Bayes for Flexibility in Urgent Times

Frank E Harrell Jr

Department of Biostatistics
Vanderbilt University School of Medicine
Nashville TN
Expert Statistical Advisor
Office of Biostatistics
Office of Translational Sciences
Center for Drug Evaluation and Research, FDA

PHASTAR Life Sciences Summit 2020-07-01

Big Picture

- Efficacy is not a hypothesis; it is a matter of degree
- Hypothesis testing and associated thresholds have hurt science
- Would you rather know the chance of making an assertion of efficacy when the treatment has no effect, or the chance the treatment is effective?
- Probabilities conditioning backwards in time/information flow are not directly actionable

Problems We Face in COVID-19

- Relatively little known about the virus and its treatment
- Rapid decision making required
- Studies must be launched and concluded quickly
- DSMBs must act on information from other studies
- DSMB meetings cannot be scheduled far in advance
- A data look may be required at any time
- Best to just allow for infinitely many data looks
- See `hbiostat.org/proj/covid19`

High Level View of Statistical Approaches

- Frequentist: probability of data given an assertion is true
- Bayesian: probability assertion is true given the data
- Frequentist type I error: probability of making an assertion of efficacy over the long run of replicate studies like yours **except** that the treatment has zero effect and does no harm
- Bayesian posterior probability of efficacy: probability of true efficacy underlying the process generating **our** data (probability that an assertion of efficacy is true)
- One minus posterior probability of efficacy: probability of no effect or harm (*regulator's regret*)

What is Actionable?

What is Not Actionable

After a patient has a diagnostic test, the sensitivity and specificity of the test

What is Actionable

The probability the patient has the disease

Advantages of Bayes, *continued*

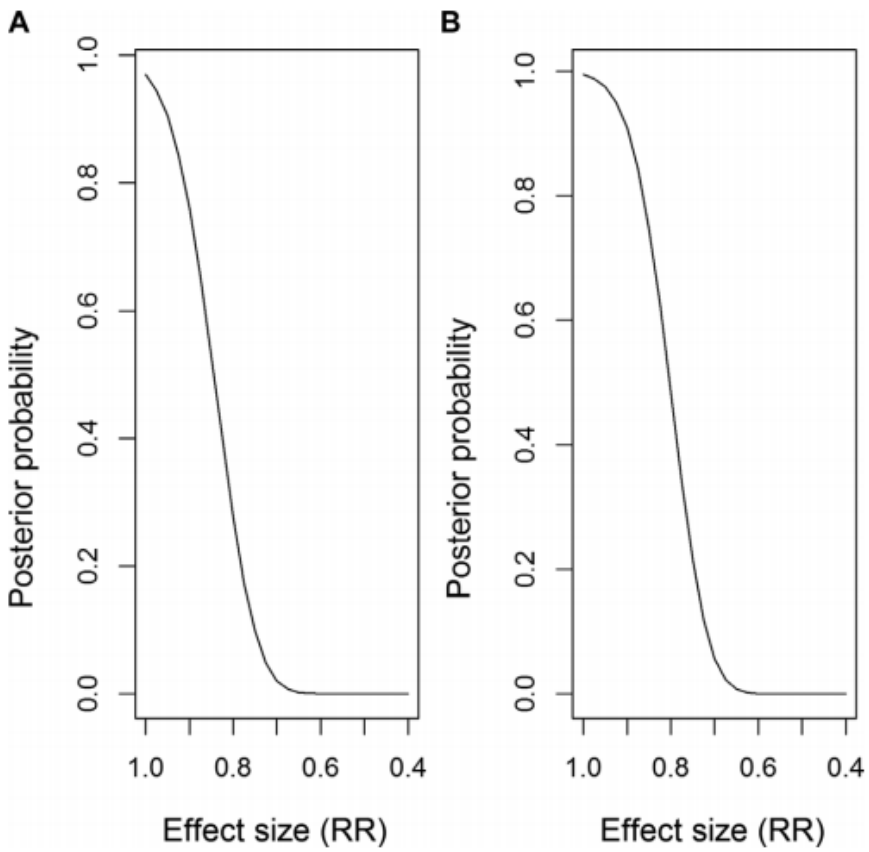


Figure 3 Posterior probability of specified effect sizes, using (A) flat prior, (B) evidence-based priors. The line shows the probability of the relative risk (RR) being lower than the values on the x-axis (ie, a bigger treatment effect). A RR <1 indicates that the primary outcome rate is smaller in the intervention arm compared with the control arm.

EG Ryan et al, *BMJ Open* 2019; 9

Examples of Posterior Probabilities

- Prob(any assertion or combination of assertions)
- Prob(efficacy > 0)
- Prob(efficacy $>$ MCID)
- Prob(non-inferiority)
- Prob(efficacy > 0) on Nov. 2: interpretation completely unaffected by:
- Prob(efficacy > 0) on Nov. 1
- Flexibility:
  Prob(hit any 2 of 4 migraine headache endpoints)

Besides The Interpretation Does It Matter That *p*-values are Backward Probabilities?

- Can't inject appropriate skepticism into the calculation
- Can't inject prior relevant information (skeptical or positive) into the calculation
- Being backwards is the **cause** of multiplicity problems: Multiplicity is caused by the chances you give data to be extreme, **not** from the chances you give assertions to be true
- Being backwards means you have to take into account how the data arose instead of just interpret the data at hand
  - Frequentist approach is cumbersome for flexible sequential designs or adaptive designs

Fully Sequential Trials: Continuous Learning with Unlimited Looks

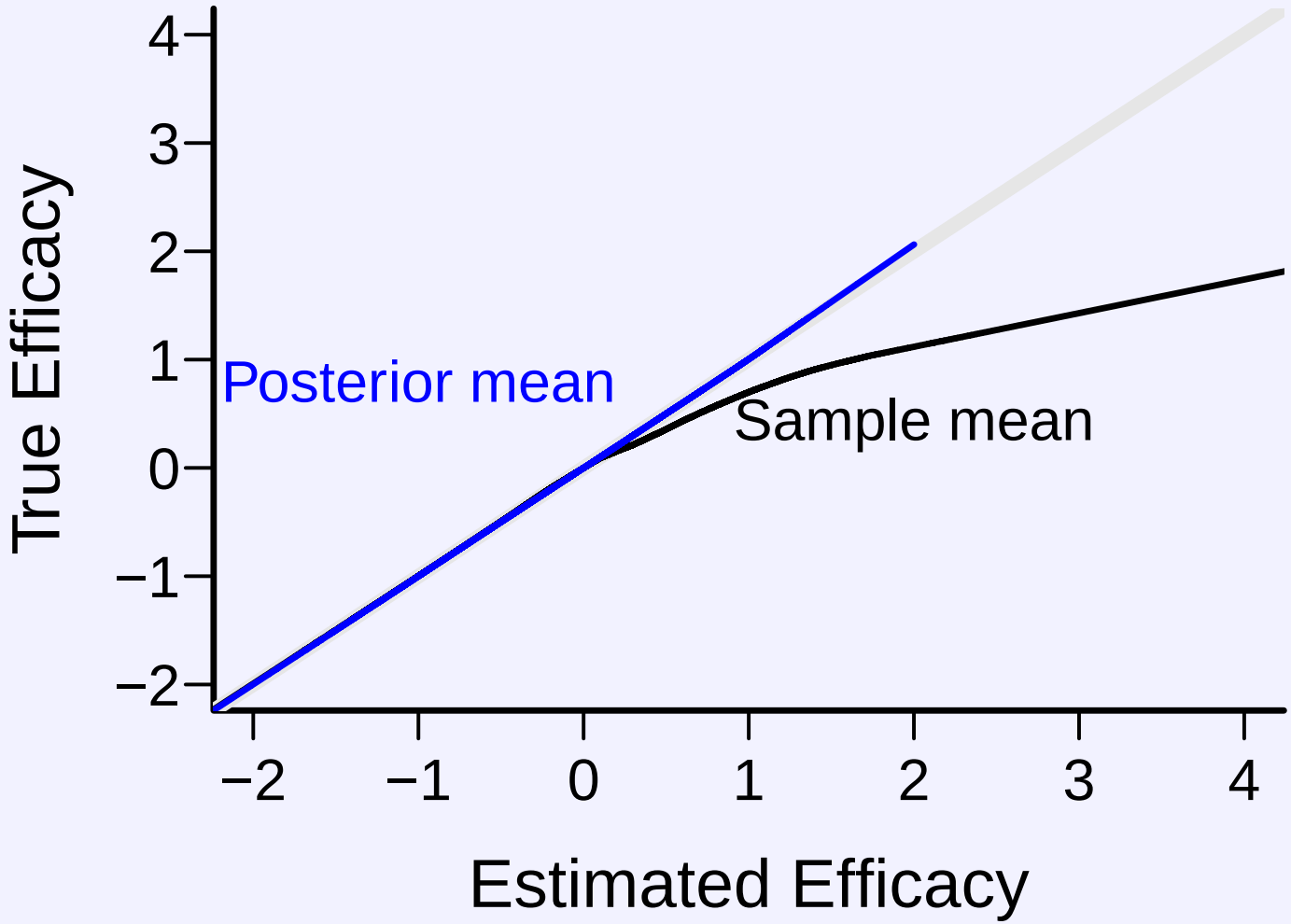
(In a Bayesian analysis) It is entirely appropriate to collect data until a point has been proven or disproven, or until the data collector runs out of time, money, or patience. - Edwards, Lindman, Savage (1963)

- Run 50,000 **different** clinical trials
  (differ on amount of efficacy)
- For each, sample one μ (true efficacy) from the prior
- Generate data ($n = 500$) under this truth
- Do analysis after $1, 2, \ldots, 500$ subjects studied (≤ 500 looks)
- Stop the instant Prob($\mu > 0$) ≥ 0.95 (efficacy) or Prob($\mu < 0.05$) ≥ 0.90 (futility)
- See `fharrell.com` for details and code
- See `hbiostat.org/proj/covid19` for a detailed sequential COVID-19 clinical trial plan

Sequential Testing Simulation, *continued*

- 20393 trials stopped early for efficacy
- 28438 trials stopped early for futility
- 1169 trials went to completion ($n = 500$)
- Average post. prob. of efficacy at stopping for efficacy: 0.961
- Of trials stopped early for efficacy, proportion with $\mu > 0$: 0.960
- Average post. prob. of futility at stopping for futility: 0.920
- Of trials stopped early for futility, proportion with $\mu < 0.05$: 0.923

Calibration of Posterior Mean at Stopping for Efficacy



Advantages of Bayesian Approach

- Computes probabilities on the actionable scale
- Is based solely on basic laws of probability; no special recipes needed
- Is flexible without hurting the science
- Allows experimentation until sufficient evidence
- Use in complex adaptive designs no more complex than use in simple static design
- Can provide simultaneous totality of evidence
- Non-inferiority involves just another posterior probability
- Evidence for non-trivial effects P(E $> \epsilon$)
- P(hitting complex efficacy targets)
- Math for incorporating external information

What is the Greatest Hesitance to Adopting Bayes?

- Fear of not preserving type I assertion probabilities (there is no type I "error")
- Type I assertion probability = long-run P(assertion of efficacy) **if efficacy = 0**
- This probability is **independent of the data**
- Contrast with P(mistake | data) = 1 - posterior P(efficacy)
- Type I error **is not regulator's regret** (approving a drug that doesn't work) but is an assertion probability **assuming the drug doesn't work**
- Analogy:
  - Judging a politician by how often he talks vs.
  - Judging him by how often he tells the truth

New Resource and Discussion Board

- Introductory Bayesian design and analysis course: `hbiostat.org/doc/bayes/course.html`
- Bayesian *t*-tests: `hbiostat.org/doc/bbr.pdf` Chapter 5
- Discussion board for the presentation you are viewing: `bit.ly/datamethods-whybayes`
- `hbiostat.org/proj/covid19`