American Society for Clinical Pharmacology and Therapeutics
Orlando, Florida
24 March 2018

Bayesian Statistics and Its Implications for Drug Development

or

pr(you are a Bayesian) > 0.50

Stephen J. Ruberg, PhD
President
Analytix Thinking, LLC

Bringing data to life.

Outline

Objective: **CLEARLY** compare and contrast Bayesian versus Frequentist paradigms [… so you **COMPLETELY** understand why the Bayesian paradigm has enormous utility (primacy?) in drug development]

1. Bayesian versus Frequentist inference
2. Designing a clinical trials versus Interpreting a clinical trial result
3. Bayes Factor versus P-values

Bringing data to life.

Bayesian versus Frequentist Inference

Data|Hypothesis vs Hypothesis|Data

Bringing data to life.

A Problem of Inference

10,000 Coins

9,999 Fair Coins (H/T)
1 Biased Coin (H/H)

<u>Problem</u>

1. I draw out one coin.
2. I will flip it repeatedly, and tell you the result.
3. You tell me when you decide whether I have the Biased Coin or not.

How Many Heads Do You Need To See?

Number of Flips	Result	Biased Coin?
1	H	Y or N
2	H	Y or N
3	H	Y or N
4	H	Y or N
5	H	Y or N
6	H	Y or N
7	H	Y or N
8	H	Y or N
9	H	Y or N
10	H	Y or N

Number of Flips	Result	Biased Coin?
11	H	Y or N
12	H	Y or N
13	H	Y or N
14	H	Y or N
15	H	Y or N
16	H	Y or N
17	H	Y or N
18	H	Y or N
19	H	Y or N
20	H	Y or N

Two Perspectives

1. What is the probability of seeing N consecutive heads ***IF*** I have a fair coin?

2. What is the probability that I selected the biased coin ***IF*** I observe N consecutive heads … [from a coin randomly drawn from a bag of 9,999 fair coins and **1 biased coin**]?

Bringing data to life.

Two Perspectives

1. Pr (observed data | coin is fair)

H_0: Coin is fair [i.e. pr(heads) = 0.50]

H_a: Coin is not fair [i.e. pr(heads) = 1.00]

Pr[N consecutive heads | fair coin] = $(0.50)^N$

Better known as the *p-value*

Frequentist perspective

Bringing data to life.

Frequentist Results

Number of Flips	Result	p-value
1	H	0.500000000
2	H	0.250000000
3	H	0.125000000
4	H	0.062500000
5	H	0.031250000
6	H	0.015625000
7	H	0.007812500
8	H	0.003906250
9	H	0.001953125
10	H	0.000976563

Number of Flips	Result	p-value
11	H	0.000488281
12	H	0.000244141
13	H	0.000122070
14	H	0.000061035
15	H	0.000030518
16	H	0.000015259
17	H	0.000007629
18	H	0.000003815
19	H	0.000001907
20	H	0.000000954

Two Perspectives

2. Pr (coin is biased | observed data)

If we have $P(A|B)$,

we want to obtain the conditional probability $P(B|A)$

Bayes Theorem (1763)*

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

*As formulated by Laplace (1812)

Bayesian Results

Number of Flips	Result	Pr(Biased Coin)
1	H	**0.000200**
2	H	**0.000400**
3	H	**0.000799**
4	H	**0.001598**
5	H	**0.003190**
6	H	**0.006360**
7	H	**0.012639**
8	H	**0.024968**
9	H	**0.048711**
10	H	**0.092897**

Number of Flips	Result	Pr(Biased Coin)
11	H	**0.170001**
12	H	**0.290600**
13	H	**0.450333**
14	H	**0.621006**
15	H	**0.766198**
16	H	**0.867624**
17	H	**0.929121**
18	H	**0.963258**
19	H	**0.981285**
20	H	**0.990554**

A Problem of Inference

100 Coins

99 Fair Coins (H/T)
1 Biased Coin (H/H)

Problem

1. I draw out one coin.
2. I will flip it repeatedly, and tell you the result.
3. You tell me when you decide whether I have the Biased Coin or not.

The Results

Number of Flips	Prior = 1/10,000 Pr(Biased Coin)	Prior = 1/100 Pr(Biased Coin)
1	0.000200	0.019802
2	0.000400	0.038835
3	0.000799	0.074766
4	0.001598	0.139130
5	0.003190	0.244275
6	0.006360	0.392638
7	0.012639	0.563877
8	0.024963	0.721127
9	0.048711	0.837971
10	0.092897	0.911843

Number of Flips	Prior = 1/10,000 Pr(Biased Coin)	Prior = 1/100 Pr(Biased Coin)
11	0.170001	0.953889
12	0.290600	0.976400
13	0.450333	0.988059
14	0.621006	0.993994
15	0.766198	0.996988
16	0.867624	0.998492
17	0.929121	0.999245
18	0.963258	0.999622
19	0.981285	0.999811
20	0.990554	0.999906

The Results

# of Flips	p-value	Prior = 1/10,000 Pr(Biased Coin)	Prior = 1/100 Pr(Biased Coin)
1	0.500000	0.000200	0.019802
2	0.250000	0.000400	0.038835
3	0.125000	0.000799	0.074766
4	0.062500	0.001598	0.139130
5	0.031250	0.003190	0.244275
6	0.015625	0.006360	0.392638
7	0.0078125	0.012639	0.563877
8	0.0039063	0.024963	0.721127
9	0.0019531	0.048711	0.837971
10	0.0009766	0.092897	0.911843

# of Flips	p-value	Prior = 1/10,000 Pr(Biased Coin)	Prior = 1/100 Pr(Biased Coin)
11	0.0004882	0.170001	0.953889
12	0.0002441	0.290600	0.976400
13	0.0001220	0.450333	0.988059
14	0.0000610	0.621006	0.993994
15	0.0000305	0.766198	0.996988
16	0.0000153	0.867624	0.998492
17	0.0000076	0.929121	0.999245
18	0.0000038	0.963258	0.999622
19	0.0000019	0.981285	0.999811
20	0.0000010	0.990554	0.999906

Note: The p-value never changes regardless of your prior knowledge!!!!

Conclusion

For the same level of evidence

in the current experiment,

different inferences are made

about the

probability of the hypothesis being true

(or false)

based on prior knowledge

!!!!!!!!!!

Bringing data to life.

CT Design versus Interpretation

The Diagnostic Test Analogy

Bringing data to life.

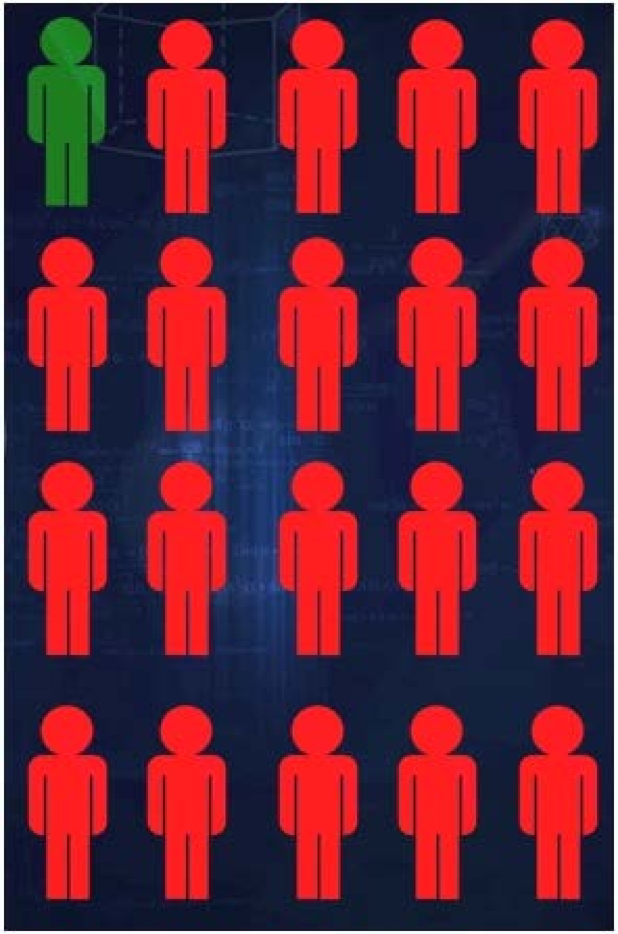
Interpreting results

A diagnostic test is analogous to a clinical trial design **<u>and interpretation</u>**.

Provides a conceptual perspective on the frequentist and Bayesian approach to **understanding what we know and how well we know it.**

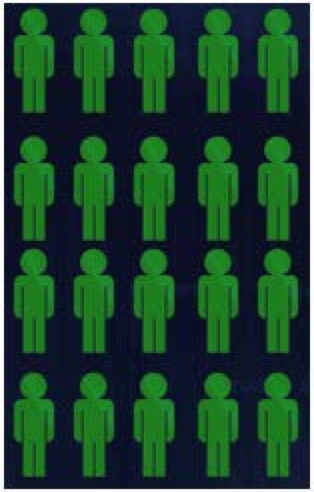
Bringing data to life.

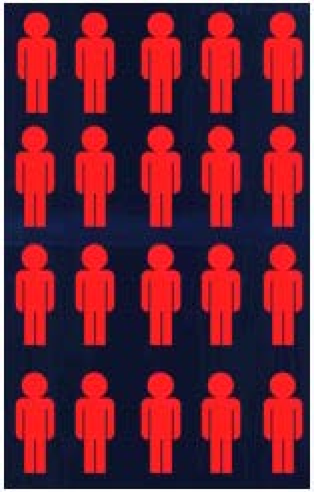
Population



5% of Population have ALK gene

Patients





Diagnostic Test

95% Sensitivity

19 +'s
1 -

95% Specificity

1 +
19 -'s

Individual Patient

ALK(-)?

ALK(+)?

Sample

Diagnostic Test

Result

+

Pr(Patient is ALK+) = ?

DIAGNOSTIC DECISION-MAKING

Developing/Designing the "Assay"

Conditional Probability

		Patient Characteristic	
		Positive	**Negative**
Diagnostic Test	**Positive**	**True Positive** **95%** **(Sensitivity)**	**False Positive** **5%**
	Negative	**False Negative** **5%**	**True Negative** **95%** **(Specificity)**

Prob (diagnostic test is positive *IF* the patient has the characteristic)

DIAGNOSTIC DECISION-MAKING

Interpreting an Observed Result

Conditional Probability

		Patient Characteristic *(Unknown Truth)*		
		Positive	**Negative**	
Diagnostic Test	**Positive**	True Positive 95%	False Positive 5%	Positive Predictive Value
	Negative	False Negative 5%	True Negative 95%	Negative Predictive Value

Prob (patient has the characteristic *IF* the diagnostic test is positive)

DIAGNOSTIC DECISION-MAKING

Underlying Prevalence for ALK gene is 5%

		Have the ALK Gene		
		Positive (5%)	**Negative (95%)**	
Diagnostic Test	**Positive**	**95** True Positive 95%	**95** False Positive 5%	**50%** Positive Predictive Value
	Negative	**5** False Negative 5%	**1805** True Negative 95%	**99.7%** Negative Predictive Value
		100	**1900**	**2000**

DIAGNOSTIC DECISION-MAKING

Underlying Prevalence for XYZ gene is 50%

		Have the ALK Gene		
		Positive (50%)	**Negative (50%)**	
Diagnostic Test	**Positive**	True Positive 95% **950**	False Positive 5% **50**	Positive Predictive Value **95%**
	Negative	False Negative 5% **50**	True Negative 95% **950**	Negative Predictive Value **95%**
		1000	**1000**	**2000**

DIAGNOSTIC DECISION-MAKING

KEY MESSAGES

Sensitivity and Specificity are the focus of ***assay design and development***

The **Positive (Negative) Predictive Values** are the focus of ***interpreting results*** (assay outputs)

THE PPV (NPV) ARE DEPENDENT ON THE UNDERLYING PREVALENCE OF THE CHARACTERISTIC (e.g. disease/marker status)

Bringing data to life.

The Clinical Trial Analogy

The diagnostic test is the clinical trial

The patient characteristic is whether the treatment meets its Critical Success Factors (unknown truth)

Sensitivity and (1-Specificity) are analogous to power and significance level of the hypothesis test for the CT

The PPV (NPV) is analogous to the "Bayesian posterior probability" that the treatment meets (fails) the CSF

THE PPV (NPV) ARE DEPENDENT ON THE PRIOR PROBABILITY OF THE TREATMENT MEETING THE CSF

THE CLINICAL TRIAL ANALOGY

Entering Ph 2 ⇒ Pr(drug meets CSFs) = 20%

"Prior"

Unknown →

Rigorous Ph 2 Trial Design

		Meets CSFs Yes (20%)	*Meets CSFs* No (80%)	
CT Result	Positive	True Positive 80% **320**	False Positive 10% **160**	Positive Predictive Value **66.7%**
	Negative	False Negative **80**	True Negative **1440**	Negative Predictive Value **94.7%**
		400	**1600**	**2000**

↑ Observed

"Posterior"

THE CLINICAL TRIAL ANALOGY

Entering Ph 2 ⇒ Pr(drug meets CSFs) = 20%

"Prior"

Unknown →

		Meets CSFs		
		Yes (20%)	No (80%)	
CT Result	Positive	True Positive 95% 380	False Positive 5% 80	82.6% Positive Predictive Value
	Negative	False Negative 5% 20	True Negative 95% 1520	98.7% Negative Predictive Value
		400	1600	2000

↑ Observed

Very Rigorous Ph 2 Trial Design

"Posterior"

Bayes Factor versus P-Values

Quantifying What We Know

The Nature of the P-Value

<u>Statistical Hypothesis Testing</u>

- Define H_0 (the null)
- Define test statistic, S(X)
  - S(x) is the value of the statistic given the data
- $F^{-1}[S(x)]$ transforms S(x) into a p-value, $p \in (0, 1)$
  - p contains no more information about H_0 than S(x)
  - p maps one-to-one and onto (0, 1)
- **Thus, the p-value is a statistic,**

NOT a probability !!

Todd A. Kuffner & Stephen G. Walker (2018): Why are p-Values Controversial?,
The American Statistician, DOI: 10.1080/00031305.2016.1277161

Bayes Factor

$p_{0,pri}$ = prior probability that H_0 is false

Convert this to the odds of H_0 vs H_1

- $O_{0,pri} = (1-p_{0,pri})/p_{0,pri}$

Let p = observed p-value for test of H_0

Bringing data to life.

Bayes Factor

Multiply $O_{0,pri}$ by Bayes factor* [$-e \times p \times \ln(p)$] to get a bound on the posterior odds

- $O_{0,post} \geq O_{0,pri} \times [-e \times p \times \ln(p)]$

Convert back to probability scale

- Posterior probability for H_0 being false is

$$\mathbf{p_{0,post} \leq 1/(1+O_{0,post})}$$

***Sellke et al (2001) Calibration of p Values for Testing Precise Null Hypotheses. The American Statistician, February 2001, Vol. 55, No. 1, pp 62-71.**

Bringing data to life.

Interpreting a Clinical Trial Results
(Using Bayes Factor)

If your prior is 30% probability of success (i.e. H_0 being false) entering Phase 2 …

Observed Phase 2 P-Value	Upper Bound on Posterior Probability for H_0* Being False
0.20	.329
0.10	.406
0.05	.513
0.01	.774

*Using Bayes factor for converting p-values into posterior probabilities

Using Bayes Factor
for Clinical Drug Development

If your prior is 30% probability of success (H_0 being false) entering Phase 2 …

And you want to exit Phase 2 with an 70% probability of success (in Phase 3) …

Then you need* …

- 1 study with a p-value of 0.016
- 2 studies each with p-values of 0.05**

Bringing data to life.

***Using Bayes factor for converting p-values into posterior probabilities**
****Successive application of Bayes factor**

Using Bayes Factor
for Clinical Biomarker Identification

100 potential biomarkers

- Prior probability of success (H_0 is false) = 0.20
- Prior on H_0 is true (none are predictive) = 0.80
- Uniform prior per biomarker = 0.20/100 = **0.002**

Observed p-value = 0.0001 for one biomarker

- Bonferroni adjusted p-value $\leq$ 0.01

Bayesian posterior pr(H_0 is false) $\leq$ 0.44.

Berger, JO, Wang X, Shen L (2014) A Bayesian Approach to Subgroup Identification, Journal of Biopharmaceutical Statistics, 24:1, 110-129, DOI: 10.1080/10543406.2013.856026

Bringing data to life.

FDA Approval

FDA wants to be sure that H_0 is false

Substantial evidence

- Consider two p-values of 0.05 for two Ph 3 trials

Prior Probability Against H_0 Entering Phase 3	Posterior Probability for H_0 Being False with Two p-values of 0.05 (≤)
0.65	.918
0.70	.933
0.75	.948
0.80	.960

FDA Approval

FDA wants to be sure that H_0 is false

Substantial evidence

- Consider one small p-values from a single Ph 3 trial

Prior Probability Against H_0 Entering Phase 3	Posterior Probability for H_0 Being False with Two p-values of 0.05 (≤)	Single P-value for 95% Posterior Probability of H_0 Being False (≤)
0.65	.918	0.007
0.70	.933	0.010
0.75	.948	0.013
0.80	.960	0.019

Conclusion

Two perspectives

1. Pr(data | hypothesis is true) **FREQUENTIST**
2. Pr(hypothesis is true | data) **BAYESIAN**

For a dataset / outcome of a study:

- Frequentist p-values are always the same
- Bayesian probabilities depend on your prior knowledge/probability

Bringing data to life.

Summary

Significance level and power are important elements of study design

Bayesian posterior probabilities are the most appropriate measures for interpretation of study outcomes

Bayesian perspective answers the question of interest.

Bringing data to life.

pr (I THANK YOU) = 0.9999

pr (YOU THANK ME) = ??

depends on your prior

Bringing data to life.