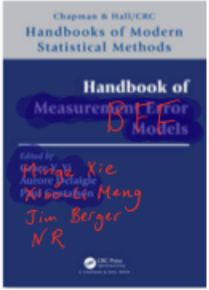
Distributions for parameters

Nancy Reid

University of Toronto

The First Workshop on BFF Inference and Statistical Foundations (BFF 2014)

November 10 – November 14, 2014



- randomized clinical trial to compare two treatments for septic shock
- estimated hazard ratio 0.75 [0.55, 1.02] after adjusting for confounders
- 2-sided p-value 0.06 34.9% vs 43.4% unadjusted

- Discussion: " a peripheral perfusion-targeted resuscitation strategy did not result in a significantly lower 28-day mortality when compared with a lactate level-targeted strategy"

- Abstract: "Among patients with septic shock, a resuscitation strategy targeting normalization of capillary refill time, compared with a strategy targeting serum lactate levels, did not reduce all-cause 28-day mortality."

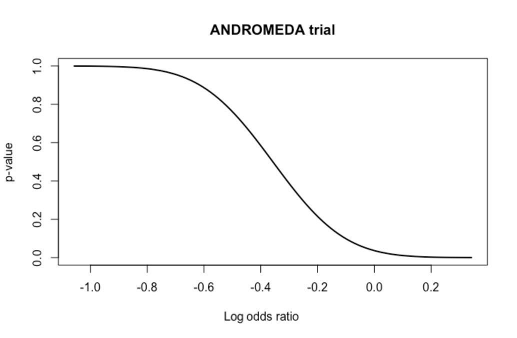
Spiegelhalter, 2019

	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



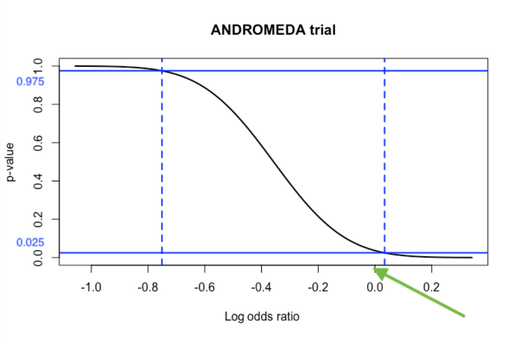
A p-value function

	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



90% confidence interval: $[-0.688, -0.030]$

95% confidence interval: $[-0.751, 0.034]$

99% confidence interval: $[-0.825, 0.107]$

Confidence distributions

SOME PROBLEMS CONNECTED WITH STATISTICAL INFERENCE

By D. R. Cox

Birkbeck College, University of London[1]

1. Introduction. This paper is based on an invited address given to a joint meeting of the Institute of Mathematical Statistics and the Biometric Society at Princeton, N. J., 20th April, 1956. It consists of some general comments, few of them new, about statistical inference.

Since the address was given publications by Fisher [11], [12], [13], have produced a spirited discussion [7], [21], [24], [31] on the general nature of statistical methods. I have not attempted to revise the paper so as to comment point by point on the specific issues raised in this controversy, although I have, of course, checked that the literature of the controversy does not lead me to change the opinions expressed in the final form of the paper. Parts of the paper are controversial; these are not put forward in any dogmatic spirit.

2. Inferences and decisions. A statistical inference will be defined for the

- "... the method of confidence intervals, as usually formulated, gives only one interval at some preselected level of probability"
- "... in ... simple cases ... there seems no reason why we should not work with confidence distributions for the unknown parameter
- "These can either be defined directly, or ... introduced in terms of the set of all confidence intervals"

… confidence distributions

Efron 1993

Biometrika (1993), **80**, 1, *pp.* 3–26
Printed in Great Britain

Bayes and likelihood calculations from confidence intervals

BY BRADLEY EFRON

Department of Statistics, Stanford University, Stanford, California 94305-4065, U.S.A.

SUMMARY

Recently there has been considerable progress on setting good approximate confidence intervals for a single parameter θ in a multi-parameter family. Here we use these frequentist results as a convenient device for making Bayes, empirical Bayes and likelihood inferences about θ. A simple formula is given that produces an approximate likelihood function $L_x^\dagger(\theta)$ for θ, with all nuisance parameters eliminated, based on any system of approximate

- "assigns probability 0.05 to θ between upper endpoints of 0.90 and 0.95 CIs, …"
- "Of course this is logically incorrect, but it has powerful intuitive appeal"
- "… no nuisance parameters [this] is exactly Fisher's fiducial distribution"

Seidenfeld 1992; Zabell 1992

Fiducial distribution

528 *Dr Fisher, Inverse probability*

Inverse Probability. By R. A. FISHER, Sc.D., F.R.S., Gonville and Caius College; Statistical Dept., Rothamsted Experimental Station.

[*Received* 23 July, *read* 28 July 1930.]

$$\mathrm{d}f = -\frac{\partial}{\partial\theta}F(Y,\theta)\mathrm{d}\theta$$

fiducial distribution of θ for a given statistic Y

"It is not to be lightly supposed that men of the mental calibre of Laplace and Gauss … could fall into error on a question of prime theoretical importance, without an uncommonly good reason"

- y has density from a location model — $f(\cdot)$ known

$$f(y - \theta)$$

- equivalently — generating model

$$y = \theta + e, \quad e \sim f(\cdot)$$

- observe $y = y^o$,

$$\theta = y^o - e$$

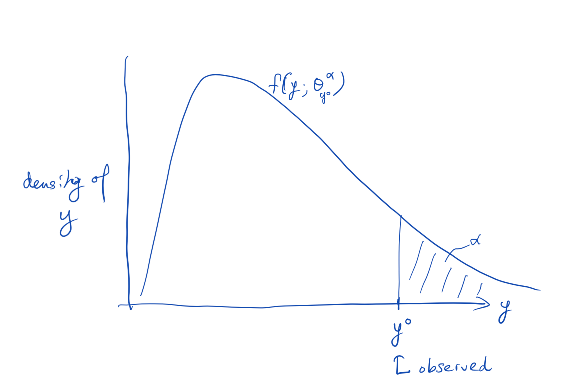
- variation in e induces variation in θ — probabilities for e generate probabilities for θ

- in general (continuous) models, fix a percentile point $p = F(y, \theta)$
- this creates a relationship between y and θ

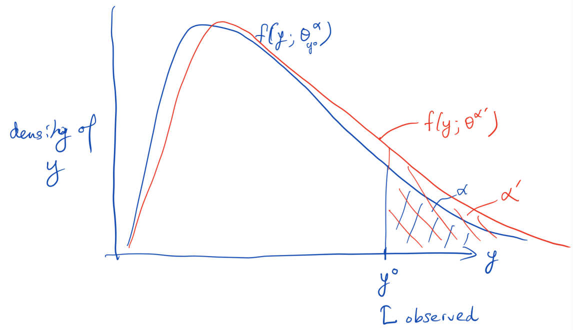
$$\frac{\partial}{\partial y}F(y, \theta) + \frac{\partial}{\partial \theta}F(y, \theta) = 0 \qquad \mathrm{d}f = -\frac{\partial}{\partial \theta}F(y, \theta)\mathrm{d}\theta$$

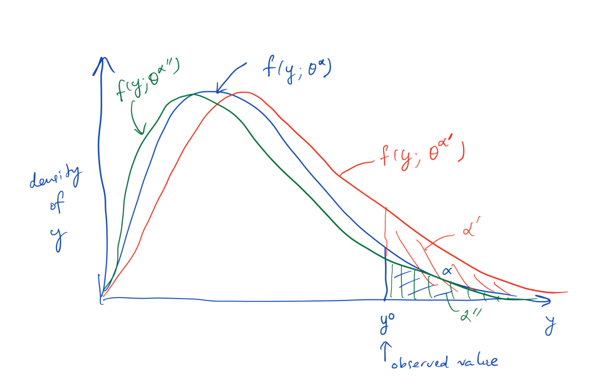
... fiducial/confidence distribution

Efron 1993

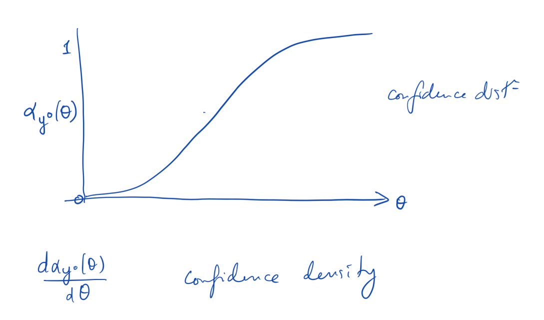


… fiducial/confidence distribution





$$\theta_{y^\circ}(\alpha) : \alpha \longrightarrow \theta$$



$$\alpha_{y^\circ}(\theta) : \theta \longrightarrow \alpha$$

inverse function

Distributions for parameters

- significance function $p(\theta) = \mathrm{pr}(y \geq y^o \mid \theta)$
- confidence distribution $\alpha_{y^o}(\theta) = \theta_{y^o}^{-1}(\alpha)$
- fiducial density $df = -(\partial F/\partial\theta)(Y;\theta)d\theta$

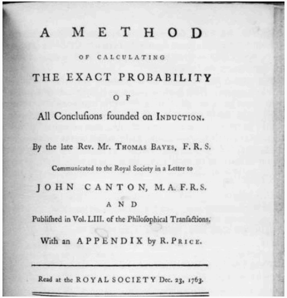
These are not "real" probability distributions

Don't obey the rules of probability calculus

Don't adapt easily to multi-parameter models

Bayesian posteriors

- introduce probability function $\pi(\theta)$ on parameter space
- consider statistical model as conditional model for data, given parameter



A METHOD
OF CALCULATING
THE EXACT PROBABILITY
OF
All Concluſions founded on INDUCTION.
By the late Rev. Mr. THOMAS BAYES, F. R. S.
Communicated to the Royal Society in a Letter to
JOHN CANTON, M. A. F. R. S.
AND
Publiſhed in Vol. LIII. of the Philoſophical Tranſactions,
With an APPENDIX by R. PRICE.
Read at the ROYAL SOCIETY Dec. 23, 1763.

- use rules of conditional probability to derive

$$\pi(\theta \mid y^o) = f(y^o; \theta)\pi(\theta) \Big/ m(y^o)$$

- directly extends to complex data y_n and complex parameter spaces

$$\Pr(\theta \in A \mid y_n^o) = \int_A \pi(\theta \mid y_n^o) d\theta$$

e.g. $y_n^o = (y_1^o, \ldots, y_n^o)$

ANDROMEDA, revisited

Zampieri et al 2020

	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

- 28-day mortality, Cox proportional hazards model
- adjustment for 5 baseline covariates
- estimated hazard ratio 0.75 (0.55, 1.02)

- Bayesian re-analysis based on logistic regression
- focus on posterior probability $\beta < 0$

log odds ratio

- equivalently P(hazard ratio $< 1 \mid$ data)

- added random effect for center, used default priors for covariates, change to logistic regression

ANDROMEDA, revisited

- under a range of normal priors for the log odds-ratio
- the posterior probability that the odds-ratio is less than 1 — treatment is beneficial
- ranges from 0.94 to 0.99 — most pessimistic to most optimistic prior

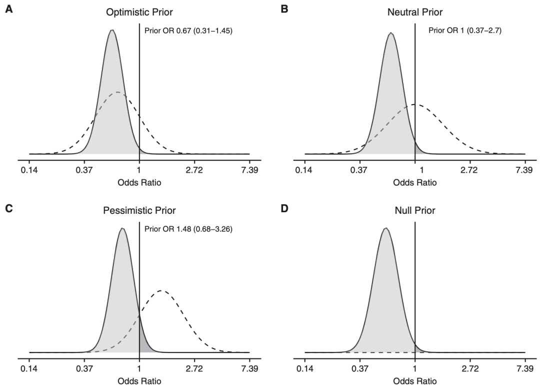


Figure 1. (A–D) Prior distributions for the odds ratio (OR) of the intervention (dashed lines). Posterior distributions of the ORs are shown by the solid lines. The light gray areas indicate the areas associated with benefit for peripheral perfusion–targeted resuscitation (i.e., OR < 1) and the dark gray areas the areas associated with harm (i.e., OR > 1). The text inside each frame reports the median and lower and upper 95% credible limits for the priors of the effect of the intervention for 28-day mortality.

see also van Zwet et al. 2021
used empirical prior
posterior prob 0.91

Why do we want distributions on parameters?

- inference is intuitive
- combines easily with decision theory
- de-emphasizes point estimation and arbitrary cut-offs

- "it's tempting to conclude that μ is more likely to be near the middle of this interval, and if outside, not very far outside"

Cox 2006

- "assigns probability 0.05 to θ lying between the upper endpoints of the 0.90 and 0.95 confidence intervals, etc."

Efron 1993

- all inference statements ~~are~~ **seem to be** probability statements about unknowns

Nature of Probability

Cox 2006; R & Cox 2015

- probability to describe physical haphazard variability — aleatory/empirical
    - probabilities represent features of the “real” world in somewhat idealized form
    - subject to empirical test and improvement
    - conclusions of statistical analysis expressed in terms of interpretable parameters
    - enhanced understanding of the data generating process

- probability to describe the uncertainty of knowledge — epistemic
    - measures rational, supposedly impersonal, degree of belief, given relevant information — Jeffreys
    - measures a particular person’s degree of belief, subject typically to some constraints of self-consistency — Ramsey, de Finetti, Savage

BFF Handbook

- posterior distribution 1763
- fiducial probability 1930
- confidence distribution 1958
- structural probability 1964
- significance function 1991
- belief functions 1967

- objective Bayes, empirical Bayes Berger, Efron
- generalized fiducial inference Hannig et al., Lang, Vovk, Taraldsen & Lindqvist
- confidence distributions and confidence curves, data fusion Thornton & Xie, Hector et al.
- approximate significance functions Davidson & R
- inferential models Shafer, Gong, Liu & Martin

high-dimensional inference, model selection, computation

Objective Bayes

- there are many proposals for priors meant to be non-informative
- examples include reference, default, matching, vague, ... priors
- a popular choice is Jeffreys' prior $\pi(\theta) \propto |i(\theta)|^{1/2}$ — expected Fisher information

- what interpretation do we put on the posterior distribution? — empirical? epistemic?

- we may avoid the need for a different version of probability by appeal to a notion of calibration — Cox 2006, R & Cox 2015
- as with other measuring devices, within this scheme of repetition, probability is defined as a hypothetical frequency
- it is unacceptable if a procedure yielding high-probability regions in some non-frequency sense are poorly calibrated

- such procedures, used repeatedly, give misleading conclusions

Bayesian Analysis, V1(3) 2006

... objective Bayes

- there are many proposals for priors meant to be non-informative
- examples include reference, default, matching, vague, ... priors
- a popular choice is Jeffreys' prior $\pi(\theta) \propto |i(\theta)|^{1/2}$ expected Fisher information
- some versions may not be correctly calibrated e.g. Jeffreys'
- requires checking in each example
- calibrated versions must be targetted on the parameter of interest Fraser 2011
- only in very special cases can calibration be achieved for more than one parameter in the model, from the same prior
- the simplicity of a fully Bayesian approach to inference is lost

Gelman 2008; PMM LW

... objective Bayes

- the simplicity of a fully Bayesian approach to inference is lost — Gelman 2008
- for example

$$\pi(\psi \mid y) = \int_{\psi(\theta)=\psi} \pi(\theta \mid y) d\theta, \quad \text{for any } \psi : \Theta \searrow \Psi$$

lower dimension

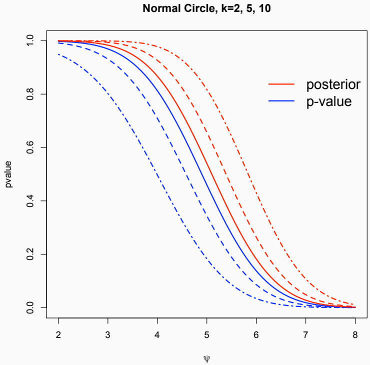
- the prior can have unexpected influence on the posterior
- even if they are seemingly noninformative — objective Bayes fails

- Stein's example: — sequence model

$$\begin{aligned} y_i &\sim N(\theta_i, 1/n), \quad i = 1, \ldots, k \\ \pi(\theta_i) &\propto 1 \\ \pi(\theta \mid y) &\propto N(y, I_k/n) \end{aligned}$$

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

  $s(\psi) = \mathrm{pr}\{\chi_k^2(n||y||^2) \geq n\psi^2\}$



global-local shrinkage priors (horseshoe) shrink the posterior in the right direction
reference and targetted priors do the same

- calibrated posterior distributions must be targetted on the parameter of interest

- matching priors set out this requirement explicitly

  defined by calibration of posterior quantiles

- reference priors are also targetted

  although with a different goal than calibration

- vague priors, hierarchical priors, weakly informative priors, ... are not (usually) targetted on a particular parameter of interest

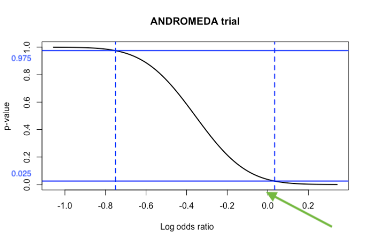
- ANDROMEDA trial used default priors on regression coefficients, except odds-ratio

  and a selection of log-normal priors on odds-ratio

Targetting on parameter

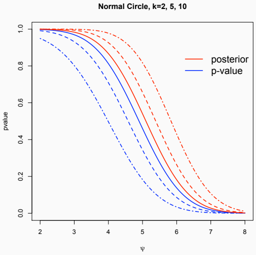
Example 1: 2×2 table

Based on conditional distribution of odds-ratio, given marginal totals



Example 2: $y_i \sim N(\theta_i, 1/n)$

Based on marginal distribution of Σy_j^2



BFF Handbook

- posterior distribution 1763
- fiducial probability 1930
- confidence distribution 1958
- structural probability 1964
- significance function 1991
- belief functions 1967

- objective Bayes, empirical Bayes Berger, Efron
- generalized fiducial inference Hannig et al., Lang, Vovk, Taraldsen & Lindqvist
- confidence distributions and confidence curves, data fusion Thornton & Xie, Hector et al.
- approximate significance functions Davidson & R
- inferential models Shafer, Gong, Liu & Martin

Confidence Distribution

Thornton & Xie, Xie & Singh; Hjort & Schweder

- any function $H : \mathcal{Y} \times \Theta \to (0, 1)$ which is
- a cumulative distribution function of θ for any $y \in \mathcal{Y}$
- and has correct coverage: $H(Y, \theta) \sim U(0, 1)$ — $Y \sim f(\cdot; \theta)$

- sufficiently general to encompass bootstrap distribution and many standard likelihood quantities
- recent examples include robust meta-analysis and identification of change-points
  Xie et al. 2011; Cunen et al. 2017; Hannig & Xie 2012
- can serve as a bridge between BFF approaches — Thornton & Xie 2022
- “as long as one can create confidence regions of all levels”

- avoids the problem of targetting by assuming a method for this has already been developed — often likelihood ratio statistic or deviance

Generalized Fiducial

Hannig et al 2016

Location model:

- data-generating equation $Y_i = \theta + e_i$; with fiducial inversion $e_i = y_i^o - \theta$
- $e_i - \bar{e} = y_i^o - \bar{y}^o, \quad i = 1, \ldots, n$ is known once the data is available
- relevant distribution is conditional on $e_i - \bar{e}$
- enables application of one-dimensional inversion

General model:

- $y = G(u, \theta), \quad u$ has known distribution — $y_1, \ldots, y_n$
- given observed data y^o find the inverse $\theta = Q_{y^o}(U)$ — Efron 1993
- generalized fiducial distribution of θ: $Q_{y^o}(U^*)$ — U^* independent copy of U

- inverse only exists if θ and Y have same dimension
- Hannig introduces an auxiliary function to adjust for dimension reduction

$$r(\theta; y^o) \propto f(y^o, \theta) J(y^o, \theta)$$

- $$r(\theta; y^o) = \frac{f(y^o, \theta)\, J(y^o, \theta)}{\int f(y^o, \theta)\, J(y^o, \theta) d\theta}$$

$y = G(u, \theta)$

- $$J(y^o, \theta) = D\left\{\frac{\partial}{\partial\theta} G(u, \theta)\right\}\Big|_{u^o}$$

D for generalized Determinant

- if we used ordinary determinant we would have

$|X^T X|^{1/2}$

$$J(y^o, \theta) = \left|\frac{dy}{d\theta}\right|$$

- well adapted to simulation, and applied in many interesting and complex models
- no guarantees on calibration, beyond first-order; needs to be checked in each case

Significance function

- focus on parameter of interest

many arguments point to the need for this

- $y \in \mathbb{R}^n, \quad \theta \in \mathbb{R}^p, \quad \psi \in \mathbb{R}$

- specifies a particular path to targetting, via two-stage dimension reduction

- $n \downarrow p$ using an approximate location model

$$df = -\frac{\partial}{\partial\theta}F(y,\theta)d\theta = -\frac{\partial}{\partial\theta}F(y,\theta)\frac{f(y^o,\theta)}{f(y^o,\theta)} = \overbrace{f(y^o,\theta)}^{\text{Likelihood}}\left.\frac{dy}{d\theta}\right|_{y^o}$$

- $p \downarrow 1$ using an approximate exponential model

- asymptotic theory applied to this exponential model gives a ?unique? pivotal quantity for inference about parameter of interest ψ
- which is a modification of the likelihood ratio statistic

r^*_ψ

So much jargon!

- marginal posteriors are not guaranteed to be calibrated
- i.e. to have guaranteed performance in sampling from the model

- although they will agree to first order, i.e. normal approximations

- fiducial, confidence, significance distributions for parameters are not guaranteed to be calibrated
- but the necessity for the user to target the parameter of interest is perhaps clearer at the outset

- e.g. confidence distribution approach asks user to identify, from the model, a quantity that measures the parameter of interest

focus parameter, Hjort & Schweder, 2016

Nature of Probability?

- Bayes / objective Bayes — epistemic / empirical
- generalized fiducial inference — empirical
- confidence distributions and confidence curves — epistemic / empirical
- approximate significance functions — empirical

ANDROMEDA

Zampieri et al 2020

Table 1. Odds Ratio, 95% Credible Interval, Probability That the Odds Ratio Is below Given Thresholds, and Absolute Difference between Groups

	28-d Outcome			90-d Outcome			
Prior	OR (95% Credible Interval)	Probability OR < 1 (Probability OR < 0.8)	Absolute Difference (95% Credible Interval)*	OR (95% Credible Interval)	Probability OR < 1 (Probability OR < 0.8)	Absolute Difference (95% Credible Interval)*	Reason for Prior Use
Optimistic	0.61 (0.41 to 0.90)	99% (92%)	−9% (−17% to −1%)	0.69 (0.47 to 1.01)	97% (79%)	−7% (−16% to 2%)	Considers an OR of 0.67 for the intervention (slightly more conservative than the effect size ANDROMEDA-SHOCK was powered to detect), while considering that there is still a 15% probability that the intervention was harmful
Neutral	0.65 (0.43 to 0.96)	98% (85%)	−7% (−16% to 1%)	0.74 (0.50 to 1.08)	94% (66%)	−5% (−14% to 4%)	Has a mean OR of 1 (i.e., absence of effect) and 50% probability of benefit and 50% of harm from the intervention
Pessimistic	0.74 (0.50 to 1.09)	94% (66%)	−5% (−13% to 3%)	0.83 (0.57 to 1.21)	83% (42%)	−3% (−11% to 6%)	Opposite values of the optimistic prior; considers a very pessimistic scenario in which the intervention is harmful but still acknowledges a 15% chance that the intervention might be beneficial
Null	0.59 (0.38 to 0.92)	98% (91%)	−8% (−17% to 1%)	0.69 (0.45 to 1.07)	95% (74%)	−6% (−15% to 4%)	No prior information is considered

Definition of abbreviation: OR = odds ratio.
*Refers to a simple model adjusted only for study arm and not for all predictors.

- initial analysis: "Observed hazard ratio of 0.75 was not statistically significantly different from 1 at level 0.05"
- $p = 0.06$, 95% confidence interval $(0.55, 1.02)$

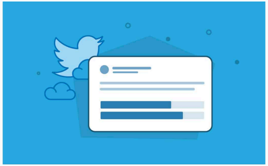
- translation: "new therapy has no benefit"

- second analysis: "Posterior probability that odds ratio is less than one is 0.98"
- posterior credible interval $(0.43, 0.96)$

van Zwet et al. 2020

- translation: "new therapy is better"

- is more study needed?



Discussion

- dichotomizing conclusions based on p-values is not a good idea
- statistical science is more nuanced than that
- science rarely advances on the basis of a single study

- posterior distributions need to be treated with care
- they can depend heavily on the prior, even when it seems uninformative

- several current versions of fiducial inference: confidence, significance, generalized fiducial
- "making a Bayesian omelette without cracking the Bayesian eggs"

Needs in applications

- something that works
- gives 'sensible' answers
- not too sensitive to model assumptions
- computable in reasonable time
- provides interpretable parameters

Role of Foundations

Cox & R, 2015

- avoid apparent discoveries based on spurious patterns
- shed light on the structure of the problem
- obtain calibrated inferences about interpretable parameters
- provide a realistic assessment of precision
- understand when/why methods work/fail

Thank you!


PLACE BONAVENTURE


THE 2022 ISBA
WORLD MEETING
June 26th - July 1st, Hotel Bonaventure, Montreal, Canada