Fundamental Advantages of Bayes in Drug Development

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

NISS-Merck Virtual Meet-up 2020-04-27

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

- Need a formal way to insert extra-study information
  - skepticism
  - trustworthy evidence / past data
- Frequentist paradigm requires a certain design rigidity
- Freq. approach conservative when want to learn continuously
  - also requires complex adjustments to point estimates if stop early
  - p-value is a function of a cutoff/stopping rule, not just data
- Each design requires a one-off freq. adjustment
  - adaptive trials use standard Bayesian machinery with **NO** modification

Problems, *continued*

- Multiplicity mess
  - Do we really believe that $A - B$ should be discounted because we compared C with D?
- p-values are not directly actionable
  - Prob(assertion of efficacy | no efficacy)
  - need Prob(efficacy) = Prob($\Delta > 0$)
  - 1 - Prob(efficacy) = Prob($\Delta \leq 0$) = Prob(inefficacy): a **real** worry for a regulator
- p-values use backwards time/information order—the cause of multiplicity problems and makes it harder to assess totality of evidence
- Multiplicity as we know it is **sampling** multiplicity not **parameter** multiplicity

High Level View of Analysis of Effectiveness

- Account for uncertainties
- Use right amount of skepticism or optimism on the most relevant scale, inserted at proper point in logic flow
- Use relevant prior information
  - If not sure of relevance, incorporate P(relevance) (e.g. extrapolation from adults to children)
  - If very unsure of relevance, be skeptical about potential efficacy unless you trust 'experts'
- Use data efficiently
- Compute pertinent probabilities and interval estimates
- Decision making under uncertainty best done with probabilistic thinking

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*)

Example Bayesian and Frequentist Statements

"Negative" Study

Frequentist : There was little evidence against the null hypothesis that A=B in mean SBP (p=0.4)

Bayesian : Under prior . . . , the probability that B<A in mean SBP is 0.67
Under prior . . . , B probably (0.67) $\downarrow$ SBP
The prob. that A and B are similar ($\pm$3 mm Hg) is 0.53

"Positive" Study

Frequentist : There is evidence against the hypothesis that A=B (p=0.02)

Bayesian : Under prior . . . , B probably (0.985) $\downarrow$ SBP
The prob. that B lowers SBP by $>$ 3mmHg is 0.81

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 Bayesian Approach

- Computes probabilities on the actionable scale
- Is based solely on basic laws of probability
- Is flexible without hurting the science; encourages learning from data
- *p*-values require complex, controversial adjustments for multiple looks & adaptation
- No need to customize stat tools for specific sequential / adaptive designs
- Non-inferiority involves just another posterior probability
- Math for incorporating external information

Advantages of Bayes, *continued*

- Difficult to know what to believe or how to act given Prob(assert efficacy | efficacy=0)
- Uses a forward predictive mode
- Optimum decision: maximize expected utility
  - utility function very hard to specify
  - expected utility needs posterior probability distribution
  - $\Rightarrow$ utilities are not known until the decision point
    We should still state results so as to lead to optimum decision (i.e., as posterior probs)
- Can quantify evidence for **all posssible magnitudes of efficacy**

Advantages of Bayes, *continued*

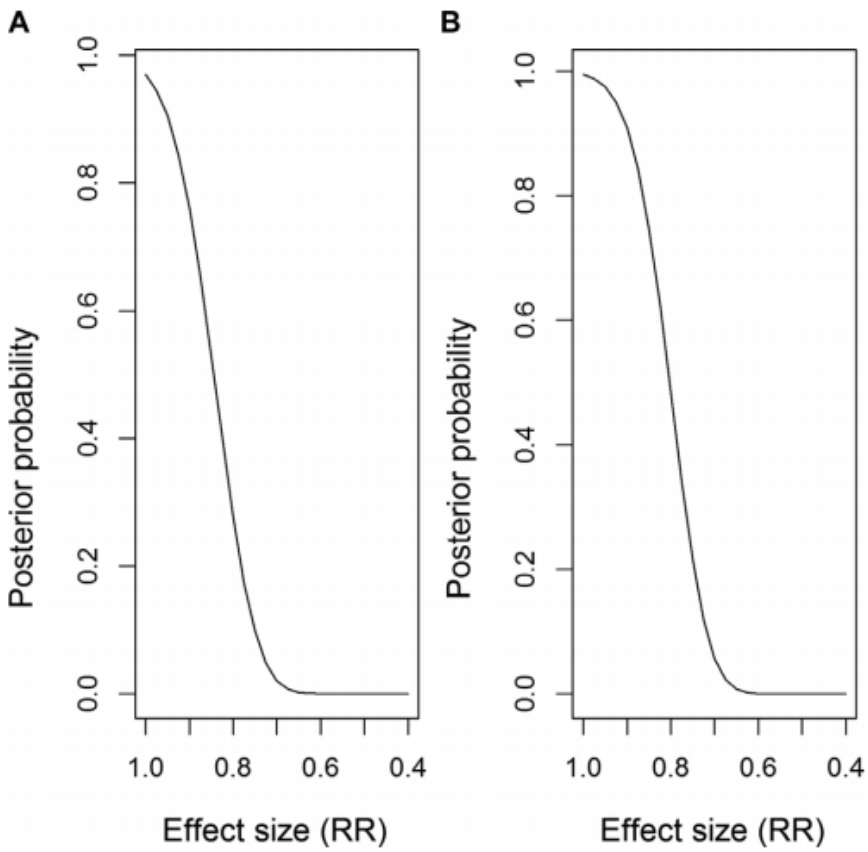

Figure 3 Posterior probability of specified effect sizes, using (A) flat prior, (B) evidence-based priors. The line shows the probability of the relative risk (RR) being lower than the values on the x-axis (ie, a bigger treatment effect). A RR <1 indicates that the primary outcome rate is smaller in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

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

What Probabilities Do We Need?

- Should be probabilities for something uncertain
- It's mainly about estimation and prediction
- Really interested in forward probabilities
- Let
  E = efficacy (difference, log ratio; higher is good)
  S = safety (e.g., SAE risk difference; higher is bad)

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)

Forward Probabilities Define Their Own Error Probabilities

Probability	Meaning
1 - $P(E > 0) = P(E \leq 0)$	no benefit, or worse (regulator's regret)
1 - $P(E >$ trivial$)$	trivial effect or harm
1 - $P(S \leq 0) = P(S > 0)$	safety signal
1 - $P(E > -3) = P(E \leq -3)$	inferiority

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

There Has To Be A Downside

- Bayesian modeling replaces endless arguments with one argument: the choice of prior
- A forward probability (Bayesian posterior prob.) cannot be calculated without starting somewhere
  Just as disease risk cannot be computed from sensitivity & specificity; need background risk (prevalence)

You can't compute a current probability without having a starting probability

Subtle Advantages of Bayesian Approach

- Regarding effect evidence presentation, the frequentist approach is unable to pre-specify
- Trade a restrictive design pre-specification for evidential pre-specification
- Enforcement of evidential discipline
- Prior is a pre-specified likelihood averaging function—a kind of restraint on how one quantifies evidence

From Steve Goodman *(personal communication)*

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

Two Endpoint Inference Example

- Treatments: A, B, $n = 1500$
- Outcomes:
  - DS: death or stroke w/in 1y
    - binary logistic model adjusted for SBP_0
    - B:A log OR: 0.8
    - prior is normal with mean 0 and SD so that P(OR < 0.5) $= 0.05$
  - SBP at 1y
    - linear model adjusted for $SBP_0, \sigma = 7$
    - B:A 3mmHg difference in SBP
    - prior is normal with mean 0 and SD so that P(SBP reduction > 10) $= 0.1$
  - Data generated so that SBP and DS are correlated
- Residual standard deviation prior: flat on $(0, \infty)$; multivariate normal for regression coefficients

Results of Bayesian Analysis for Two Endpoints

DS similarity: OR $\in [0.85, \frac{1}{0.85}]$

$$
\begin{aligned}
&\text{P(SBP reduced at least 2 mmHg)} &&= 0.999\\
&\text{P(B:A OR for DS} < 1) &&= 0.908\\
&\text{P(SBP reduced by 2 and OR} < 1) &&= 0.908\\
&\text{P(SBP reduced by 2 or OR} < 1) &&= 1.000\\
&\text{P(DS Non-inferiority)} &&= 0.948\\
&\text{P(DS similar)} &&= 0.363\\
&\text{E(\# targets achieved)} &&= 1.908
\end{aligned}
$$

Misperceptions About Bayes

- Bayes requires you to borrow information
  - It doesn't
- The need for a prior makes it subjective
  - Priors can be mutually-agreeable skeptical distributions or agreed-upon mixture of skeptical prior and posterior from other studies (probability of applicability)
  - Frequentist paradigm
    - uses subjective "intent to analyze"
    - requires subjective interpretation at the end
    - can only subjectively use extra-study information
    - requires arbitrary choice of multiplicity adjustment
  - For Bayes, subjectivity is encapsulated in the prior

Misperceptions About Bayes, *continued*

- Bayes is more complicated
  - Only the computations; interpretation is simpler
- Bayes can get by with lower N
  - N may $\uparrow$
  - $N \downarrow$ if trust historical data or test more often
- Bayes lowers the bar
  - The bar can be anywhere desired
  - When no relevant prior data are available, best to start with some skepticism against large E

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**

Summary

- Would a decision maker rather have
  - Prob(someone's data more extreme than mine $|$effect=0) or
  - Prob(effect $> \epsilon|$ my current data) ?
- Bayesian approach is a flexible forward predictive one that maximizes learning from data
- Bayes is a unified, consistent approach not requiring one-off solutions

  (by considering the parameter space instead of the sample space)
- It solves multiple longstanding problems with the frequentist approach while introducing only two challenges:
  - choice of prior
  - computational
- Bayesian results are directly actionable and formalize the use of extra-study information and consideration of totality of evidence

Summary, *continued*

- Efficacy is not a hypothesis; evidence is not dichotomous
- Hypothesis acceptance/rejection invite the use of arbitrary thresholds
- Arbitrary thresholds on post. prob. also problematic
  Totality of evidence is the key!
- Clear actionable conclusions:
  - With prior . . . treatment B probably (0.93) lowered the hazard of stroke when compared to treatment A
  - With priors . . . treatment B probably (0.99) improved ≥ 3 of 4 migraine endpoints compared to treatment A

Summary, *continued*

Bayes answers: What is our current judgment or what do we believe now that we have these data?

(can also tell us what to do, if we have a utility function)

You can't quantify current evidence without a starting probability which may represent general skepticism

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`