The Supreme Court Matrixx Decision: Is Significance Significant?

Danny Kaplan

Macalester College, Saint Paul, Minnesota

Joint Statistics Meeting, Aug 3, 2011

Examples where the Law is Statistically Ill-Informed

- Spurious forensics. No calibration of common tests, e.g. eyewitness testimony,
- Prosecutor's Fallacy. Confusion between false-positive rate and "probability of innocence". A 10% false positive rate does not imply a 90% probability of guilt.
- No double jeopardy, but multiple tests.

The Washington Post

DNA sets free D.C. man imprisoned in 1981 student slaying

By Keith L. Alexander
Washington Post Staff Writer
Wednesday, December 16, 2009; B01

A District man who was incarcerated for 28 years in the rape and murder of a Georgetown University student in Rock Creek Park was ordered released Tuesday by a D.C. Superior Court judge after DNA evidence revealed that another man committed the crime.

Donald Eugene Gates, now 58, had maintained his innocence from the start. He was to board a bus from a prison in Arizona on Tuesday afternoon and head to a new home -- and a new life -- in his home town of Akron, Ohio.

Although the judge's ruling frees Gates, it does not exonerate him. There will be a separate hearing to make that determination after more DNA testing is completed.

"This is very exciting and beautiful," Gates said as he tried to figure out how to operate the cellphone belonging to his Arizona-based attorney. Gates said he was trying to "process everything" now that he had been released from a life sentence.

At Tuesday's hearing, senior Judge Fred B. Ugast angrily criticized government officials who relied heavily on the testimony of an FBI analyst during Gates's trial. The analyst incorrectly linked two hairs from an African American male to Gates. The hairs were found on the body of Catherine Schilling, 21, a white college student who had had been shot five times in the head in 1981. Semen was found on her body.

A 1997 review by the Justice Department discredited the work of that FBI analyst, Michael P. Malone, and 13 other analysts, finding that they had made false reports and performed inaccurate tests.

WORKSHEET 17-12a: STATISTICAL ANALYSIS

(Based on Selection Rates)

This Worksheet may be used to find the statistical significance of a difference in selection rates.

At the top of the sheet, enter the SCRR page 17 problem number and, beside "Issue," enter the names of the particular groups whose selection rates you are

comparing, the type of selection decision and for what job(s); i.e., Female vs. Male Hires for Management Trainee. Then complete this page as follows:

(a) Step #1: Find the difference in selection rates by subtracting the lower rate from the higher.

(b) Step #2: For the value of one Standard Deviation (S.D.):

Multiply the result of (a) by the result of (b) and find the square root. This is one Standard Deviation.

(c) Step #3: For the number of Standard Deviations of the difference in selection rates shown in Step #1, divide the result of Step #1 by the result of Step #2.

Generally speaking, a result of 2 or more Standard Deviations is considered statistically significant.

Two Principles that Guide the Decision

"Reasonable Investor" and "Bright-Line Rule"

Bright-Line Rule

A rule that leaves little or no room for varying interpretation: producing predictable and consistent results. Compare: *fine line*.

Examples of Bright-Line Rules

- Statutory rape: depends on age of victim and accused.
- Police can detain occupants of a residence while conducting a search for contraband.

Bright-Line Rules in Statistics

- $p < 0.05$

Most educators would resist the bright-line application of $p < 0.05$ as a decision rule.

- 0.05 is conventional, but arbitrary.
- Adjustments for multiple comparisons, *post hoc* tests
- Confounders.
- One-sided v. two-sided. (FDA tends to oppose one-tailed tests.) Dubey, S.D. (1991) "Some thoughts on the one-sided and two-sided tests" *J. Biopharm. Stat.* **1**:139-150 S-C Chow and J-P Liu (2004) *Design and Analysis of Clinical Trials: Concepts and Methodologies*, 2e, p.77

See *The Cult of Statistical Significance*, Steve Ziliak

Court's Reasoning in Dismissing Matrixx's Bright-Line Rule

"The Court declined to adopt a bright-line rule for determining materiality in Basic, observing that '[a]ny approach that designates a single fact or occurrence as always determinative of an inherently fact-specific finding such as materiality, must necessarily be overinclusive or underinclusive.' "

"Given that medical professionals and regulators act on the basis of evidence of causation that is not statistically significant, it stands to reason that in certain cases **reasonable investors** would as well."

The Reasonable Investor

15 U. S. C. §78j(b). SEC Rule 10b-5: Employment of Manipulative and Deceptive Practices,

U.S. Supreme Court, TSC Industries v. Northway, Inc. 426 U.S. 438 (1976) "an omitted fact is material if there is a substantial likelihood that a reasonable shareholder would consider it important in deciding how to vote."

"so obviously important to an investor, that **reasonable minds cannot differ on the question of materiality**"

The Keynesian Beauty Contest

"[P]rofessional investment may be likened to those newspaper competitions in which the competitors have to pick out the six prettiest faces from a hundred photographs, the prize being awarded to the competitor whose choice most nearly corresponds to the average preferences of the competitors as a whole; so that each competitor has to pick, not those faces which he himself finds prettiest, but those which he thinks likeliest to catch the fancy of the other competitors, all of whom are looking at the problem from the same point of view. **It is not a case of choosing those which, to the best of one's judgment, are really the prettiest**, nor even those which average opinion genuinely thinks the prettiest. We have reached the third degree where we devote our intelligences to anticipating what average opinion expects the average opinion to be. And there are some, I believe, who practise the fourth, fifth and higher degrees."
— J. M. Keynes (1936) *General Theory of Employment Interest and Money*, Chap. 12

One Goal: Avoid Information Overload

We were "careful not to set too low a standard of materiality," for fear that management would "... bury the shareholders in an avalanche of trivial information.' " 485 U. S., at 231 (quoting TSC Industries, 426 U. S., at 448449)."

Information Overload

ARCHIVES OF
INTERNAL MEDICINE

SEARCH:

Welcome | My Account | E-mail A

HOME | CURRENT ISSUE | PAST ISSUES | TOPIC COLLECTIONS | CME | PHYSICIAN JOBS | SUBMIT | SUBSCRIBE | HELP | INFORMATION FOR:

Vol. 171 No. 10, May 23, 2011 TABLE OF CONTE

Research Letters

HEALTH CARE REFORM

A Quantitative Analysis of Adverse Events and "Overwarning" in Drug Labeling

Jon Duke, MD; Jeff Friedlin, DO; Patrick Ryan, MEng

Product labeling is a primary source of drug safety information for physicians. However, the effectiveness of labeling in communicating advers drug events (ADEs) may be diminished by the problem of "overwarning," in which excessively long and complex lists of potential reactions can result in information overload.[1-2] The Food and Drug Administration (FDA) highlighted this issue in 2006 as they unveiled new labeling guidelin specifically discouraging the inclusion of "exhaustive lists of every reported adverse event, no matter how infrequent or minor."[3-4(p2)] Yet, at present, there are no baseline data on overwarning, nor are there benchmarks against which the success of the FDA's interventions can be measured. The goal of our study was to address this gap by producing comprehensive quantitative data on ADE labeling patterns. We further

Clusters and Anecdotes

- Auto-immune reaction to breast implants.
- Cancer and electric power lines
- Autism and vaccination

Lessons from Cancer Clusters?

In the late eighties, public-health departments were receiving between thirteen hundred and sixteen hundred reports of feared cancer clusters, or "cluster alarms," each year. Last year, in Massachusetts alone, the state health department responded to between three thousand and four thousand cluster alarms.

Raymond Richard Neutra, California's chief environmental health investigator and an expert on cancer clusters, points out that among hundreds of exhaustive, published investigatins of residential clusters in the United States, not one has convincingly identified an underlying environmental cause.

"The reality is that they're an absolute, total, and complete waste of taxpayer dollars," says Alan Bender, an epidemiologist with the Minnesota Deptment of Health, which investigated more than a thousand cancer clusters in the state between 1984 and 1995.

— Atul Gawande (1999) "The Cancer-Cluster Myth" *The New Yorker*, Feb. 8: 34-37.

The "Reasonable Statistician"

- An aspirational definition to use for discussion:

  > The beliefs and understanding typical of a person who has passed a university-level introductory statistics course or a social-science methods course.

I suspect that this is a more stringent definition than would be accepted by a court.

But would even this aspirational standard fix the problem?

What would such a "Reasonable Statistician" Know?

- That "data beat anecdotes"?
- That it's hard to detect effects from clusters?
- That you don't accept the null hypothesis?
- Sensible ways to evaluate risks?
- How to interpret basic medical statistics?

Conditional Probabilities

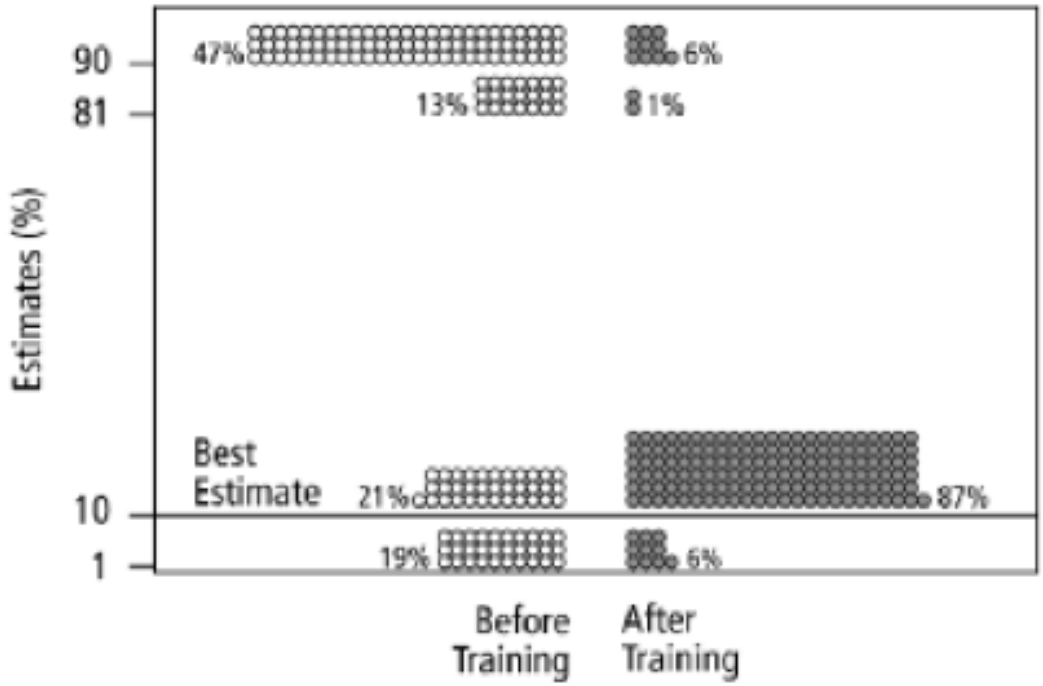


Fig. 2. Estimates by 160 gynecologists of the probability that a woman has breast cancer given a positive mammogram, before and after receiving training in how to translate conditional probabilities into natural frequencies.

[From Gigerenzer *et al.* (2008) "Helping Doctors and Patients Make Sense of Health Statistics"]

Risk and Benefit

Arguing that the information about side-effects would have cut into revenues substantially ...

> *"Consumers likely would have viewed Zicams risk as substantially outweighing its benefit."* — *the Matrixx decision*

Benefit is relieving symptoms of a cold, and risk is that of permanently losing one's sense of smell.
I think that most anybody would think the benefit not as big as the harm. In the same way, driving to the grocery store has a benefit of alleviating hunger, and a risk of dying in a car accident. But it would be wrong to say that dying "outweighs" eating. The risks, as probabilities, need to be taken into account.

"The courts have resolved the vexing problem of the proper valuation of life by ignoring it." — Richard A. Posner in "Economic Analysis of Law", judge on the United States Court of Appeals for the Seventh Circuit in Chicago and a Senior Lecturer at the University of Chicago Law School.

How are We Doing? The Bad News

Little evidence that stat-course graduates are prepared to navigate the issues presented in Matrixx.

- We define "statistics" very narrowly, along the lines of "techniques for research workers."
- They might well accept the Null, and we spend far too much time on the 2nd decimal point of a p-value and not enough on
  - Confounding and Adjusting
  - Causation
  - Limitations of p-values.
- Very little about risk.
- Very little about conditional probabilities.
- Many students emerge with a belief that "You can make the statistics say anything you want."

How are we Doing? The Good News.

- Interest in statistics is growing, e.g., AP enrollments.
- Movement away from mathematical statistics as the introduction.
- GAISE Guidelines
- Textbooks to draw on:
  - Moore, *Statistics: Concepts and Controversies*
  - Woloshin et al. *Know Your Chances: Understanding Health Statistics*
  - The classics: Freedman, Purves, Pisani or *Statistics: A Guide to the Unknown*

Suggestion: A Course Correction is Needed

- Orient introductory statistics much more toward decision-making.
- Teach about the situations and methods that people encounter reading the news
  - "... after adjusting for ..."
  - The anecdotal nature of clusters.
  - Causal inference when there is no experimentation.
- Step forward as the stewards of quantitative reasoning and literacy, not just statistical methods.
- Fill the vacuum in teaching about computation.

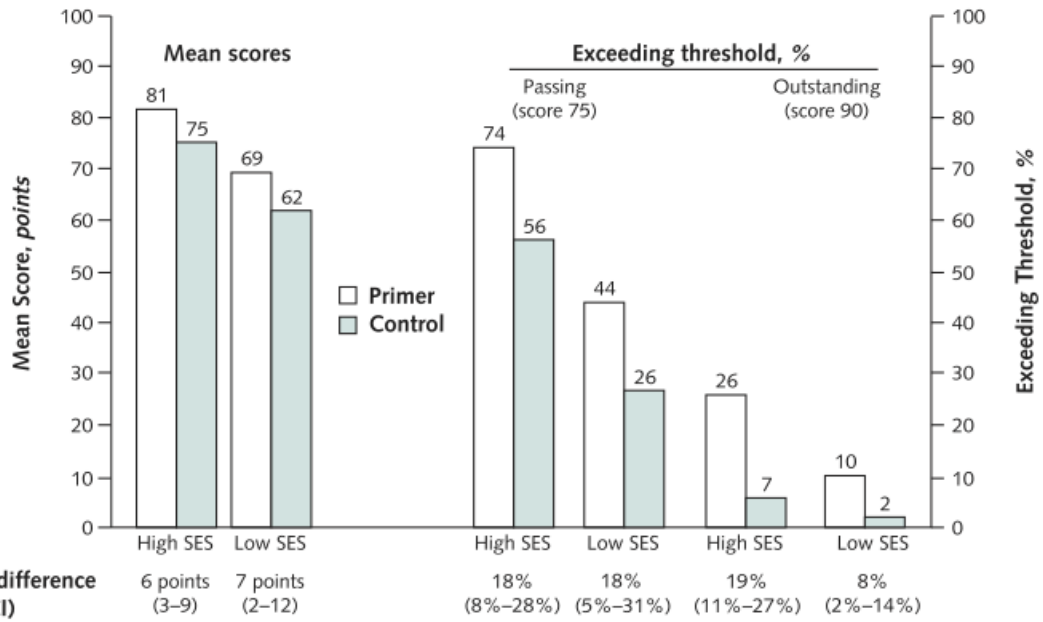
Teaching about Risk

Annals of Internal Medicine | A

The Effectiveness of a Primer to Help People Understand Risk

Two Randomized Trials in Distinct Populations

Steven Woloshin, MD, MS; Lisa M. Schwartz, MD, MS; and H. Gilbert Welch, MD, MPH



Conclusion

If we are to expect statistically literate results from a system that bases judgment on the "reasonable man," we need to ensure that people are statistically literate.

This is our job as statistical educators.

We need to define our educational goals, in substantial part, in terms of the needs of civic life:

- Conditional probability
- Evaluation of risk
- The nature of evidence and reasons to be skeptical
- Adjusting for confounders

We need to define success in terms of practical decision making, not arithmetic calculations.