# What Does a Bayesian Owe a Frequentist?

Frank E Harrell Jr

Department of Biostatistics
Vanderbilt University School of Medicine

ACDRS Washington DC 2016-04-21

# Background



- As measures of statistical evidence $P$-values have many problems
- Attempt to use one evidence measure for two conflicting ideas:
  - Measure strength of the observed evidence
  - Measure how often believing such things would be wrong
- Not possible to measure both with the same statistic
- Only one is relevant before data collected
- Only one is relevant after data collected

Blume 2011

# Background, *continued*



- Backwards time order/information flow
- Analogy with sensitivity and specificity in diagnostic testing
  - Pretend that diagnosis is all-or-nothing
  - Condition on ultimate diagnosis
  - Pretend that sens and spec do not vary with subject characteristics
  - Must adjust for workup bias
  - When using Bayes rule to get Prob(disease) adjustment cancels out
  - In cohort study can directly estimate Prob(disease) bypassing sens & spec

# Background, *continued*



- Multiplicity mess; frequentist approach has no principled, prescriptive strategy
- Evidence for A vs. B discounted for comparing C to D
- Complexity of $P$-value adjustment in sequential testing; hard to adjust point estimates and CLs for early termination
- Frequentists interpret results by inferring "what would have occurred following results that were not observed at analyses that were never performed" (Emerson 1995)
- Frequentists: Could we have gotten here another way?
- Bayesians: Given we are here what is the evidence for $X$?

# A Classic



*J. R. Statist. Soc.* A (1994)
**157**, *Part* 3, *pp.* 357–416

## Bayesian Approaches to Randomized Trials

By DAVID J. SPIEGELHALTER†,

*Medical Research Council Biostatistics Unit, Cambridge, UK*

LAURENCE S. FREEDMAN

*National Cancer Institute, Bethesda, USA*

and MAHESH K. B. PARMAR

*Medical Research Council Cancer Trials Office, Cambridge, UK*

[*Read before* The Royal Statistical Society *at a meeting organized by the* Medical Section *on Wednesday, February 16th, 1994, the President*, Professor D. J. Bartholomew, *in the Chair*]

# Skepticism



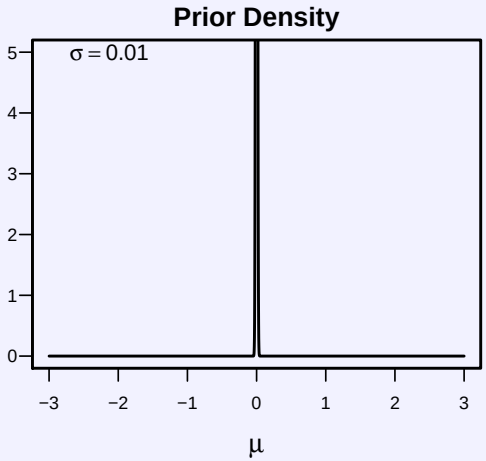
"theoretical possibility of 'sampling to a foregone conclusion', in which asymptotically we are guaranteed at some point to obtain a significant result even if the null hypothesis is true (McPherson, 1974). Cornfield (1996) argued that, if you are worried about this, it must reflect consideration of the null hypothesis as having a distinct probability of being true. It follows that we should put a lump of probability (however small) on the null hypothesis, and then the phenomenon will not occur." (Spiegelhalter, Freedman, and Parmar 1994)

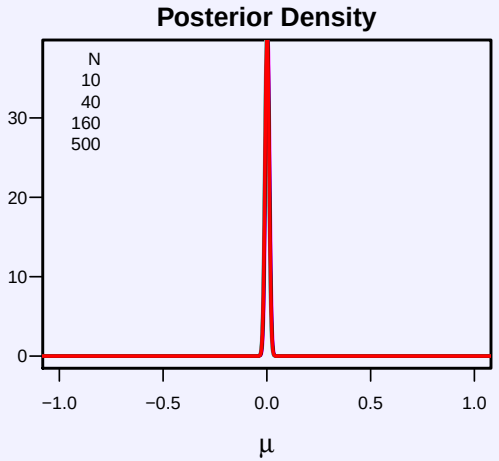
- Skepticism must be incorporated through the prior, not through creating different cutoffs for the posterior
- Skepticism is only about **this** treatment and not about previously failed treatments
- Extreme skepticism can be overwhelmed by extreme evidence
- Community of priors (Kass & Greenhouse, 1989)

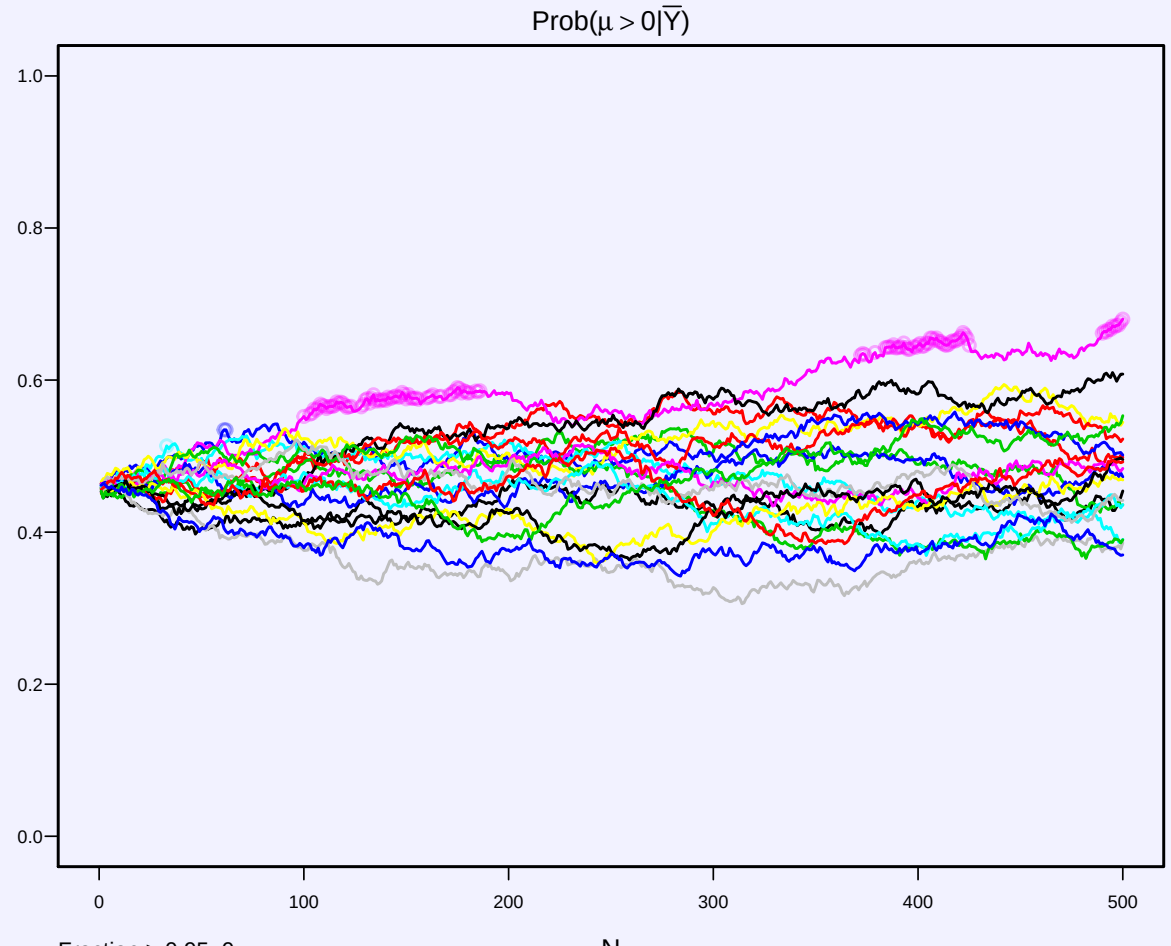
# Simulations



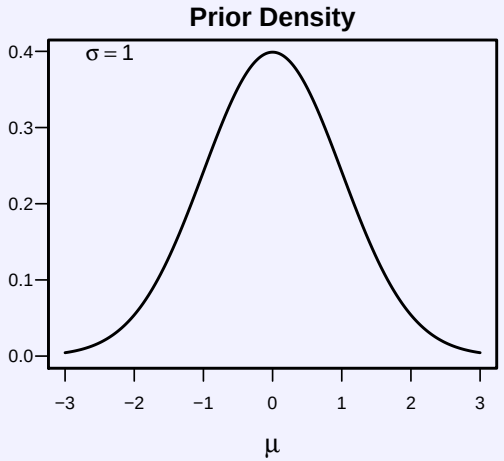
- One-arm study, $Y \sim n(0, 1)$, final $N = 500$
- Analysis after each subject has $Y$ measured (500 looks)
- Efficacy: $\mu > 0$
- Priors for $\mu$: $(w, 1 - w)$ mixtures of $n(0, \sigma_1^2)$ and $n(0, \sigma_2^2)$
- 1000 trial simulations for each set; $\mu = 0$ for all
- Posterior densities for first 10
- Posterior paths for first 20

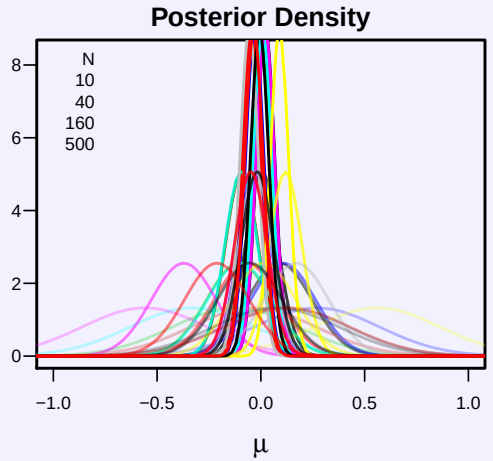


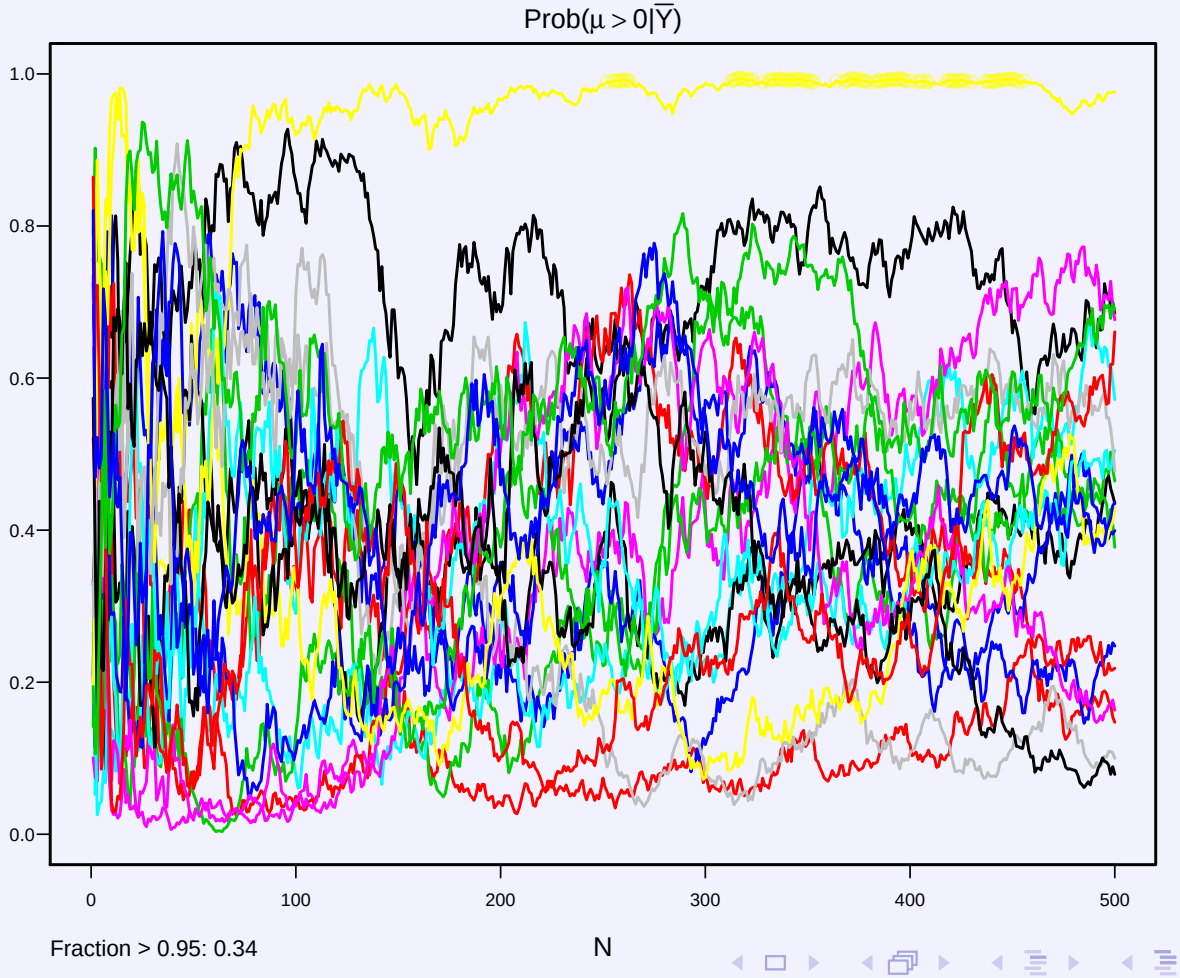



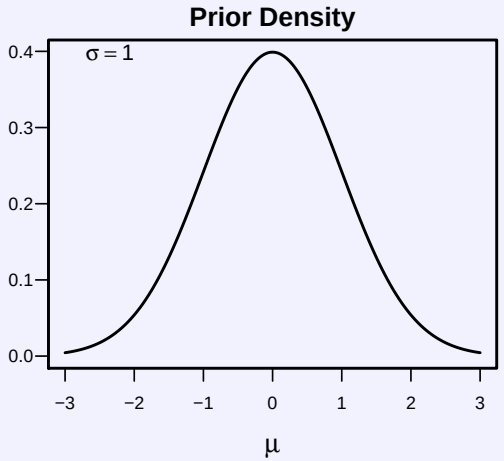

Fraction > 0.95: 0

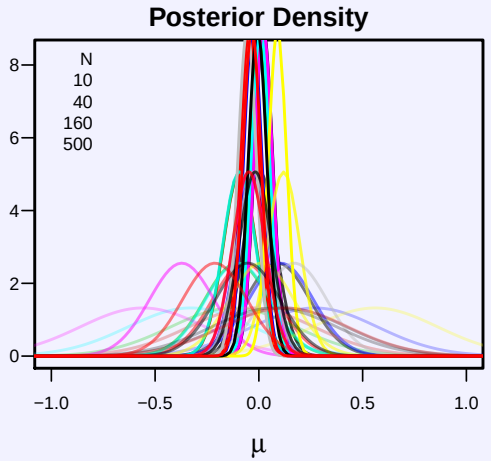


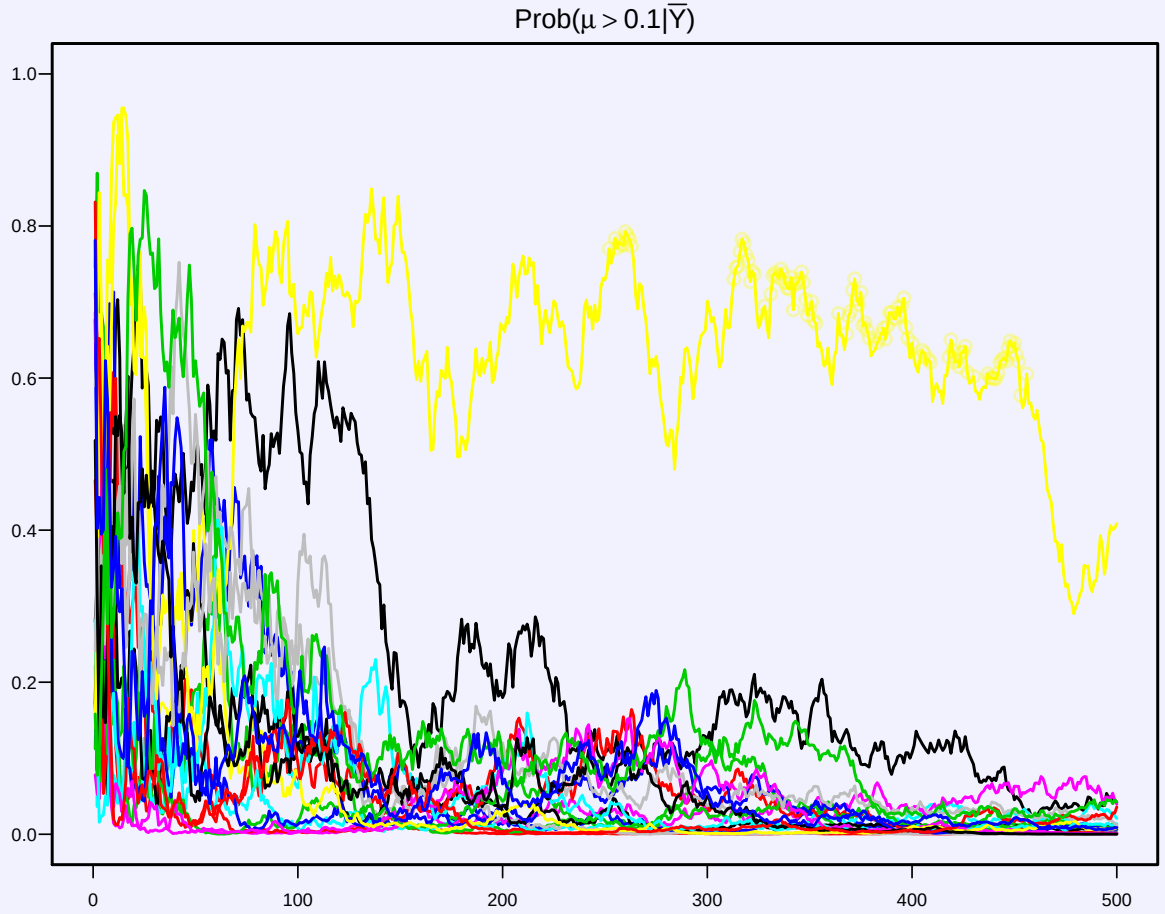



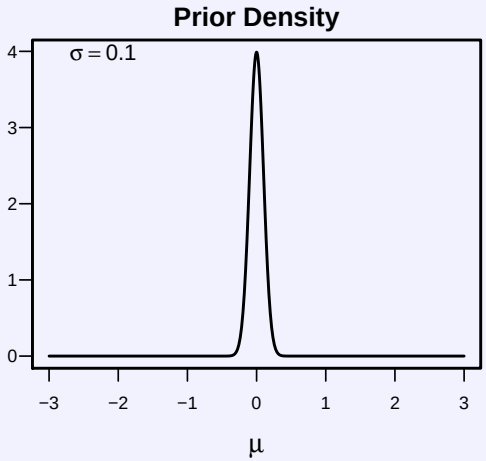

Fraction > 0.95: 0.34

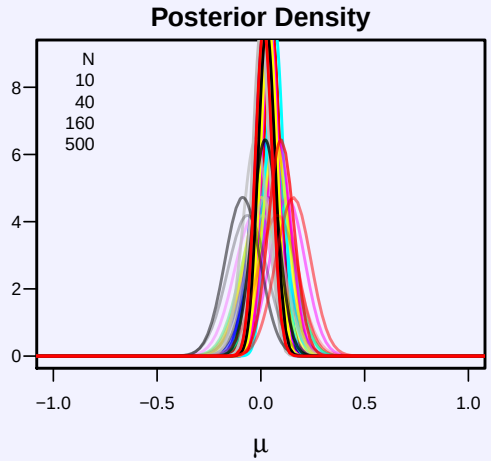


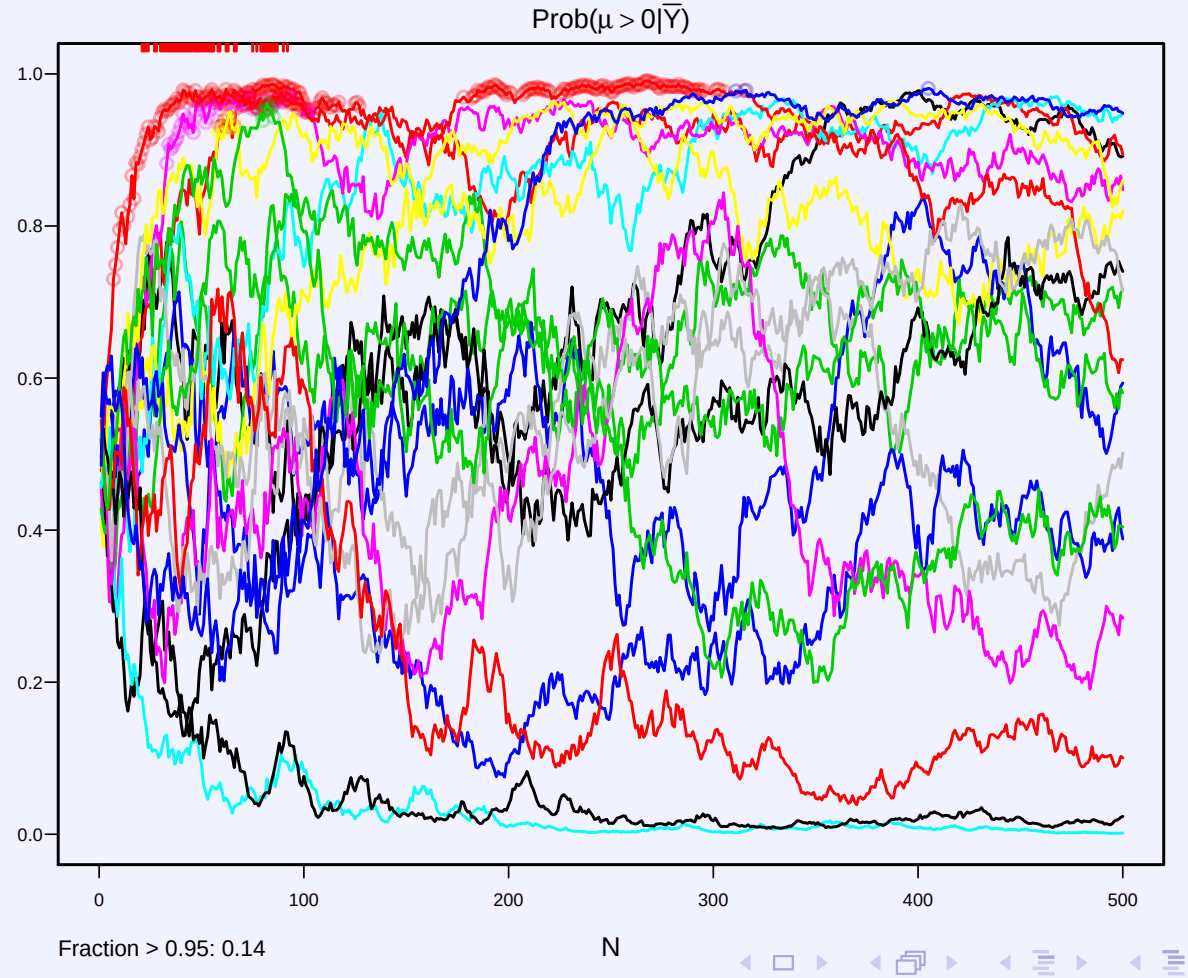



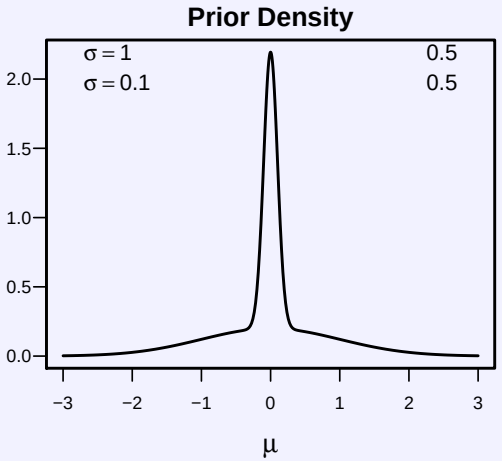



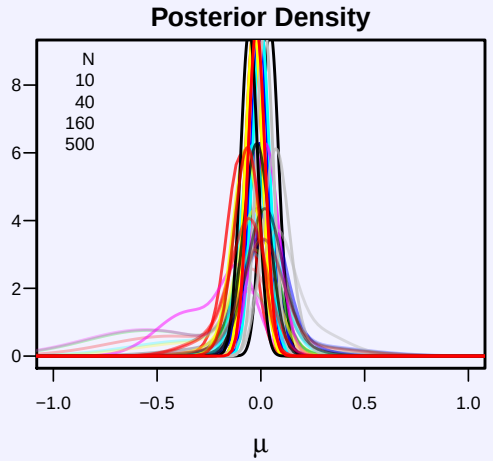

Fraction > 0.95: 0.1

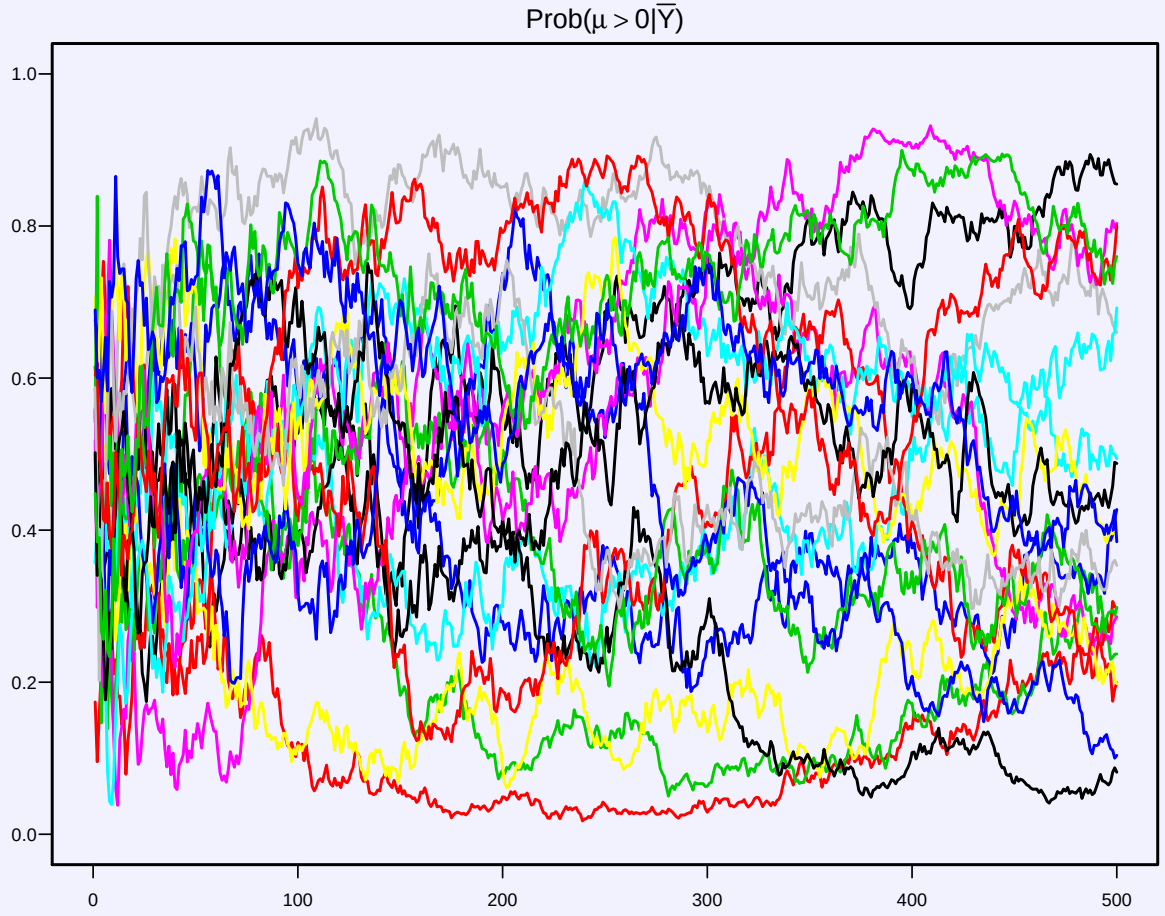

Prior Density
σ = 0.1
μ


Posterior Density
N
10
40
160
500
μ

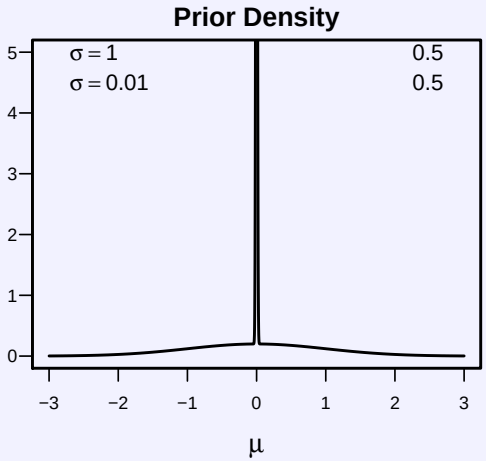

Prob(μ > 0|Ȳ)
Fraction > 0.95: 0.14
N

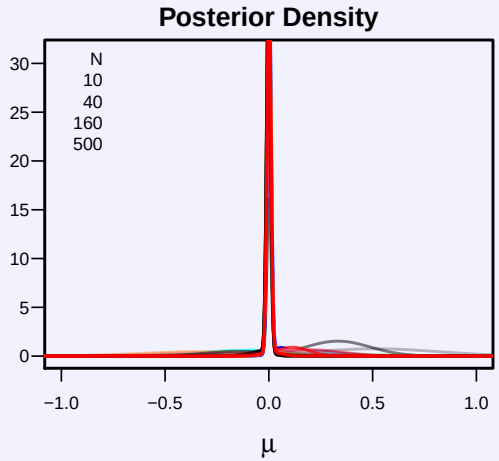


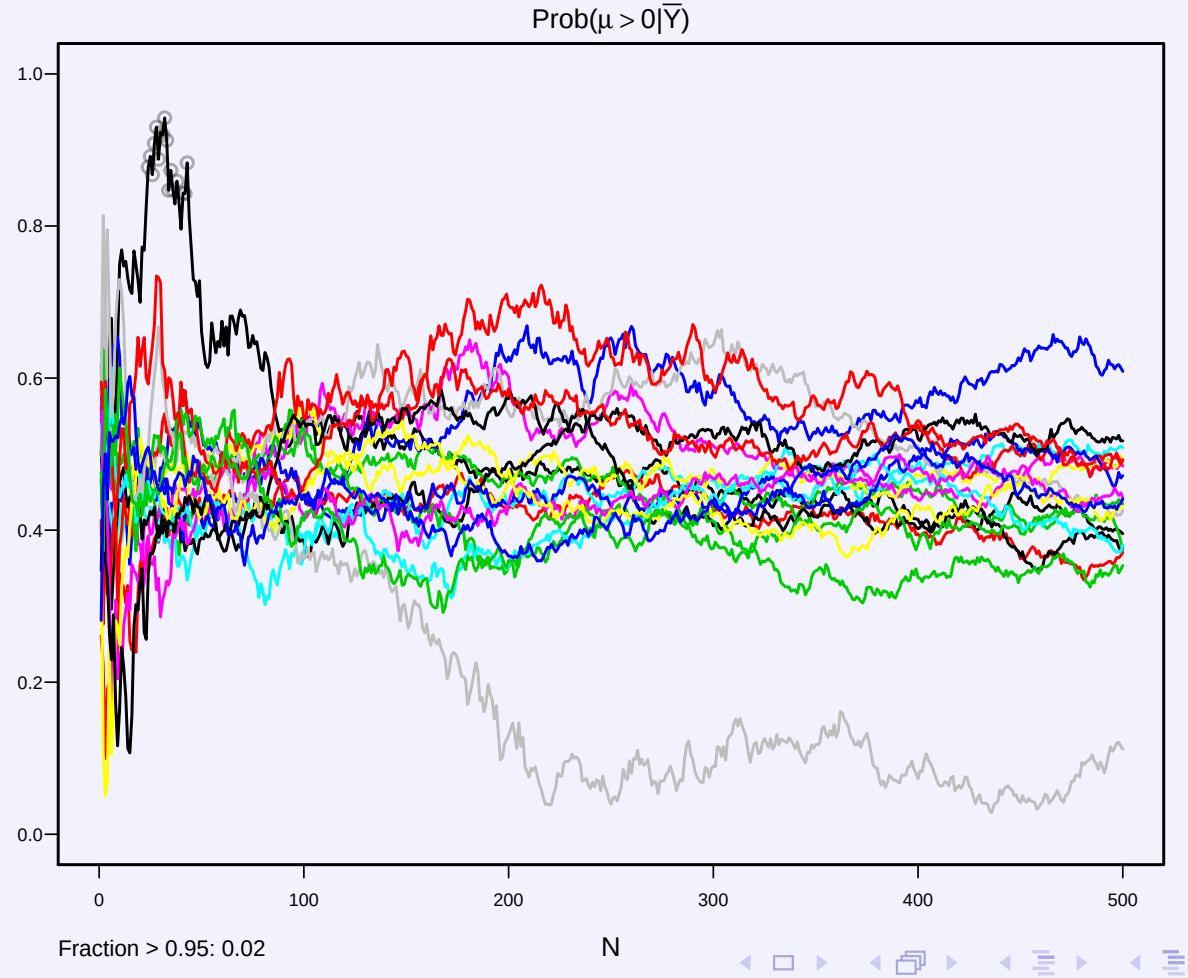



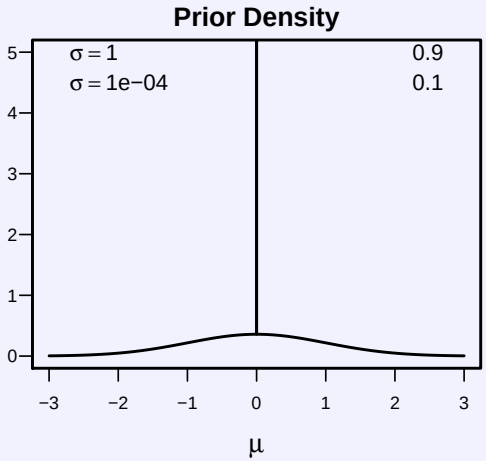



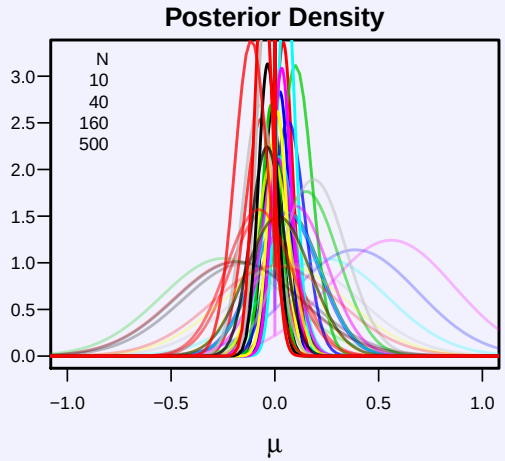

Fraction > 0.95: 0.17

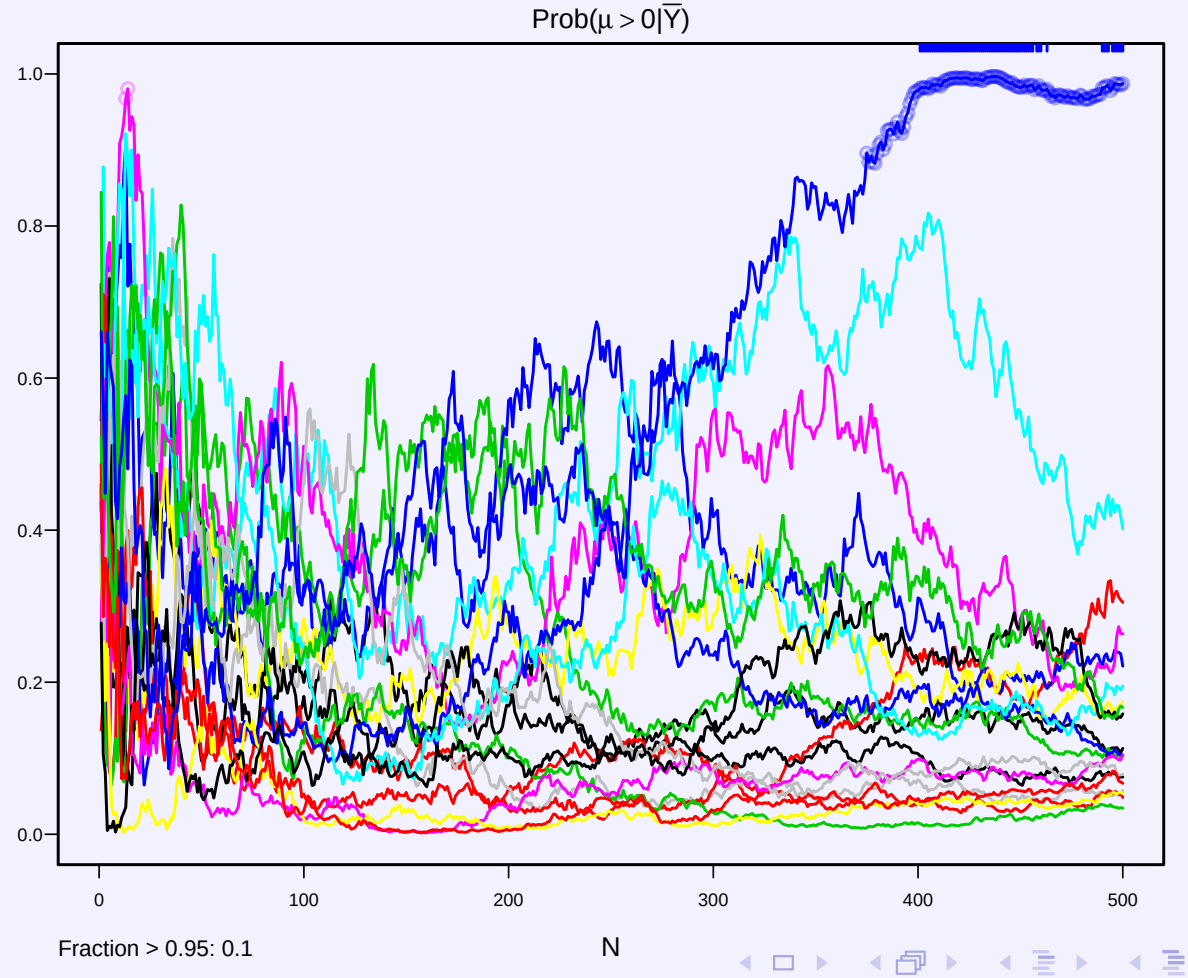

Prior Density
σ = 1
σ = 0.01
0.5
0.5
μ

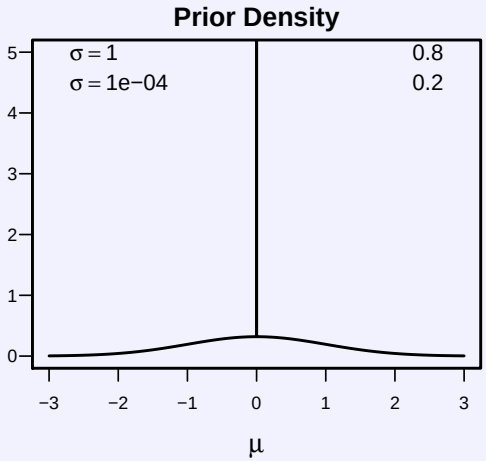

Posterior Density
N
10
40
160
500
μ

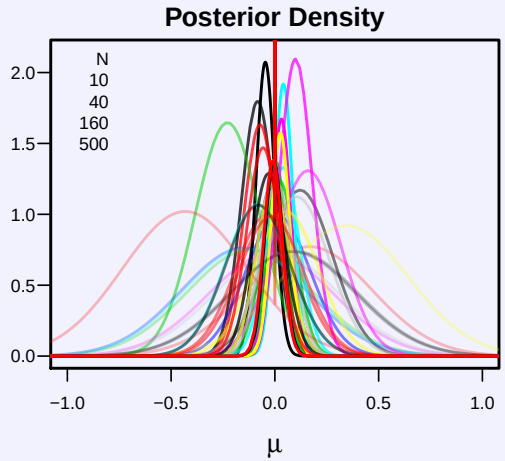

Prob(μ > 0|Ȳ)
Fraction > 0.95: 0.02
N

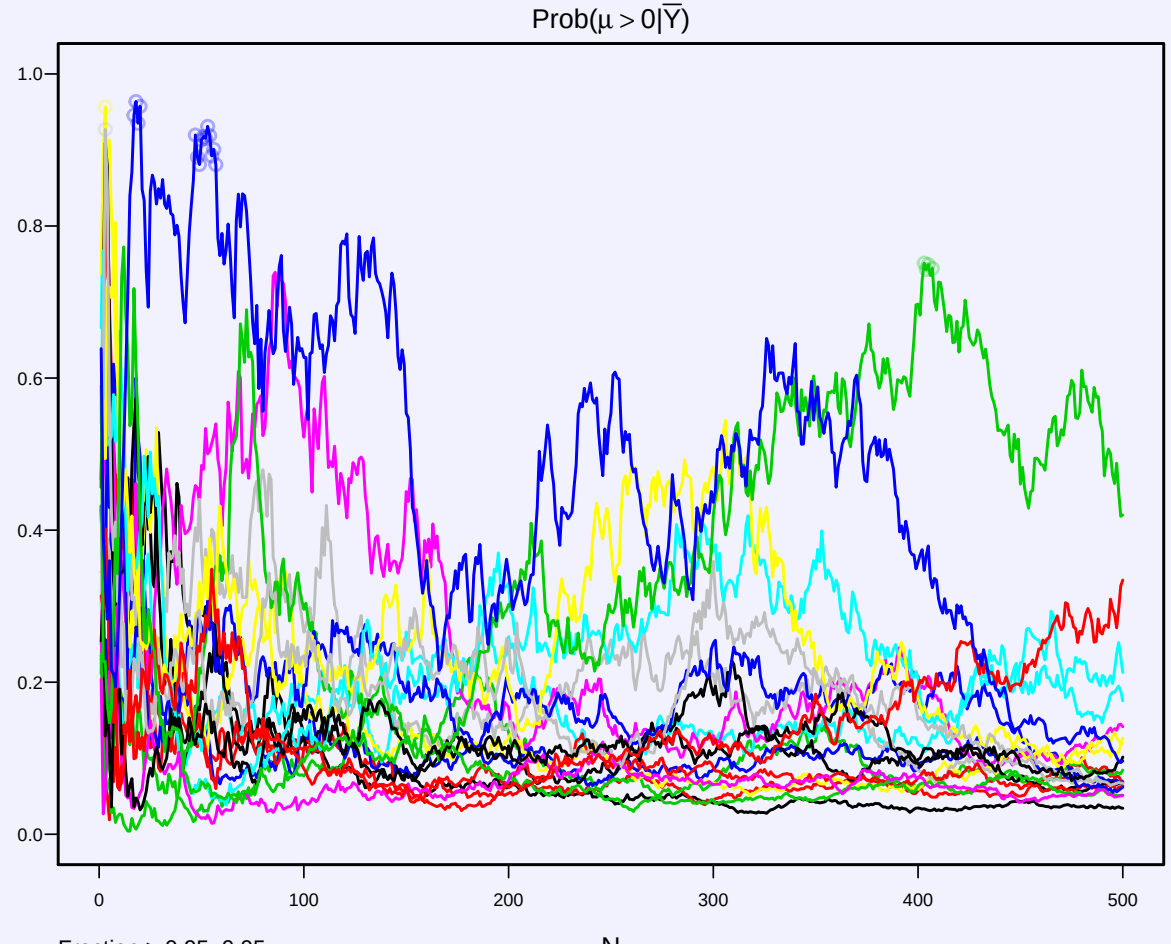

Prior Density
σ = 1
σ = 1e−04
0.9
0.1
μ


Posterior Density
N
10
40
160
500
μ


Prob(μ > 0|Ȳ)
Fraction > 0.95: 0.1
N


Prior Density
σ = 1
σ = 1e−04
0.8
0.2
μ


Posterior Density
N
10
40
160
500
μ


Prob(μ > 0|Ȳ)
N
Fraction > 0.95: 0.05

# BUT . . .



- A posterior probability of efficacy of 0.96 after the first 125 subjects **does not** need to be reinterpreted just because some other trial *may* achieve a probability $> 0.95$ when $\mu = 0$.
- Posterior probabilities are meaningful and represent a martingale process

---

- Probabilities of high posterior probs over imagined repetitions are not relevant
- If frequentists insist on this, Bayesians can solve the problem by having $\text{Prob}(\mu = 0) > 0$ or computing posterior $\text{Prob}(\mu > c|Y)$ where $c > 0$

# What Does a Frequentist Owe a Bayesian?



- Demonstrate an error in a particular posterior probability when
  - the frequentist specifies the prior
  - the Bayesian and frequentist agree on the statistical model
- Allowance for Bayesian to show performance consistent with a special belief in $H_0$

# What Does a Bayesian Owe a Frequentist?



- Use of the frequentist's effective prior to demonstrate consistency
- Posterior probabilities of meaningful assertions/events
- Posterior probabilities are well constructed
- Answering "What is the evidence **now**?"
- **But** the Bayesian is not obligated to care about "populations" of trials

# Moving Past CDRH Bayesian Guidance Document



- Demanding evaluation of frequentist properties of Bayesian procedures requires
  - a huge time investment
  - bringing the sample space back into consideration
    - Example: planned for looks that didn't happen because recruitment faster than anticipated
- Persisting with frequentist thinking prevents the full power of Bayesian approaches from being realized
  - $\infty$ data looks
  - posterior probabilities of compound events , e.g.
    - superiority on mortality or $\uparrow$ walking distance by $\geq 30m$
    - efficacy on any 2 of 3 endpoints
  - adaptation, simplicity, interpretability, . . .

# Making Progress



Progress is made when we learn from the past but are not afraid to break from it

Mixing of old and new paradigms can create confusion more than provide comfort

Rather than viewing evidence through a (usually conservative) frequentist eye, the FDA can incentivize better science by allowing the use of highly flexible methods that will

- motivate more randomized trials to be launched and faster with ability to change protocol during execution Please no more co-primary endpoints and closed testing procedures!
- provide more meaningful results
- allow non-promising studies to be terminated faster

# References


Blume, J. D. (2011). "Likelihood and its evidential framework". In: *Handbook of the Philosophy of Science: Philosophy of Statistics*. Ed. by D. M. Gabbay and J. Woods. San Diego: North Holland, pp. 493–511.


Emerson, S. S. (1995). "Stopping a clinical trial very early based on unplanned interim analysis: A group sequential approach". In: *Biometrics* 51, pp. 1152–1162.


Spiegelhalter, D. J., L. S. Freedman, and M. K. B. Parmar (1994). "Bayesian approaches to randomized trials". In: *J Roy Stat Soc A* 157, pp. 357–416.