Advantages of a wholly Bayesian approach to assessing efficacy in early drug development: a case study

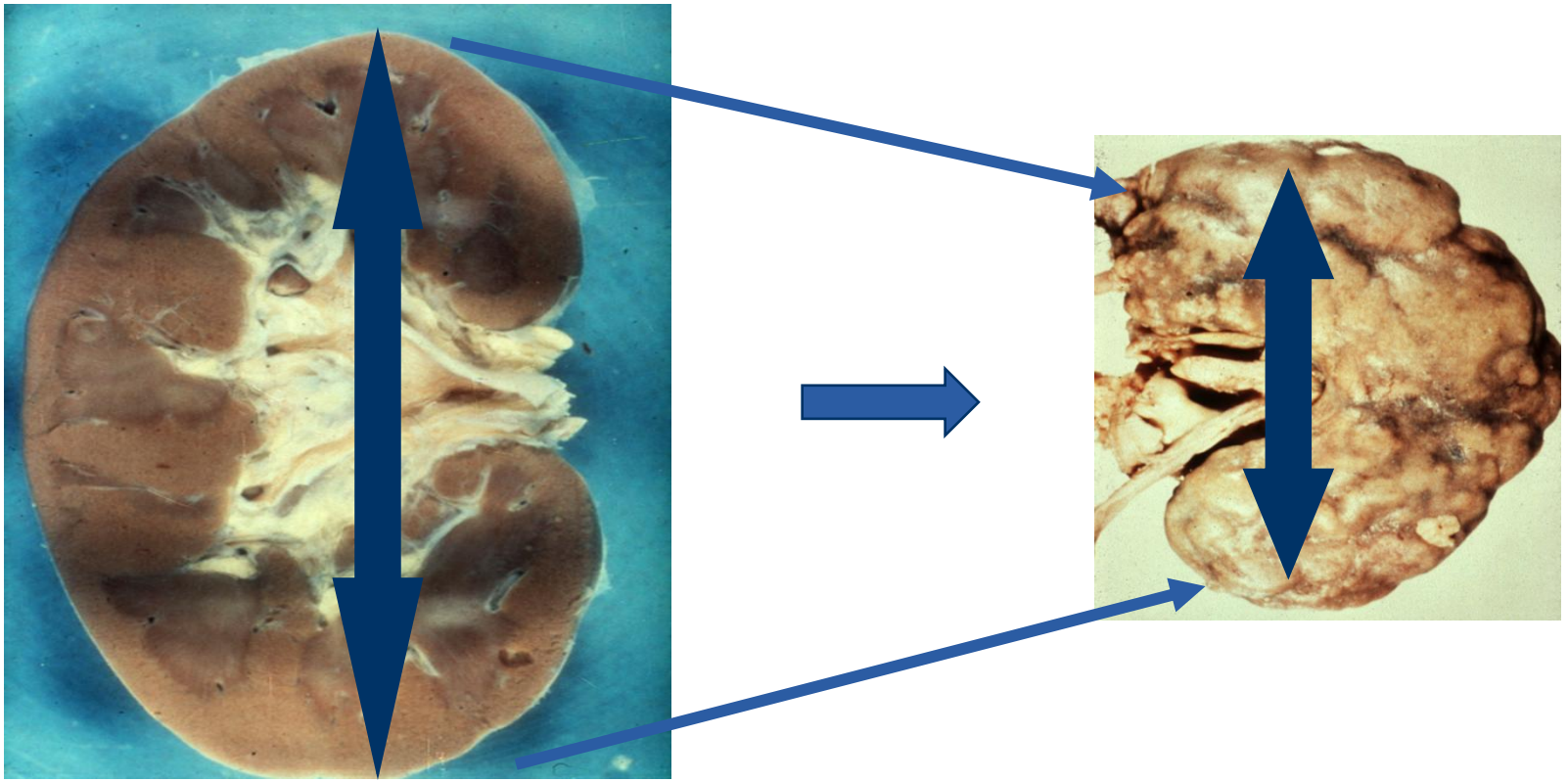
Phil Woodward, Ros Walley, Claire Birch, Jem Gale

Outline

- **Background**
- **Prior information**
- **Decision criteria**
- **Theory: normal prior with normal likelihood**
- **Assessment of study design**
  - Approximate posterior distribution for treatment effect
  - Design characteristics
  - Impact of study design on beliefs as to treatment effect
- **Interim analysis**
- **Interactions with ethics boards and regulators**
- **Conclusions**
  - Can we make better decisions using informative treatment priors?

Background: Chronic Kidney Disease

- 20 million Americans - 1 in 9 US adults - have chronic kidney disease (CKD).
- Diabetes is the fastest growing risk factor for CKD, and almost 40% of new dialysis patients have diabetes.
- CKD can be detected by increases in urine albumin, serum creatinine and BUN.
- CV disease is the major cause of death for all people with CKD.



Background to study

- **Proof of concept study for diabetic nephropathy**
  3 month duration plus follow-up → parallel group study
  All subjects remain on standard of care

- **Primary endpoint: urinary albumin creatinine ratio**
  Very variable
  Work on log scale

- **Bayesian design allows for relevant probability statements to be made at the end of the study**

- **In addition, informative prior for placebo response (standard of care)**
  large published studies
  reduced the required sample size
  led to choice of unequal randomisation 3:1 active: placebo

- **Interim analysis to allow early stopping for futility, based on predictive probabilities**

Pfizer

Prior information

- **Two uses of priors:**
  - Design priors
    - to assess the study design only e.g. unconditional probability of success
  - Analysis priors
    - for use in analysis of the data (should be included in assessment of design)

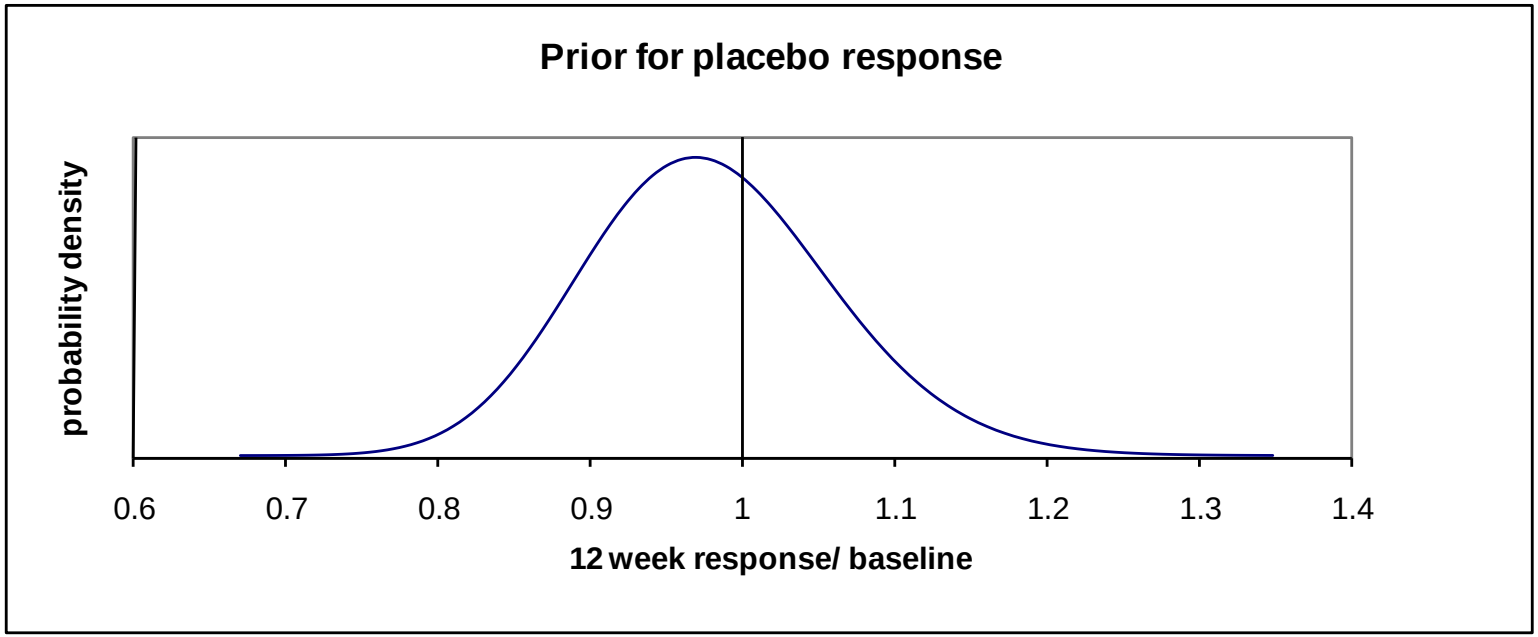
- **In this example,**
  - Design priors: treatment effect and variance
  - Analysis prior: placebo response

- **Found from**
  - Published studies and internal data
  - Eliciting views from experts

- **Sensitivity to priors will be assessed**

Prior for placebo response

- Used both for design and in analysis
- Obtained by elicitation:
  Expected to be between [0.85, 1.05]
  75% distribution set to be within this range
- Consistent with the literature
- Expected to be equivalent to ~100 placebo subjects



Prior for placebo response

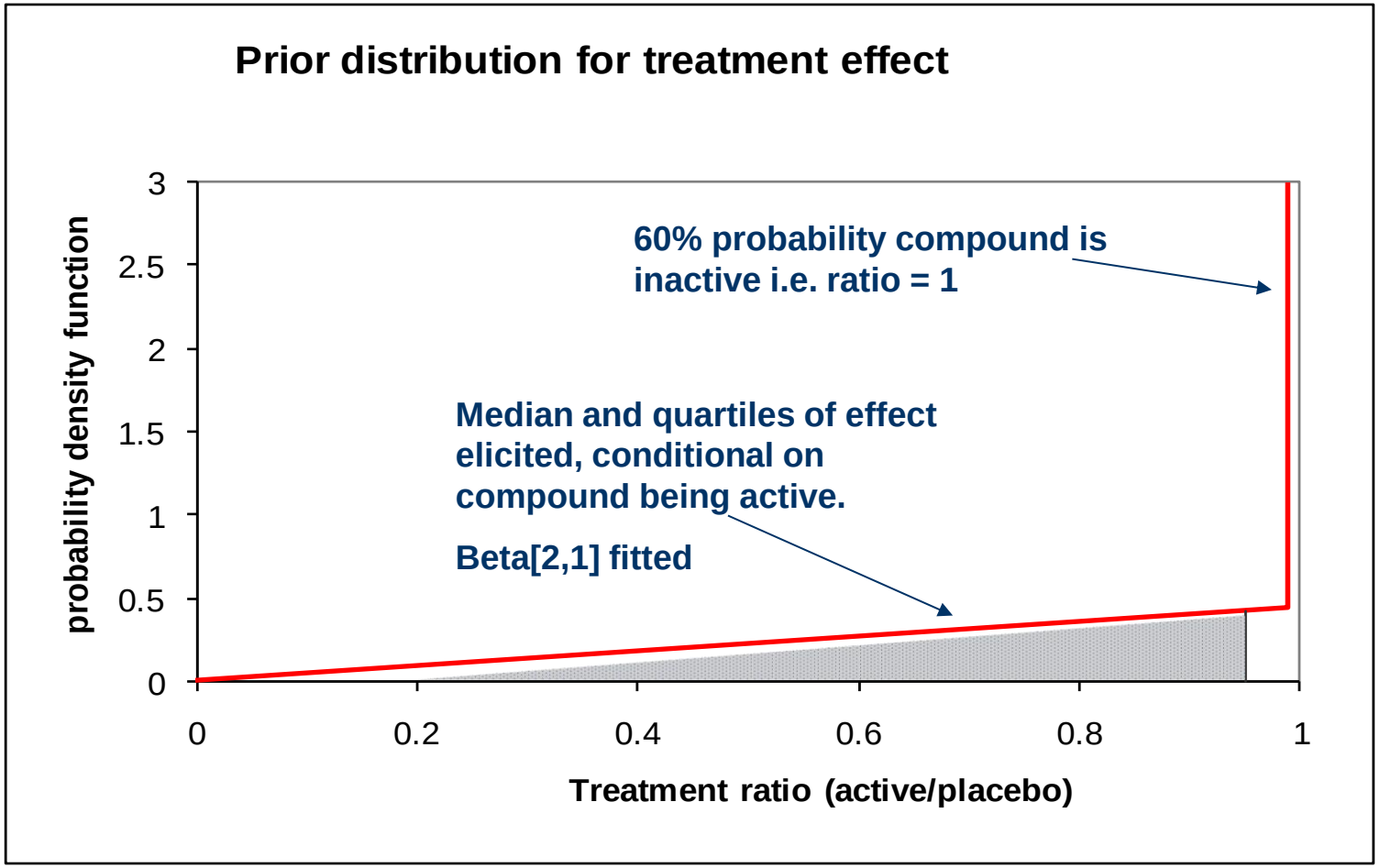
observed placebo mean response

- **Empirical criticism of priors**
  - George Box suggested a Bayesian p-value
    - Prior predictive distribution for future observation
    - Compare actual observation with predictive dist.
    - Calculate prob. of observing more extreme
    - Measure of conflict between prior and data
  - But what should you do if conflict occurs?
    - At least report this fact
    - Greater emphasis on analysis with a vaguer prior
  - Robust prior approach
    - Formally model doubt using a mixture prior

or could use a heavy tailed distribution e.g. t_4

Prior for treatment effect

- Used only to assess the design
- Elicited from experts



Decision criteria

- ***In terms of 12 week data:***

  Criterion 1. At least 90% sure that the treatment ratio (active/placebo) < 1

  Criterion 2. At least 67% sure the treatment ratio < 0.8

Will revisit these statements in light of using "flat" analysis prior for treatment effect

- ***In terms of n-fold reduction from baseline data:***

  Using the following notation for the posterior estimates on the log scale:

  δ treatment difference, calculated as

  log(active) – log(placebo)

  μ_δ posterior mean for δ

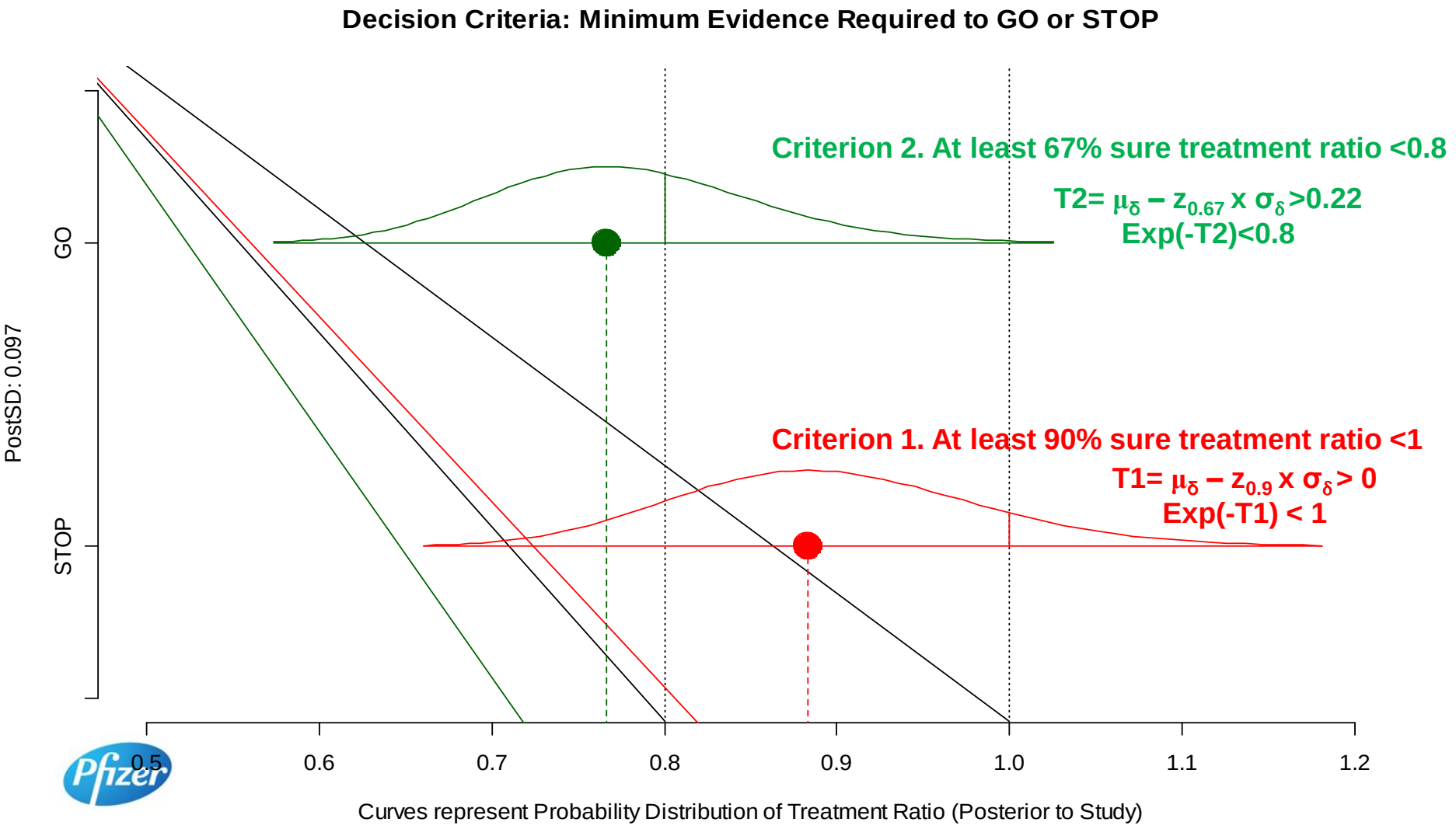
  σ_δ posterior standard deviation of δ

  $T1 = \mu_\delta - z_{0.9} \cdot \sigma_\delta$; Criterion 1. T1 > 0

  $T2 = \mu_\delta - z_{0.67} \cdot \sigma_\delta$; Criterion 2. T2 > -ln(0.8) = 0.22

Pfizer

Illustration of Decision Criteria



Notation and assumptions

- **Working on natural log scale and assuming**
  known variance
  no covariates (in assessing design and interim analysis only)
  data are normally distributed, with independent errors

- **Model:**
  Placebo $x_{1j} = \gamma + \varepsilon_{ij}$; $j=1 \ldots n_1$, $\varepsilon_{1j} \sim N(0, \sigma^2)$
  Active $x_{2j} = \gamma + \delta + \varepsilon_{ij}$; $j=1 \ldots n_2$, $\varepsilon_{2j} \sim N(0, \sigma^2)$

- **Sample means:**
  Placebo mean, $\bar{x}_1$
  Active mean, $\bar{x}_2$

Model was actually "outlier robust", mixture of Normals with 5% weight given to highly dispersed "outlier" distribution

- **Priors**
  Uninformative prior for δ, $p(\delta) \propto 1$
  Informative prior for placebo response, $\gamma \sim N(g, \sigma^2/m)$

Pfizer

Posterior distribution for δ

- **Posterior distribution for δ: normally distributed with**
  mean **and variance**

$$\bar{x}_2 - \frac{\left(\frac{g}{\sigma^2/m} + \frac{\bar{x}_1}{\sigma^2/n_1}\right)}{\left(m/\sigma^2\right) + \left(n_1/\sigma^2\right)}$$

$$\left(\sigma^2/n_2\right) + \left\{\left(m/\sigma^2\right) + \left(n_1/\sigma^2\right)\right\}^{-1}$$

- **Change notation for variance of prior for mean from σ^2/m to ω^2**
- **Mean for posterior distribution for δ can be expressed**

$$\bar{x}_2 - \underbrace{\frac{\left(\frac{1}{\sigma^2/n_1}\right)}{\left(1/\omega^2\right) + \left(n_1/\sigma^2\right)}}_{k_1} \bar{x}_1 - \underbrace{\frac{\left(\frac{g}{\omega^2}\right)}{\left(1/\omega^2\right) + \left(n_1/\sigma^2\right)}}_{k_2}$$

Probability of success

- ***In the analysis at end of the study***
  - We will assess criteria of the form:

$$T = \mu_\delta - z_\alpha.\sigma_\delta > \Delta$$

  - Approx. equivalent to using:

$$\bar{x}_2 - k_1.\bar{x}_1 - k_2 - z_\alpha.\sigma_\delta > \Delta$$

- ***At the design stage:***
  - We know the predictive distributions of $\bar{x}_1$ and $\bar{x}_2$, conditional on γ, δ and σ , so can estimate the probability of success:

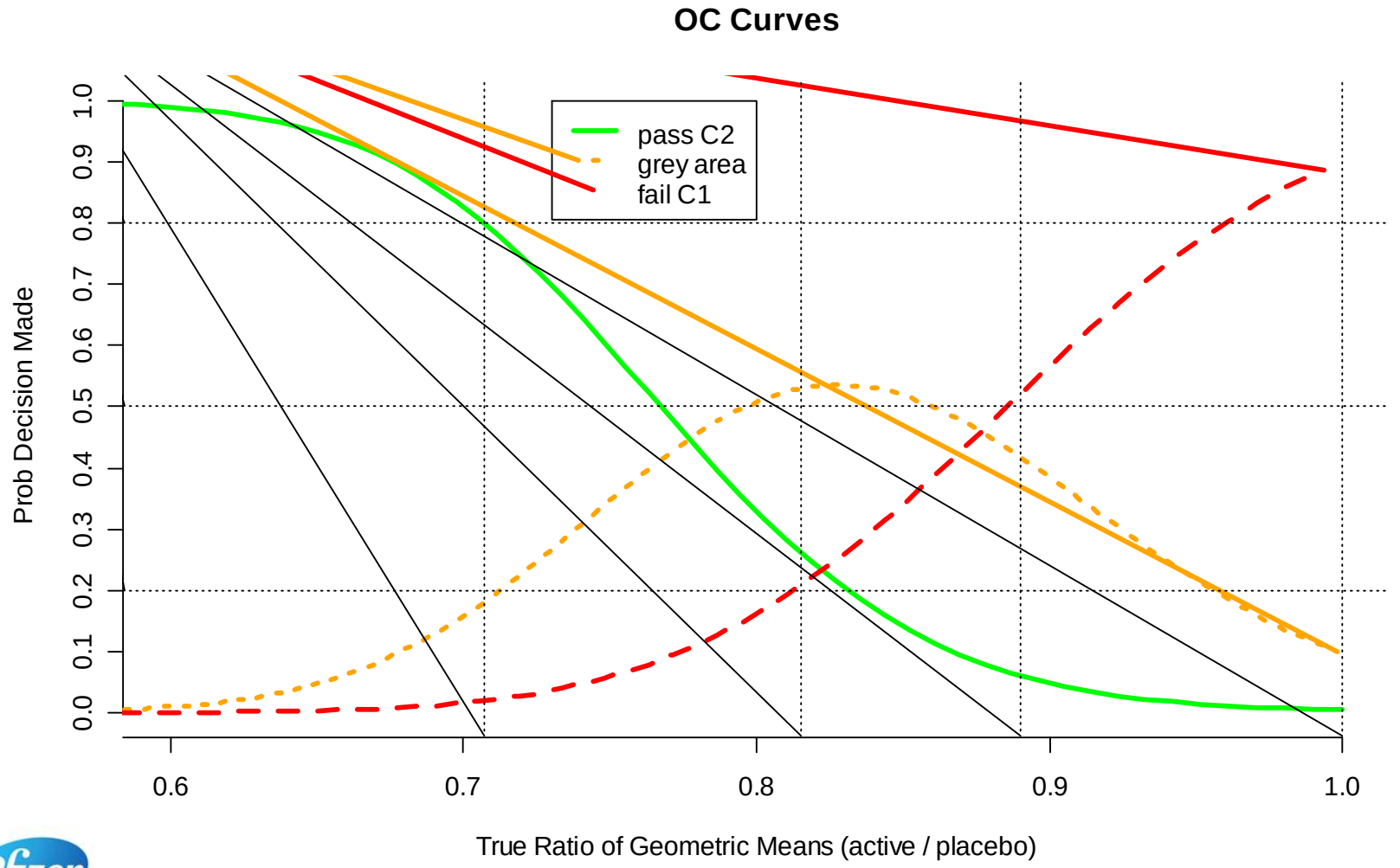
$$P(\bar{x}_2 - k_1.\bar{x}_1) > \Delta + k_2 + z_\alpha.\sigma_\delta$$

  - To obtain unconditional probabilities, by simulation we integrate the conditional probabilities with respect to the design priors.

$$P(\text{success} \mid \delta, \sigma) = \int P(\text{success} \mid \gamma, \delta \text{ and } \sigma) p(\gamma)\, d\gamma$$

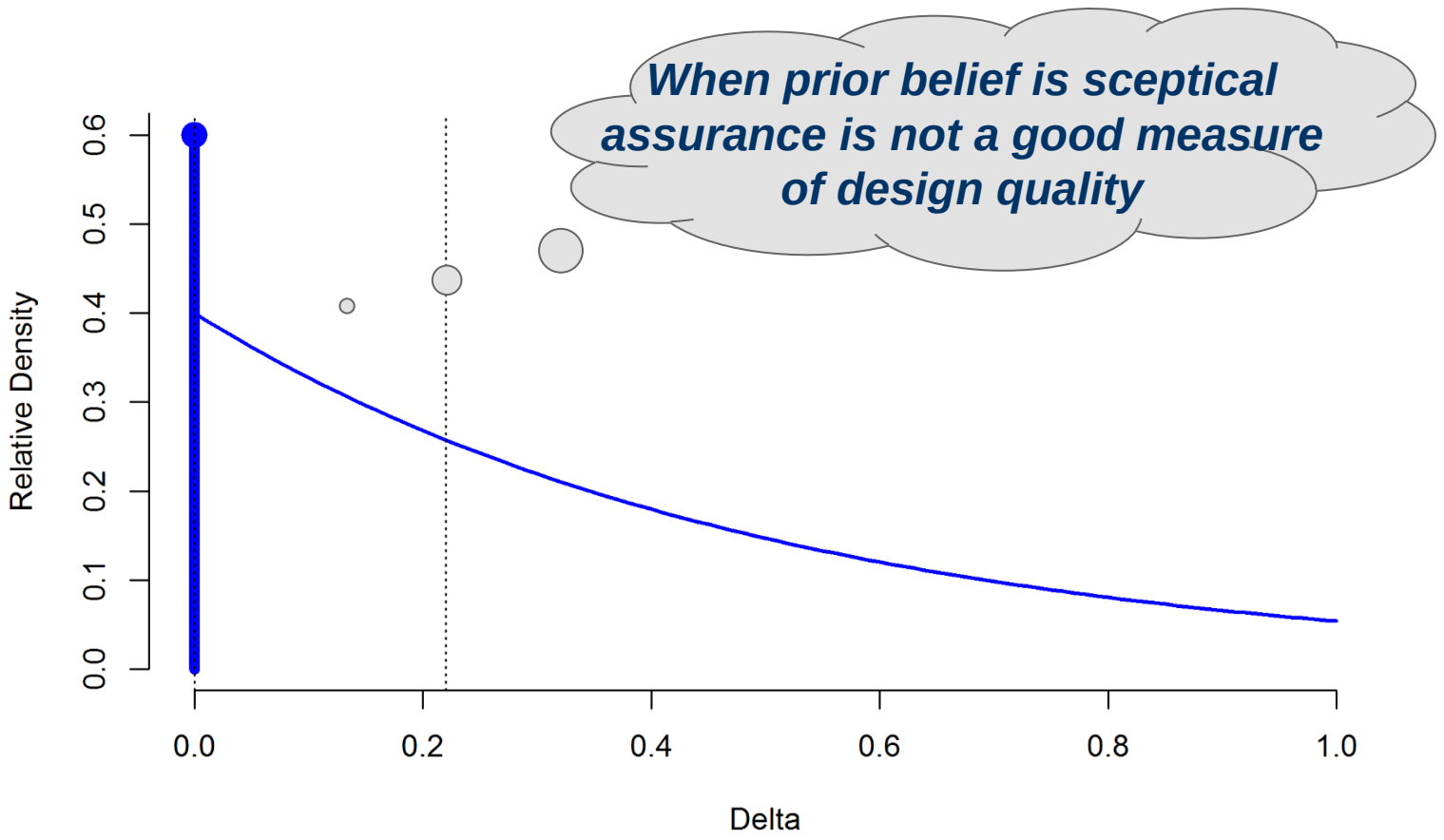
$$P(\text{success}) = \int\int\int P(\text{success} \mid \gamma, \delta \text{ and } \sigma)\, p(\gamma)\, p(\delta)\, p(\sigma)\, d\gamma\, d\delta\, d\sigma$$

Design Characteristics

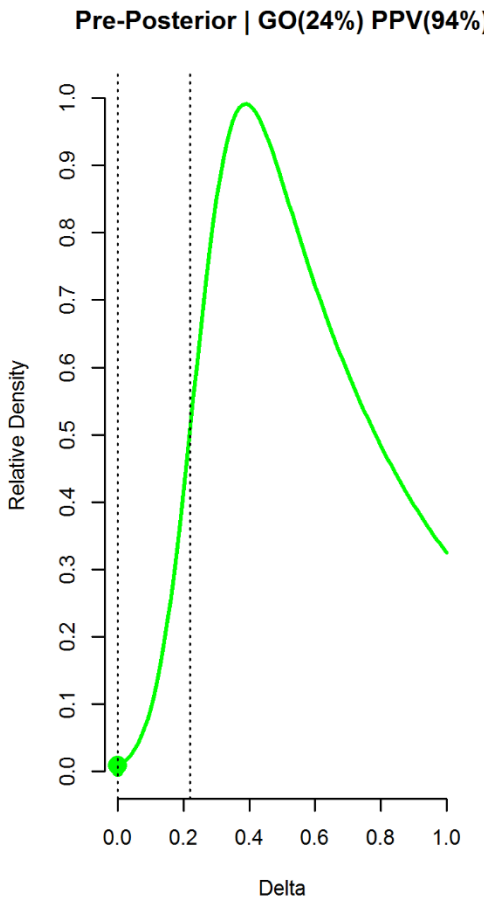


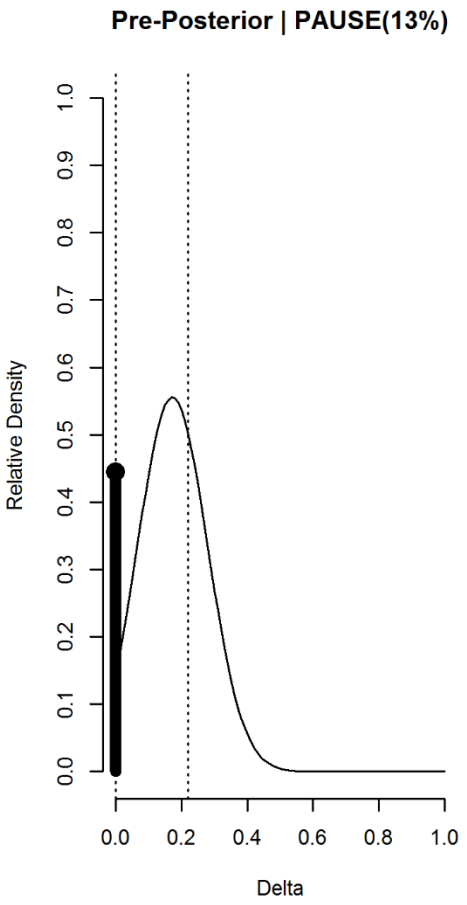
These probabilities are conditional on δ but not on γ or σ

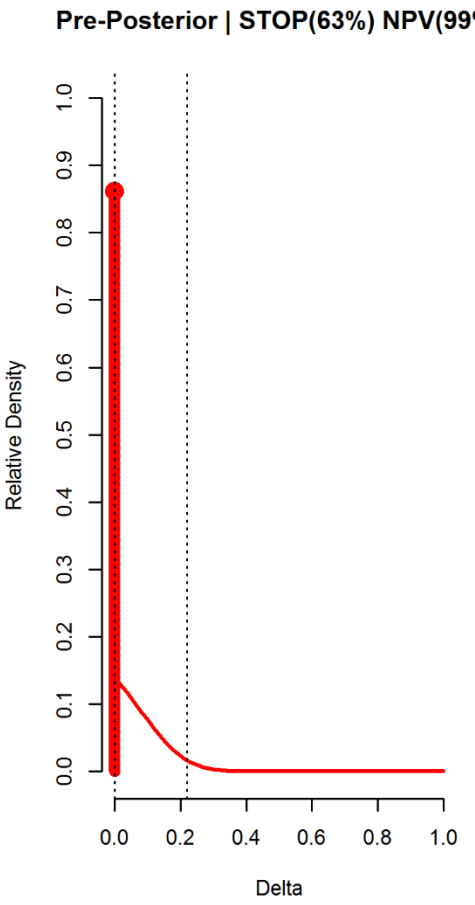
Design treatment prior (delta on log-scale)



Impact of study design on beliefs as to treatment ratio







Interim Analysis

- Proposal: To carry out an internal analysis when 25% subjects have completed, analysing end of treatment data
- Stopping rule: Stop at interim if the predictive probability of passing criterion 1 (lower hurdle) is less than 20%
- Potential saving: At the end of the interim, we estimate there will be 50 subjects left to recruit
- Implication: If stop decision at interim, small probability after all subjects have completed that we will just pass criterion 1.

Posterior distributions conditional on interim

- Observed placebo data: mean, $\bar{y}_1$, and no. observations, r_1
- Remaining placebo data: mean, $\bar{z}_1$, and no. observations, s_1

- Prior placebo data: mean g, and equivalent no. of observations, $m=\sigma^2/\omega^2$
  Assume m known

- Posterior distribution for placebo mean **will be** normally distributed with
  mean $\dfrac{m.g + r_1.\bar{y}_1 + s_1.\bar{z}_1}{m + r_1 + s_1}$
  and variance $\dfrac{\sigma^2}{(m + r_1 + s_2)}$

$$= \frac{m.g + r_1.\bar{y}_1}{m + r_1 + s_1} + \frac{s_1}{m + r_1 + s_1} \cdot \bar{z}_1$$

Known at interim

Known at interim

Predictive Distribution

Normally distributed with mean $(mg + r_1y_1)/(m + r_1)$ *and variance*

$$\frac{\sigma^2}{s_1} + \frac{\sigma^2}{m + r_1}$$

Assessing probability of success at interim

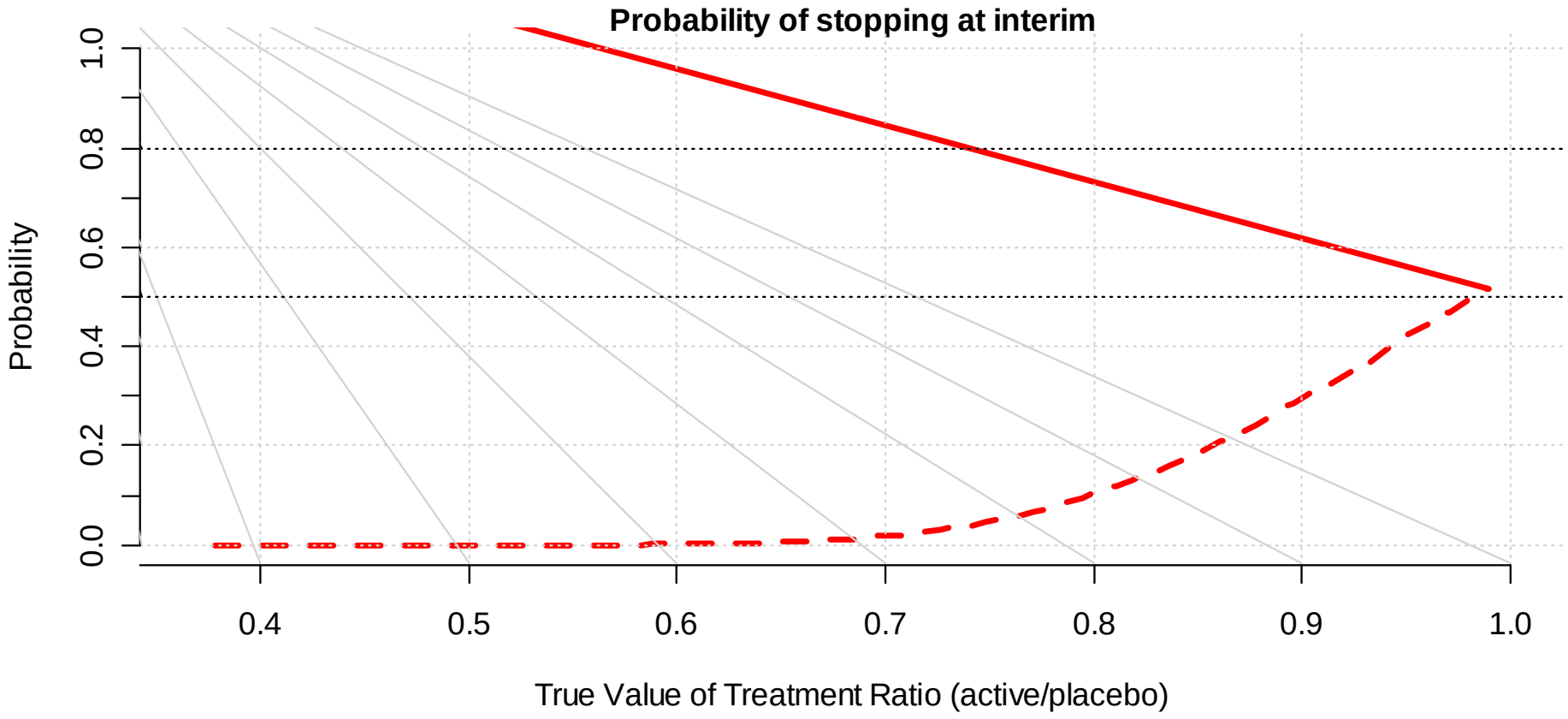
- Similarly can construct a posterior distribution for the active mean conditional on the data.

- Recall at the end of the study we will assess criteria of the form:

$$T = \mu_\delta - z_\alpha.\sigma_\delta > \Delta$$

- From the joint distribution of these means we can compute the predictive distribution for the treatment difference, δ, conditional on the interim data, and thus calculate the probability this criterion will be satisfied

Probability of Stopping at Interim



Interactions with Ethics Boards and Regulators

- **Non-standard approach → anticipate additional questions as well as standard ones e.g. method of randomisation**
- **No need to panic!**
  Problem with translation
  More information → informed view
  No delay over and above other questions
- **Lack of understanding versus wanting more detail**
  e.g. functional forms of priors
- **Whole power curve versus power at minimally clinically relevant difference**
- **Analogy with frequentist approach**
- **Level of detail**
  No need to include priors that are used just to assess the design and give unconditional probabilities of success

Pfizer

Bayesian thinking
confidence is not “chance”

Small p-values are interpreted as evidence of real effect

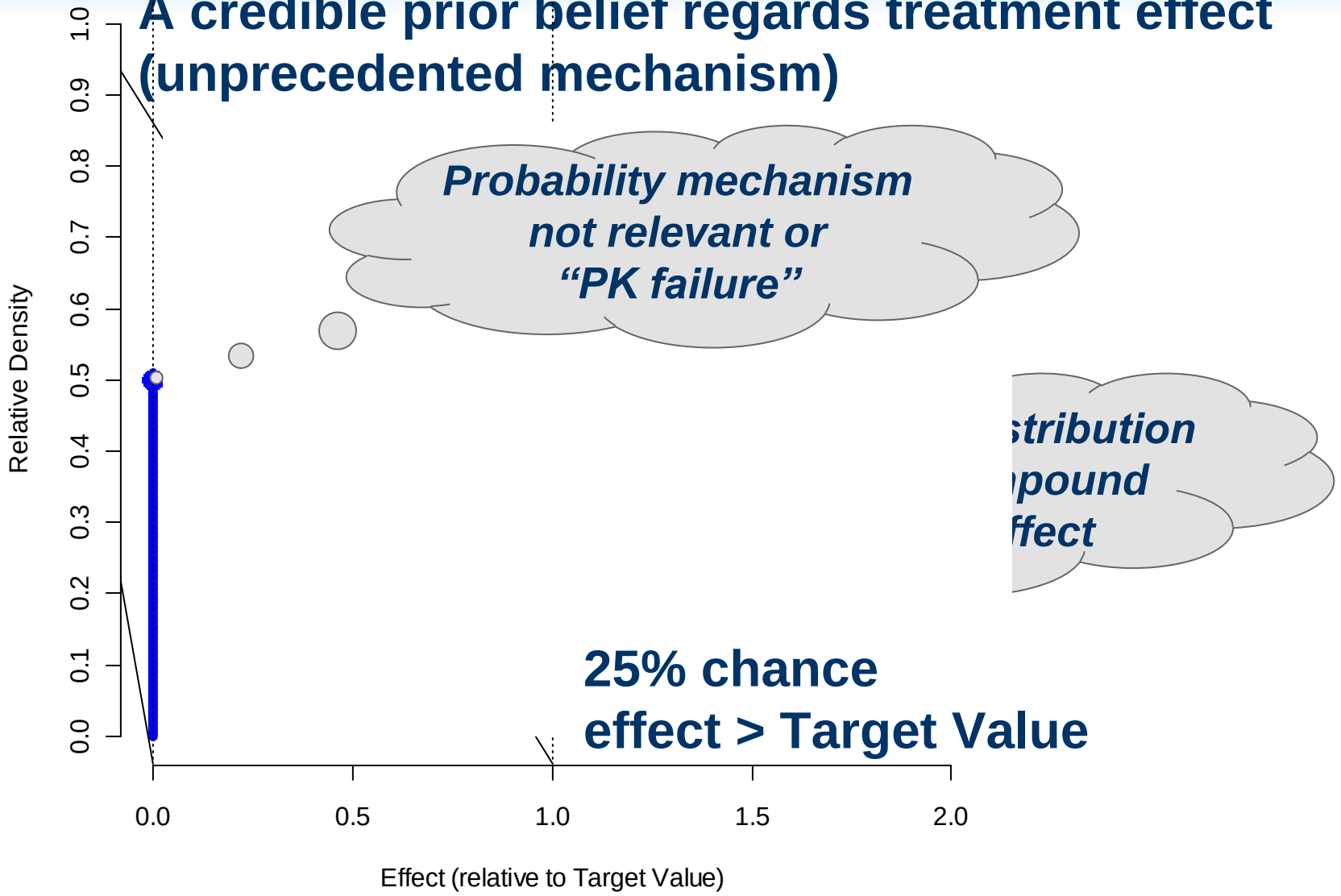
But how much confidence do they provide in ED studies?

Are statements like “90% confidence effect > 0” understood?

Does it matter?

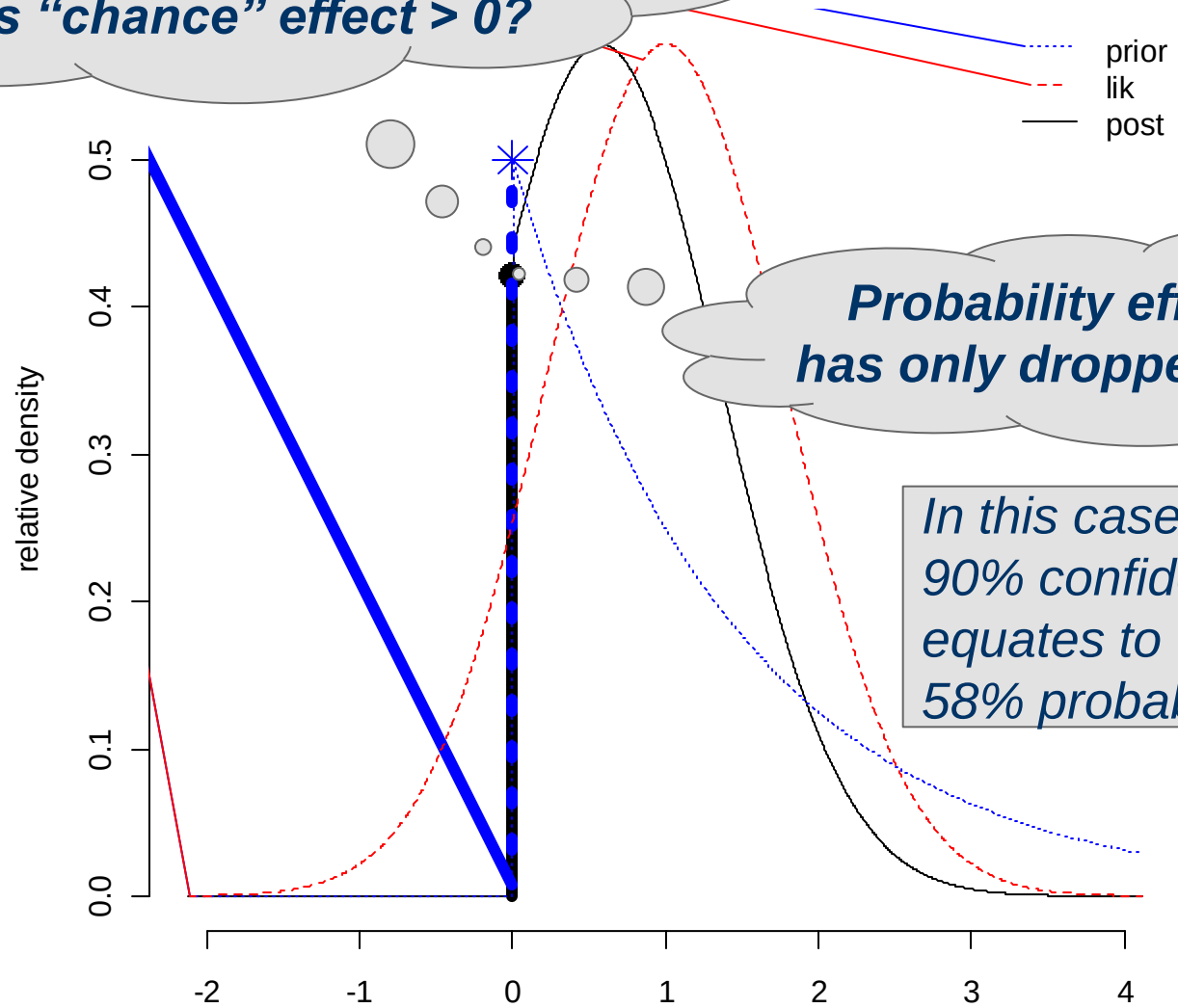
Bayesian thinking
confidence is not “chance”

A credible prior belief regards treatment effect (unprecedented mechanism)



Bayesian thinking
confidence is not “chance”

Suppose observed effect = Target
90% confident effect > 0
What is “chance” effect > 0?

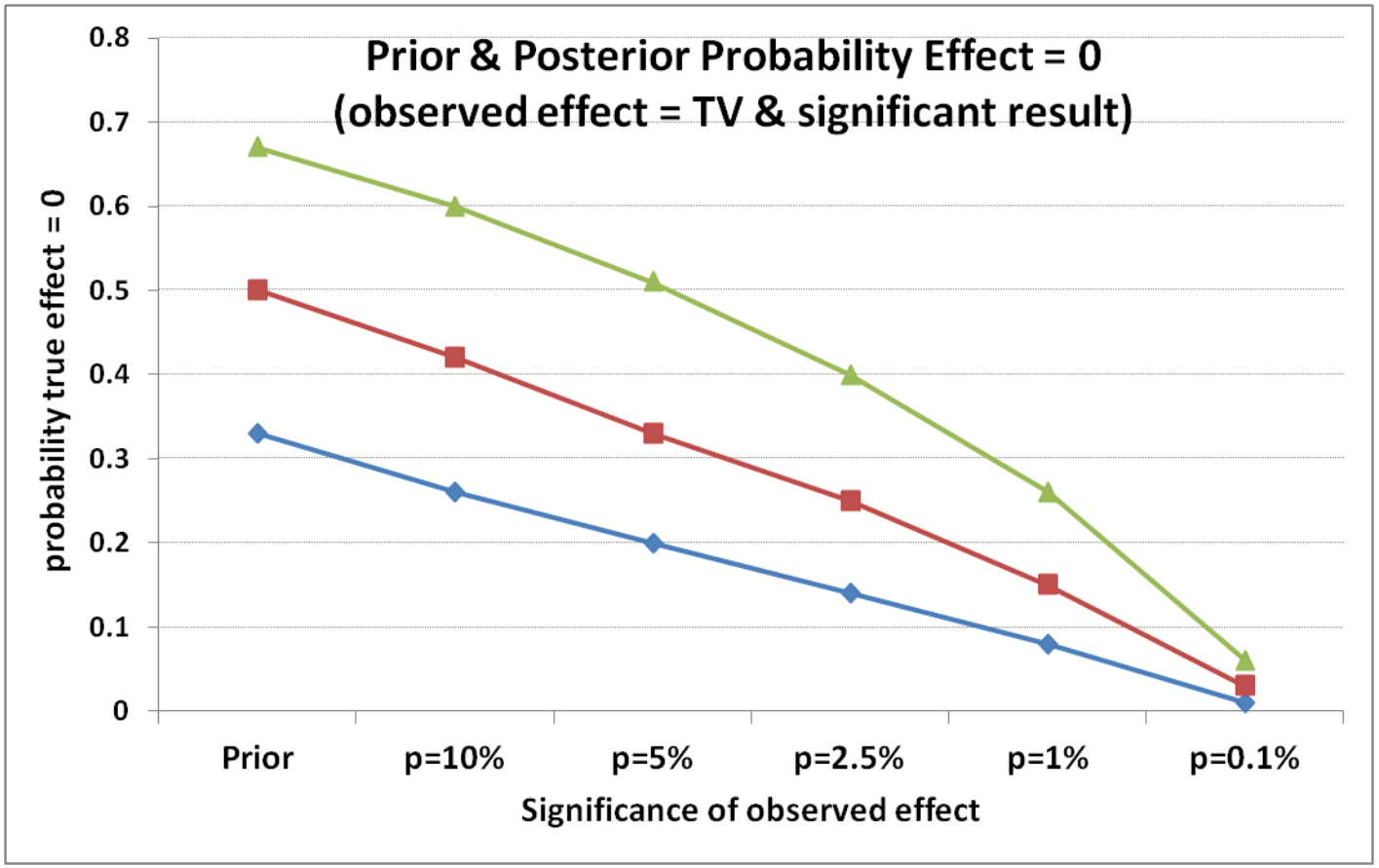


Probability effect = 0
has only dropped to 0.42

In this case
90% confidence effect > 0
equates to
58% probability effect > 0

Bayesian thinking
confidence is not “chance”

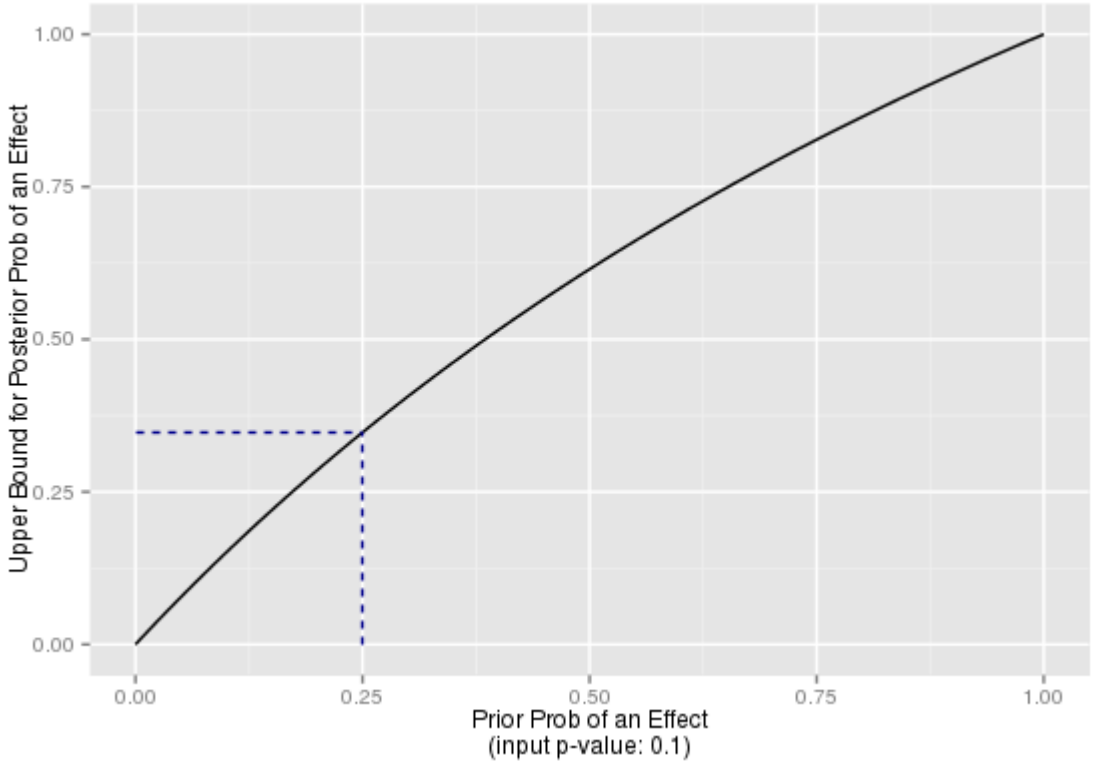
“Extraordinary claims require extraordinary evidence”



Bayesian thinking
confidence is not “chance”

Calibration of *p Values for Testing Precise Null Hypotheses*
T.Sellke, M.J.Bayarri, and J.O.Berger
The American Statistician, February 2001, Vol. 55, No. 1

They showed that “confidence” was optimistic no matter what shape the prior distribution of non-zero effects.



Conclusions

- **Sample size**
  - High variability of primary endpoint ⇒ low power or large sample size
  - Published data can be used for an informative prior, reducing sample size

- **Utility of interim analysis.**
  - Resource saving if stopping
  - Accelerate future work if interim analysis suggests compound efficacious
  - Probability of being able to make stop or accelerate decision

- **Bayesian framework**
  - Novel approach ⇒ Education (team, management, ethics/regulators)
  - At design stage
    - Incorporation of priors
    - Unconditional probabilities of success
    - Flexibility in selecting decision criteria
    - Leads to more thorough thinking
  - At end of study
    - Flexible decision criteria and probability statements

Pfizer