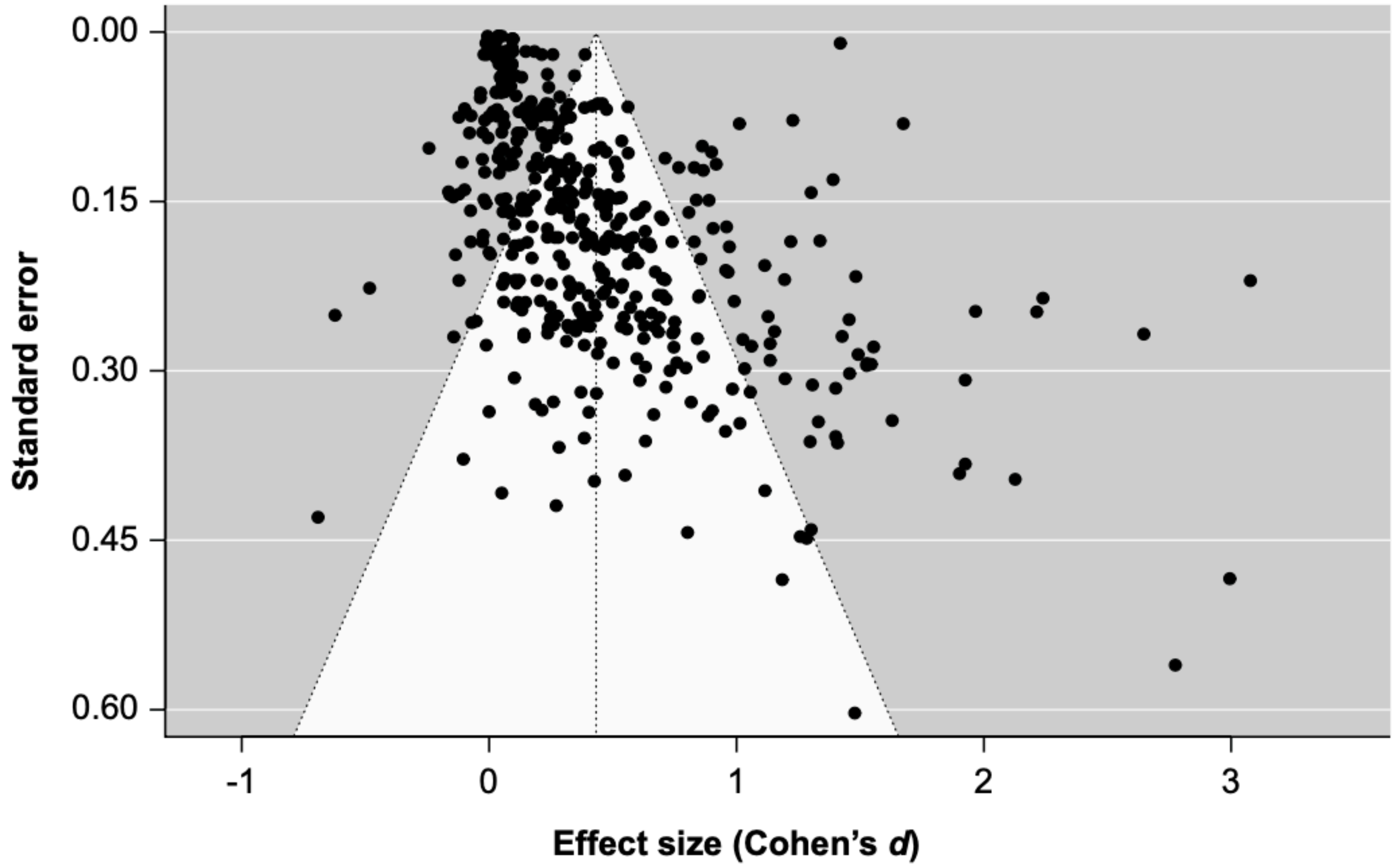


Fig. 3. Funnel plot displaying each observation as a function of its effect size and SE. In the absence of publication bias, observations should scatter

Standard error increases with effect size

The letters

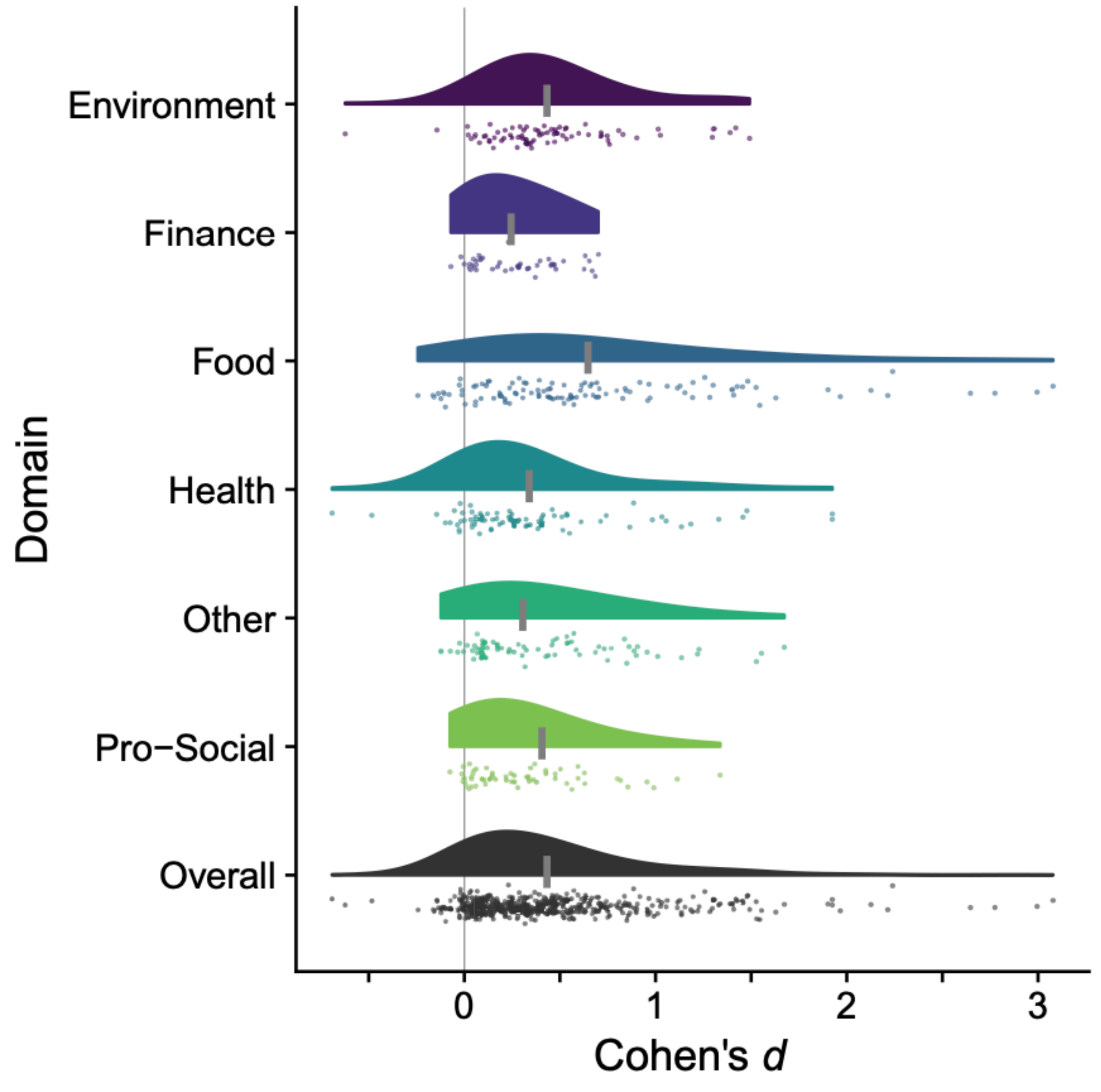
1. Maier et al. — publication bias not correctly taken into consideration

   "A newly-proposed bias-correction technique — robust Bayesian meta-analysis avoids an 'all-or-none' debate over whether or not publication bias is 'severe' "

2. Szaszi et al. — the average effect size is not very informative, given the variation between studies

   "Even after adjusting for publication bias, the effects … vary considerab

3. Bakdash & Maurisch — estimated effects in studies are right-skewed

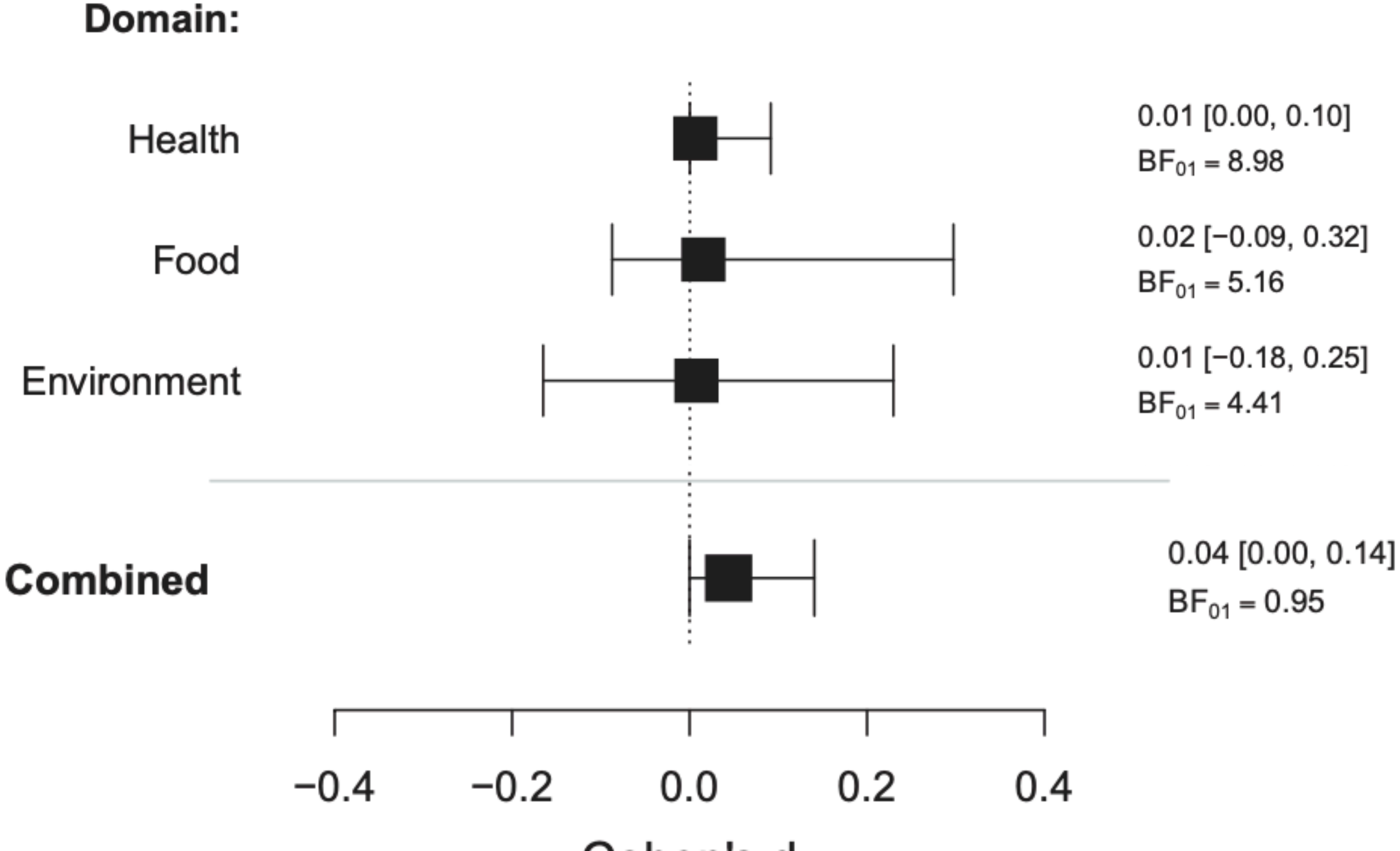


Foundations

1. Maier et al. — publication bias not correctly taken into consideration

“A newly-proposed bias-correction technique — robust Bayesian meta-analysis avoids an ‘all-or-none’ debate over whether or not publication bias is ‘severe’ ”

Figure 1 Meier et al.



Model-averaged mean effect size estimates with posterior credibility intervals and Bayes factors

Szaszi et al applied various non-Bayesian adjustments for bias with similar results

Science and sports

Borg et al. 2022

Tweet

Thomas Lumley
@tslumley

"the journal Science and Medicine in Football, for example, introduced registered reports three years ago but has yet to receive a single submission."

outsideonline.com
Why That New “Science-Backed” Supplement Probably Doesn’t Work
A deep dive into the sports science literature shows why you should be wary of results that seem too good to be true

8:57 PM · Aug 4, 2022 · Twitter Web App

Outside

Why That New “Science-Backed” Supplement Probably Doesn’t Work

A deep dive into the sports science literature shows why you should be wary of results that seem too good to be true

OSFPREPRINTS Add a Preprint Search Support Dona

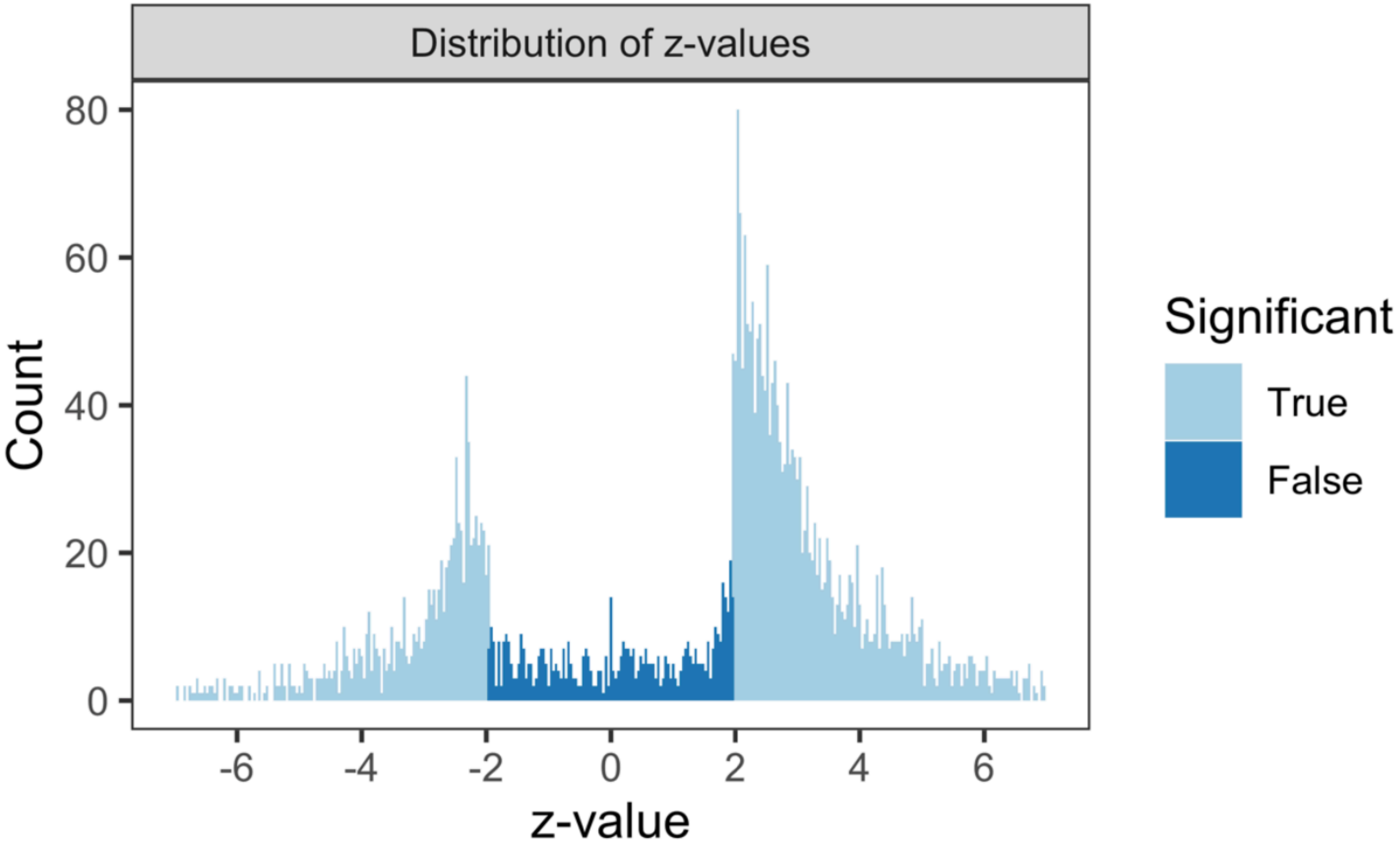
The Bias for Statistical Significance in Sport and Exercise Medicine

AUTHORS
David N Borg, Adrian Barnett, Aaron R Caldwell, Nicole White, Ian Stewart

AUTHOR ASSERTIONS
Conflict of Interest: Yes
Public Data: Available
Preregistration: No

The missing middle

Borg et al. Figure 3



Another nudge

Science Ullman & Chrysler 2022

A dynamic message sign displays traffic deaths above the H-1 Freeway on the Liliha Street overpass in Hawaii.

BEHAVIORAL SCIENCE

How safe are safety messages?

Highway fatalities increased in response to certain messages

- "… research by Hall and Madsen suggests that, contrary to expectations, displaying traffic fatality numbers in traffic safety messages is associated with an increase in fatalities downstream"
- "… seems inconsistent with other research that has found … mostly ineffective"
- "… the issue may be one of excessive salience or cognitive overload"

A natural experiment

Hall & Madsen 2022

***Science* April 22 2022**

RESEARCH

RESEARCH ARTICLE

TRAFFIC SAFETY

Can behavioral interventions be too salient? Evidence from traffic safety messages

Jonathan D. Hall[1,2] and Joshua M. Madsen[3]*

	Crashes per hour (%)		
	3 km	5 km	10 km
	(1)	(2)	(3)
Campaign week × post	1.13 (0.86)	1.52 (0.68)**	1.35 (0.60)**
Campaign week	0.35 (0.63)	−0.27 (0.48)	−0.32 (0.43)
Observations	61,697,666	61,697,666	61,697,666
Adjusted R^2	0.02	0.03	0.08
Rain and interactions	Yes	Yes	Yes
S-Y-M-D-H FE	Yes	Yes	Yes
Holiday FE	Yes	Yes	Yes

- data from Texas; messages about fatalities posted 1 week of four
- researchers were able to use the other weeks as "controls"
- with adjustments for weather, time of day, etc.,
- concluded that accidents increased by roughly 1.5% in weeks when messages displayed
- small but "statistically significant"

?

Diet and health

NY Times July 12 2022

The New York Times Account

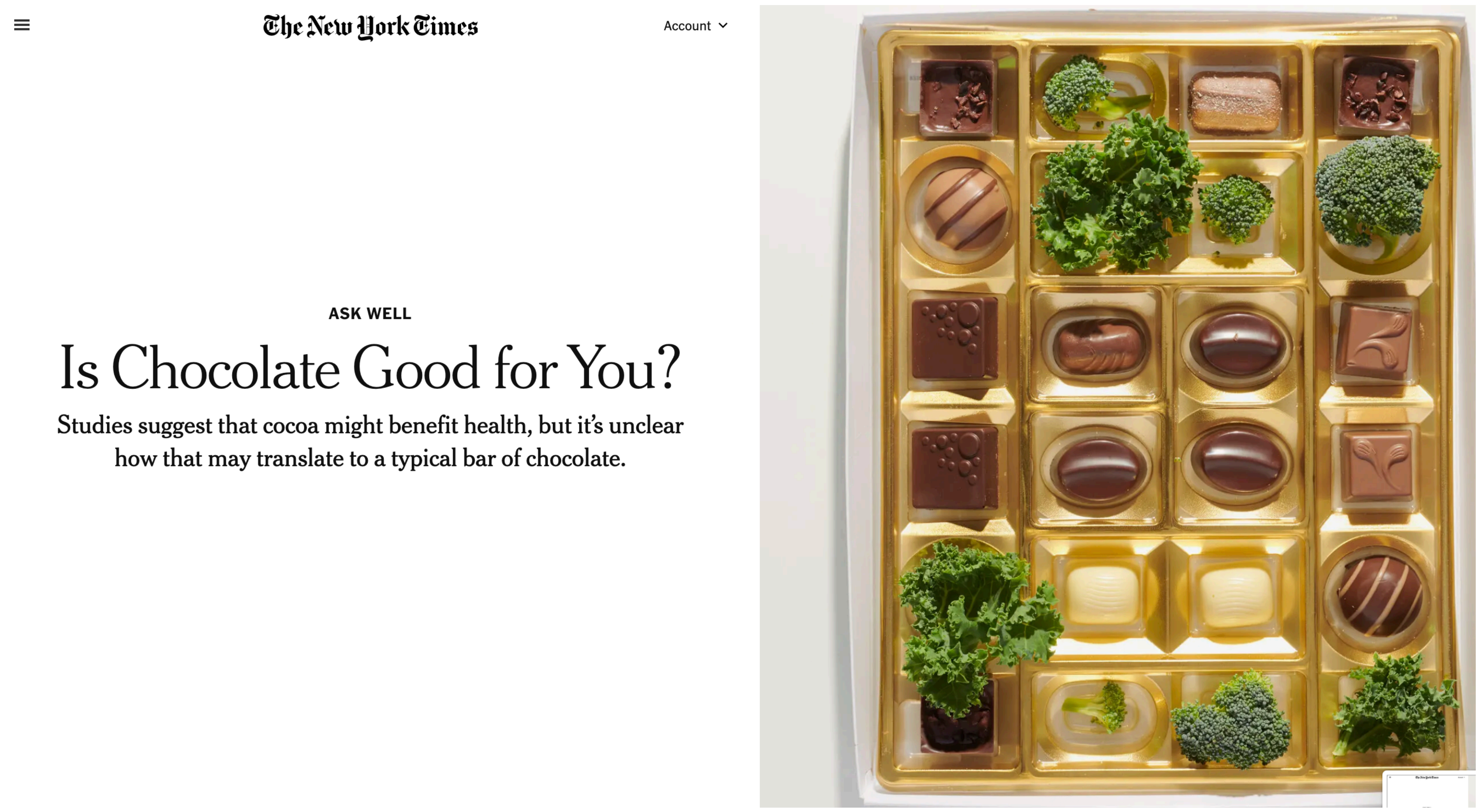
ASK WELL

Is Chocolate Good for You?

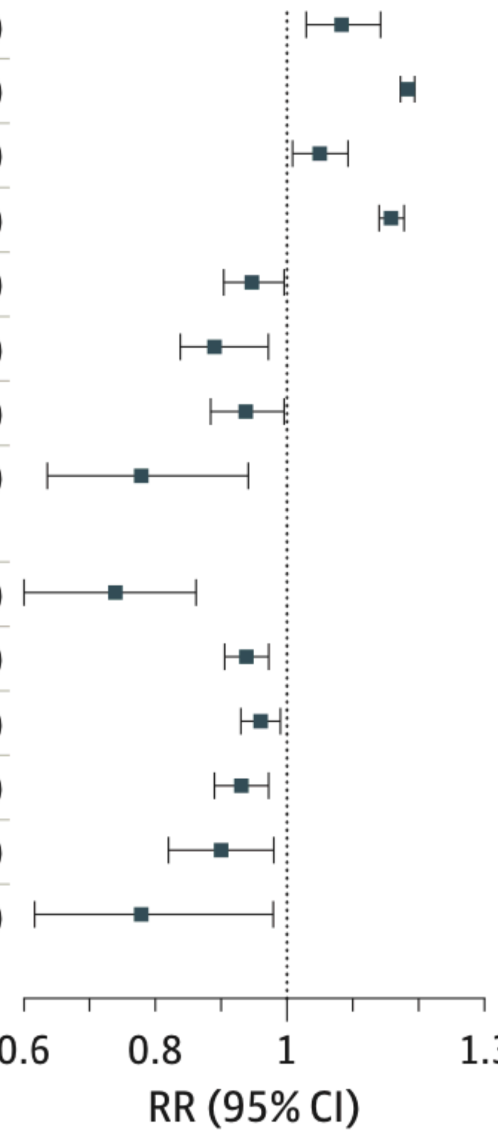
Studies suggest that cocoa might benefit health, but it's unclear how that may translate to a typical bar of chocolate.

NY Times

- refers to:
  - a review of several (small) meta-analyses Miller et al 2022
  - a large randomized trial
    Sessa et al 2022



Bechthold et al,[47] 2019	Processed meats	CHD	3	151 373	6659	150 g/wk	1.08 (1.03-1.14)	0
Zeraatkar et al,[46] 2019	Processed meats	Stroke	6	254 742	13 113	150 g/wk	1.18 (1.17-1.19)	40.2
Bechthold et al,[47] 2019	Processed meats	Ischemic stroke	5	NR	NR	150 g/wk	1.05 (1.01-1.09)	18.0
Zeraatkar et al,[46] 2019	Processed meats	Diabetes	NR (17)	758 540	NR	150 g/wk	1.16 (1.14-1.18)	92.0
Bechthold et al,[47] 2019	Fish/seafood	CHD	15	479 657	14 056	300 g/wk	0.95 (0.90-1.00)	40.0
Jayedi et al,[48] 2021	Fish/seafood	MI	11	398 221	8468	300 g/wk	0.89 (0.84-0.97)	64.5
Bechthold et al,[47] 2019	Fish/seafood	Stroke	14 (15)	370 844	11 326	300 g/wk	0.94 (0.88-1.00)	25.0
Jayedi et al,[48] 2021	Fish/seafood	CHD in patients with diabetes	3	8464	NR	300 g/wk	0.78 (0.64-0.94)	0
Gijsbers et al,[64] 2016	Yogurt	Diabetes	9	438 140	36 125	244 g/d	0.74 (0.60-0.86)	73.3
Ren et al,[50] 2019	Chocolate	CVD	12 (18)	369 599	19 530	10 g/d	0.94 (0.91-0.97)	50.4
Morze et al,[55] 2020	Chocolate	CHD	7	416 185	19 812	10 g/d	0.96 (0.93-0.99)	29.0
Larsson et al,[51] 2016	Chocolate	MI	4	109 118	7267	10 g/d	0.93 (0.89-0.97)	0
Morze et al,[55] 2020	Chocolate	Stroke	7	275 070	9087	10 g/d	0.90 (0.82-0.98)	59.0
Ren et al,[50] 2019	Chocolate	Hemorrhagic stroke	4	155 072	NR	10 g/d	0.78 (0.62-0.98)	0



The COSMOS trial

Sessa et al. 2022

The New York Times

Account

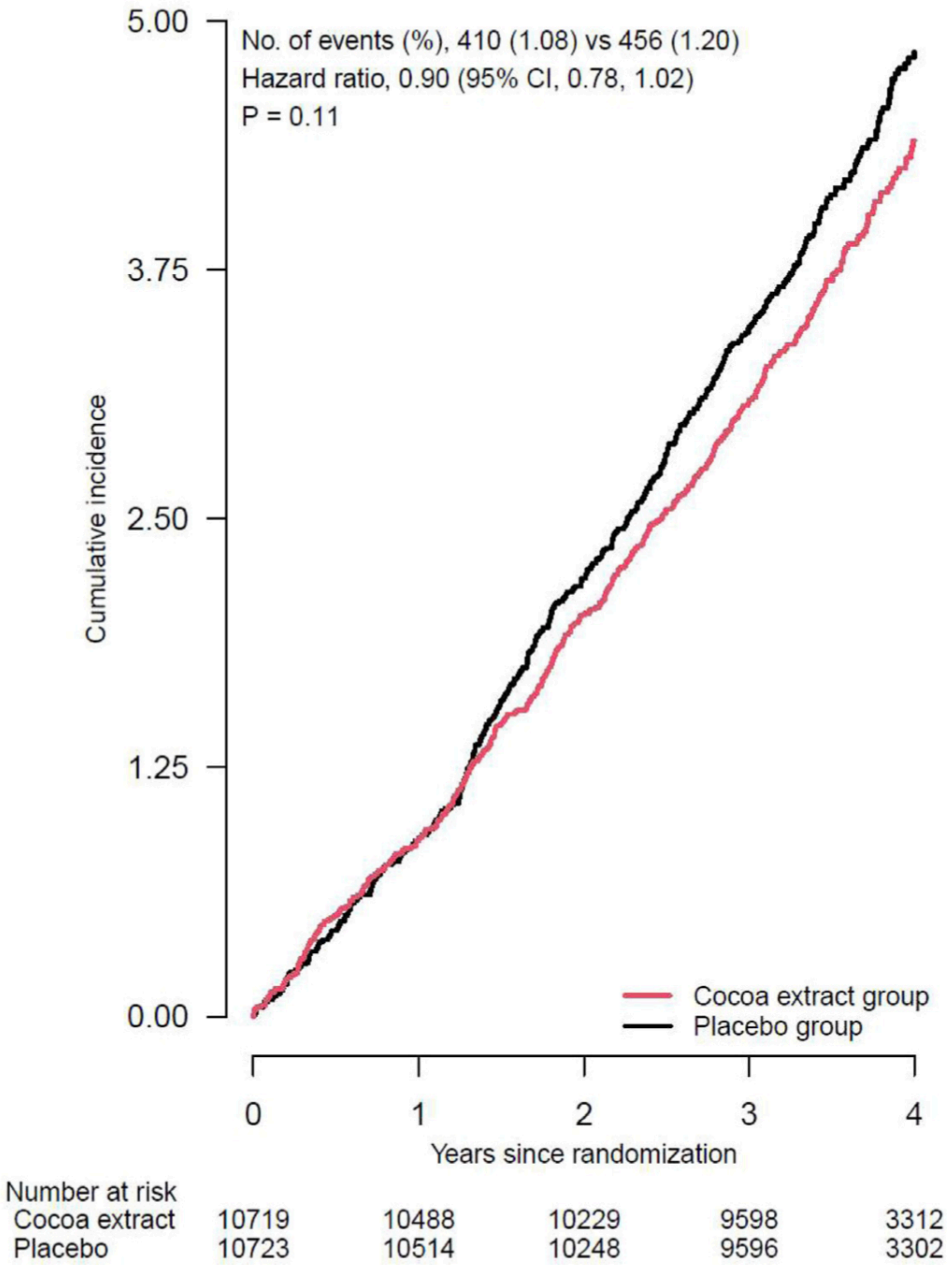
ASK WELL

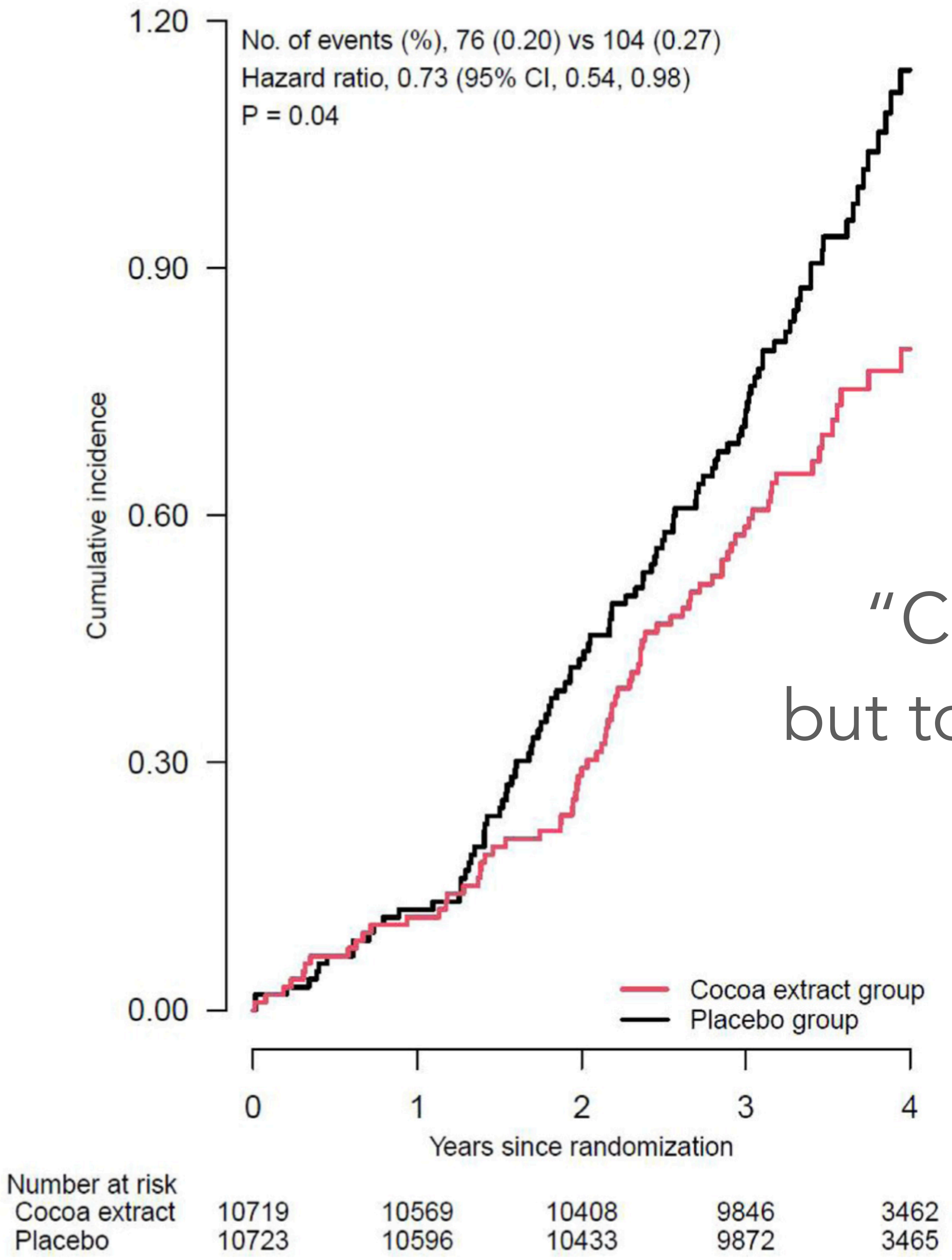
Is Chocolate Good for You?

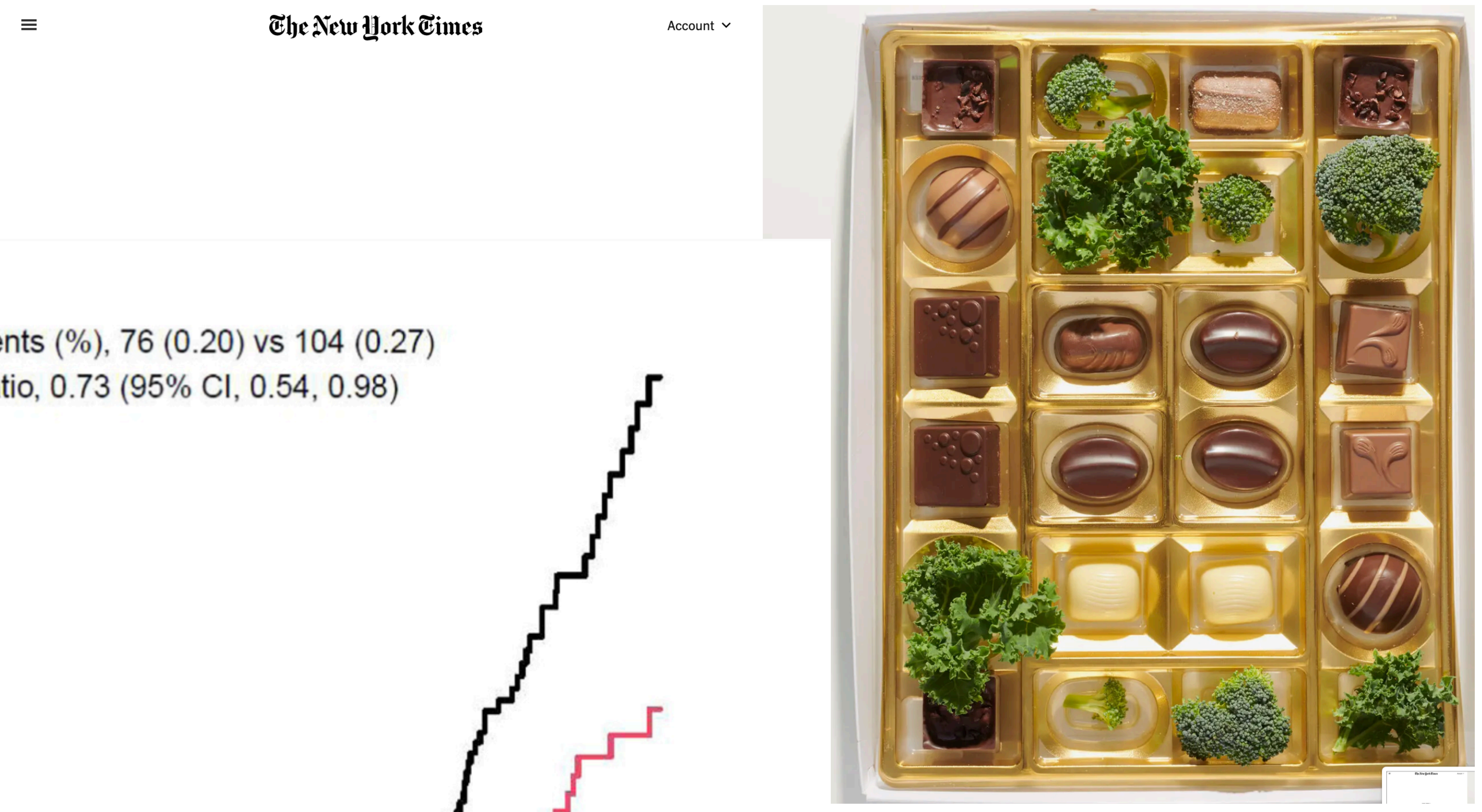
Studies suggest that cocoa might benefit health, but it's unclear how that may translate to a typical bar of chocolate.

- randomized, double-blind, placebo controlled trial
- 21,442 US adults (convenience sample)
- tested a cocoa extract supplement (not chocolate)
- limitations carefully noted in discussion
- "there was no statistically significant effect on the primary outcome"
- "however, cocoa … significantly reduced CVD death by 27%

The COSMOS trial







"Chocolate is a wonderful treat, but to perceive it as a health food, I think it has its limitations"

Some stories never die

March 2010

“Researchers find those who eat 7.5 grams a day have a lower risk of heart disease”

An ounce of (dark chocolate) prevention

Or less: Researchers find those who eat 7.5 grams a day have a lower risk of heart disease

BY KATE KELLAND

Easter eggs may be good for you, but only if you eat small ones made from cocoa-rich dark chocolate, according to the latest in a string of scientific studies to show potential health benefits of chocolate.

German researchers studied more than 19,300 people over a decade and found those who ate the most chocolate – an average of 7.5 grams a day – had lower blood pressure and a 39 per cent lower risk of having a heart attack or stroke than those who ate the least amount of chocolate – an average of 1.7 grams a day.

But the difference between the two groups was just less than six grams of chocolate a day, less than one small square of an average 100-gram bar, they wrote in a study in the European Heart Journal to be published today.

ELISE AMENDOLA/ASSOCIATED PRESS

Hospital Zurich said basic science had now demonstrated “quite convincingly” that dark chocolate with a cocoa content of at least 70 per cent reduces some kinds of stress and can improve blood flow and blood pressure.

But he said: "Before you rush to add dark chocolate to your diet, be aware that 100 grams … contains roughly 500 calories.

"You may want to subtract an equivalent amount of calories by cutting back on other foods

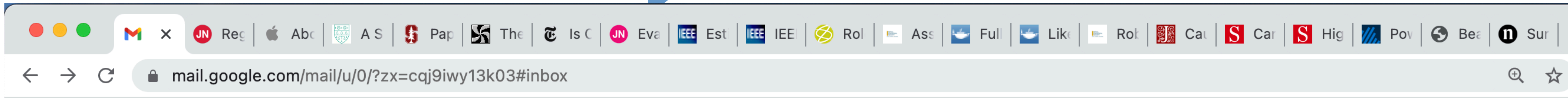
August 2005

ONTARIO EDITION • TORONTO WEATHER: AFTERNOON THUNDERSHOWERS, HIGH 32. MAP AND DETAILS, S8

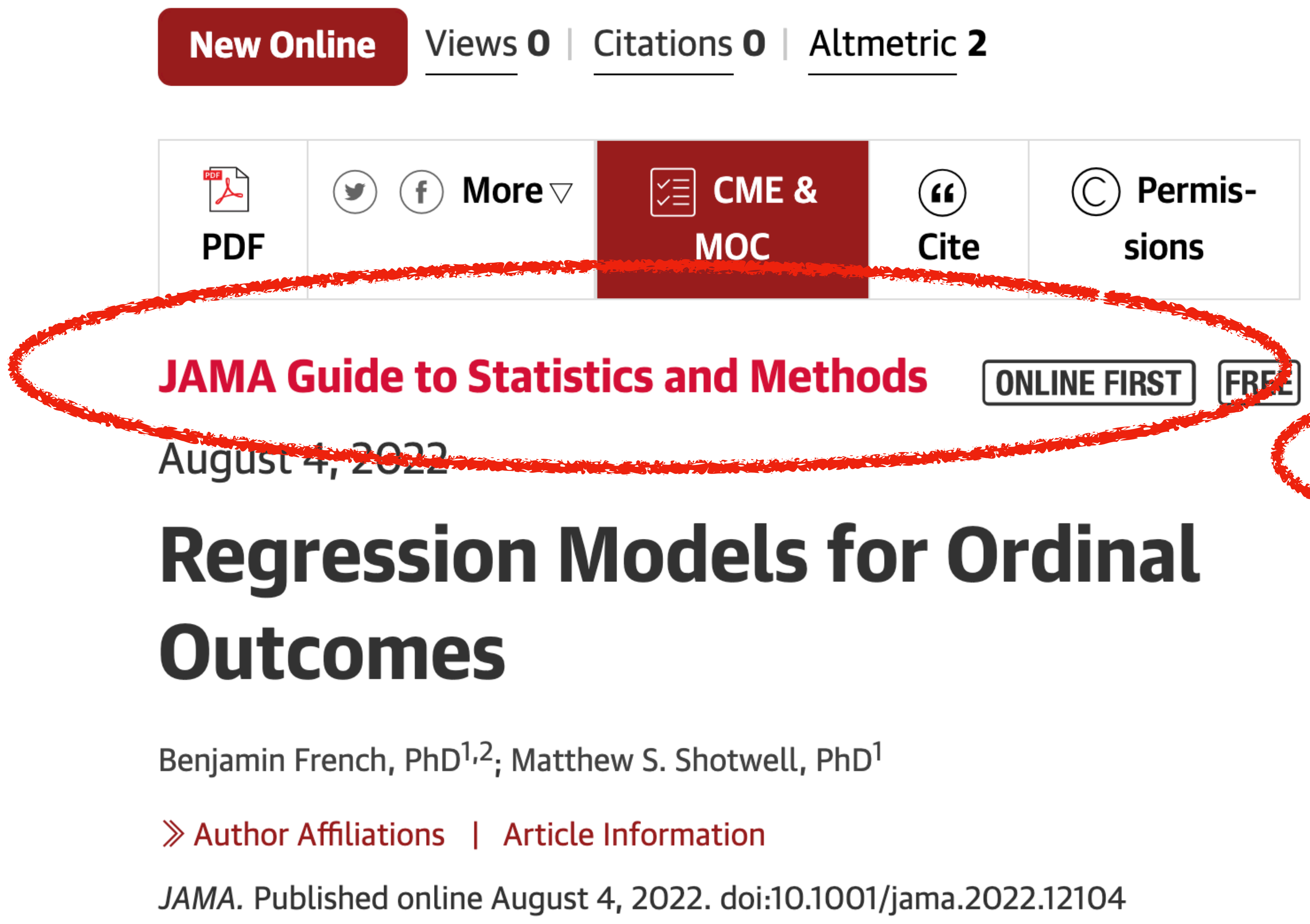
Can chocolate save your life?

EATING JUST THE RIGHT KIND CAN HELP WARD OFF DISEASE, LESLIE BECK WRITES. A11

Statistics in science and society



- Haphazard examples from hundreds of similar stories
- P-values are everywhere
- But statistical issues of sampling, bias, reproducibility, etc. much more prominent
- For example:



Statistics in data science

Kapoor & Narayanan

Leakage and the Reproducibility Crisis in ML-based Science

We argue that there is a reproducibility crisis in ML-based science. We compile evidence of this crisis across fields, identify data leakage as a pervasive cause of reproducibility failures, conduct our own reproducibility investigations using in-depth code-review, and propose a solution.

PRINCETON UNIVERSITY

Draft paper

July 28 online workshop

- blog post emphasizes data leakage — overlap between train and test sets, features proxy for outcome, test set has different dust’s
- “There is a much better known reproducibility crisis in research that uses traditional statistical methods.”

Thanks are due

Christian Genest

Erica Moodie

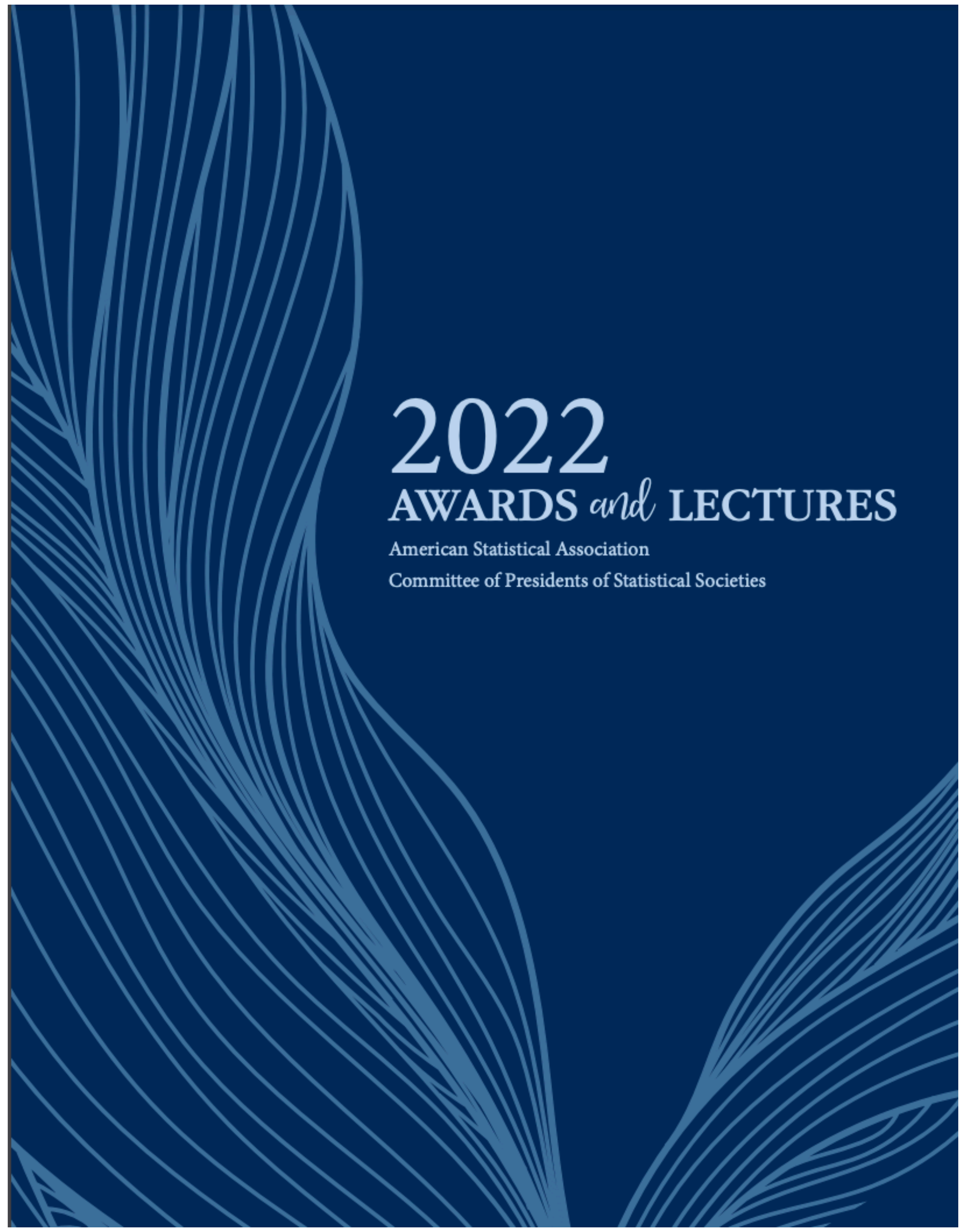
Heather Battey

Yanbo Tang

It's the friends you make




THANK YOU


2022
AWARDS and LECTURES
American Statistical Association
Committee of Presidents of Statistical Societies