Why Bayes for Clinical Trials?

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

DIA Bayesian Scientific Working Group 2019-09-20

Updated 2019-09-22
Why Bayes for Clinical Trials?

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
Why Bayes for Clinical Trials?

Background
Freq & Bayes
Types of Probabilities
Needed Probabilities
Bayes
Multiplicity
Fully Sequential Trials
Using Prior Data
Two-Endpoint Inference
Summary

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
Why Bayes for Clinical Trials?

Background

Types of Probabilities

Needed Probabilities

Bayes

Multiplicity

Fully Sequential Trials

Using Prior Data

Two-Endpoint Inference

Summary

**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**
  - $\text{Prob}(\text{assertion of efficacy} \mid \text{no efficacy})$
  - need $\text{Prob}(\text{efficacy}) = \text{Prob}(\Delta > 0)$
  - $1 - \text{Prob}(\text{efficacy}) = \text{Prob}(\Delta \leq 0) = \text{Prob}(\text{inefficacy})$: 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
Why Bayes for Clinical Trials?

Background
Freq & Bayes
Types of Probabilities
Needed Probabilities
Bayes
Multiplicity
Fully Sequential Trials
Using Prior Data
Two-Endpoint Inference
Summary

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 $P(\text{relevance})$ (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
Why Bayes for Clinical Trials?

Background

Freq & Bayes

Types of Probabilities

Needed Probabilities

Bayes

Multiplicity

Fully Sequential Trials

Using Prior Data

Two-Endpoint Inference

Summary

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)
**p-value**

Over an indefinitely long series of exact replications of our trial but with a different treatment effect ($0$) inserted

$$\text{Prob(their data more impressive than ours)}$$

- $p$-values and the $p < 0.05$ rule of thumb came into use before the computing revolution
- Assuming the null hypothesis is true greatly simplified the model, often requiring only manual calculations
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) ↓ 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) ↓ SBP
The prob. that $B$ lowers SBP by $> 3$mmHg is 0.81
High Level View of Bayes

- Compute probabilities of things you don’t know assuming things you do
- The chance that an assertion is true is more actionable than the chance of making the assertion given it’s false
- It is important to be able to compute probabilities of non-trivial effects and simultaneous probabilities about multiple endpoints
- Posterior probabilities are perfectly calibrated independent of the analysis frequency and stopping rule
- A Bayesian approach to the simplest study design can without modification handle complex sequential or adaptive designs
The benefits of obtaining direct, simply stated evidence about effects of interest, formally incorporating extra-study data and handling complex designs, are worth the price of having a prior distribution to anchor probability calculations.
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
- Much more . . .
Advantages of Bayes, continued

- Difficult to know what to believe or how to act given \( \text{Prob} \text{(assert efficacy | efficacy=0)} \)
- 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)
Why Bayes for Clinical Trials?

**Background**

**Freq & Bayes**

**Types of Probabilities**

**Needed Probabilities**

**Bayes**

**Multiplicity**

Fully Sequential Trials

Using Prior Data

Two-Endpoint Inference

Summary

---

**Types of Probabilities**

- **Conceptualization**
  - Frequency
  - Quantified judgments, evidence-based belief, subjective belief

- **Assumed information**
  - $\Delta = 0$
  - Data

- **Time and information flow**
  - Backward
    - $P(\text{present} \mid \text{future})$
    - $P(\text{known} \mid \text{unknown})$
  - Forward
    - $P(\text{future} \mid \text{present})$
    - $P(\text{unknown} \mid \text{known})$

---

**Conceptualization**

- Frequency
- Quantified judgments, evidence-based belief, subjective belief

**Assumed information**

- $\Delta = 0$
- Data

**Time and information flow**

- Backward
  - $P(\text{present} \mid \text{future})$
  - $P(\text{known} \mid \text{unknown})$
- Forward
  - $P(\text{future} \mid \text{present})$
  - $P(\text{unknown} \mid \text{known})$
### Types of Probabilities, continued

<table>
<thead>
<tr>
<th>Type</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Forecast</strong></td>
<td></td>
</tr>
<tr>
<td>Backward</td>
<td>$P(\text{current state}</td>
</tr>
<tr>
<td>Forward</td>
<td>$P(\text{future event}</td>
</tr>
<tr>
<td><strong>Diagnosis</strong></td>
<td></td>
</tr>
<tr>
<td>Backward</td>
<td>$P(\text{positive test}</td>
</tr>
<tr>
<td></td>
<td>sensitivity</td>
</tr>
<tr>
<td>Forward</td>
<td>$P(\text{disease}</td>
</tr>
<tr>
<td><strong>Disease Incidence</strong></td>
<td></td>
</tr>
<tr>
<td>Backward</td>
<td>$P(\text{AA}</td>
</tr>
<tr>
<td>Forward</td>
<td>$P(\text{diabetes}</td>
</tr>
<tr>
<td><strong>General</strong></td>
<td></td>
</tr>
<tr>
<td>Backward</td>
<td>$P(\text{observed data}</td>
</tr>
<tr>
<td>Forward</td>
<td>$P(\text{assertion } X \text{ true}</td>
</tr>
<tr>
<td><strong>Inference</strong></td>
<td></td>
</tr>
<tr>
<td>$p$-value</td>
<td>$P(\text{data in general more extreme}</td>
</tr>
<tr>
<td>Posterior Prob.</td>
<td>$P(\text{effect}</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Probability</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - P(E &gt; 0) = P(E ≤ 0)</td>
<td>no benefit, or worse (regulator’s regret)</td>
</tr>
<tr>
<td>1 - P(E &gt; trivial)</td>
<td>trivial effect or harm</td>
</tr>
<tr>
<td>1 - P(S ≤ 0) = P(S &gt; 0)</td>
<td>safety signal</td>
</tr>
<tr>
<td>1 - P(E &gt; −3) = P(E ≤ −3)</td>
<td>inferiority</td>
</tr>
</tbody>
</table>
What is a P-Value?

- Backward probability:
  \[ P(\text{data in general more extreme} \mid H_0) \]
- Is \textbf{not} the probability of \( H_0 \)
- Is misinterpreted much of the time

“The absurdity of the common backwards interpretation might be appreciated by pondering how the P value, which is a probability deduced from a set of assumptions (the statistical model), can possibly refer to the probability of those assumptions.” (Greenland et al 2016)
Why Bayes for Clinical Trials?

Background
Freq&Bayes
Types of Probabilities
Needed Probabilities
Bayes
Multiplicity
Fully Sequential Trials
Using Prior Data
Two-Endpoint Inference
Summary

Statistical Significance and the Dichotomization of Evidence

Blakeley B. McShane and David Gal

aKellogg School of Management, Northwestern University, Evanston, IL; bCollege of Business Administration, University of Illinois at Chicago, Chicago, IL

ABSTRACT
In light of recent concerns about reproducibility and replicability, the ASA issued a Statement on Statistical Significance and p-values aimed at those who are not primarily statisticians. While the ASA Statement notes that statistical significance and p-values are “commonly misused and misinterpreted,” it does not discuss and document broader implications of these errors for the interpretation of evidence. In this article, we review research on how applied researchers who are not primarily statisticians misuse and misinterpret p-values in practice and how this can lead to errors in the interpretation of evidence. We also present new data showing, perhaps surprisingly, that researchers who are primarily statisticians are also prone to misuse and misinterpret p-values thus resulting in similar errors. In particular, we show that statisticians tend to interpret evidence dichotomously based on whether or not a p-value crosses the conventional 0.05 threshold for statistical significance. We discuss implications and offer recommendations.
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
Why Bayes for Clinical Trials?

Background
Freq & Bayes
Types of Probabilities
Needed Probabilities
Bayes
Multiplicity
Fully Sequential Trials
Using Prior Data
Two-Endpoint Inference
Summary

$p$-values, continued

- $p$-value is a measure or surprise if no effect
- $p$-value is just “the degree to which the data are embarrassed by the null hypothesis” (Maxwell 2004)
- Can never provide evidence in favor of $H_0$, e.g., tell us a drug is not effective
- Contrast with Bayesian:
  \[ P(\Delta SBP \text{ between -3 and +3 mmHg}) = 0.53 \]
Why Bayes for Clinical Trials?

Background
Freq&Bayes
Types of Probabilities
Needed Probabilities
Bayes
Multiplicity
Fully Sequential Trials
Using Prior Data
Two-Endpoint Inference
Summary

What Is So Good About Forward Probabilities?

- Their interpretation does not depend on how one arrived
- They are perfectly calibrated independently of stopping rules
- Once one has a Bayesian model for the simplest design (e.g., 2-arm parallel group RCT) one injects that analysis into highly complex adaptive situations with no modification
Example Frequentist & Bayesian Analysis

- 2-arm study
- Unknown mean difference in SBP $\Delta$
- Frequentist & Bayesian 2-sample $t$-tests

<table>
<thead>
<tr>
<th>Frequentist</th>
<th>Bayesian</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p=0.03$</td>
<td>$P(\Delta &gt; 0) = 0.98$</td>
</tr>
<tr>
<td>unexpectedness</td>
<td>Prior: $</td>
</tr>
<tr>
<td>C.I. based on repeated RCTs</td>
<td>Credible interval for $\Delta$ for this 1-of-a-kind study</td>
</tr>
<tr>
<td>Assumes fixed sample size, 1 look</td>
<td>Interpretation independent of previous looks, adaptations</td>
</tr>
</tbody>
</table>

$P(\Delta > 5) = 0.86$
Efficacy targets: SBP, 6-minute walk test (6MWD, meters)

**Inaccurate Frequentist Statement**
Treatment B did not improve 6MWD (p=0.21)

**Unclear Frequentist Statement**
There was a statistically significant reduction in SBP with B (p=0.02) and no statistically significant increase in 6MWD (p=0.21)

**Honest Frequentist Statement**
If B has no effect on SBP, the chance of obtaining data more impressive than observed in the trial is p=0.02. The trial was uninformative on 6MWD (p=0.21).
Honest Bayesian Statements

- The prior assumption about the true B-A difference on SBP was that there was only a 0.05 probability of there being an absolute difference exceeding 10mmHg and only a 0.05 chance of an absolute difference in 6MWD exceeding 150m.
- Under these priors, treatment B probably (0.996) reduces SBP and probably (0.84) improves 6MWD.
- The probability that B improves either outcome is 0.999, and the chance it improves both is 0.81.
- The probability that B improves SBP more than 5mmHg is 0.8, and that it improves 6MWD by more than 50m is 0.3.
Bayesian modeling replaces endless arguments with one argument: the choice of prior.

A forward probability (Bayesian posterior prob.) cannot be calculated without starting somewhere - an anchor point. Just as disease risk cannot be computed from sensitivity & specificity; need background risk (prevalence).

You can’t compute a current probability without having an anchor probability.
Essence of Bayesian Models and Inference

- Probabilities quantify evidence
- Before data, describe state of knowledge with a prior distribution for a quantity of interest (e.g., treatment effect)
  - Pre-study evidence; pre-data evidence; extra-study information
  - General skepticism
  - Anchor
- Use Bayes’ theorem for rationally updating beliefs in light of new data
- Uses only basic rules of probability—\textbf{No}
  - large sample theory, central limit theorem
  - closed testing procedures
  - multiplicity adjustments
  - discussion of 1-tailed vs. 2-tailed tests
  - approximations to get probabilities, confidence intervals

Why Bayes for Clinical Trials?

Background

Freq&Bayes

Types of Probabilities

Needed Probabilities

Bayes

Multiplicity

Fully Sequential Trials

Using Prior Data

Two-Endpoint Inference

Summary
Bayesian Approach in a Nutshell

- Most arguments need to be completed before analysis.
- Better to have medical reviewers interject their skepticism in the prior when the stat plan is being finalized.
  - ...just as sponsors cannot change SAP after seeing results.
  - Avoid “I’ll know it after I see it” which is biased by observed results.
- Select data model (as with frequentist).
- Select prior.
- Update prior with data using Bayes’ rule.
- At any point current state of knowledge/evidence summarized by the posterior probability distribution.
- Used to construct credible intervals and PP of various assertions (often of form P(effect > 0)).
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*)
Why Bayes for Clinical Trials?

Types of Prior Distributions

- Flat (non-informative)
- Interval: rule out impossible values (e.g., hazard ratio < 0.1), otherwise flat
- Optimistic: favors $E > 0$
- Pessimistic: favors $E = 0$ — center prior at $E = 0$
- Very Pessimistic: favors $E < 0$ — center prior at $E < 0$
Suggestions for Choosing Prior

No pertinent trustworthy prior info: slightly skeptical prior; equal chance of benefit and detriment

Highly relevant prior info: take prior as posterior from previous study, with some discounting. e.g.: strong, large phase 2 study posterior $= \text{phase 3 prior}$

Probably relevant prior info: more discounting; prior $= \text{mixture of skeptical distribution and posterior}$

Process: iterative between sponsor and reviewers, before study
Caused by multiple chances we give data to be extreme, **not** by chances we give assertions to be true

If stick with chance an assertion is true, probabilities are perfectly calibrated no matter how often the assertion is assessed or what is the stopping rule

Example: simulated 1-arm clinical trial with a continuous normal outcome, $\sigma = 1$, $n = 500$
(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 \( \mu \) (true efficacy) from the prior
- Generate data (\( n = 500 \)) under this truth
- Do analysis after 1, 2, \ldots, 500 subjects studied (\( \leq 500 \) looks)
- Stop the instant \( \text{Prob}(\mu > 0) \geq 0.95 \) (efficacy) or \( \text{Prob}(\mu < 0.05) \geq 0.90 \) (futility)
- See fharrell.com for details and code
Skeptical Prior: Mixture of Two Normals
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
Why Bayes for Clinical Trials?

Background

Freq & Bayes

Types of Probabilities Needed

Bayes

Multiplicity

Fully Sequential Trials

Using Prior Data

Two-Endpoint Inference

Summary

Calibration of Posterior Mean at Stopping for Efficacy

- Estimated Efficacy vs. True Efficacy
- Sample mean vs. Posterior mean

Graph showing the calibration of posterior mean at stopping for efficacy.
Why Bayes for Clinical Trials?

Example Using Prior Data

- Suppose adult trial had $P(E > 0) = 0.98$
- Take prior for pediatrics study to be a weighted mixture of a skeptical prior $f(E)$ and the adult study posterior $g(E)$
- $\text{prior} = (1 - a)f(E) + ag(E)$
- $a = P(\text{applicability of adult results})$

![Graph showing the relationship between the applicability of adults and the efficacy in children.](image)
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 \( P(SBP \text{ reduction} > 10) = 0.1 \)
  - Data generated so that SBP and DS are correlated
- **Residual standard deviation prior:** flat on \((0, \infty)\); multivariate normal for regression coefficients
Why Bayes for Clinical Trials?

Background

Types of Probabilities

Needed Probabilities

Bayes

Multiplicity

Fully Sequential Trials

Using Prior Data

Two-Endpoint Inference

Summary

- Stan and R rstan package
- Total of 20,000 posterior draws from 4 chains
- No-U-turn sampler
- Run time: 10m
Results of Bayesian Analysis for Two Endpoints

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

- $P(\text{SBP reduced at least 2 mmHg}) = 0.999$
- $P(\text{B:A OR for DS < 1}) = 0.908$
- $P(\text{SBP reduced by 2 and OR < 1}) = 0.908$
- $P(\text{SBP reduced by 2 or OR < 1}) = 1.000$
- $P(\text{DS Non-inferiority}) = 0.948$
- $P(\text{DS similar}) = 0.363$
- $E(\# \text{targets achieved}) = 1.908$
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 ↑
  - $N$ ↓ 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 $E$
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 \( P(E > \epsilon) \)
- \( P(\text{hitting complex efficacy targets}) \)
- Math for incorporating external information
What is the Greatest Hesitance to Adopting Bayes?

- Fear of not preserving type I “error”
- But type I error probability = long-run \( P(\text{assertion of efficacy}) \) if efficacy = 0
- This probability is \textbf{independent of the data}
- Contrast with \( P(\text{mistake} | \text{data}) = 1 - \text{posterior } P(\text{efficacy}) \)
- Type I error \textbf{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
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 $\geq 3$ of 4 migraine endpoints compared to treatment A
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 an anchor probability which may represent general skepticism
New Resource and Discussion Board

- Introductory Bayesian design and analysis course: hbiostat.org/doc/bayes/course.html
- Discussion board for the presentation you are viewing: https://discourse.datamethods.org/t/discussion-why-bayes-for-clinical-trials
Interactive Demonstration of Priors/Posteriors

Why Bayes for Clinical Trials?

rpsychologist.com/d3/bayes

Acknowledgements: Current and Former FDA Colleagues

- Lisa LaVange
- Jonathan Jarow
- Telba Irony
- Estelle Russek-Cohen
- John Scott
- Daniel Rubin
- Edward Bein
- Tom Permutt
- Paul Schuette
- Several other CDER OB statisticians
Why Bayes for Clinical Trials?

Background

Freq&Bayes

Types of Probabilities

Needed Probabilities

Bayes

Multiplicity

Fully Sequential Trials

Using Prior Data

Two-Endpoint Inference

Summary

Cohen 1994

The following is almost but not quite the reasoning of null hypothesis rejection:

If the null hypothesis is correct, then this datum (D) can not occur.
   It has, however, occurred.
   Therefore the null hypothesis is false.

   If this were the reasoning of $H_0$ testing, then it would be formally correct.
   . . . But this is not the reasoning of NHST. Instead, it makes this reasoning probabilistic, as follows:

If the null hypothesis is correct, then these data are highly unlikely.
   These data have occurred.
   Therefore, the null hypothesis is highly unlikely.

   By making it probabilistic, it becomes invalid. . . . the syllogism becomes formally incorrect and leads to a conclusion that is not sensible:

   If a person is an American, then he is probably not a member of Congress.
   (TRUE, RIGHT?)
   This person is a member of Congress.
   Therefore, he is probably not an American. (Pollard & Richardson, 1987)