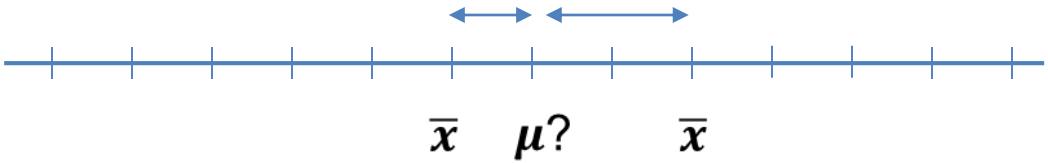
Probability and Mathematical Statistics in Data Science

Lecture 37: Integrity in Research

Data Integrity : The Quality of Our Data

$\bar{x}$: sample mean (known estimate of truth)

μ: population mean (unknown truth)



- Statistics (calculated from our data) are estimators of some unknown truth in the population.
- We try minimize the distance of our estimator of truth, the sample statistic, from the truth itself by collecting quality data.
- We learnt to reason with the space between what we know (the sample statistic) and what we want to know (the truth).

Data Integrity : The Quality of Our Data

- We have learned that the pursuit of truth involves making every effort to minimize error and bias in our sample data.

- We can only be confident our statistical analysis is accurate and valid if the data collected is of the utmost quality adhering to the necessary assumptions and conditions.

- No matter how knowledgeable we are in the methods and techniques of statistics analysis - if our data is of poor quality, then our statistical analysis will be of poor quality.

Meaning of Integrity

integrity [in-**teg**-ri-tee] SHOW IPA

SYNONYMS | EXAMPLES | WORD ORIGIN

SEE MORE SYNONYMS FOR *integrity* ON THESAURUS.COM

noun

1. adherence to moral and ethical principles; soundness of moral character; honesty.
2. the state of being whole, entire, or undiminished:
   to preserve the integrity of the empire.
3. a sound, unimpaired, or perfect condition:
   the integrity of a ship's hull.

"Integrity is doing the right thing even when no one is watching" – C.S. Lewis

Dishonesty in scientific research

Implications for biomedical research

In research and in the practice of medicine, there are lots of opportunities for conflicts of interest and motivated reasoning. These opportunities are not due to bad people, they are just a by-product of how the academic research system and the rewards for certain kinds of research outcomes have been created (just as there were opportunities due to the structure of the financial system and bonuses for certain kinds of investment gains). No system design is perfect, and selfish motives are an evolutionary factuality. This is why society has developed a system of moral values and principles of conduct that are taught to children from the moment they are born.

For example, the academic research system rewards statistically significant research findings with prestigious publications, grants, and promotions. Statistically nonsignificant research findings, on the other hand, are almost entirely disregarded, despite the fact that we sometimes learn more from them. Consequently, the system sets up a conflict of interest when, after thousands of dollars of research funding and hundreds of hours of work, one faces null effects (3). Tampering with data and misreporting of experimental procedures and results seems like a severe and rare reaction, but even the tendency to underreport negative and overreport positive data, which may appear less severe and therefore more acceptable, is a troublesome practice with potentially harmful consequences for the biomedical research community.

The New York Times

Many Psychology Findings Not as Strong as Claimed, Study Says

Many Psychology Findings Not as Strong as Claimed, Study Says

The past several years have been bruising ones for the credibility of the social sciences. A star social psychologist was caught fabricating data, leading to more than 50 retracted papers. A top journal published a study supporting the existence of ESP that was widely criticized. The journal Science pulled a political science paper on the effect of gay canvassers on voters' behavior because of concerns about faked data.

Now, a painstaking yearslong effort to reproduce 100 studies published in three leading psychology journals has found that more than half of the findings did not hold up when retested. The analysis was done by research psychologists, many of whom volunteered their time to double-check what they considered important work. Their conclusions, reported Thursday in the journal Science, have confirmed the worst fears of scientists who have long worried that the field needed a strong correction.

Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect

The term *psi* denotes anomalous processes of information or energy transfer that are currently unexplained in terms of known physical or biological mechanisms. Two variants of psi are *precognition* (conscious cognitive awareness) and *premonition* (affective apprehension) of a future event that could not otherwise be anticipated through any known inferential process. Precognition and premonition are themselves special cases of a more general phenomenon: the anomalous retroactive influence of some future event on an individual's current responses, whether those responses are conscious or nonconscious, cognitive or affective. This article reports 9 experiments, involving more than 1,000 participants, that test for retroactive influence by "time-reversing" well-established psychological effects so that the individual's responses are obtained before the putatively causal stimulus events occur. Data are presented for 4 time-reversed effects: precognitive approach to erotic stimuli and precognitive avoidance of negative stimuli; retroactive priming; retroactive habituation; and retroactive facilitation of recall. The mean effect size (*d*) in psi performance across all 9 experiments was 0.22, and all but one of the experiments yielded statistically significant results. The individual-difference variable of stimulus seeking, a component of extraversion, was significantly correlated with psi performance in 5 of the experiments, with participants who scored above the midpoint on a scale of stimulus seeking achieving a mean effect size of 0.43. Skepticism about psi, issues of replication, and theories of psi are also discussed.

Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect

- This study was looking for evidence in favor of **precognition** - foreknowledge of a paranormal kind of event – the ability of humans to predict something that has not happened yet

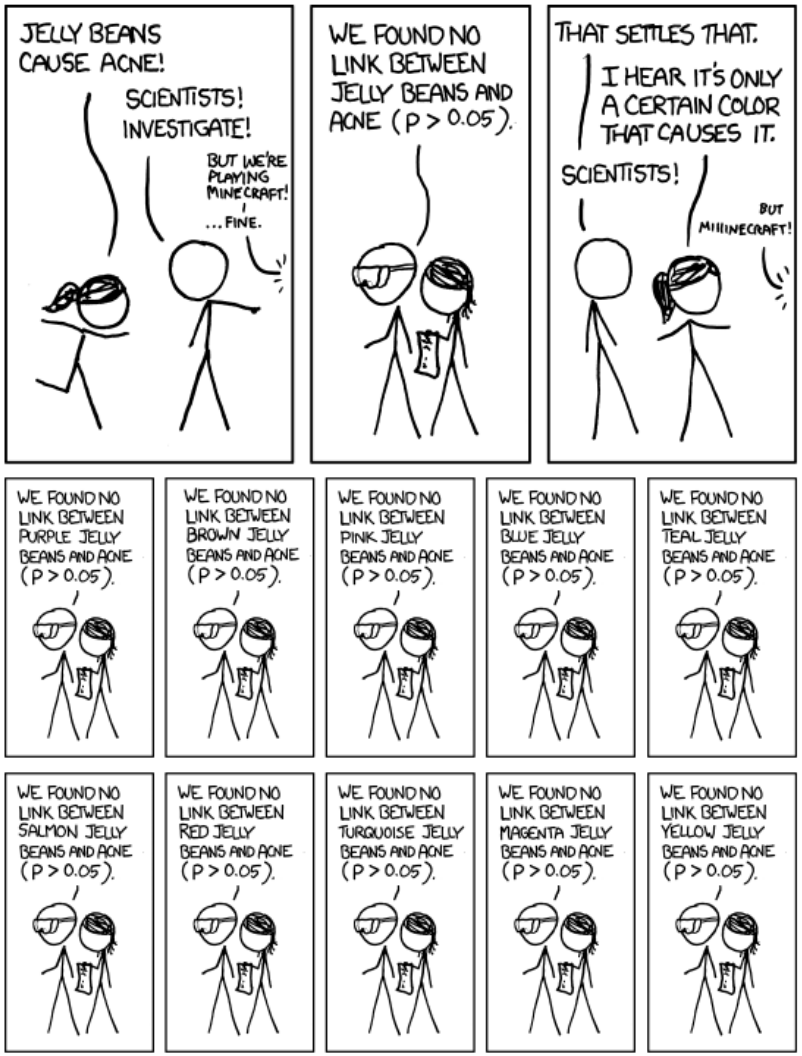
- This study conducted nine experiments on over a 1000 subjects

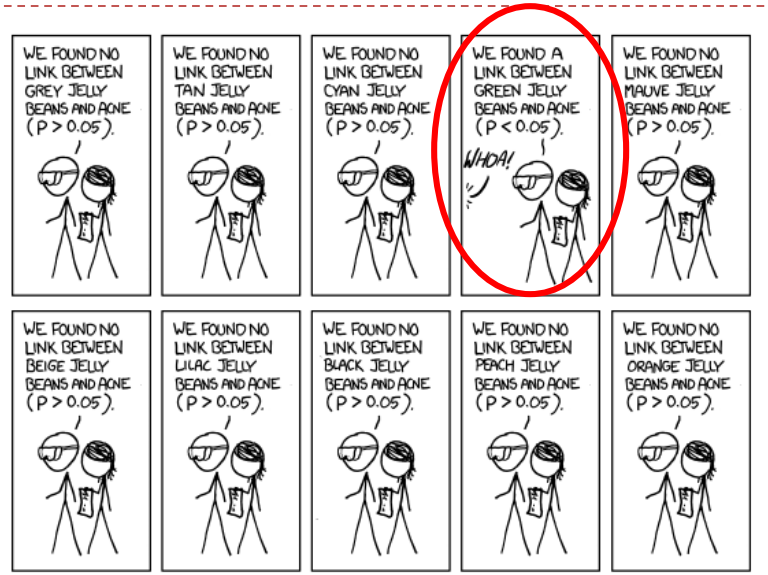
- The researcher found that 8 out of 9 experiments conducted rejected the null hypothesis of no precognition in favor of the alternative that precognition exists.

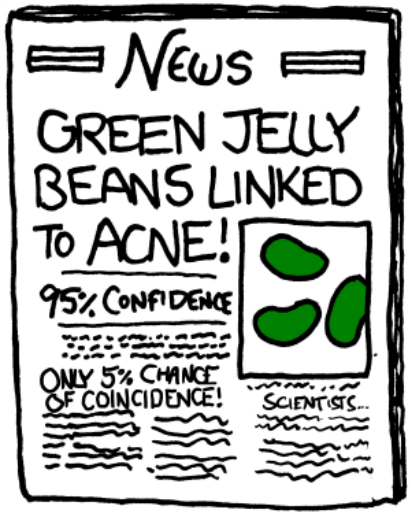
The Reasoning of Hypothesis Testing

- The first step in any hypothesis test is to clearly define your null and alternative hypotheses as follows.
  - **Null Hypothesis:** No Precognition Exists
  - **Alternative Hypothesis**: Precognition Exists
- The researcher should then run an experiment that tests for precognition collect the data, and analyze the results
- This is **not what the researcher did**. Instead, he went on a **exploratory expedition** until he found a statistically significant result.

The Reasoning of Hypothesis Testing







http:xkcd.com/882/

The Reasoning of Hypothesis Testing

- Every time we conduct an experiment where the null hypothesis is actually true, there is a 5% chance we will reject the null – **a false positive**
- As we conduct more experiments (where the null is in fact true), we increase our chances of a false positive
- When conducting multiple experiments in this way, the researcher **needs to adjust their p-values** (upwards) for the fact that false positives are more likely to occur
- In this study, the adjustment would have resulted in at least some (if not all) of the experiments failing to reject the null

The Reasoning of Hypothesis Testing

- For his first experiment, the researcher recruited 100 students from the psychology department at Cornell University volunteered and were given extra credit or $5 to participate
- The participants sat in front of a computer screen with **two curtains** appearing on the screen
- For each of the **36 trials** of the experiment, the participant had to guess which of the curtains would reveal an erotic image
- From the student's point of view, the experiment was a test of clairvoyance
- However, the specific picture was not placed behind one of the curtains until after the guess was made, making it a **test of precognition**

The Reasoning of Hypothesis Testing

- **Null Hypothesis:** Percentage of erotic images guessed correctly equal to 50%
- **Alternative Hypothesis**: Percentage of erotic images guessed correctly is greater than 50% (one-sided alternative)
- The average proportion of times that all 100 participants guessed correctly, 0.531 (or 53.1%), resulting in a p-value = 0.01
- With 3600 replications (or sample size), it is easier to declare statistical significance for even the smallest departure from 0.50.
- The fact that these sort of experiments are also prone to **systematic bias** - inaccuracies in the measurement process, could be the primary reason for the statistically significant departure from 0.50.

The Reasoning of Hypothesis Testing

- In their critique of this study titled **"Why Psychologists Must Change the Way They Analyze Their Data: The Case of Psi,"** the researchers allowed for the possibility of such small departures from 50% when analyzing the results of this research.

- By using **alternative statistical analysis** tools (that are more sensitive to such small effect sizes), the researchers found little to no evidence for precognition in eight of the nine experiments conducted.

- It would not be surprising if the ninth experiment was statistically significant by chance due to sampling variation.

Amy Cuddy Takes a Stand

source : https://www.nytimes.com/

Amy Cuddy Takes a Stand – Sept 2014

The TED conference has made a star of many unlikely people, but perhaps no one more so than Amy Cuddy, a social psychologist and associate professor at Harvard Business School, whose talk promises personal transformation with nary a pill, cleanse or therapy bill.

Her rousing presentation in 2012 at TED Global on what she calls "power poses" is among the most viewed TED Talks of all time (it is No. 2; Sir Ken Robinson's "How Schools Kill Creativity" is No. 1). In its wake, Ms. Cuddy, 42, has attracted lucrative speaking invitations from around the world, a contract from Little, Brown & Co. for a book to be published next year, and an eclectic army of posture-conscious followers.

source : https://www.nytimes.com/

New York Times Magazine – Oct. 2017

FEATURE

When the Revolution Came for Amy Cuddy

As a young social psychologist, she played by the rules and won big: an influential study, a viral TED talk, a prestigious job at Harvard. Then, suddenly, the rules changed.

Power Posing: Brief Nonverbal Displays Affect Neuroendocrine Levels and Risk Tolerance

Abstract

Humans and other animals express power through open, expansive postures, and they express powerlessness through closed, contractive postures. But can these postures actually cause power? The results of this study confirmed our prediction that posing in high-power nonverbal displays (as opposed to low-power nonverbal displays) would cause neuroendocrine and behavioral changes for both male and female participants: High-power posers experienced elevations in testosterone, decreases in cortisol, and increased feelings of power and tolerance for risk; low-power posers exhibited the opposite pattern. In short, posing in displays of power caused advantaged and adaptive psychological, physiological, and behavioral changes, and these findings suggest that embodiment extends beyond mere thinking and feeling, to physiology and subsequent behavioral choices. That a person can, by assuming two simple 1-min poses, embody power and instantly become more powerful has real-world, actionable implications.

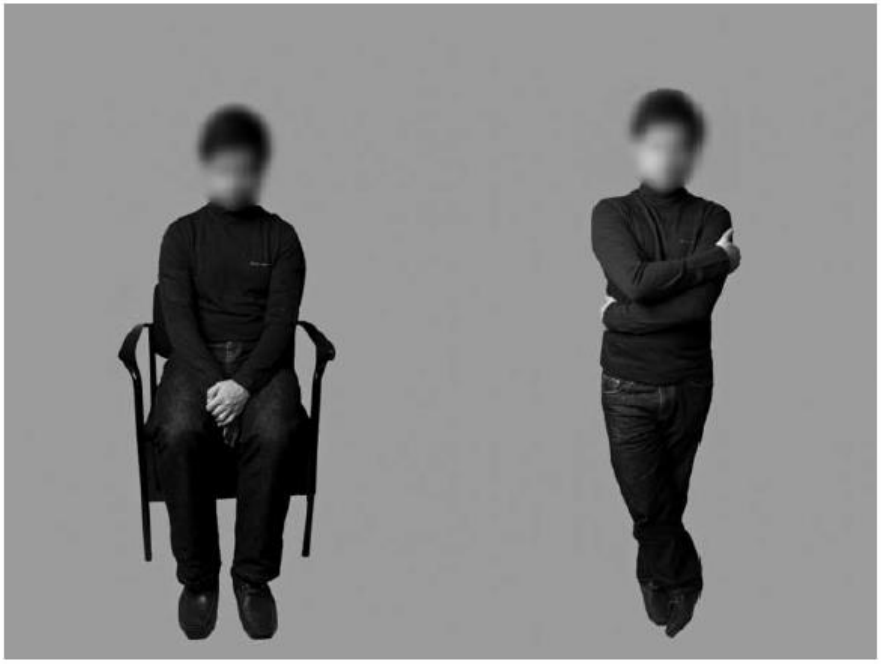
Power Posing: Brief Nonverbal Displays Affect Neuroendocrine Levels and Risk Tolerance

Method

Participants and overview of procedure

Forty-two participants (26 females and 16 males) were randomly assigned to the high-power-pose or low-power-pose condition. Participants believed that the study was about the science of physiological recordings and was focused on how placement of electrocardiography electrodes above and below the heart could influence data collection. Participants' bodies were posed by an experimenter into high-power or low-power poses. Each participant held two poses for 1 min each. Participants' risk taking was measured with a gambling task; feelings of power were measured with self-reports. Saliva samples, which were used to test cortisol and testosterone levels, were taken before and approximately 17 min after the power-pose manipulation.

Power Posing: Brief Nonverbal Displays Affect Neuroendocrine Levels and Risk Tolerance



Results

One-way analyses of variance examined the effect of power pose on postmanipulation hormones (Time 2), controlling for baseline hormones (Time 1). As hypothesized, high-power poses caused an increase in testosterone compared with low-power poses, which caused a decrease in testosterone, $F(1, 39) = 4.29$, $p < .05$; $r = .34$ (Fig. 3). Also as hypothesized, high-power poses caused a decrease in cortisol compared with low-power poses, which caused an increase in cortisol, $F(1, 38) = 7.45$, $p < .02$; $r = .43$ (Fig. 4).

Also consistent with predictions, high-power posers were more likely than low-power posers to focus on rewards—86.36% took the gambling risk (only 13.63% were risk averse). In contrast, only 60% of the low-power posers took the risk (and 40% were risk averse), $\chi^2(1, N = 42) = 3.86$, $p < .05$; $\Phi = .30$. Finally, high-power posers reported feeling significantly more "powerful" and "in charge" ($M = 2.57$, $SD = 0.81$) than low-power posers did ($M = 1.83$, $SD = 0.81$), $F(1, 41) = 9.53$, $p < .01$; $r = .44$. Thus, a simple 2-min power-pose manipulation was enough to significantly alter the physiological, mental, and feeling states of our participants. The implications of these results for everyday life are substantial.

Power Posing: Brief Nonverbal Displays Affect Neuroendocrine Levels and Risk Tolerance

Observed (Expected) Counts			
	Pose		
Risk Taking	HPP	LPP	Total
Yes	18 (15.5)	13 (15.5)	31
No	3 (5.5)	8(5.5)	11
Total	21	21	42

- 86.36% of HPP or 18 subjects took gambling risk
- 60% of LPP or 13 subjects took gambling risk
- The *χ2 test of independence resulted in a χ2 statistic of 3.86, with in a* ***p-value = 0.0495****!*

Q. *Why should we be* skeptical about this result? How do you think the researcher might have obtain this statistically significant result?

My position on "Power Poses"

Dana Carney, a coauthor in the study who has since distanced herself from the research, said that some p-hacking did occur when analyzing the data. The use of the abbreviation "DV" is assumed here to mean "dependent variable".

4. The data are flimsy. The effects are small and barely there in many cases.
5. Initially, the primary DV of interest was risk-taking. We ran subjects in chunks and checked the effect along the way. It was something like 25 subjects run, then 10, then 7, then 5. Back then this did not seem like p-hacking. It seemed like saving money (assuming your effect size was big enough and p-value was the only issue).

10. The self-report DV was p-hacked in that many different power questions were asked and those chosen were the ones that "worked."

Where do I Stand on the Existence of "Power Poses"

1. I do not have any faith in the embodied effects of "power poses." I do not think the effect is real.
2. I do not study the embodied effects of power poses.
3. I discourage others from studying power poses.
4. I do not teach power poses in my classes anymore.
5. I do not talk about power poses in the media and haven't for over 5 years (well before skepticism set in)

Assessing the Robustness of Power Posing

Table 10 .1: Results from Replication Study: Assessing the Robustness of Power Posing

Measurement	Sample Statistic	Confidence Interval	P-value
Risk-Taking	Mean difference	[−0.085, 0.019]	0.215
Testosterone	Mean difference	[−9.801, 1.647]	0.162
Cortisol	Mean difference	[−0.078, 0.022]	0.272

The researchers used a sample size of 200 participants and found no statistically significant effect size for the three main outcomes of the original research, as shown in Table 10.1.

Merck Agrees to Settle Vioxx Suits for $4.85 Billion

By Alex Berenson

Nov. 9, 2007

Three years after withdrawing its pain medication Vioxx from the market, Merck has agreed to pay $4.85 billion to settle 27,000 lawsuits by people who claim they or their family members suffered injury or died after taking the drug, according to two lawyers with direct knowledge of the matter.

The settlement, one of the largest ever in civil litigation, comes after nearly 20 Vioxx civil trials over the last two years from New Jersey to California. After losing a $253 million verdict in the first case, Merck has won most of the rest of the cases that reached juries, giving plaintiffs little choice but to settle.

Scientists Again Defend Study on Vioxx

With a crucial personal-injury trial over Vioxx set to begin in New Jersey next week, the debate heated up again yesterday about whether Merck understated the drug's risks in a journal article in November 2000.

In an letter published online by The New England Journal of Medicine, 11 scientists who were co-authors of the article said they stood by its original conclusions, despite heavy criticism from the editors of the journal.

The trial confirmed that Vioxx seemed to be safer on the stomach, but it also showed that more patients taking Vioxx than naproxen died and that many more suffered heart attacks. As published, the article reported that 17 patients taking Vioxx and 4 taking naproxen had heart attacks during the trial.

In fact, 20 patients on Vioxx suffered heart attacks, but the final three heart attacks were reported after Feb. 10, 2000, a cut-off date specified by Merck for reporting cardiovascular problems.

Analysis of 2 x 2 Tables: Risk and Relative Risk

- Another much used statistic that can be calculated from the a two-by-two table is known as relative risk.

- **Risk** is a term used in epidemiology (the study and analysis of health outcomes and diseases in populations) defined as the probability that an event (such as heart attack or death) will occur.

- **Relative risk** is a value that compares one group's risk of a disease or outcome relative to another group.

- **Example:** If the relative risk of a disease is 4, then the risk is four times as great for one group relative to another group

Analysis of 2 x 2 Tables: Risk and Relative Risk

Group	Heart Attack	No Heart Attack	Total
High Cholesterol	2	98	100
Low Cholesterol	1	99	100
Total	3	197	200

Q. How do we calculate the risk of heart attack for HC and LC patients?

Relative Risk = Risk for HC patients/Risk for LC patients
= 0.02/0.01 = 2

Relative Risk = Risk for LC patients/Risk for HC patients
= 0.01/0.02 = 0.5

Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis

Methods We randomly assigned 8076 patients who were at least 50 years of age (or at least 40 years of age and receiving long-term glucocorticoid therapy) and who had rheumatoid arthritis to receive either 50 mg of rofecoxib daily or 500 mg of naproxen twice daily. The primary end point was confirmed clinical upper gastrointestinal events (gastroduodenal perforation or obstruction, upper gastrointestinal bleeding, and symptomatic gastroduodenal ulcers).

Results Rofecoxib and naproxen had similar efficacy against rheumatoid arthritis. During a median follow-up of 9.0 months, 2.1 confirmed gastrointestinal events per 100 patient-years occurred with rofecoxib, as compared with 4.5 per 100 patient-years with naproxen (relative risk, 0.5; 95 percent confidence interval, 0.3 to 0.6; $P<0.001$). The respective rates of complicated confirmed events (perforation, obstruction, and severe upper gastrointestinal bleeding) were 0.6 per 100 patient-years and 1.4 per 100 patient-years (relative risk, 0.4; 95 percent confidence interval, 0.2 to 0.8; $P=0.005$). The incidence of myocardial infarction was lower among patients in the naproxen group than among those in the rofecoxib group (0.1 percent vs. 0.4 percent; relative risk, 0.2; 95 percent confidence interval, 0.1 to 0.7); the overall mortality rate and the rate of death from cardiovascular causes were similar in the two groups.

Q. What important statistical value is missing from the results related to myocardial infarctions (or heart attacks)?

Q. Do you notice a difference in the way relative risk was calculated across the results presented?

https://www.ncbi.nlm.nih.gov/pubmed/11087881

Vioxx and Heart Attacks

The relative risk was 0.4 for complicated confirmed events, calculated as follows:

Relative Risk = Risk on Vioxx/Risk on Aleve

= 0.006/0.014

= 0.4

It is accepted (and proper) practice to list the primary drug of interest (Vioxx) first and then the comparator drug (Aleve)

Vioxx and Heart Attacks

The risk of heart attack on Vioxx was 0.004 or 0.4 percent, representing the 17 patients who got heart attacks. The risk of heart attack on Aleve was 0.001 or 0.1 percent, representing the 4 patients who got heart attacks.

The relative risk for heart attacks was presented as 0.2, calculated as follows:

Relative Risk = Risk on Aleve/Risk on Vioxx

= 0.001/0.004

= 0.25

Questions to Ask about this Study?

- Why cause confusion by not consistently calculating relative risk?
- Why would Merck present the relative risk as 0.2 (instead of 0.25) making Vioxx look worse when to comes to the risk of heart attack than it actually was?
- Why did Merck present the p-values for the primary analysis and not for the analysis of heart attacks?
- Why did peer review at the *New England Journal of Medicine* not question how the results were presented?
- How come the FDA did not catch the problems with how the results were presented?
- **Did Merck know the drug was causing heart attacks before the drug was brought to market?**

REPORTS & MULTIMEDIA / CASE STUDY

Merck Manipulated the Science about the Drug Vioxx

Published Oct 12, 2017

Scientists from the pharmaceutical giant Merck skewed the results of clinical trials in favor of the arthritis drug, Vioxx, to hide evidence that the drug increased patients' risk of heart attack.

Sacklers Directed Efforts to Mislead Public About OxyContin, Court Filing Claims

A filing in a Massachusetts lawsuit contains dozens of internal Purdue Pharma documents suggesting the family was far more involved than the company has long contended.

693

Sacklers Directed Efforts to Mislead Public About OxyContin, Court Filing Claims

Members of the Sackler family, which owns the company that makes OxyContin, directed years of efforts to mislead doctors and patients about the dangers of the powerful opioid painkiller, a court filing citing previously undisclosed documents contends.

When evidence of growing abuse of the drug became clear in the early 2000s, one of them, Richard Sackler, advised pushing blame onto people who had become addicted.

"We have to hammer on abusers in every way possible," Mr. Sackler wrote in an email in 2001, when he was president of the company, Purdue Pharma. "They are the culprits and the problem. They are reckless criminals."

www.nytimes.com

THE PROMOTION AND MARKETING OF OXYCONTIN: COMMERCIAL TRIUMPH: PUBLIC HEALTH TRAGEDY

The researcher discusses the massive marketing campaign conducted by Purdue Pharma. Also, when discussing the origins of the FDA approval of Oxycontin, he states the following:

> *The FDA's medical review officer, in evaluating the efficacy of OxyContin in Purdue's 1995 new drug application, concluded that OxyContin* ***had not been shown to have a significant advantage*** *over conventional, immediate-release oxycodone taken 4 times daily other than a reduction in frequency of dosing.*

Q. Why was a drug approved (that was highly addictive) if it was found to be no better than what was already on the market?

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2622774/

Purdue Pharma Pleads Guilty to Criminal Charges for Opioid Sales

The Justice Department announced an $8 billion settlement with the company. Members of the Sackler family will pay $225 million in civil penalties but criminal investigations continue.

www.nytimes.com

Purdue Pharma Pleads Guilty to Criminal Charges for Opioid Sales – Nov. 30th, 2020

But state attorneys general from Massachusetts, New York and North Carolina, among others, have raised questions about just how much of an effect the settlement will have with respect to holding the Sackler family to account. Purdue was keen to settle its federal legal troubles under a Trump administration, which it sensed would cut a better deal than a new Biden administration.

The $225 million that the Sacklers would pay as part of their civil settlements is small relative to the family's net worth, estimated to be at least $13 billion, much of it generated from sales of OxyContin.

www.nytimes.com

McKinsey Proposed Paying Pharmacy Companies Rebates for OxyContin Overdoses – Dec 2020

In a 2017 presentation, according to the records, which were filed in court on behalf of multiple state attorneys general, McKinsey laid out several options to shore up sales. One was to give Purdue's distributors a rebate for every OxyContin overdose attributable to pills they sold.

The presentation estimated how many customers of companies including CVS and Anthem might overdose. It projected that in 2019, for example, 2,484 CVS customers would either have an overdose or develop an opioid use disorder. A rebate of $14,810 per "event" meant that Purdue would pay CVS $36.8 million that year.

www.nytimes.com

McKinsey Proposed Paying Pharmacy Companies Rebates for OxyContin Overdoses

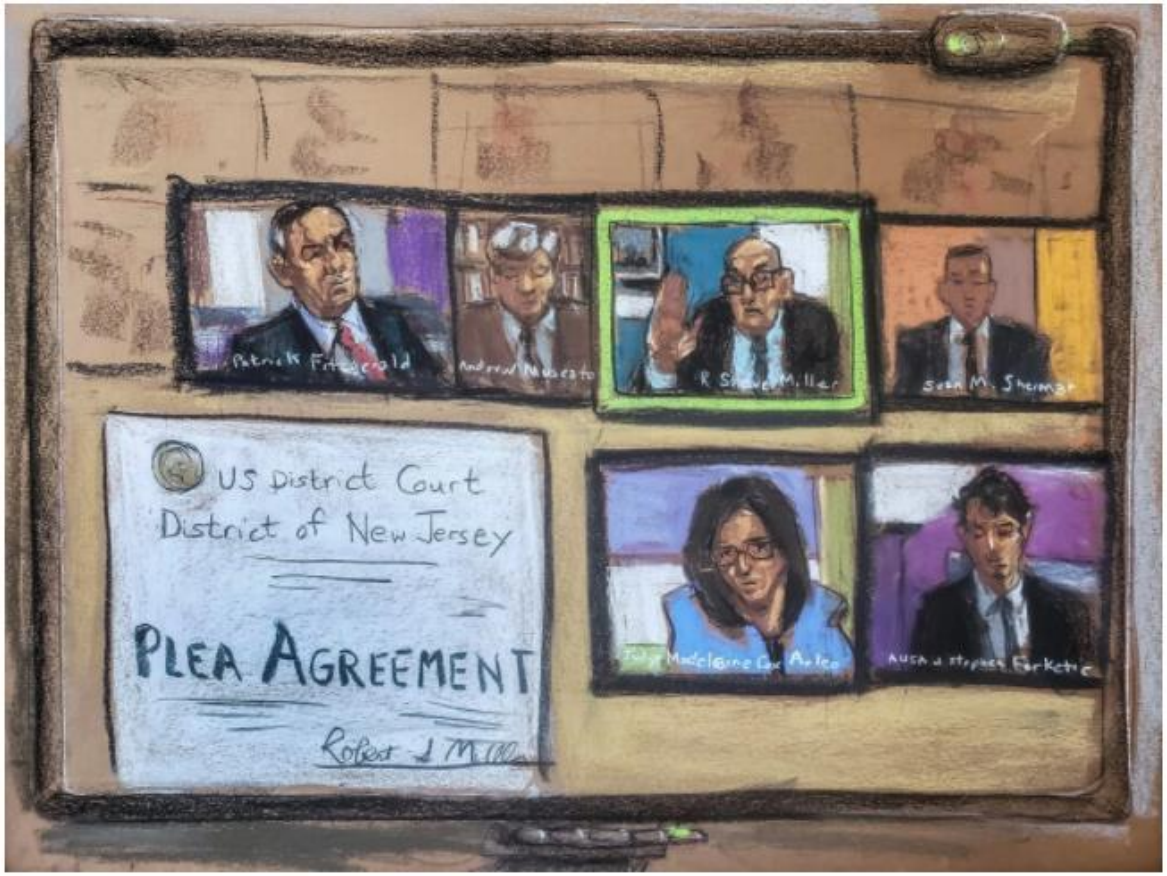
The federal settlement with Purdue comes as states and municipalities seek compensation from opioid makers for helping fuel a health crisis that has killed more than 450,000 Americans since 1999. Purdue is now seeking bankruptcy protection, as are other manufacturers.

"This is the banality of evil, M.B.A. edition," Anand Giridharadas, a former McKinsey consultant who reviewed the documents, said of the firm's work with Purdue. "They knew what was going on. And they found a way to look past it, through it, around it, so as to answer the only questions they cared about: how to make the client money and, when the walls closed in, how to protect themselves."

www.nytimes.com

The Sacklers' Last Poison Pill

A trapdoor in the Justice Department deal with the family's company might enable them to escape a full accounting of their part in the opioid epidemic.



The Sacklers' Last Poison Pill – Dec. 5th, 2020

Now, we have learned that the committee, pressured by the Sackler legal team, has postponed the hearing to January. But January may be too late. By waiting, the House Oversight Committee may miss the opportunity to weigh in before advances in Purdue's bankruptcy case possibly allow the Sacklers — one of America's richest families, who took in billions in revenue from sales of OxyContin — to escape with little public scrutiny or accountability.

By then, a bankruptcy plan to reorganize Purdue will probably have been proposed. If, as expected, the plan seeks to release the Sacklers from liability, it will become practically impossible to uncover the full truth about the Sacklers' role in the opioid crisis.

Integrity and Leadership

"May I stress the need for courageous, intelligent, and dedicated leadership... Leaders of sound integrity. Leaders not in love with publicity, but in love with justice. Leaders not in love with money, but in love with humanity. Leaders who can subject their particular egos to the greatness of the cause."

– Martin Luther King

Martin Luther King's Last Speech: "I've Been To The Mountaintop"

- Martin Luther King used his natural intelligence and leadership skills to move forward a cause he truly believed in. He had the courage and commitment to followed his path to truth until he reached the mountaintop.