# And the Bayesians and the frequentists shall lie down together... 

Keith Winstein

MIT CSAIL

October 16, 2013


## Axioms of Probability (1933)

S: a finite set (the sample space)
A: any subset of $S$ (an event)
$P(A)$ : the probability of $A$ satisfies

- $P(A) \in \mathbb{R}$
- $P(A) \geq 0$
- $P(S)=1$
- $P(A \cup B)=P(A)+P(B)$ if $A \cap B=\emptyset$

If $S$ infinite, axiom becomes: for an infinite sequence of disjoint subsets $A_{1}, A_{2}, \ldots$,

$$
P\left(\bigcup_{i=1}^{\infty} A_{i}\right)=\sum_{i=1}^{\infty} P\left(A_{i}\right)
$$

## Some Theorems

- $P(\bar{A})=1-P(A)$
- $P(\emptyset)=0$
- $P(A) \leq P(B)$ if $A \subset B$
- $P(A) \leq 1$
- $P(A \cup B)=P(A)+P(B)-P(A \cap B)$
- $P(A \cup B) \leq P(A)+P(B)$


## Joint \& Conditional Probability

- For $A, B \subseteq S, P(A \cap B)$ is joint probability of $A$ and $B$.
- The conditional probability of $A$ given $B$ in:

$$
P(A \mid B)=\frac{P(A \cap B)}{P(B)}
$$

- $A$ and $B$ are independent iff $P(A \cap B)=P(A) P(B)$.
- $A, B$ independent $\rightarrow P(A \mid B)=P(A)$.


## Bayes' Theorem

We have:

- $P(A \mid B)=\frac{P(A \cap B)}{P(B)}$
- $P(B \mid A)=\frac{P(A \cap B)}{P(A)}$

Therefore: $P(A \cap B)=P(A \mid B) P(B)=P(B \mid A) P(A)$

Bayes' Theorem:

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

## On the islands of Ste. Frequentiste and Bayesienne...

On the islands of Ste. Frequentiste and Bayesienne...


The king has been poisoned!

## A letter goes out. . .

Dear Governor: Attached is a blood test for proximity to the poison. It has a $0 \%$ rate of false negative and a $\mathbf{1 \%}$ rate of false positive. Jail those responsible.

But remember the nationwide law: You must be 95\% certain to send a citizen to jail.

## On Ste. Frequentiste:

Test has a 0\% rate of false negative and a $1 \%$ rate of false positive. You must be $95 \%$ certain to send a citizen to jail.

- $P($ Positive $\mid$ Guilty $)=1$
- $P($ Negative $\mid$ Guilty $)=0$
- $P($ Positive $\mid$ Innocent $)=0.01$
- $P($ Negative $\mid$ InNocent $)=0.99$

How to interpret the law?
"We must be $95 \%$ certain" $\rightarrow$

## On Ste. Frequentiste:

Test has a 0\% rate of false negative and a $1 \%$ rate of false positive. You must be $95 \%$ certain to send a citizen to jail.

- $P($ Positive $\mid$ Guilty $)=1$
- $P($ Negative $\mid$ Guilty $)=0$
- $P($ Positive $\mid$ Innocent $)=0.01$
- $P($ Negative $\mid$ InNocent $)=0.99$

How to interpret the law?
"We must be $95 \%$ certain" $\rightarrow \mathbf{P}($ Jail $\mid$ InNOCENT $) \leq \mathbf{0 . 0 5}$

## On Ste. Frequentiste:

Test has a $0 \%$ rate of false negative and a $1 \%$ rate of false positive. You must be $95 \%$ certain to send a citizen to jail.

- $P($ Positive $\mid$ Guilty $)=1$
- $P($ Negative $\mid$ Guilty $)=0$
- $P($ Positive $\mid$ InNocent $)=0.01$
- $P($ Negative $\mid$ InNocent $)=0.99$

How to interpret the law?
"We must be $95 \%$ certain" $\rightarrow \mathbf{P}($ Jail $\mid$ InNOCENT $) \leq \mathbf{0 . 0 5}$
Can Positive $\rightarrow$ Jail? Yes.

## On Isle Bayesienne:

Test has a $0 \%$ rate of false negative and a $1 \%$ rate of false positive. You must be $95 \%$ certain to send a citizen to jail.

- $P($ Positive $\mid$ Guilty $)=1$
- $P($ Negative $\mid$ Guilty $)=0$
- $P($ Positive $\mid$ InNocent $)=0.01$
- $P($ Negative $\mid$ InNocent $)=0.99$

How to interpret the law?
"We must be $95 \%$ certain" $\rightarrow$

## On Isle Bayesienne:

Test has a $0 \%$ rate of false negative and a $1 \%$ rate of false positive. You must be $95 \%$ certain to send a citizen to jail.

- $P($ Positive $\mid$ Guilty $)=1$
- $P($ Negative $\mid$ Guilty $)=0$
- $P($ Positive $\mid$ Innocent $)=0.01$
- $P($ Negative $\mid$ InNocent $)=0.99$

How to interpret the law?
"We must be $95 \%$ certain" $\rightarrow \mathbf{P}($ InNocent $\mid$ Jail $) \leq \mathbf{0 . 0 5}$

## Isle Bayesienne: the need for prior assumptions

- "We must be $95 \%$ certain" $\rightarrow \mathbf{P}($ InNocent $\mid$ Jail $) \leq \mathbf{0 . 0 5}$
- Can Positive $\rightarrow$ Jail?
- Apply Bayes' theorem

$$
P(\text { InNocent } \mid \text { Positive })=\frac{P(\text { Positive } \mid \text { InNocent }) P(\text { InNOcent })}{P(\text { Positive })}
$$

## Isle Bayesienne: the need for prior assumptions

- "We must be $95 \%$ certain" $\rightarrow \mathbf{P}($ InNocent $\mid$ Jail $) \leq \mathbf{0 . 0 5}$
- Can Positive $\rightarrow$ Jail?
- Apply Bayes' theorem

$$
P(\text { InNocent } \mid \text { Positive })=\frac{P(\text { Positive } \mid \text { InNOCENT }) P(\text { InNOcent })}{P(\text { Positive })}
$$

## Isle Bayesienne: the need for prior assumptions

- "We must be $95 \%$ certain" $\rightarrow \mathbf{P}($ InNocent $\mid$ Jail $) \leq \mathbf{0 . 0 5}$
- Can Positive $\rightarrow$ Jail?
- Apply Bayes' theorem

$$
P(\text { InNocent } \mid \text { Jail })=\frac{P(\text { Positive } \mid \text { InNOCENT }) P(\text { InNOCENT })}{P(\text { Positive })}
$$

## Isle Bayesienne: the need for prior assumptions

- "We must be $95 \%$ certain" $\rightarrow \mathbf{P}($ InNocent $\mid$ Jail $) \leq \mathbf{0 . 0 5}$
- Can Positive $\rightarrow$ Jail?
- Apply Bayes' theorem

$$
P(\text { Innocent } \mid \text { Jail } \quad)=\frac{P(\text { Positive } \mid \text { InNocent }) P(\text { InNocent })}{P(\text { Positive })}
$$

## Isle Bayesienne: the need for prior assumptions

- "We must be $95 \%$ certain" $\rightarrow \mathbf{P}($ InNOCENT $\mid$ Jail $) \leq \mathbf{0 . 0 5}$
- Can Positive $\rightarrow$ Jail?
- Apply Bayes' theorem

$$
P(\text { InNocent } \mid \text { Jail })=\frac{P(\text { Positive } \mid \text { InNOCENT }) P(\text { InNOcent })}{P(\text { Positive })}
$$

## Isle Bayesienne: the need for prior assumptions

- "We must be $95 \%$ certain" $\rightarrow \mathbf{P}($ InNOCENT $\mid$ Jail $) \leq \mathbf{0 . 0 5}$
- Can Positive $\rightarrow$ Jail?
- Apply Bayes' theorem
$P($ InNocent $\mid$ Jail $)=\frac{(0.01)}{P(\text { Positive })}$


## Isle Bayesienne: the need for prior assumptions

- "We must be $95 \%$ certain" $\rightarrow \mathbf{P}($ InNOCENT $\mid$ Jail $) \leq \mathbf{0 . 0 5}$
- Can Positive $\rightarrow$ Jail?
- Apply Bayes' theorem
$P($ InNOCENT $\mid$ Jail $)=\frac{(0.01) \quad P(\text { InNOCENT })}{P(\text { Positive })}$


## Isle Bayesienne: the need for prior assumptions

- "We must be $95 \%$ certain" $\rightarrow \mathbf{P}($ InNOCENT $\mid$ Jail $) \leq \mathbf{0 . 0 5}$
- Can Positive $\rightarrow$ Jail?
- Apply Bayes' theorem
$P($ InNocent $\mid$ Jail $)=\frac{(0.01)}{P(\text { Positive })}$


## Isle Bayesienne: the need for prior assumptions

- "We must be $95 \%$ certain" $\rightarrow \mathbf{P}($ InNOCENT $\mid$ Jail $) \leq \mathbf{0 . 0 5}$
- Can Positive $\rightarrow$ Jail?
- Apply Bayes' theorem



## Isle Bayesienne: the need for prior assumptions

- "We must be $95 \%$ certain" $\rightarrow \mathbf{P}($ InNOCENT $\mid$ Jail $) \leq \mathbf{0 . 0 5}$
- Can Positive $\rightarrow$ Jail?
- Apply Bayes' theorem

$$
\begin{aligned}
& P(\text { InNOCENT } \mid \text { Jail } \quad= \frac{(0.01)}{P(\text { InNOCENT })} \\
&+\mathrm{P}(\text { Positive } \mid \text { InNOCENT }) \\
& \mathrm{P}(\text { InNOCENT }) \\
&+ \text { GuILTY }) \\
& \mathrm{P}(\text { GuILTY })
\end{aligned}
$$

## Isle Bayesienne: the need for prior assumptions

- "We must be $95 \%$ certain" $\rightarrow \mathbf{P}($ InNOCENT $\mid$ Jail $) \leq \mathbf{0 . 0 5}$
- Can Positive $\rightarrow$ Jail?
- Apply Bayes' theorem

$$
\begin{aligned}
P(\text { InNocent } \mid \text { Jail } \quad= & (0.01) \\
\hline & \mathrm{P} \text { (Positive } \mid \text { InNocent }) \\
& +\mathrm{P} \text { (InNocent) }) \\
\hline \text { (Positive } \mid \text { Guilty }) & \mathrm{P} \text { (GuILTY) })
\end{aligned}
$$

## Isle Bayesienne: the need for prior assumptions

- "We must be $95 \%$ certain" $\rightarrow \mathbf{P}($ InNOCENT $\mid$ Jail $) \leq \mathbf{0 . 0 5}$
- Can Positive $\rightarrow$ Jail?
- Apply Bayes' theorem

$P($ InNOCENT $\mid$ Jail $)=\frac{(0.01)}{P(\text { InNOCENT })}$| $1-(0.99)$ |
| :--- |
| $P($ InNOCENT $)$ |

## Isle Bayesienne: the need for prior assumptions

- "We must be $95 \%$ certain" $\rightarrow \mathbf{P}$ (InNOCENT $\mid$ Jail $) \leq \mathbf{0 . 0 5}$
- Can Positive $\rightarrow$ Jail?
- Apply Bayes' theorem


## Isle Bayesienne: the need for prior assumptions

- "We must be $95 \%$ certain" $\rightarrow \mathbf{P}($ InNOCENT $\mid$ Jail $) \leq \mathbf{0 . 0 5}$
- Can Positive $\rightarrow$ Jail?
- Apply Bayes' theorem

$$
\begin{aligned}
& P \text { (InNOCENT } \mid \text { Jail } \\
& \text { (0.01) } \\
& \text { 1-(0.99) P(InNocent) }
\end{aligned}
$$

## Isle Bayesienne: the need for prior assumptions

- "We must be $95 \%$ certain" $\rightarrow \mathbf{P}($ InNOCENT $\mid$ Jail $) \leq \mathbf{0 . 0 5}$
- Can Positive $\rightarrow$ Jail?
- Apply Bayes' theorem
$P$ (InNOCENT $\mid$ Jail

$$
)=\begin{array}{cc}
(0.01) & P(\text { InNocent }) \\
\hline 1-(0.99) & P(\text { InNOCENT })
\end{array}
$$

- $P($ Innocent $)=? ? ?$


## Isle Bayesienne: the need for prior assumptions

- "We must be $95 \%$ certain" $\rightarrow \mathbf{P}($ InNOCENT $\mid$ Jail $) \leq \mathbf{0 . 0 5}$
- Can Positive $\rightarrow$ Jail?
- Apply Bayes' theorem
$P$ (InNOCENT $\mid$ Jail

$)=$| $(0.01)$ | $P($ InNOCENT $)$ |
| :--- | :--- |
| $1-(0.99)$ | $P($ INNOCENT $)$ |

- $\mathbf{P}($ Innocent $)=0.9 \rightarrow$


## Isle Bayesienne: the need for prior assumptions

- "We must be $95 \%$ certain" $\rightarrow \mathbf{P}($ InNOCENT $\mid$ Jail $) \leq \mathbf{0 . 0 5}$
- Can Positive $\rightarrow$ Jail?
- Apply Bayes' theorem
$P$ (InNOCENT $\mid$ Jail

$)=$| $(0.01)$ | $P($ InNOCENT $)$ |
| :--- | :--- |
| $1-(0.99)$ | $P($ InNOCENT $)$ |

- $\mathbf{P}($ Innocent $)=0.9 \rightarrow \mathbf{P}($ Innocent $\mid$ Jail $) \approx 0.08$!!


## On the islands of Ste. Frequentiste and Bayesienne...

- More than $1 \%$ of Ste. Frequentiste goes to jail.
- On Isle Bayesienne, 10\% are assumed guilty, but nobody goes to jail.
- The disagreement wasn't about math or how to interpret $P()$.
- What was it about?


## The islanders' concerns

- Frequentist cares about the rate of jailings among innocent people. Concern: overall rate of false positive
- Bayesian cares about the rate of innocence among jail inmates. Concern: rate of error among positives
- The Bayesian had to make an assumption about the overall probability of innocence.


## Quantifying uncertainty

## Quantifying uncertainty



## Quantifying uncertainty



## Quantifying uncertainty



## Quantifying uncertainty



Inference procedure

## Quantifying uncertainty



## Quantifying uncertainty



## Jewel's Cookies

Cookie jars A, B, C, D have 100 cookies each, but different numbers of chocolate chips per cookie:

| $P($ chips $\mid$ jar $)$ | $\mathbf{A}$ | $\mathbf{B}$ | $\mathbf{C}$ | $\mathbf{D}$ |
| :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1 | 12 | 13 | 27 |
| $\mathbf{1}$ | 1 | 19 | 20 | 70 |
| $\mathbf{2}$ | 70 | 24 | 0 | 1 |
| $\mathbf{3}$ | 28 | 20 | 0 | 1 |
| $\mathbf{4}$ | 0 | 25 | 67 | 1 |
| total | $100 \%$ | $100 \%$ | $100 \%$ | $100 \%$ |

## Quantifying cookie jar uncertainty



Underlying parameters

Inference procedure

## Quantifying cookie jar uncertainty



## Cookie jar

 A, B, C or D
## Experiment <br> 

Inference procedure

## Quantifying cookie jar uncertainty



## Cookie jar

 A, B, C or D
## Sample 1 cookie <br> 

Inference procedure


## Quantifying cookie jar uncertainty



## Sample 1 cookie

Inference procedure

## Quantifying cookie jar uncertainty



## Sample 1 cookie

## Uncertainty interval



## Quantifying cookie jar uncertainty



## Frequentist inference

A 70\% confidence interval method includes the correct jar with at least $70 \%$ probability in the worst case, no matter what.

## Making 70\% confidence intervals

Cookie jars A, B, C, D have 100 cookies each, but different numbers of chocolate chips per cookie:

| $P($ chips $\mid$ jar $)$ | $\mathbf{A}$ | $\mathbf{B}$ | $\mathbf{C}$ | $\mathbf{D}$ |
| :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1 | 12 | 13 | 27 |
| $\mathbf{1}$ | 1 | 19 | 20 | 70 |
| $\mathbf{2}$ | 70 | 24 | 0 | 1 |
| $\mathbf{3}$ | 28 | 20 | 0 | 1 |
| $\mathbf{4}$ | 0 | 25 | 67 | 1 |
| coverage |  |  |  |  |

## Making 70\% confidence intervals

Cookie jars A, B, C, D have 100 cookies each, but different numbers of chocolate chips per cookie:

| $P($ chips $\mid$ jar $)$ | $\mathbf{A}$ | $\mathbf{B}$ | $\mathbf{C}$ | $\mathbf{D}$ |
| :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1 | 12 | 13 | 27 |
| $\mathbf{1}$ | 1 | 19 | 20 | 70 |
| $\mathbf{2}$ | 70 | 24 | 0 | 1 |
| $\mathbf{3}$ | 28 | 20 | 0 | 1 |
| $\mathbf{4}$ | 0 | 25 | 67 | 1 |
| coverage | $70 \%$ |  |  |  |

## Making 70\% confidence intervals

Cookie jars A, B, C, D have 100 cookies each, but different numbers of chocolate chips per cookie:

| $P($ chips $\mid$ jar $)$ | $\mathbf{A}$ | $\mathbf{B}$ | $\mathbf{C}$ | $\mathbf{D}$ |
| :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1 | 12 | 13 | 27 |
| $\mathbf{1}$ | 1 | 19 | 20 | 70 |
| $\mathbf{2}$ | 70 | 24 | 0 | 1 |
| $\mathbf{3}$ | 28 | 20 | 0 | 1 |
| $\mathbf{4}$ | 0 | 25 | 67 | 1 |
| coverage | $70 \%$ | $25 \%$ |  |  |

## Making 70\% confidence intervals

Cookie jars A, B, C, D have 100 cookies each, but different numbers of chocolate chips per cookie:

| $P($ chips $\mid$ jar $)$ | $\mathbf{A}$ | $\mathbf{B}$ | $\mathbf{C}$ | $\mathbf{D}$ |
| :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1 | 12 | 13 | 27 |
| $\mathbf{1}$ | 1 | 19 | 20 | 70 |
| $\mathbf{2}$ | 70 | 24 | 0 | 1 |
| $\mathbf{3}$ | 28 | 20 | 0 | 1 |
| $\mathbf{4}$ | 0 | 25 | 67 | 1 |
| coverage | $70 \%$ | $49 \%$ |  |  |

## Making 70\% confidence intervals

Cookie jars A, B, C, D have 100 cookies each, but different numbers of chocolate chips per cookie:

| $P($ chips $\mid$ jar $)$ | $\mathbf{A}$ | $\mathbf{B}$ | $\mathbf{C}$ | $\mathbf{D}$ |
| :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1 | 12 | 13 | 27 |
| $\mathbf{1}$ | 1 | 19 | 20 | 70 |
| $\mathbf{2}$ | 70 | 24 | 0 | 1 |
| $\mathbf{3}$ | 28 | 20 | 0 | 1 |
| $\mathbf{4}$ | 0 | 25 | 67 | 1 |
| coverage | $70 \%$ | $69 \%$ |  |  |

## Making 70\% confidence intervals

Cookie jars A, B, C, D have 100 cookies each, but different numbers of chocolate chips per cookie:

| $P($ chips $\mid$ jar $)$ | $\mathbf{A}$ | $\mathbf{B}$ | $\mathbf{C}$ | $\mathbf{D}$ |
| :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1 | 12 | 13 | 27 |
| $\mathbf{1}$ | 1 | 19 | 20 | 70 |
| $\mathbf{2}$ | 70 | 24 | 0 | 1 |
| $\mathbf{3}$ | 28 | 20 | 0 | 1 |
| $\mathbf{4}$ | 0 | 25 | 67 | 1 |
| coverage | $70 \%$ | $88 \%$ |  |  |

## Making 70\% confidence intervals

Cookie jars A, B, C, D have 100 cookies each, but different numbers of chocolate chips per cookie:

| $P($ chips $\mid$ jar $)$ | $\mathbf{A}$ | $\mathbf{B}$ | $\mathbf{C}$ | $\mathbf{D}$ |
| :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1 | 12 | 13 | 27 |
| $\mathbf{1}$ | 1 | 19 | 20 | 70 |
| $\mathbf{2}$ | 70 | 24 | 0 | 1 |
| $\mathbf{3}$ | 28 | 20 | 0 | 1 |
| $\mathbf{4}$ | 0 | 25 | 67 | 1 |
| coverage | $70 \%$ | $88 \%$ | $67 \%$ |  |

## Making 70\% confidence intervals

Cookie jars A, B, C, D have 100 cookies each, but different numbers of chocolate chips per cookie:

| $P($ chips $\mid$ jar $)$ | $\mathbf{A}$ | $\mathbf{B}$ | $\mathbf{C}$ | $\mathbf{D}$ |
| :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1 | 12 | 13 | 27 |
| $\mathbf{1}$ | 1 | 19 | 20 | 70 |
| $\mathbf{2}$ | 70 | 24 | 0 | 1 |
| $\mathbf{3}$ | 28 | 20 | 0 | 1 |
| $\mathbf{4}$ | 0 | 25 | 67 | 1 |
| coverage | $70 \%$ | $88 \%$ | $87 \%$ |  |

## Making 70\% confidence intervals

Cookie jars A, B, C, D have 100 cookies each, but different numbers of chocolate chips per cookie:

| $P($ chips $\mid$ jar $)$ | $\mathbf{A}$ | $\mathbf{B}$ | $\mathbf{C}$ | $\mathbf{D}$ |
| :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1 | 12 | 13 | 27 |
| $\mathbf{1}$ | 1 | 19 | 20 | 70 |
| $\mathbf{2}$ | 70 | 24 | 0 | 1 |
| $\mathbf{3}$ | 28 | 20 | 0 | 1 |
| $\mathbf{4}$ | 0 | 25 | 67 | 1 |
| coverage | $70 \%$ | $88 \%$ | $87 \%$ | $70 \%$ |

## Bayesian inference

A 70\% credible interval has at least 70\% conditional probability of including the correct jar, given the observation and the prior assumptions.

## Uniform prior

Our prior assumption: jars $A, B, C$, and $D$ have equal probability.

## Conditional probabilities

| $P($ chips $\mid$ jar $)$ | $\mathbf{A}$ | $\mathbf{B}$ | $\mathbf{C}$ | $\mathbf{D}$ |
| :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1 | 12 | 13 | 27 |
| $\mathbf{1}$ | 1 | 19 | 20 | 70 |
| $\mathbf{2}$ | 70 | 24 | 0 | 1 |
| $\mathbf{3}$ | 28 | 20 | 0 | 1 |
| $\mathbf{4}$ | 0 | 25 | 67 | 1 |
| total | $100 \%$ | $100 \%$ | $100 \%$ | $100 \%$ |

## Conditional probabilities

| $P($ chips $\mid$ jar $)$ | A | B | C | $\mathbf{D}$ |
| :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1 | 12 | 13 | 27 |
| $\mathbf{1}$ | 1 | 19 | 20 | 70 |
| $\mathbf{2}$ | 70 | 24 | 0 | 1 |
| $\mathbf{3}$ | 28 | 20 | 0 | 1 |
| $\mathbf{4}$ | 0 | 25 | 67 | 1 |
| total | $100 \%$ | $100 \%$ | $100 \%$ | $100 \%$ |

## Joint probabilities under uniform prior

| $P($ chips $\cap$ jar $)$ | A | B | C | D | $P$ (chips $)$ |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | $1 / 4$ | $12 / 4$ | $13 / 4$ | $27 / 4$ | $13.25 \%$ |
| $\mathbf{1}$ | $1 / 4$ | $19 / 4$ | $20 / 4$ | $70 / 4$ | $27.50 \%$ |
| $\mathbf{2}$ | $70 / 4$ | $24 / 4$ | $0 / 4$ | $1 / 4$ | $23.75 \%$ |
| $\mathbf{3}$ | $28 / 4$ | $20 / 4$ | $0 / 4$ | $1 / 4$ | $12.25 \%$ |
| $\mathbf{4}$ | $0 / 4$ | $25 / 4$ | $67 / 4$ | $1 / 4$ | $23.25 \%$ |
| total $P($ jar $)$ | $25 \%$ | $25 \%$ | $25 \%$ | $25 \%$ | $100 \%$ |

## Joint probabilities under uniform prior

| $P($ chips $\cap$ jar $)$ | A | B | C | D | $P($ chips $)$ |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | $1 / 4$ | $12 / 4$ | $13 / 4$ | $27 / 4$ | $13.25 \%$ |
| $\mathbf{1}$ | $1 / 4$ | $19 / 4$ | $20 / 4$ | $70 / 4$ | $27.50 \%$ |
| $\mathbf{2}$ | $70 / 4$ | $24 / 4$ | $0 / 4$ | $1 / 4$ | $23.75 \%$ |
| $\mathbf{3}$ | $28 / 4$ | $20 / 4$ | $0 / 4$ | $1 / 4$ | $12.25 \%$ |
| $\mathbf{4}$ | $0 / 4$ | $25 / 4$ | $67 / 4$ | $1 / 4$ | $23.25 \%$ |
| total $P($ jar $)$ | $25 \%$ | $25 \%$ | $25 \%$ | $25 \%$ | $100 \%$ |

## Posterior probabilities under uniform prior

|  | A | B | C | D | P(chips) |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | $1 / 4$ | $12 / 4$ | $13 / 4$ | $27 / 4$ | $13.25 \%$ |
| $\mathbf{1}$ | $1 / 4$ | $19 / 4$ | $20 / 4$ | $70 / 4$ | $27.50 \%$ |
| $\mathbf{2}$ | $70 / 4$ | $24 / 4$ | $0 / 4$ | $1 / 4$ | $23.75 \%$ |
| $\mathbf{3}$ | $28 / 4$ | $20 / 4$ | $0 / 4$ | $1 / 4$ | $12.25 \%$ |
| $\mathbf{4}$ | $0 / 4$ | $25 / 4$ | $67 / 4$ | $1 / 4$ | $23.25 \%$ |

## Posterior probabilities under uniform prior

|  | A | B | C | D | P(chips) |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1.9 | 22.6 | 24.5 | 50.9 | $100 \%$ |
| $\mathbf{1}$ | $1 / 4$ | $19 / 4$ | $20 / 4$ | $70 / 4$ | $27.50 \%$ |
| $\mathbf{2}$ | $70 / 4$ | $24 / 4$ | $0 / 4$ | $1 / 4$ | $23.75 \%$ |
| $\mathbf{3}$ | $28 / 4$ | $20 / 4$ | $0 / 4$ | $1 / 4$ | $12.25 \%$ |
| $\mathbf{4}$ | $0 / 4$ | $25 / 4$ | $67 / 4$ | $1 / 4$ | $23.25 \%$ |

## Posterior probabilities under uniform prior

|  | A | B | C | D | P(chips) |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1.9 | 22.6 | 24.5 | 50.9 | $100 \%$ |
| $\mathbf{1}$ | $1 / 4$ | $19 / 4$ | $20 / 4$ | $70 / 4$ | $27.50 \%$ |
| $\mathbf{2}$ | $70 / 4$ | $24 / 4$ | $0 / 4$ | $1 / 4$ | $23.75 \%$ |
| $\mathbf{3}$ | $28 / 4$ | $20 / 4$ | $0 / 4$ | $1 / 4$ | $12.25 \%$ |
| $\mathbf{4}$ | $0 / 4$ | $25 / 4$ | $67 / 4$ | $1 / 4$ | $23.25 \%$ |

## Posterior probabilities under uniform prior

|  | A | B | C | D | P(chips) |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1.9 | 22.6 | 24.5 | 50.9 | $100 \%$ |
| $\mathbf{1}$ | 0.9 | 17.3 | 18.2 | 63.6 | $100 \%$ |
| $\mathbf{2}$ | $70 / 4$ | $24 / 4$ | $0 / 4$ | $1 / 4$ | $23.75 \%$ |
| $\mathbf{3}$ | $28 / 4$ | $20 / 4$ | $0 / 4$ | $1 / 4$ | $12.25 \%$ |
| $\mathbf{4}$ | $0 / 4$ | $25 / 4$ | $67 / 4$ | $1 / 4$ | $23.25 \%$ |

## Posterior probabilities under uniform prior

|  | A | B | C | D | P(chips) |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1.9 | 22.6 | 24.5 | 50.9 | $100 \%$ |
| $\mathbf{1}$ | 0.9 | 17.3 | 18.2 | 63.6 | $100 \%$ |
| $\mathbf{2}$ | $70 / 4$ | $24 / 4$ | $0 / 4$ | $1 / 4$ | $23.75 \%$ |
| $\mathbf{3}$ | $28 / 4$ | $20 / 4$ | $0 / 4$ | $1 / 4$ | $12.25 \%$ |
| $\mathbf{4}$ | $0 / 4$ | $25 / 4$ | $67 / 4$ | $1 / 4$ | $23.25 \%$ |

## Posterior probabilities under uniform prior

|  | A | B | C | D | P(chips) |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1.9 | 22.6 | 24.5 | 50.9 | $100 \%$ |
| $\mathbf{1}$ | 0.9 | 17.3 | 18.2 | 63.6 | $100 \%$ |
| $\mathbf{2}$ | 73.7 | 25.3 | 0.0 | 1.1 | $100 \%$ |
| $\mathbf{3}$ | $28 / 4$ | $20 / 4$ | $0 / 4$ | $1 / 4$ | $12.25 \%$ |
| $\mathbf{4}$ | $0 / 4$ | $25 / 4$ | $67 / 4$ | $1 / 4$ | $23.25 \%$ |

## Posterior probabilities under uniform prior

|  | A | B | C | D | P(chips) |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1.9 | 22.6 | 24.5 | 50.9 | $100 \%$ |
| $\mathbf{1}$ | 0.9 | 17.3 | 18.2 | 63.6 | $100 \%$ |
| $\mathbf{2}$ | 73.7 | 25.3 | 0.0 | 1.1 | $100 \%$ |
| $\mathbf{3}$ | $28 / 4$ | $20 / 4$ | $0 / 4$ | $1 / 4$ | $12.25 \%$ |
| $\mathbf{4}$ | $0 / 4$ | $25 / 4$ | $67 / 4$ | $1 / 4$ | $23.25 \%$ |
|  |  |  |  |  |  |

## Posterior probabilities under uniform prior

|  | A | B | C | D | P(chips $)$ |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1.9 | 22.6 | 24.5 | 50.9 | $100 \%$ |
| $\mathbf{1}$ | 0.9 | 17.3 | 18.2 | 63.6 | $100 \%$ |
| $\mathbf{2}$ | 73.7 | 25.3 | 0.0 | 1.1 | $100 \%$ |
| $\mathbf{3}$ | 57.1 | 40.8 | 0.0 | 2.0 | $100 \%$ |
| $\mathbf{4}$ | $0 / 4$ | $25 / 4$ | $67 / 4$ | $1 / 4$ | $23.25 \%$ |

## Posterior probabilities under uniform prior

|  | A | $\mathbf{B}$ | $\mathbf{C}$ | $\mathbf{D}$ | P(chips $)$ |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1.9 | 22.6 | 24.5 | 50.9 | $100 \%$ |
| $\mathbf{1}$ | 0.9 | 17.3 | 18.2 | 63.6 | $100 \%$ |
| $\mathbf{2}$ | 73.7 | 25.3 | 0.0 | 1.1 | $100 \%$ |
| $\mathbf{3}$ | 57.1 | 40.8 | 0.0 | 2.0 | $100 \%$ |
| $\mathbf{4}$ | $0 / 4$ | $25 / 4$ | $67 / 4$ | $1 / 4$ | $23.25 \%$ |

## Posterior probabilities under uniform prior

|  | A | B | C | D | P(chips) |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1.9 | 22.6 | 24.5 | 50.9 | $100 \%$ |
| $\mathbf{1}$ | 0.9 | 17.3 | 18.2 | 63.6 | $100 \%$ |
| $\mathbf{2}$ | 73.7 | 25.3 | 0.0 | 1.1 | $100 \%$ |
| $\mathbf{3}$ | 57.1 | 40.8 | 0.0 | 2.0 | $100 \%$ |
| $\mathbf{4}$ | 0.0 | 26.9 | 72.0 | 1.1 | $100 \%$ |

## Posterior probabilities under uniform prior

| $P($ jar $\mid$ chips $)$ | A | B | C | D | $P($ chips $)$ |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1.9 | 22.6 | 24.5 | 50.9 | $100 \%$ |
| $\mathbf{1}$ | 0.9 | 17.3 | 18.2 | 63.6 | $100 \%$ |
| $\mathbf{2}$ | 73.7 | 25.3 | 0.0 | 1.1 | $100 \%$ |
| $\mathbf{3}$ | 57.1 | 40.8 | 0.0 | 2.0 | $100 \%$ |
| $\mathbf{4}$ | 0.0 | 26.9 | 72.0 | 1.1 | $100 \%$ |

## 70\% credible intervals

| $P($ jar $\mid$ chips $)$ | A | B | C | $\mathbf{D}$ | credibility |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1.9 | 22.6 | 24.5 | 50.9 |  |
| $\mathbf{1}$ | 0.9 | 17.3 | 18.2 | 63.6 |  |
| $\mathbf{2}$ | 73.7 | 25.3 | 0.0 | 1.1 |  |
| $\mathbf{3}$ | 57.1 | 40.8 | 0.0 | 2.0 |  |
| $\mathbf{4}$ | 0.0 | 26.9 | 72.0 | 1.1 |  |

## 70\% credible intervals

| $P($ jar $\mid$ chips $)$ | A | B | C | $\mathbf{D}$ | credibility |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1.9 | 22.6 | 24.5 | 50.9 | $51 \%$ |
| $\mathbf{1}$ | 0.9 | 17.3 | 18.2 | 63.6 |  |
| $\mathbf{2}$ | 73.7 | 25.3 | 0.0 | 1.1 |  |
| $\mathbf{3}$ | 57.1 | 40.8 | 0.0 | 2.0 |  |
| $\mathbf{4}$ | 0.0 | 26.9 | 72.0 | 1.1 |  |

## $70 \%$ credible intervals

| $P($ jar $\mid$ chips $)$ | A | B | C | $\mathbf{D}$ | credibility |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1.9 | 22.6 | 24.5 | 50.9 | $75 \%$ |
| $\mathbf{1}$ | 0.9 | 17.3 | 18.2 | 63.6 |  |
| $\mathbf{2}$ | 73.7 | 25.3 | 0.0 | 1.1 |  |
| $\mathbf{3}$ | 57.1 | 40.8 | 0.0 | 2.0 |  |
| $\mathbf{4}$ | 0.0 | 26.9 | 72.0 | 1.1 |  |

## $70 \%$ credible intervals

| $P($ jar $\mid$ chips $)$ | A | B | C | D | credibility |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1.9 | 22.6 | 24.5 | 50.9 | $75 \%$ |
| $\mathbf{1}$ | 0.9 | 17.3 | 18.2 | 63.6 | $64 \%$ |
| $\mathbf{2}$ | 73.7 | 25.3 | 0.0 | 1.1 |  |
| $\mathbf{3}$ | 57.1 | 40.8 | 0.0 | 2.0 |  |
| $\mathbf{4}$ | 0.0 | 26.9 | 72.0 | 1.1 |  |

## $70 \%$ credible intervals

| $P($ jar $\mid$ chips $)$ | A | B | C | D | credibility |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1.9 | 22.6 | 24.5 | 50.9 | $75 \%$ |
| $\mathbf{1}$ | 0.9 | 17.3 | 18.2 | 63.6 | $82 \%$ |
| $\mathbf{2}$ | 73.7 | 25.3 | 0.0 | 1.1 |  |
| $\mathbf{3}$ | 57.1 | 40.8 | 0.0 | 2.0 |  |
| $\mathbf{4}$ | 0.0 | 26.9 | 72.0 | 1.1 |  |

## 70\% credible intervals

| $P($ jar $\mid$ chips $)$ | A | B | C | D | credibility |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1.9 | 22.6 | 24.5 | 50.9 | $75 \%$ |
| $\mathbf{1}$ | 0.9 | 17.3 | 18.2 | 63.6 | $82 \%$ |
| $\mathbf{2}$ | 73.7 | 25.3 | 0.0 | 1.1 | $74 \%$ |
| $\mathbf{3}$ | 57.1 | 40.8 | 0.0 | 2.0 |  |
| $\mathbf{4}$ | 0.0 | 26.9 | 72.0 | 1.1 |  |

## 70\% credible intervals

| $P($ jar $\mid$ chips $)$ | A | B | C | D | credibility |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1.9 | 22.6 | 24.5 | 50.9 | $75 \%$ |
| $\mathbf{1}$ | 0.9 | 17.3 | 18.2 | 63.6 | $82 \%$ |
| $\mathbf{2}$ | 73.7 | 25.3 | 0.0 | 1.1 | $74 \%$ |
| $\mathbf{3}$ | 57.1 | 40.8 | 0.0 | 2.0 | $57 \%$ |
| $\mathbf{4}$ | 0.0 | 26.9 | 72.0 | 1.1 |  |

## 70\% credible intervals

| $P($ jar $\mid$ chips $)$ | A | B | C | D | credibility |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1.9 | 22.6 | 24.5 | 50.9 | $75 \%$ |
| $\mathbf{1}$ | 0.9 | 17.3 | 18.2 | 63.6 | $82 \%$ |
| $\mathbf{2}$ | 73.7 | 25.3 | 0.0 | 1.1 | $74 \%$ |
| $\mathbf{3}$ | 57.1 | 40.8 | 0.0 | 2.0 | $98 \%$ |
| $\mathbf{4}$ | 0.0 | 26.9 | 72.0 | 1.1 |  |

## 70\% credible intervals

| $P($ jar $\mid$ chips $)$ | A | B | C | D | credibility |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1.9 | 22.6 | 24.5 | 50.9 | $75 \%$ |
| $\mathbf{1}$ | 0.9 | 17.3 | 18.2 | 63.6 | $82 \%$ |
| $\mathbf{2}$ | 73.7 | 25.3 | 0.0 | 1.1 | $74 \%$ |
| $\mathbf{3}$ | 57.1 | 40.8 | 0.0 | 2.0 | $98 \%$ |
| $\mathbf{4}$ | 0.0 | 26.9 | 72.0 | 1.1 | $72 \%$ |

## Confidence \& credible intervals together

| confidence | A | B | C | $\mathbf{D}$ | credibility |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1 | 12 | 13 | 27 | $\mathbf{0 \%}$ |
| $\mathbf{1}$ | 1 | 19 | 20 | 70 | $99 \%$ |
| $\mathbf{2}$ | 70 | 24 | 0 | 1 | $99 \%$ |
| $\mathbf{3}$ | 28 | 20 | 0 | 1 | $\mathbf{4 1 \%}$ |
| $\mathbf{4}$ | 0 | 25 | 67 | 1 | $99 \%$ |
| coverage | $70 \%$ | $88 \%$ | $87 \%$ | $70 \%$ |  |


| credible | A | B | C | D | credibility |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1 | 12 | 13 | 27 | $75 \%$ |
| $\mathbf{1}$ | 1 | 19 | 20 | 70 | $82 \%$ |
| $\mathbf{2}$ | 70 | 24 | 0 | 1 | $74 \%$ |
| $\mathbf{3}$ | 28 | 20 | 0 | 1 | $98 \%$ |
| $\mathbf{4}$ | 0 | 25 | 67 | 1 | $72 \%$ |
| coverage | $98 \%$ | $\mathbf{2 0} \%$ | $100 \%$ | $\mathbf{9 7 \%}$ |  |

## Correlation of error



## Correlation of error

## Experiment <br> 

## Correlation of error



## Criticism of frequentist style

## Essay

## Why Most Published Research Findings Are False

John P. A. loannidis

## Summary

There is increasing concern that most current published research findings are false. The probability that a research clain is true may depend on study power and bias the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probedin each scientific field. In this framework, a res earch finding
factors that influence this problem and some corollaries thereof.

## Modeling the Framework for False Positive Findings

> It can be proven that most claimed research findings are false.

yet ill-founded strategy of claiming conclusive research findiness solelv on
is characteristic of the vary a lot depending o field targets highly like or searches for only on true relationships amo and millions of hypoth be postulated. Let us a for computational sim circumscribed fields wl is only one true relatio many that can be hypa the power is similar to

Why Most Published Research Findings Are False, Ioannidis JPA, PLoS Medicine Vol. 2, No. 8, e124 doi:10.1371/journal.pmed. 0020124

## Criticism of Bayesian style

## ECONOMETRICA

Volume 47
November, 1979
Number 6

## THE IMPOSSIBILITY OF BAYESIAN GROUP DECISION MAKING WITH SEPARATE AGGREGATION OF BELIEFS AND VALUES

By Aanund Hylland and Richard Zeckhauser ${ }^{1}$

Bayesian theory for rational individual decision making under uncertainty prescribes that the decision maker define independently a set of beliefs (probability assessments for the states of the world) and a system of values (utilities). The decision is then made by maximizing expected utility. We attempt to generalize the model to group decision making. It is assumed that the group's belief depends only on individual beliefs and the group's values only on individual values, that the belief aggregation procedure respects unanimity, and that the entire procedure guarantees Pareto optimality. We prove that only trivial (dictatorial) aggregation procedures for beliefs are possible.

## 1. INTRODUCTION

MANY DECISIONS MADE under uncertainty, indeed many important ones, are made by a group, be it a collection of friends, the Congress of the United States, or

## Disagreement in the real world

- Avandia: world's \#1 diabetes drug, approved in 1999.
- Sold by GlaxoSmithKline PLC
- Sales: $\$ 3$ billion in 2006
- In 2004, GSK releases results of many small studies.


## GSK releases 42 small studies

| Study | Avandia heart attacks | Control heart attacks |
| :--- | :--- | :--- |
| $49632-020$ | $2 / 391$ | $1 / 207$ |
| 49653-211 | $5 / 110$ | $2 / 114$ |
| DREAM | $15 / 2635$ | $9 / 2634$ |
| 49653-134 | $0 / 561$ | $2 / 276$ |
| 49653-331 | $0 / 706$ | $0 / 325$ |
| $\vdots$ | $\vdots$ | $\vdots$ |

## In 2007, Dr. Nissen crashes the party



## Effect of Rosiglitazone on the Risk of Myocardial Infarction and Death from Cardiovascular Causes

Steven E. Nissen, M.D., and Kathy Wolski, M.P.H.

## ABSTRACT

## background

Rosiglitazone is widely used to treat patients with type 2 diabetes mellitus, but its effect on cardiovascular morbidity and mortality has not been determined.

## methods

We conducted searches of the published literature, the Web site of the Food and Drug Administration, and a clinical-trials registry maintained by the drug manufacturer (GlaxoSmithKline). Criteria for inclusion in our meta-analysis included a study duration of more than 24 weeks, the use of a randomized control group not receiving rosiglitazone, and the availability of outcome data for myocardial infarction and death from cardiovascular causes. Of 116 potentially relevant studies, 42 trials met the inclusion criteria. We tabulated all occurrences of myocardial infarction and death from cardiovascular causes.

## results

Data were combined by means of a fixed-effects model. In the 42 trials, the mean age of the subjects was approximately 56 years, and the mean baseline glycated hemoglobin level was approximately $8.2 \%$. In the rosiglitazone group, as compared with the control group, the odds ratio for myocardial infarction was 1.43 (95\% confidence interval $[\mathrm{CI}], 1.03$ to $1.98 ; \mathrm{P}=0.03$ ), and the odds ratio for death from cardiovascular causes was 1.64 ( $95 \% \mathrm{CI}, 0.98$ to $2.74 ; \mathrm{P}=0.06$ ).

From the Cleveland Clinic, Cleveland. Address reprint requests to Dr . Nissen at the Department of Cardiovascular Medicine, Cleveland Clinic, 9500 Euclid Ave, Cleveland, OH 44195 , or at nissens @ccf. org.
This article ( $10.1056 / \mathrm{NE}$ ]Moa072761) was published at www.nejm.org on May 21, 2007.

N Engl/ Med 2007;356:2457.71.
Coppriger © 2007 Massachunets Medical Socily.

## Frequentist inference



## THE WALL STREET JOURNAL.



## Buxinese arul Firmance

T- es. ampoyers are atvitest indertimingits nrospect of he coming tove Eminlotyers who rely on ansiolled wotkars menenally on cusialled workars grenerally indurtries that need sialled worl ers complain thm it doesn't give ors complam them the fowitiry 10 recrit hom the fun they noed from atrond A1, R9

- Kerkarlan's Tracindalaunchad an overure for MGMMiraselt leat sown Hival and cituc antor nimiont in Las Vegas, a volley that has pat the whole company in play. As

4 Glaws sharos slid aftor tho Now Pngland Sournal of Medirinere hesed an analysis siggestingus-
ars of dnibetes drus Avandialives ahtgher risk of hism athacks $\mathrm{A}, \mathrm{D}$,
a EMI agreed to be bought by or efine bimy firm ferra firma or $\$ 4.74$ billion, but the mustic fimphidaing may not be over. 83
-Lowe's posted a $12 \%$ profit
irmp and cut its full-year cutlons wat ssid it will ineon un un sporros.
sive stort-opening compaignias

Warld-Wile


Iwo Jima Letters
Of Young Japanese
Are Home at Last

An Amerlean's Sommenit,
They Had Sat on a Shulf;
\$ Solving a Funily Mystery


## MEDICAL DETECTIVE

## Sequel for Vioxx Critic: Attack on Diabetes Pill

Glaxo Shares Plunge
As Dr. Nissen Sees Risk
To Heart From Avandia
by Anna Wide Mathews
An analysis linking the widely used diabetes drug Avandiato higher risk of
heart attacks represents a serious blow to GlaxoSmithkine PLC and underscores how outside crities have been empowered to challenge big-selling drugs after the outcry over the withdrawn painkiller Vioxx.

Glaxo rang up more than $\$ 3$ bir lion in world-wide sales of Avandia last year. Its share
price fell more than price fell more than $7 \%$ after the New Medticine released the analysis by prominent ardiologist Steven Nissen gist Steven Nissen
of the Cleveland Clinic who helped
 C Clinic, who helped raise early safety concerns about ple on Avandia have a $43 \%$ higher ple on Avandia have a $43 \%$ higher
chance of suffering a heart attack.
Glaxo said it "strongly disagrees"
with his miclusione which enme from

## Drug in Demand

Sales of GlaxoSmithKline's Avandia,
in billions of pounds:

 Averdanct and Avindy
Sours the campur
and Drug Administration should have acted faster to alert the public about possible risk fromavandia. Glaxoperalso showed a potential danger It also showed a potential danger, It FDA in September 2005 and a more complete oneinAugust 2006. The findings weren't reflected on the U.S. label, which is supposed to give a combel, which is supposed to give a com-
prehensive review of the drug's risks. prehensive review of the drug's risks.
Robert Meyer, head of the FDA offlce that oversees diabetes drugs, said the agency is still working on its analysis. "We have other data that suegests we

## GlaxoSmithKline loses $\$ 12$ billion

Avandia worldwide sales


## Bayesian inference disagrees, for risk ratio

P.D.F. on Avandia's risk ratio for heart attack

(Joint work with Joshua Mandel, Children's Hospital Informatics Program)

## Or does it? Here, risk difference model

P.D.F. on Avandia's risk difference for heart attack

(Joint work with Joshua Mandel, Children's Hospital Informatics Program)

## The TAXUS ATLAS Experiment

- Boston Scientific proposed to show that new heart stent was not "inferior" to old heart stent, with $95 \%$ confidence.
- Inferior means three percentage points more "bad" events.
- Control $7 \%$ vs. Treatment $10.5 \% \Rightarrow$ inferior
- Control $7 \%$ vs. Treatment $9.5 \% \Rightarrow$ non-inferior.


## ATLAS Results (May 2006)

May 16, 2006 - NATICK, Mass. and PARIS, May 16
/PRNewswire-FirstCall/ - Boston Scientific Corporation today announced nine-month data from its TAXUS ATLAS clinical trial. [...] The trial met its primary endpoint of nine-month target vessel revascularization (TVR), a measure of the effectiveness of a coronary stent in reducing the need for a repeat procedure.

## ATLAS Results (April 2007)

Turco et al., Polymer-Based, Paclitaxel-Eluting TAXUS Liberté Stent in De Novo Lesions, Journal of the American College of Cardiology, Vol. 49, No. 16, 2007.

Results: The primary non-inferiority end point was met with the 1 -sided $95 \%$ confidence bound of $2.98 \%$ less than the pre-specified non-inferiority margin of $3 \%(p=0.0487)$.

Statistical methodology. Student $t$ test was used to compare independent continuous variables, while chi-square or Fisher exact test was used to compare proportions.

## $p$-value

$p<0.05 \rightarrow 95 \%$ confidence interval excludes inferiority

The problem

|  | Event | No event | Total |
| :--- | :--- | :--- | :--- |
| Control | 67 | 889 | 956 |
| Treatment | 68 | 787 | 855 |
| Total | 135 | 1676 | 1811 |

## The problem

|  | Event | No event | Total |
| :--- | :--- | :--- | :--- |
| Control | 67 | 889 | 956 |
| Treatment | 68 | 787 | 855 |
| Total | 135 | 1676 | 1811 |

- With uniform prior on rates,
$\operatorname{Pr}($ inferior $\mid$ data $) \approx 0.050737979 \ldots$
- Posterior probability of non-inferiority is less than $95 \%$.


## ATLAS trial solution

- Confidence interval: approximate each binomial separately with a normal distribution. Known as Wald interval.
- $p=\int_{0.03}^{\infty} \mathcal{N}\left(\frac{i}{m}-\frac{j}{n}, \frac{i(m-i)}{m^{3}}+\frac{j(n-j)}{n^{3}}\right) \approx 0.0487395 \ldots$
- $p<0.05 \rightarrow$ success

The ultimate close call

| Wald's area $(\approx p)$ with $(m, n)=(855,956)$ |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 70 | 9.7 | 8.4 | 7.2 | 6.2 | 5.3 |
| $\bigcirc 69$ | 8.1 | 7.0 | 6.0 | 5.1 | 4.3 |
| 68 | 6.7 | 5.7 | 4.9 | 4.1 | 3.5 |
| \& 67 | 5.5 | 4.7 | 3.9 | 3.3 | 2.8 |
| 66 | 4.5 | 3.8 | 3.1 | 2.6 | 2.2 |
|  | 65 | 66 | 67 | 68 | 69 |
|  | TVR (Express) |  |  |  |  |


| confidence | A | $\mathbf{B}$ | $\mathbf{C}$ | $\mathbf{D}$ |
| :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1 | 12 | 13 | 27 |
| $\mathbf{1}$ | 1 | 19 | 20 | 70 |
| $\mathbf{2}$ | 70 | 24 | 0 | 1 |
| $\mathbf{3}$ | 28 | 20 | 0 | 1 |
| $\mathbf{4}$ | 0 | 25 | 67 | 1 |
| coverage | $70 \%$ | $88 \%$ | $87 \%$ | $70 \%$ |


| confidence | A | $\mathbf{B}$ | $\mathbf{C}$ | $\mathbf{D}$ |
| :--- | :--- | :--- | :--- | :--- |
| $\mathbf{0}$ | 1 | 12 | 13 | 27 |
| $\mathbf{1}$ | 1 | 19 | 20 | 70 |
| $\mathbf{2}$ | 70 | 24 | 0 | 1 |
| $\mathbf{3}$ | 28 | 20 | 0 | 1 |
| $\mathbf{4}$ | 0 | 25 | 67 | 1 |
| coverage | $70 \%$ | $88 \%$ | $87 \%$ | $70 \%$ |
| false positive rate | $30 \%$ | $12 \%$ | $13 \%$ | $30 \%$ |

## The Wald interval undercovers

False Positive Rate of ATLAS non-inferiority test along critical line


## Better approximation: score interval

False Positive Rate of maximum-likelihood $z$-test along critical line


## Other methods all yield failure

| Method | $p$-value or confidence bound | Result |
| :--- | :--- | :--- |
| Wald interval | $p=0.04874$ | Pass |
| $z$-test, constrained max likelihood standard error | $p=0.05151$ | Fail |

## Other methods all yield failure

| Method | $p$-value or confidence bound | Result |
| :--- | :--- | :--- |
| Wald interval | $p=0.04874$ | Pass |
| z-test, constrained max likelihood standard error | $p=0.05151$ | Fail |
| z-test with Yates continuity correction | $c=0.03095$ | Fail |
| Agresti-Caffo I4 interval | $p=0.05021$ | Fail |
| Wilson score | $c=0.03015$ | Fail |
| Wilson score with continuity correction | $c=0.03094$ | Fail |
| Farrington \& Manning score | $p=0.05151$ | Fail |
| Miettinen \& Nurminen score | $p=0.05156$ | Fail |
| Gart \& Nam score | $p=0.05096$ | Fail |
| NCSS's bootstrap method | $c=0.03006$ | Fail |
| NCSS's quasi-exact Chen | $c=0.03016$ | Fail |
| NCSS's exact double-binomial test | $p=0.05470$ | Fail |
| StatXact's approximate unconditional test of non-inferiority | $p=0.05151$ | Fail |
| StatXact's exact unconditional test of non-inferiority | $p=0.05138$ | Fail |
| StatXact's exact CI based on difference of observed rates | $c=0.03737$ | Fail |
| StatXact's approximate CI from inverted 2-sided test | $c=0.03019$ | Fail |
| StatXact's exact Cl from inverted 2-sided test | $c=0.03032$ | Fail |

## Nerdiest chart contender?

## Degree of Certainty

Medical studies define success or failure in testing a hypothesis by calculating a degree of certainty, known as the p-value. The p -value must be less than $5 \%$ for the results to be considered significant. Boston Scientific's study, which used a statistical method called a Wald Interval, produced a p-value below $5 \%$. But using 16 other methods turned up a p-value greater than $5 \%$. Here are some of the $p$-values that resulted from the data in the study, using those different methodologies.

Source: WSJ research


## Boston Scientific Stent Study Flawed

## By Kerth J, Winstein

A
heart stent manufactured by Boston Scientific Corp. and expecting approval for US. sales is backed by flawed research despite the company's claims of success in a clinical trial, according to a Wall Street Journal review of the data

Boston Scientific submitted the results of the 2006 trial to the Food and Drug Administration to gain U.S. approval for the Taxus Liberte, which already is one of the top-selling stents abroad. Coronary stents-tiny scaffolds that prop open arteries clogged by heart disease-are one of the most popular methods for treating heart patients, and have been implanted in more than 15 million people world-wide

But Boston Scientific's claím was based on a flawed statistical equation that favored the Libertestent, a Journal analysis has found. Using a number of other methods of calculation-imcluding 14 available in off-the-shelf software programs-the Liberte study would have been a failure by the common standards of statistical significance in research.

Boston Scientific isn't the only company to use the equation, known as a Wald interval, which has long been criticized

by statisticians for exaggerating the certainty of research results. Rivals Medtronic Inc. and AbbottLaboratories have used the same equation in stent studies. But in those cases, ary boos provided by the Wald equation wouldn't have changed the outcome of the study In the Libert study, the equation's shortcommgs meant the difference between success and failure in the study's main goal.

The difference also sheds light on the leeway that device makers have when designing studies for the FDA. Studies designed to satisfy the requirements of the FDA's medical-de vice branch can be less rigorous
than those aimed at winning U.S. approval for drugs. That is partly because of a 1997 federal law aimed at lessening the regulatory requirements on device makers.

The FDA declined to specifically discuss its deliberations of the Liberte, which is still under review by the agency

Boston Scientific doesn't agree that it made a mistake or that the study failed tor each statistical significance. "We used standard methociology that we discussed with the FDA up front, and then executed," said Donald Baim, Boston Scientific's chief scientific and medical officer.

Please turn to page B6

## The statistician says...

- " $n p>5$, therefore, the Central Limit Theorem applies and a Gaussian approximation is appropriate."
- "We had even more data points than we powered the study for, so there was adequate safety margin."
- "'Exact' tests are too conservative."


## StatXact calculates an "exact" test

"Other statistical applications often rely on large-scale assumptions for inferences, risking incorrect conclusions from data sets not normally distributed. StatXact utilizes Cytel's own powerful algorithms to make exact inferences. .."

## StatXact calculates an "exact" test

"Other statistical applications often rely on large-scale assumptions for inferences, risking incorrect conclusions from data sets not normally distributed. StatXact utilizes Cytel's own powerful algorithms to make exact inferences. .."

## Graphing the coverage

- Problem: hard to calculate a million "exact" p-values
- (StatXact: about 10 minutes each)
- Contribution: method for calculating whole tableau
- Calculates all p-values in time for "hardest" square
- Trick: calculate in the right order, cache partial results


## StatXact calculates an "exact" test

"Other statistical applications often rely on large-scale assumptions for inferences, risking incorrect conclusions from data sets not normally distributed. StatXact utilizes Cytel's own powerful algorithms to make exact inferences..."

## StatXact calculates an "exact" test

"Other statistical applications often rely on large-scale assumptions for inferences, risking incorrect conclusions from data sets not normally distributed. StatXact utilizes Cytel's own powerful algorithms to make exact inferences..."

Type I rate of StatXAct 8 non-inferiority test (Berger Boos-adjusted Chan)


## Animation

## Barnard's test for $\mathrm{N}=256 \times 257$



Assume a prior hypothesis. . .


Frequentists can benefit from priors too!


## Final thoughts

- Bayesian and frequentist schools have much in common.
- If stark disagreement between Bayesian and frequentist methods, probably sign of bigger problems!
- What's important: say what we're trying to infer, how we get there, what we care about.

