Styles of Inference: Bayesianness and Frequentism

Keith Winstein

keithw@mit.edu

April 8, 2011

Axioms of Probability

Let S be a finite set called the *sample space*, and let A be any subset of S, called an *event*. The *probability* $P(A)$ is a real-valued function that satisfies:

- $P(A) \geq 0$
- $P(S) = 1$
- $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \emptyset$

For infinite sample space, third axiom is that for an infinite sequence of disjoint subsets $A_1, A_2, \ldots,$

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

Some Theorems

- $P(\overline{A}) = 1 - P(A)$
- $P(\emptyset) = 0$
- $P(A) \leq P(B)$ 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

- If A and B are two events (subsets of S), then call $P(A \cap B)$ the *joint probability* of A and B.
- Define the *conditional probability of A given B* as:

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

- A and B are said to be *independent* if $P(A \cap B) = P(A)P(B)$.
- If A and B are independent, then $P(A|B) = P(A)$.

Bayes' Theorem

We have:

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

Therefore:

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

And Bayes' Theorem is:

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

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

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

The king has been poisoned!

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

The king of Ste. F & B has been poisoned! It's a conspiracy. An order goes out to the regional governors of Ste. Frequentiste and of Isle Bayesienne: find those responsible, and jail them.

Dear Governor: Attached is a blood test for proximity to the poison that killed the king. It has a 0% rate of false negative and a 1% rate of false positive. Administer it to everybody on your island, and if you conclude they're guilty, jail them.

BUT REMEMBER THE NATIONWIDE LAW: **We must be 95% certain of guilt to send a citizen to jail.**

On Ste. Frequentiste:

The test has a 0% rate of false negative and a 1% rate of false positive. **We must be 95% certain of guilt to send a citizen to jail.**

- $P(E^+|\textsc{Guilty}) = 1$
- $P(E^-|\textsc{Guilty}) = 0$
- $P(E^+|\textsc{Innocent}) = 0.01$
- $P(E^-|\textsc{Innocent}) = 0.99$

How to interpret the law?

"We must be 95% certain of guilt" $\Rightarrow P(\textsc{Jail}|\textsc{Innocent}) \leq 5\%$.

On Ste. Frequentiste:

The test has a 0% rate of false negative and a 1% rate of false positive. **We must be 95% certain of guilt to send a citizen to jail.**

- $P(E^+|\textsc{Guilty}) = 1$
- $P(E^-|\textsc{Guilty}) = 0$
- $P(E^+|\textsc{Innocent}) = 0.01$
- $P(E^-|\textsc{Innocent}) = 0.99$

How to interpret the law?
"We must be 95% certain of guilt" $\Rightarrow P(\textsc{Jail}|\textsc{Innocent}) \leq 5\%$.

Governor F.: *Ok, what if I jail everybody with a positive test result? Then* $P(\textsc{Jail}|\textsc{Innocent}) = P(E^+|\textsc{Innocent}) = 1\%$. *That's less than 5%, so we're obeying the law."*

On Isle Bayesienne:

The test has a 0% rate of false negative and a 1% rate of false positive. **We must be 95% certain of guilt to send a citizen to jail.**

How to interpret the law?

"We must be 95% certain of guilt" $\Rightarrow P(\textsc{Innocent}|\textsc{Jail}) \leq 5\%$.

On Isle Bayesienne:

The test has a 0% rate of false negative and a 1% rate of false positive. **We must be 95% certain of guilt to send a citizen to jail.**

How to interpret the law?
"We must be 95% certain of guilt" $\Rightarrow P(\textsc{Innocent}|\textsc{Jail}) \leq 5\%$.

Governor B.: *Can I jail everyone with a positive result? I'll apply Bayes' theorem...*

$$P(\textsc{Innocent}|E^+) = P(E^+|\textsc{Innocent})\frac{P(\textsc{Innocent})}{P(E^+)}$$

We need to know $P(\textsc{Innocent})$.

On Isle Bayesienne:

> The test has a 0% rate of false negative and a 1% rate of false positive. **We must be 95% certain of guilt to send a citizen to jail.**

How to interpret the law?
"We must be 95% certain of guilt" $\Rightarrow P(\textsc{Innocent}|\textsc{Jail}) \leq 5\%$.

Governor B.: *Can I jail everyone with a positive result? I'll apply Bayes' theorem...*

$$P(\textsc{Innocent}|E^+) = P(E^+|\textsc{Innocent})\frac{P(\textsc{Innocent})}{P(E^+)}$$

We need to know $P(\textsc{Innocent})$. Governor B.: *Hmm, I will assume that 10% of my subjects were guilty of the conspiracy.* $P(\textsc{Innocent}) = 0.9$.

On Isle Bayesienne:

Apply Bayes' theorem

- We know the conditional probabilities of the form $P(E^+|\textsc{Guilty})$.
- Governor knows the "overall" probability of each event $\textsc{Guilty}$ and $\textsc{Innocent}$. Since this is our estimate of the chance someone is guilty *before* a blood test, we call it the *prior probability*.
- Now calculate: $P(\textsc{Innocent}|E^+)$

On Isle Bayesienne:

Apply Bayes' theorem

- We know the conditional probabilities of the form $P(E^+|\textsc{Guilty})$.
- Governor knows the "overall" probability of each event $\textsc{Guilty}$ and $\textsc{Innocent}$. Since this is our estimate of the chance someone is guilty *before* a blood test, we call it the *prior probability*.
- Now calculate: $P(\textsc{Innocent}|E^+) \approx 8\%$. Too high!

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

Results:

- More than 1% of Ste. Frequentiste goes to jail.
- On Isle Bayesienne, 10% are guilty, but nobody goes to jail.
- The disagreement isn't about math. It isn't necessarily about philosophy. Here, the frequentist and Bayesian used tests that met different constraints and got different results.

The Constraints

- The frequentist cares about the rate of jailings among innocent people and wants it to be less than 5%. Concern: **overall rate of false positive**.
- The Bayesian cares about the rate of innocence among jail inmates and wants it to be less than 5%. Concern: **rate of error among positives**.
- The Bayesian had to make assumptions about the overall, or prior, probabilities.

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim 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 probed in each scientific field. In this framework, a research 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 findings solely 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 simp
circumscribed fields wl
is only one true relatio
many that can be hypo
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

Confidence & Credibility

- For similar reasons, frequentists and Bayesians express uncertainty differently.
- Both use *intervals*: a function that maps each possible observation to a set of parameters.
- Frequentists use **confidence intervals**. A 95% confidence interval *method* will output an interval that includes the true value at least 95% of the time.
- Bayesians use **credibility intervals**. A 95% credibility interval has 95% probability of including the true value — if drawn according to the prior.

Jewel's Cookies

Cookie jars **A**, **B**, **C**, **D** have the following distribution of cookies with chocolate chips:

$P($ *chips* $\mid$ *jar* $)$	**A**	**B**	**C**	**D**
0	1	17	14	27
1	1	20	22	70
2	70	22	20	1
3	28	20	22	1
4	0	21	22	1
total	100%	100%	100%	100%

Let's construct a **70%** confidence interval.

70% Confidence Intervals

Cookie jars **A**, **B**, **C**, **D** have the following distribution of cookies with chocolate chips:

$P(\ chips \mid jar\ )$	**A**	**B**	**C**	**D**
0	1	17	14	27
1	1	**[20**	**22**	**70]**
2	**[70**	**22**	**20]**	1
3	28	**[20**	**22]**	1
4	0	**[21**	**22]**	1
coverage	70%	83%	86%	70%

The **70%** confidence interval has at least 70% coverage for every value of the parameter.
Now assume a uniform prior and calculate $P(\ jar \cap\ chips\ )$.

Joint Probabilities

Cookie jars **A**, **B**, **C**, **D** have equal chance of being selected, and the following joint distribution of jar and chips:

$P(\ jar \cap chips\ )$	**A**	**B**	**C**	**D**	total
0	1/4	17/4	14/4	27/4	14.75%
1	1/4	20/4	22/4	70/4	28.25%
2	70/4	22/4	20/4	1/4	28.25%
3	28/4	20/4	22/4	1/4	17.75%
4	0/4	21/4	22/4	1/4	11.00%
total	25%	25%	25%	25%	

Now calculate $P(\ jar \mid chips\ )$.

$P(\ outcome\ |\theta)$

Cookie jars **A**, **B**, **C**, **D** have the following conditional probability of each jar given the number of chips:

$P(\ jar \mid chips\ )$	**A**	**B**	**C**	**D**	total
0	1.7	28.8	23.7	45.8	100%
1	0.9	17.7	19.5	61.9	100%
2	61.9	19.5	17.7	0.9	100%
3	39.4	28.2	31.0	1.4	100%
4	0.0	47.7	50.0	2.3	100%

Now let's make **70%** credibility intervals.

70% Credibility Intervals

Cookie jars **A**, **B**, **C**, **D** have the following conditional probability of each jar given the number of chips:

$P($ *jar* $\mid$ *chips* $)$	**A**	**B**	**C**	**D**	credibility
0	1.7	**[28.8]**	23.7	**[45.8]**	75%
1	0.9	17.7	**[19.5**	**61.9]**	81%
2	**[61.9**	**19.5]**	17.7	0.9	81%
3	**[39.4]**	28.2	**[31.0]**	1.4	70%
4	0.0	**[47.7**	**50.0]**	2.3	98%

Confidence & Credible Intervals (uniform prior)

$4P(\ jar \cap chips\ )$	**A**	**B**	**C**	**D**	credibility
0	1	17	14	27	**0%**
1	1	**[20**	**22**	**70]**	99%
2	**[70**	**22**	**20]**	1	99%
3	28	**[20**	**22]**	1	**59%**
4	0	**[21**	**22]**	1	98%
coverage	70%	83%	86%	70%	

$4P(\ jar \cap chips\ )$	**A**	**B**	**C**	**D**	credibility
0	1	**[17]**	14	**[27]**	75%
1	1	20	**[22**	**70]**	81%
2	**[70**	**22]**	20	1	81%
3	**[28]**	20	**[22]**	1	70%
4	0	**[21**	**22]**	1	98%
coverage	98%	**60%**	**66%**	97%	

Disagreement in the real world

- Avandia: world's #1 diabetes drug
- Approved in 1999.
- Sold by GlaxoSmithKline PLC.
- Lowers blood sugar, a lot.
- Sales: $3 billion in 2006 alone
- In 2004, GSK releases results of many small studies of Avandia.
- This enables inference.

Individually, 42 small studies are pretty lame.

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
⋮	⋮	⋮

In 2007, Dr. Nissen crashes the party

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 JUNE 14, 2007 VOL. 356 NO. 24

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 [CI], 1.03 to 1.98; P=0.03), and the odds ratio for death from cardiovascular causes was 1.64 (95% CI, 0.98 to 2.74; P=0.06).

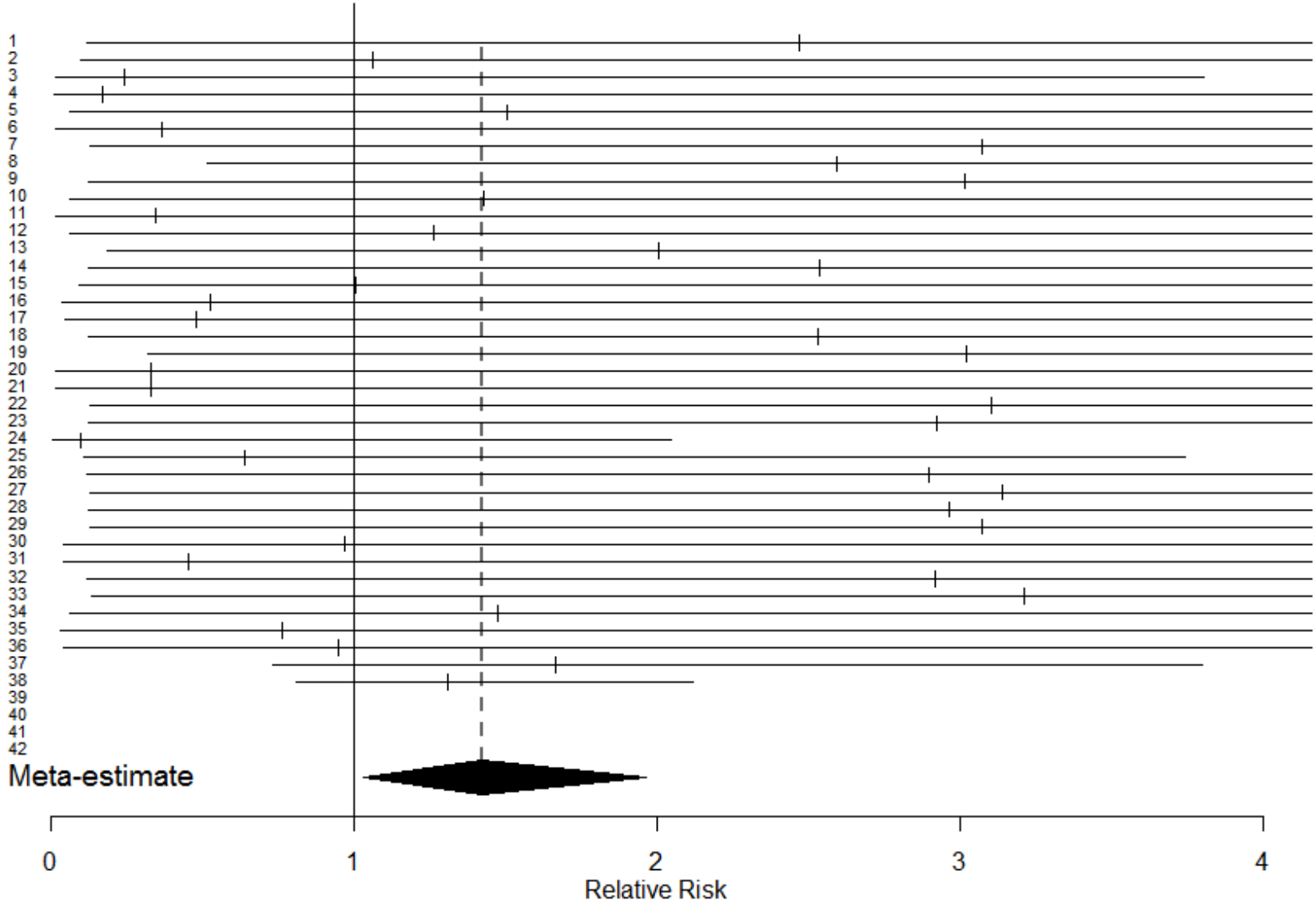
From the Cleveland Clinic, Cleveland. Address reprint requests to Dr. Nissen at the Department of Cardiovascular Medicine, Cleveland Clinic, 9500 Euclid Ave., Cleveland, OH 44195, or at nissens@ccf.org.

This article (10.1056/NEJMoa072761) was published at www.nejm.org on May 21, 2007.

N Engl J Med 2007;356:2457-71.

Copyright © 2007 Massachusetts Medical Society.

Frequentist inference



THE WALL STREET JOURNAL.

DOWJONES

TUESDAY, MAY 22, 2007 - VOL. CCXLIX NO. 119

★★★★ $1.00

DJIA 13542.88 ▼13.65 -0.1% NASDAQ 2578.79 ▲0.8% NIKKEI 17556.87 ▲0.9% DJ STOXX 50 3905.70 ▼0.3% 10-YR TREAS ▲4/32, yield 4.790% OIL $66.27 ▲$1.33 GOLD $662.90 ▲$1.90 EURO $1.3470 YEN 121.45

What's News—

Business and Finance

U.S. employers are divided over the immigration bill, undermining its prospect of becoming law. Employers who rely on unskilled workers generally support the deal, but high-tech industries that need skilled workers complain that it doesn't give them the flexibility to recruit workers with the specific skills they need from abroad. A1, A6

■ **Kerkorian's Tracinda launched** an overture for MGM Mirage's Bellagio Hotel and CityCenter project in Las Vegas, a volley that has put the whole company in play. A3

■ **Glaxo shares slid** after the New England Journal of Medicine released an analysis suggesting users of diabetes drug Avandia have a higher risk of heart attacks. A1, D2

■ **EMI agreed** to be bought by private-equity firm Terra Firma for $4.74 billion, but the music company's shares rose 9.3% in a sign bidding may not be over. A3

■ **Lowe's posted** a 12% profit drop and cut its full-year outlook but said it will keep up an aggressive store-opening campaign. A3

World-Wide

■ **Lebanon pounded a Palestinian camp in a second day of fighting.** Artillery and tank fire engulfed a refugee camp outside Tripoli in violence that has killed at least 50 combatants and an unknown number of civilians. The Lebanese military surrounding the Nahr el-Balad camp sought to crush a militant al Qaeda-inspired group holed up inside. Lebanon's worst internal violence since the 1975-1990 civil war raises fears tension could spread. In Beirut, a bomb rocked a shopping area in a mainly Sunni Muslim district wounding at least four in the second explosion in two days. A5

The U.S. is bracing for a possible showdown with Russia and China over the establishment of a U.N. court to try suspects in the killing of Lebanon's ex-prime minister Hariri.

■ **Iraq's military** is drawing up plans to cope with any quick U.S. military pullout, the defense minister said, as an American official warned the Bush administration may reconsider its support if Iraqi leaders don't make major reforms by fall. Meanwhile, several mortar shells hit the Green Zone but caused no casualties.

■ **U.S. troops raided** safe houses south of Baghdad but failed to find three soldiers missing since May 12.

■ **A Florida doctor** was convicted of

Iwo Jima Letters Of Young Japanese Are Home at Last

An American's Souvenir, They Had Sat on a Shelf; Solving a Family Mystery

By Sebastian Moffett

KOBE, Japan—After the fighting died down in the Battle of Iwo Jima, Victor Voegelin, then 19 years old, was searching for the wounded when he saw a piece of thread poking out of the ground in a blown-out gun emplacement. The U.S. Navy petty officer pulled the thread, and found it was attached to a pack of letters, along with part of a ceramic sake cup and some cigarettes. He picked it all up and put everything in his bag.

Over the decades, Mr. Voegelin looked at the letters just three or four times. He couldn't read the Japanese script, and he always wanted to send them back to Japan. As he got older, "I started thinking about these letters," says Mr. Voegelin, "and thought that people around my age might be around who would want them."

Finally spurred by the release last year of the movie "Letters From Iwo Jima," he took action. He found that the letters had belonged to Tadashi Matsukawa, a Japanese sailor who was 23 when he died. Earlier this year, he sent them to Tadashi's brother, Masaji, [illegible] the same age as Mr. Voegelin.

MEDICAL DETECTIVE

Sequel for Vioxx Critic: Attack on Diabetes Pill

Glaxo Shares Plunge As Dr. Nissen Sees Risk To Heart From Avandia

By Anna Wilde Mathews

An analysis linking the widely used diabetes drug Avandia to higher risk of heart attacks represents a serious blow to GlaxoSmithKline PLC and underscores how outside critics have been empowered to challenge big-selling drugs after the outcry over the withdrawn painkiller Vioxx.

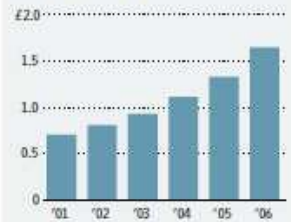
Steven Nissen

Glaxo rang up more than $3 billion in world-wide sales of Avandia last year. Its share price fell more than 7% after the New England Journal of Medicine released the analysis by prominent cardiologist Steven Nissen of the Cleveland Clinic, who helped raise early safety concerns about Vioxx. The analysis suggested that people on Avandia have a 43% higher chance of suffering a heart attack.

Glaxo said it "strongly disagrees" with his conclusions, which come from

Drug in Demand

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

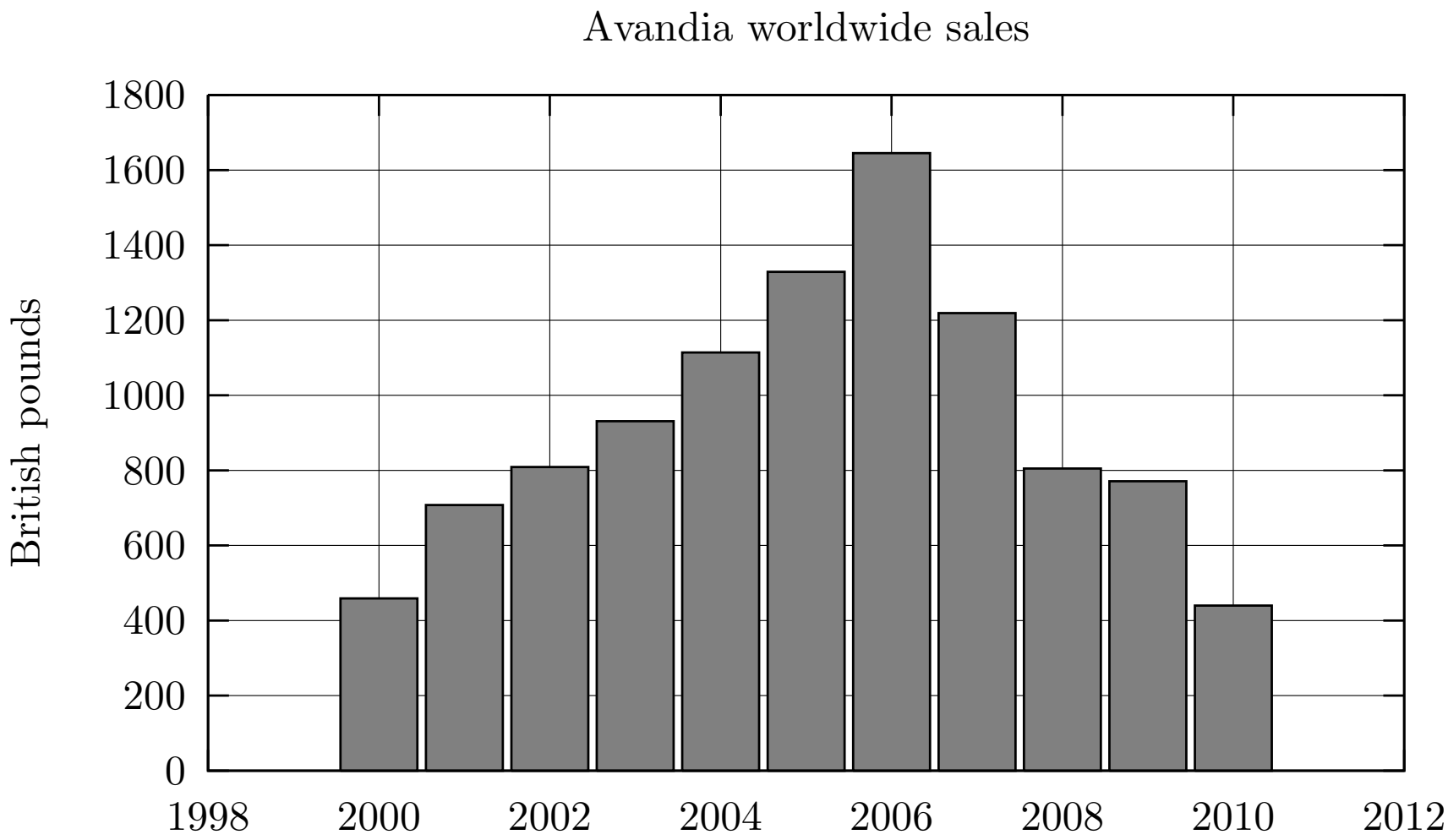


Note: £1 = $1.97 at the current rate; includes sales of Avandamet and Avandaryl
Source: the company

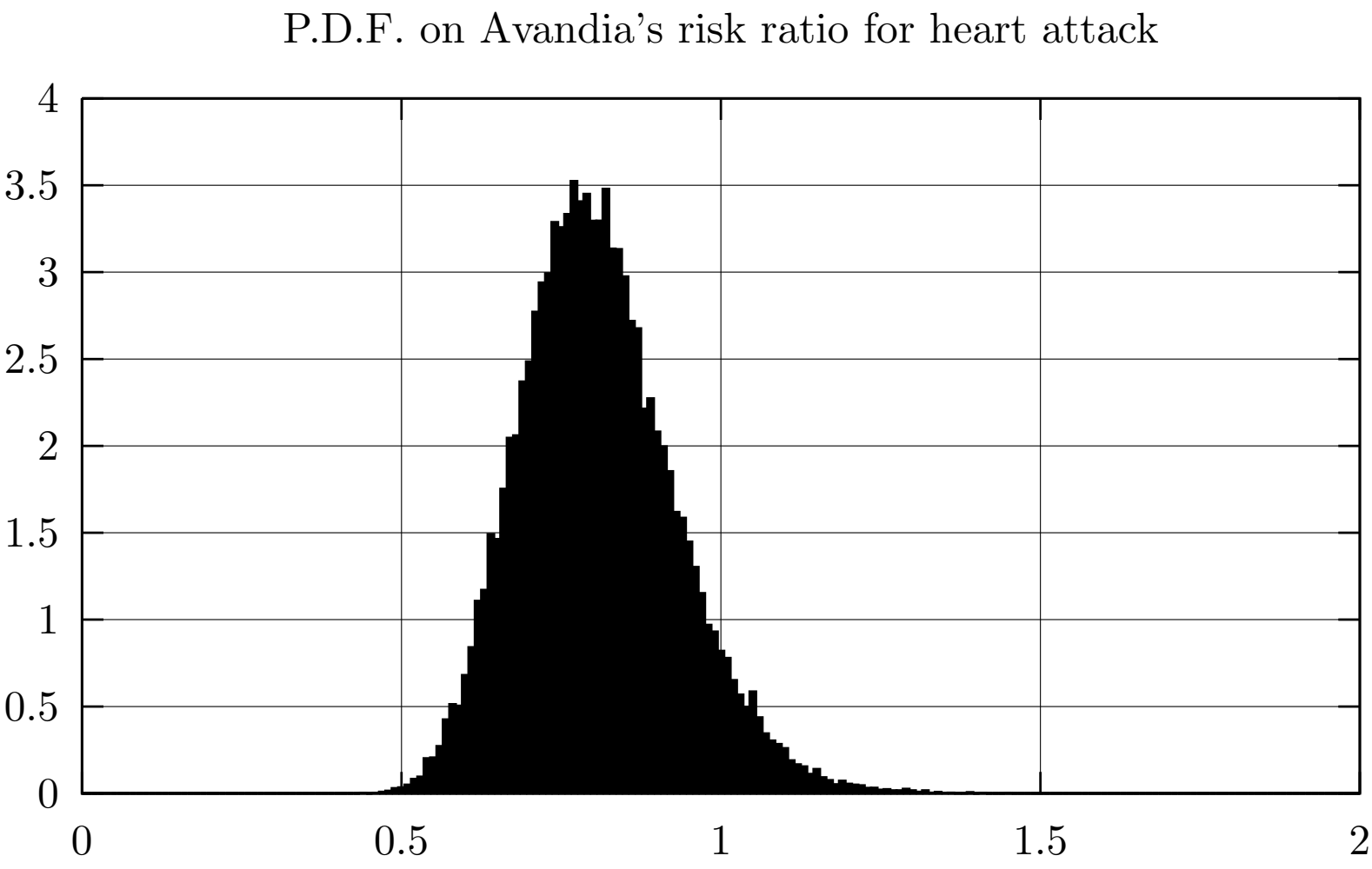
and Drug Administration should have acted faster to alert the public about possible risk from Avandia. Glaxo performed its own meta-analysis, which also showed a potential danger. It shared an early version of it with the FDA in September 2005 and a more complete one in August 2006. The findings weren't reflected on the U.S. label, which is supposed to give a comprehensive review of the drug's risks.

Robert Meyer, head of the FDA office that oversees diabetes drugs, said the agency is still working on its analysis. "We have other data that suggests we

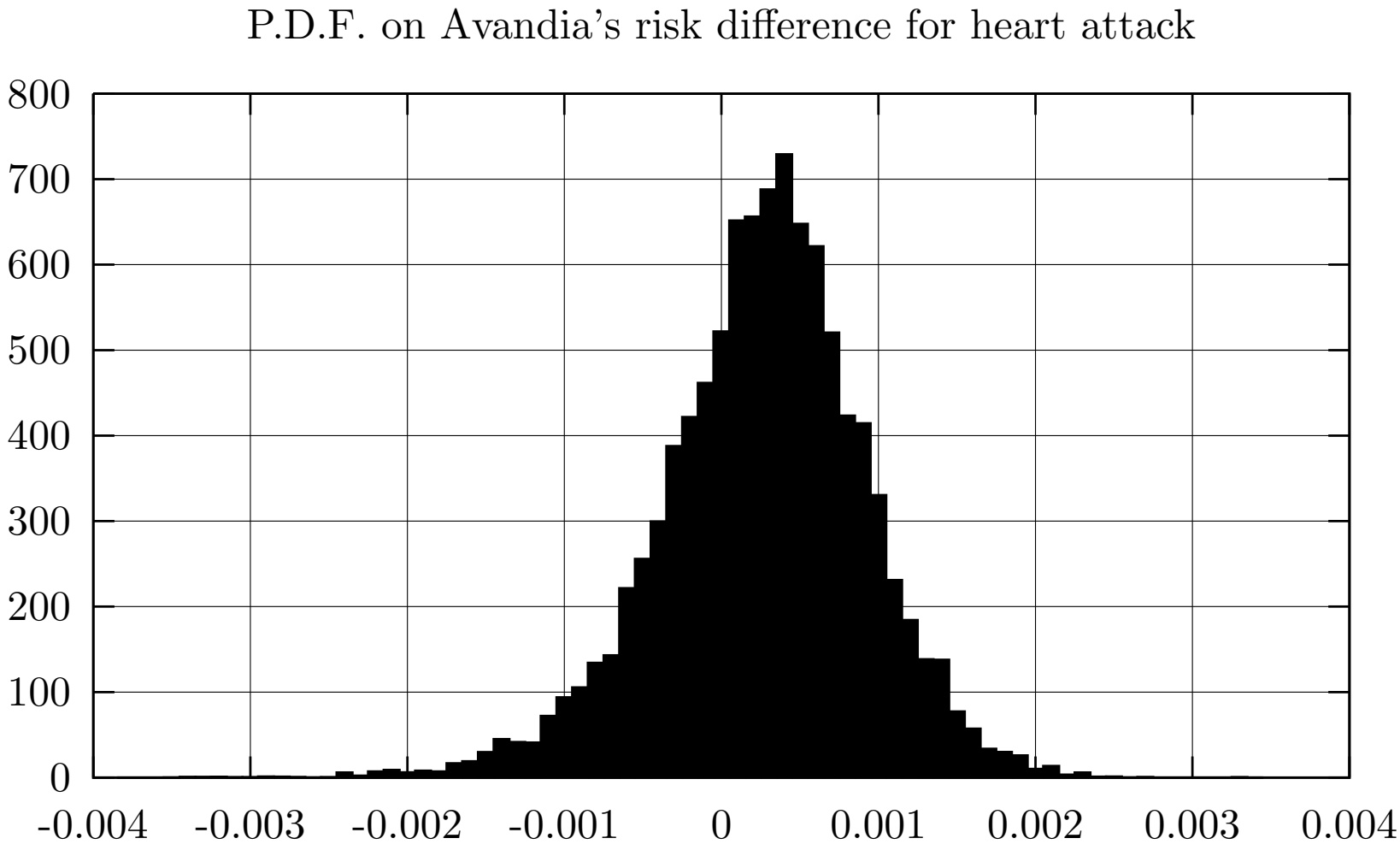
GlaxoSmithKline loses $12 billion



Bayesian inference disagrees, for risk **ratio**.



Or does it? Results depend on model. Here, risk **difference**.



The TAXUS ATLAS Experiment

- FDA asked manufacturer 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.

Bayesian Results

- Assume I know nothing about π_t and π_c *a priori*. Chosen randomly on [0,1], independently and with uniform probability.
- Then we sample: in TREATMENT, 68 heads in 855 samples In CONTROL, 67 heads in 956 samples.
- For a particular p, Pr(k heads in N flips)

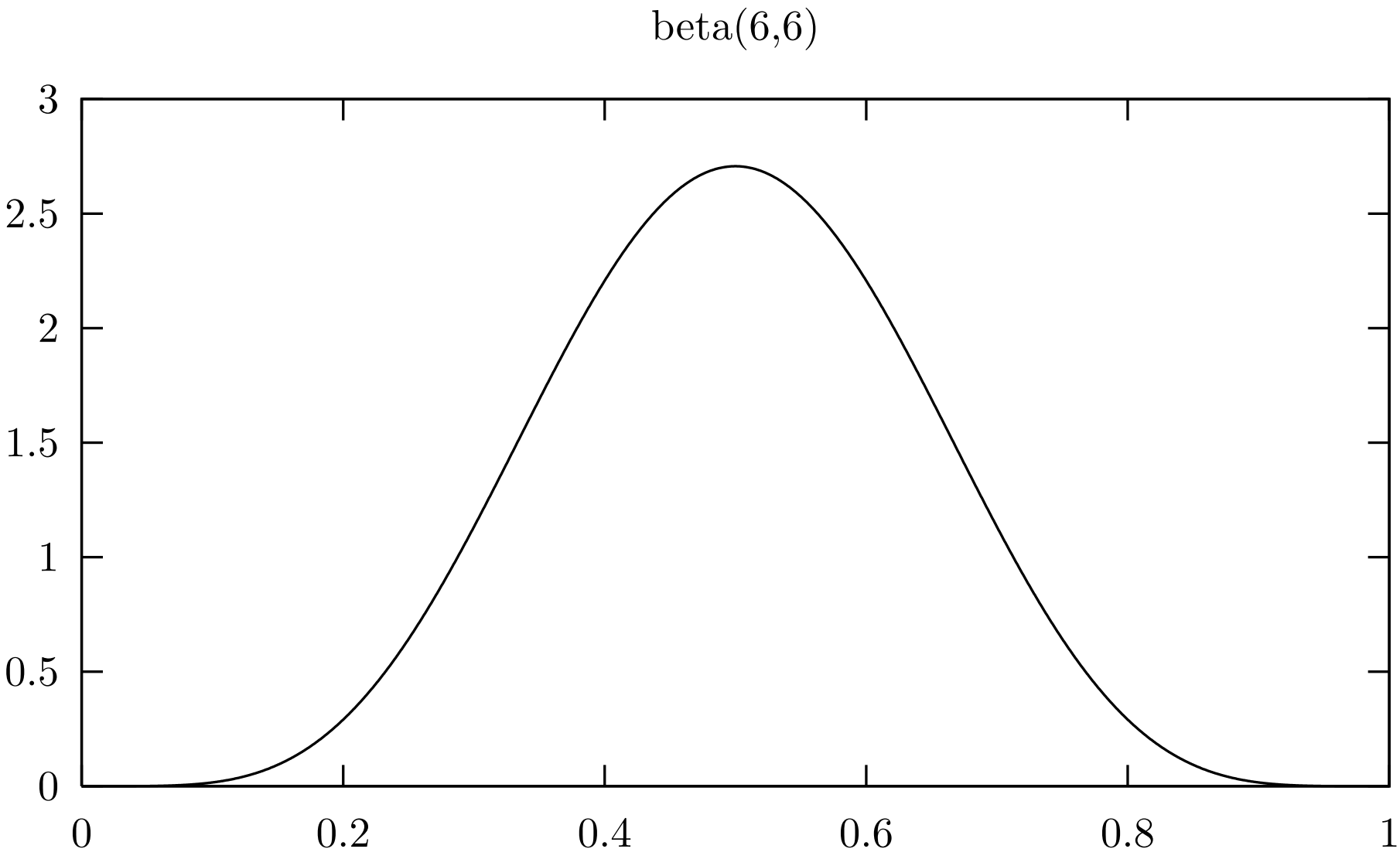
$$= \binom{N}{k} p^k (1-p)^{N-k}$$

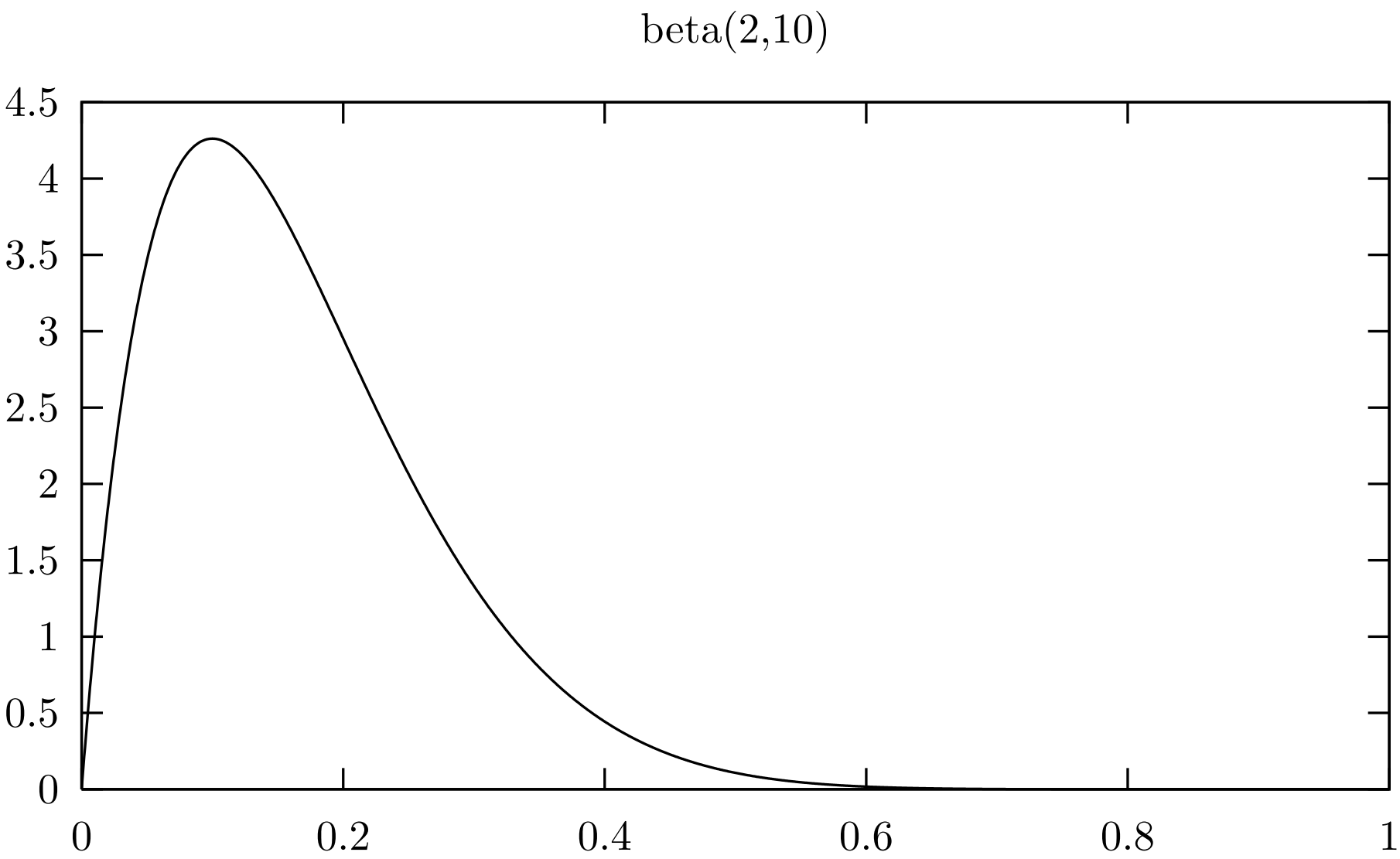
- Apply Bayes' theorem.

Bayesian Results

- Likelihood: $L_{Nk}(\pi) = \binom{N}{k}\pi^k(1-\pi)^{N-k}$
- Probability: Apply Bayes' theorem. With a uniform prior, just normalize. Result is called a Beta distribution.
- $$f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$$

  where $\alpha =$ heads observed plus one, and $\beta =$ tails observed plus one.


beta(6,6)
3
2.5
2
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1


beta(2,10)
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0
0.2
0.4
0.6
0.8
1

Bayesian Results

- $\pi_c \sim \beta(x; 68, 890)$
- $\pi_t \sim \beta(x; 69, 788)$
- Calculate probability $\pi_t - \pi_c < 0.03$:

$$\int_0^1 \int_{\min(x+0.03,1)}^1 \beta(x; 68, 890)\beta(y; 69, 788)\, dy\, dx \approx 0.050737979\ldots$$

- **Result:** Just over 5%.

ATLAS Trial Solution

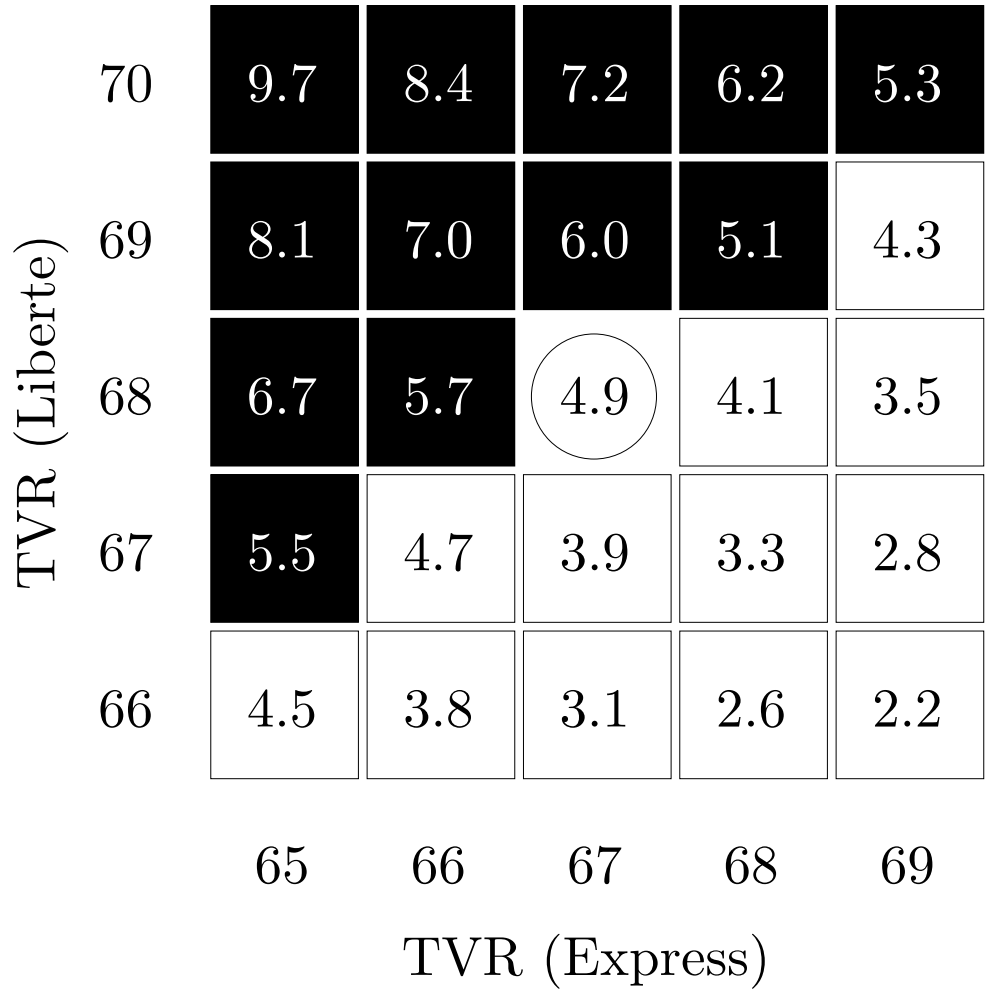
- Use a one-sided 95% confidence interval for $\pi_t - \pi_c$. If its upper limit is less than 0.03, accept. Otherwise reject.
- Confidence interval: approximate *each binomial separately* with a normal distribution. Known as Wald interval.
- Calculate the distribution of the difference, and see if less than 5% of the area exceeds 0.03.
- $$p = \int_{0.03}^{\infty} \mathcal{N}\left(\frac{i}{m} - \frac{j}{n}, \frac{i(m-i)}{m^3} + \frac{j(n-j)}{n^3}\right)$$

Published Results

- We measure 68/855 events in TREATMENT (7.95%), and 67/956 events in CONTROL (7.01%).
- Procedure: if $p < 5\%$, we reject inferiority.
- $p = \int_{0.03}^{\infty} \mathcal{N}\left(\frac{i}{m} - \frac{j}{n}, \frac{i(m-i)}{m^3} + \frac{j(n-j)}{n^3}\right) = 0.0487395\ldots$.
- Accept.

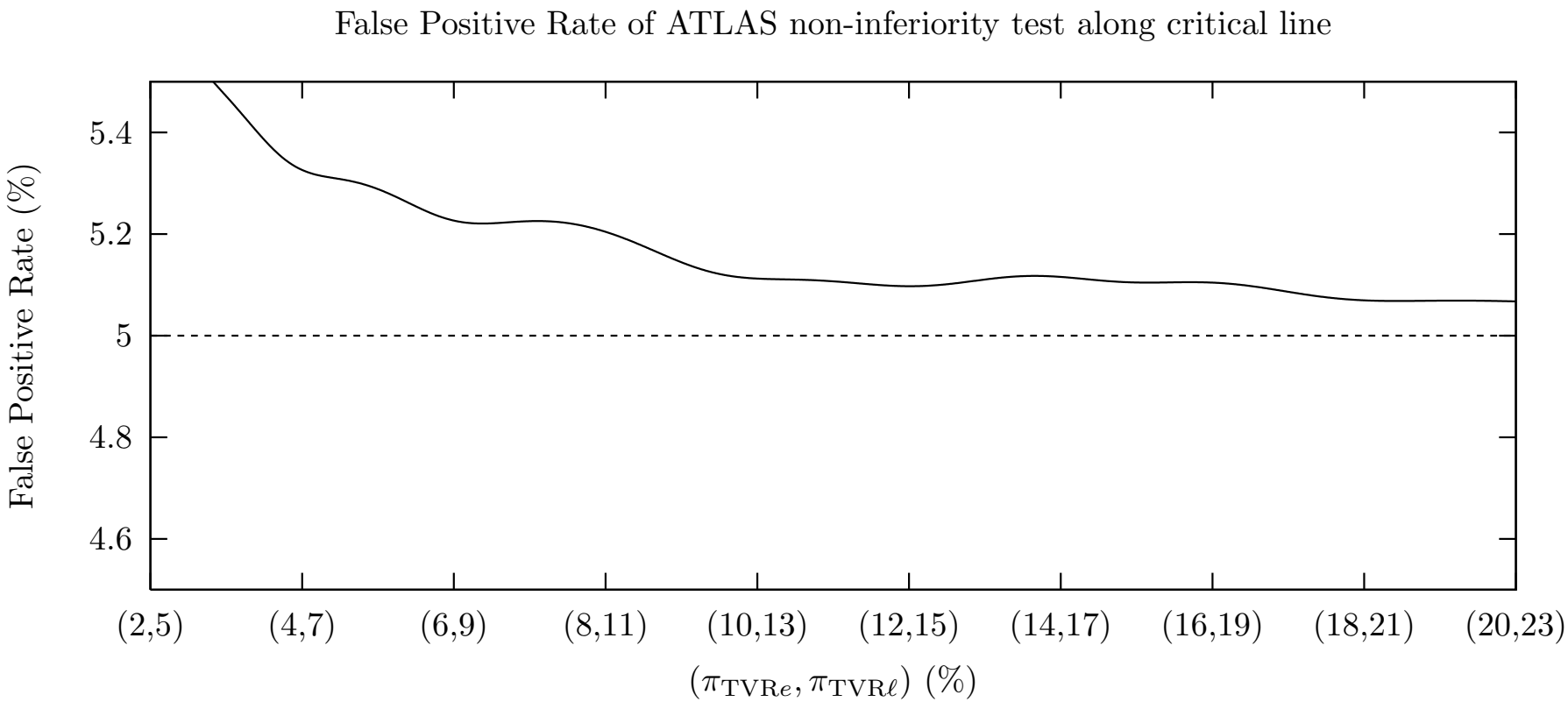
The Ultimate Close Call

Wald's area ($\approx p$) with $(m, n) = (855, 956)$



The Wald Interval Undercovers

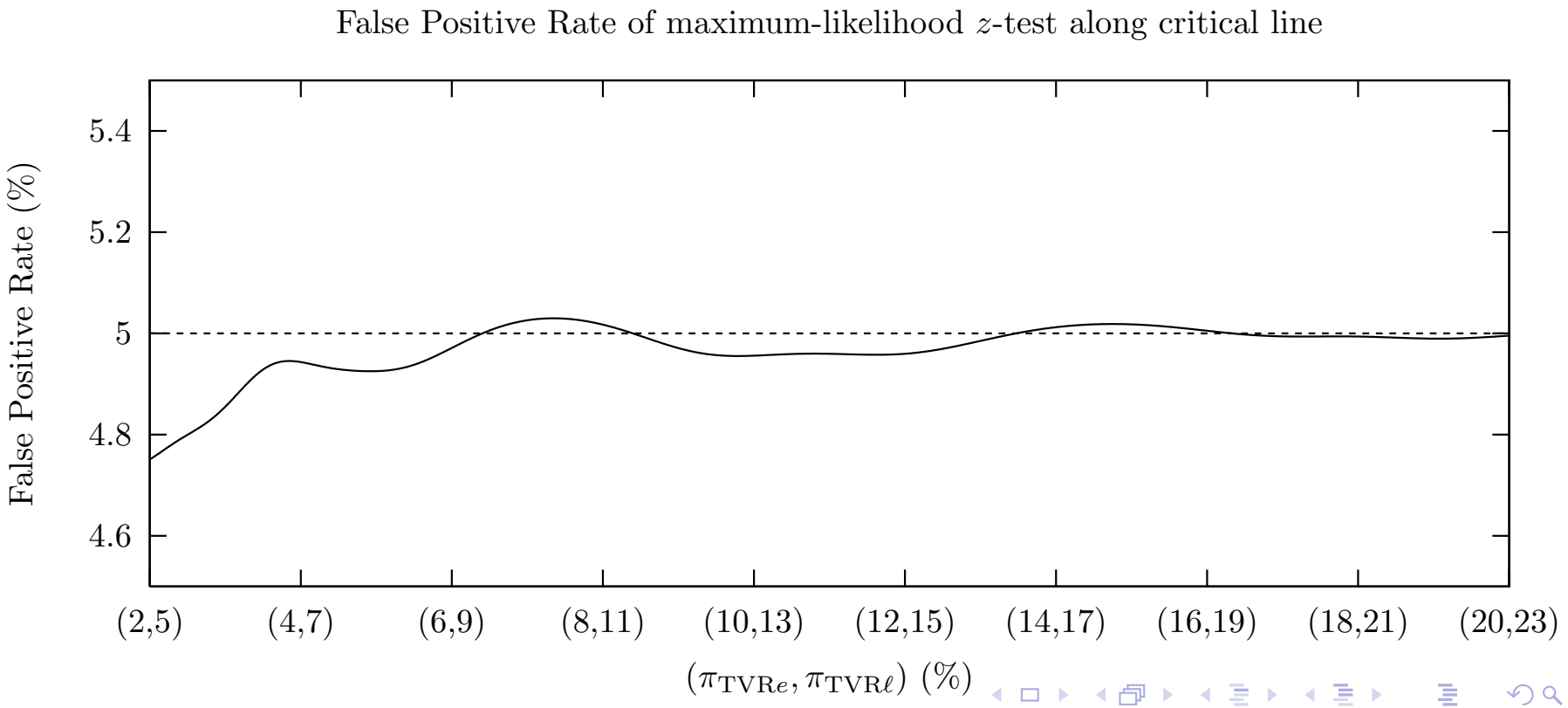
Our confidence interval doesn't have 95% coverage, so the test didn't bound the rate of false positives by 0.05. The approximation is lousy here.



One solution: constrained variance

The Wald interval approximated each binomial *separately* as a Gaussian, with variance of $\frac{i(N-i)}{N^3}$. (E.g., 7% and 8%.) But this is not consistent with H_0, which says $\pi_t > \pi_c + 0.03$.

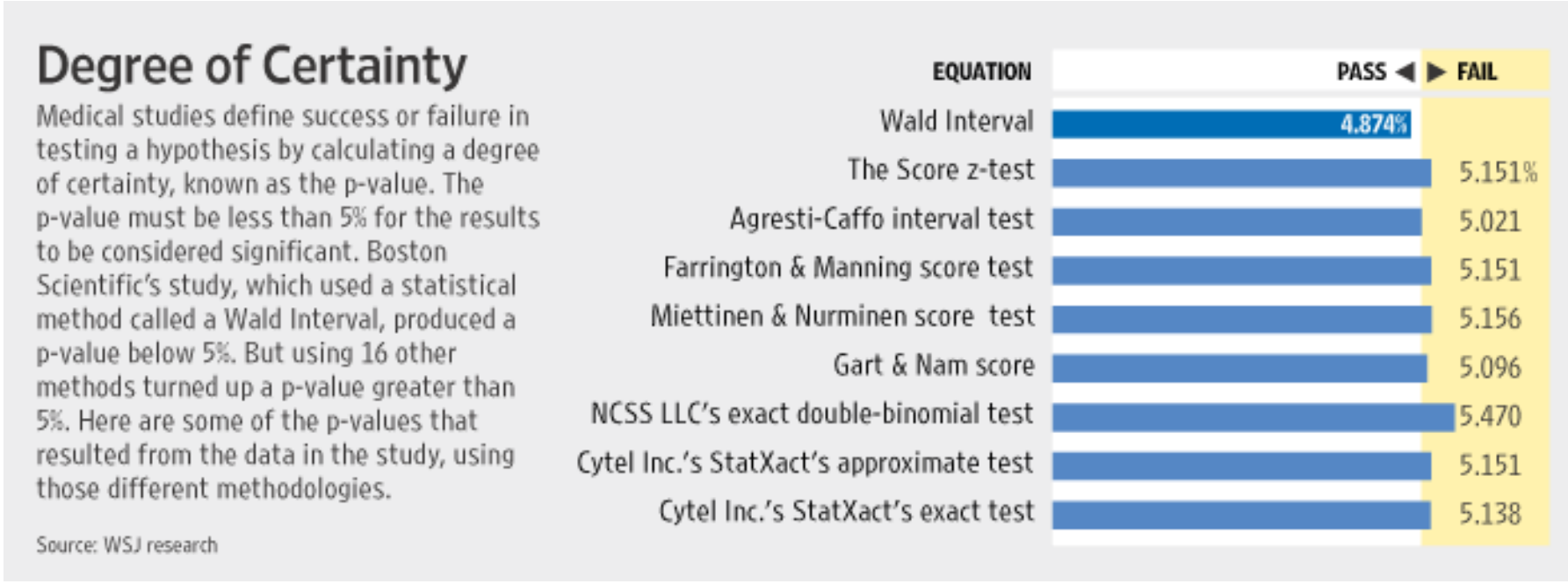
One improvement is to approximate the variances by finding the most likely pair consistent with H_0 (i.e., separated by 3 percentage points). E.g., 6% and 9%.



Every other published interval fails to exclude inferiority.

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 I_4 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 CI from inverted 2-sided test	$c = 0.03032$	Fail

Nerdiest chart contender?



Boston Scientific Stent Study Flawed

By Keith J. Winstein

A heart stent manufactured by Boston Scientific Corp. and expecting approval for U.S. 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 claim was based on a flawed statistical equation that favored the Liberte stent, a Journal analysis has found. Using a number of other methods of calculation—including 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 Abbott Laboratories have used the same equation in stent studies.

Boston Scientific

Boston Scientific is seeking FDA approval for its **Taxus Liberte stent**.

But in those cases, any boost provided by the Wald equation wouldn't have changed the outcome of the study. In the Liberte study, the equation's shortcomings 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-device 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 to reach statistical significance. "We used standard methodology 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

World's most advanced non-inferiority test

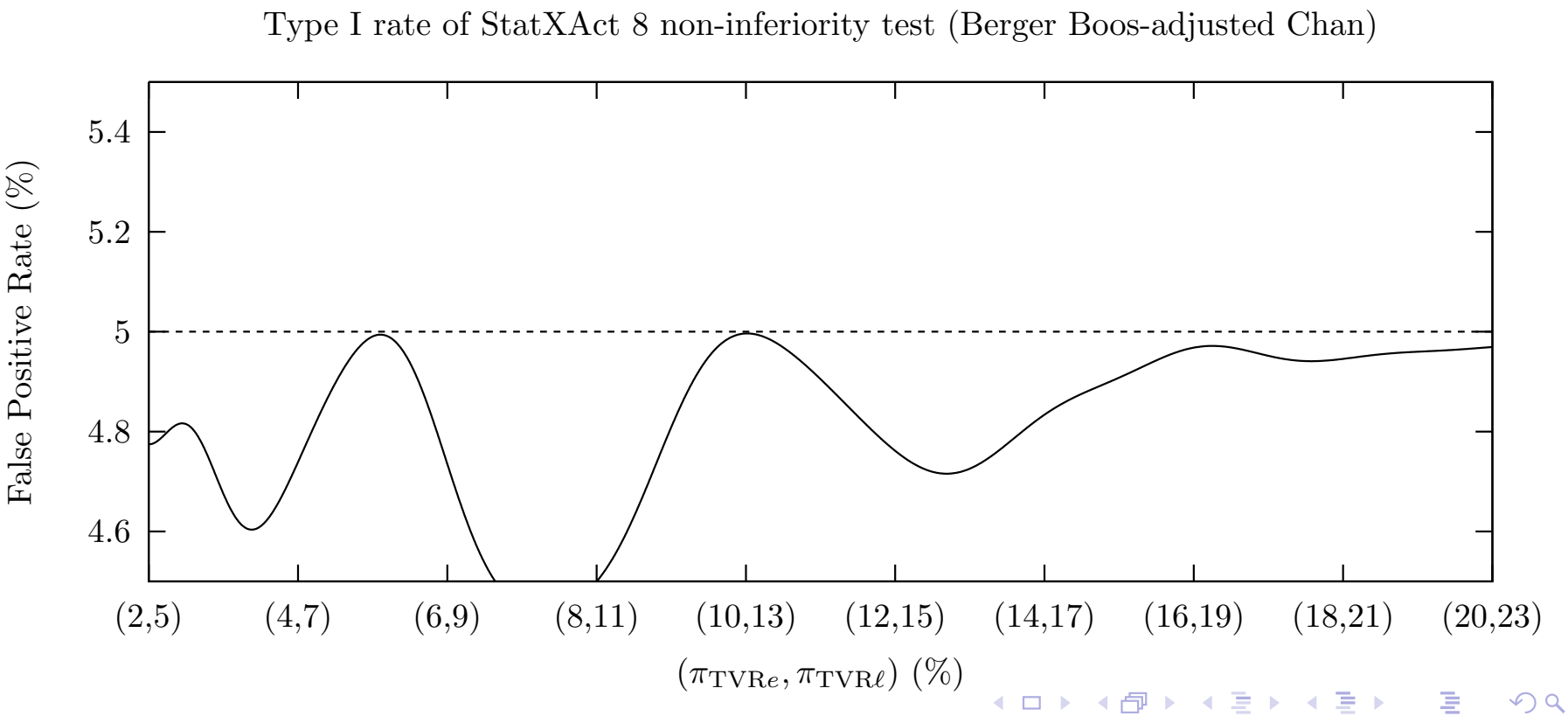
The StatXAct 8 software package sells for $1,000 and takes 15 minutes to calculate a single p-value. Made by MIT's Zoroastrian chaplain, Cyrus Mehta.

"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..."

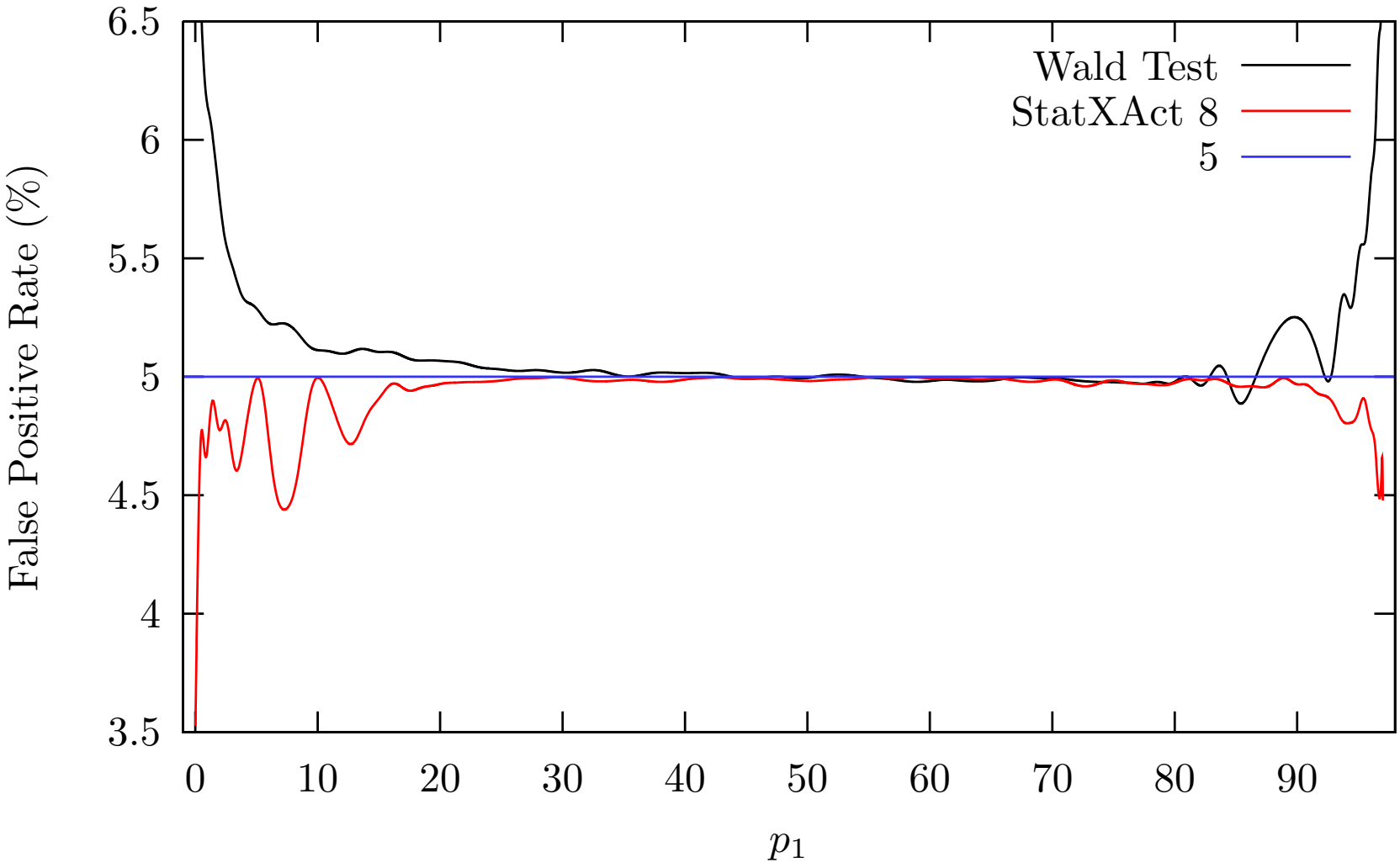
World's most advanced non-inferiority test

The StatXAct 8 software package sells for $1,000 and takes 15 minutes to calculate a single p-value. Made by MIT's Zoroastrian chaplain, Cyrus Mehta.

"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. . ."



Both tests, together



Final Thoughts

- What's important: say what you're trying to infer, how you get there, and what your criteria are.
- Don't be surprised if frequentist and Bayesian approaches differ in their results.
- Sometimes they will agree numerically but not on what the numbers mean!
- If they disagree starkly, you have bigger problems than your interpretation of probability.
- Same goes if the Bayesian answer depends heavily on the prior. If two reasonable priors give starkly disagreeing results, you don't have a good answer.