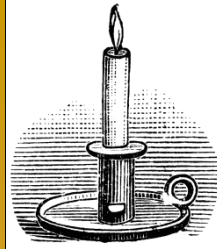




THE UNIVERSITY OF  
ALABAMA AT BIRMINGHAM

Office  
of  
Energetics



David B. Allison, Ph.D.



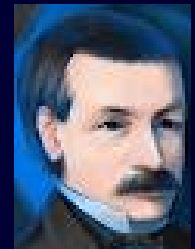
*Making data publically  
available: benefits,  
challenges, and suggestions*

# Disclosure

I have received financial and other benefits from the following entities: the Frontiers Foundation; The Federal Trade Commission; Vivus, Inc; Kraft Foods; University of Wisconsin; University of Arizona; Paul, Weiss, Wharton & Garrison LLP; Sage Publications, and additional government, non-profit and for-profit organizations with interests in obesity, nutrition, and health.

# The Problem is Not New

- Brother Gregor Mendel
- Louis Pasteur
- Sir Cyril Burt
- Dominique Cassini's suppression of Ole Roemer's 1676 demonstration of the finite speed of the light.
- Leopold Kronecker's attempted suppression of Cantor's work on infinities.



# Outline – Public Data Sharing

- Potential Benefits w/ Examples
- Current State of Affairs
- Challenges
- Possible Approaches

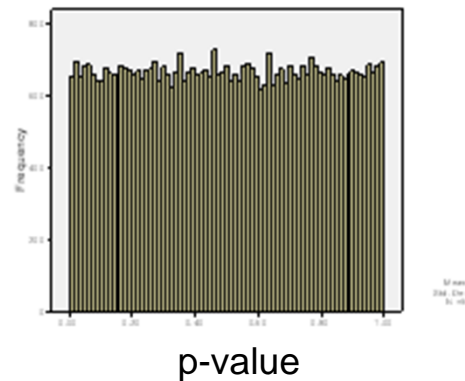
# Public Data Sharing: Potential Benefits

- Detecting Fraud
- Detecting Errors in Analysis
- Detecting Problems with the Data
- Making New Discoveries
- Trying Alternative Analyses
- Trying Analyses that Provide Alternative Insights
- Re-expressing results (e.g., in common metrics)

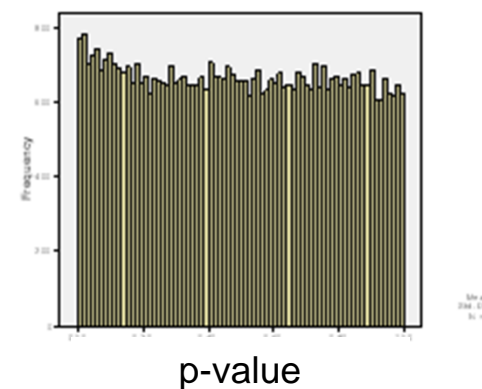
Potential Benefit: To Detect Patent  
Fraud (as Dr. Coombes Illustrated)

# Distortions via Statistical Fiddling in Obesity Trials? (Preliminary Unpublished Data)

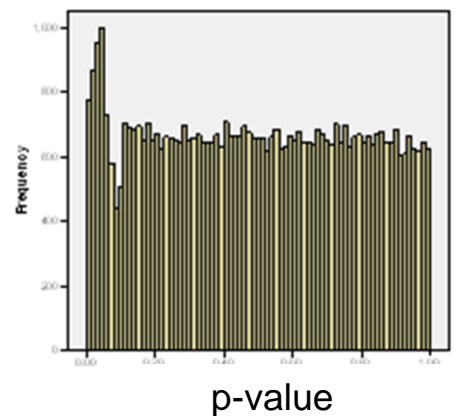
Simulated Distribution of p-values all nulls true.



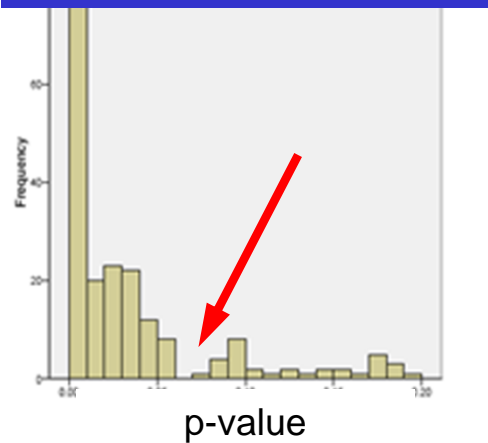
Simulated Distribution of p-values some nulls false.



Simulated p-values all nulls true + 'Fiddling'.



Real data (N=347 obesity RCTs; p for dip = .052)



## Potential Benefit: Detecting Errors

(e.g., Flegal et al able to detect an analysis error I had made)

Sources of differences in estimates of obesity-associated deaths from first National Health and Nutrition Examination Survey (NHANES I) hazard ratios<sup>1-4</sup>

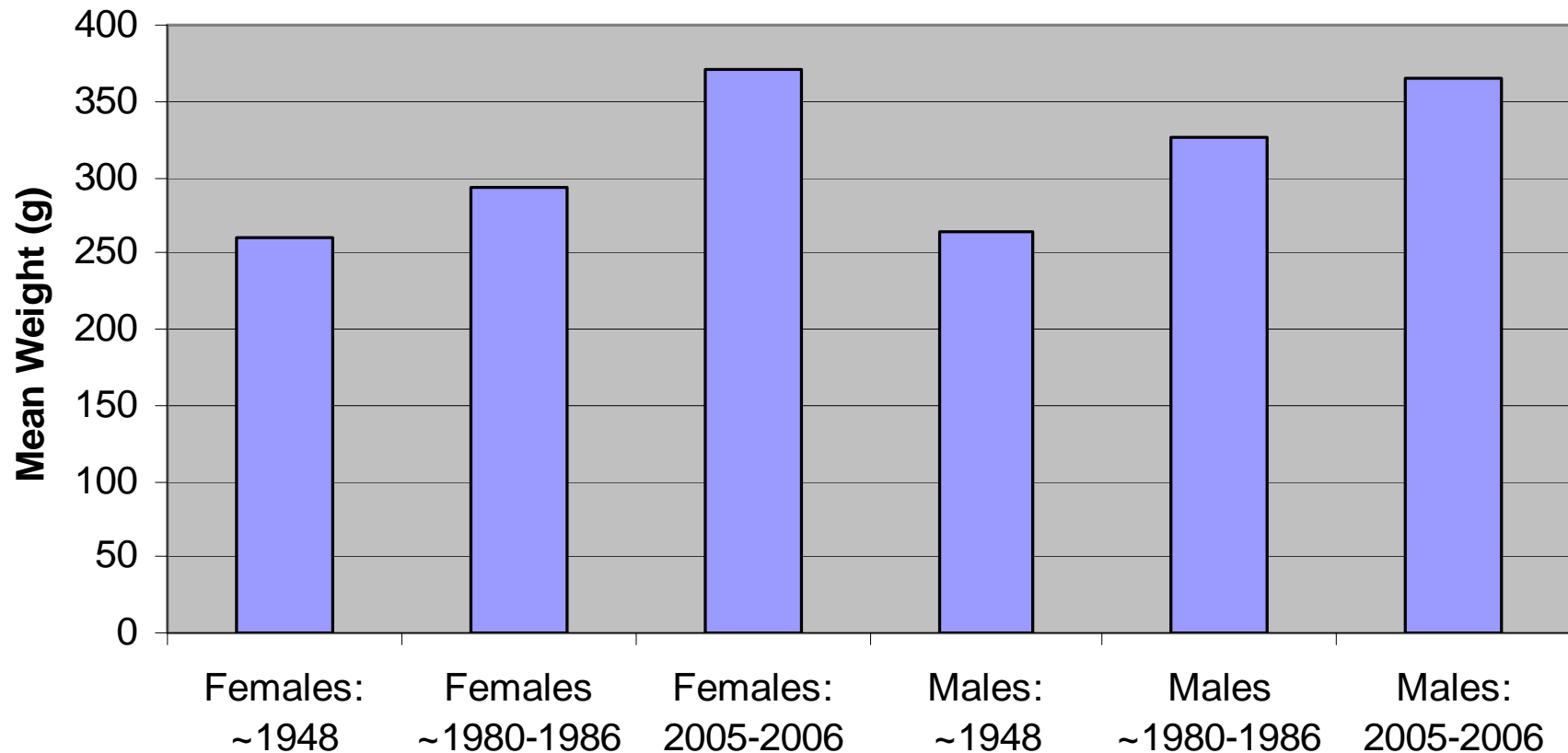
*Katherine M Flegal, Barry I Graubard, David F Williamson, and Mitchell H Gail*

**Conclusion:** Methodologic sources of bias in the reports by Allison et al and Mokdad et al include the assessment of smoking status in NHANES I and the method of calculating attributable fractions. *Am J Clin Nutr* 2010;91:519–27.



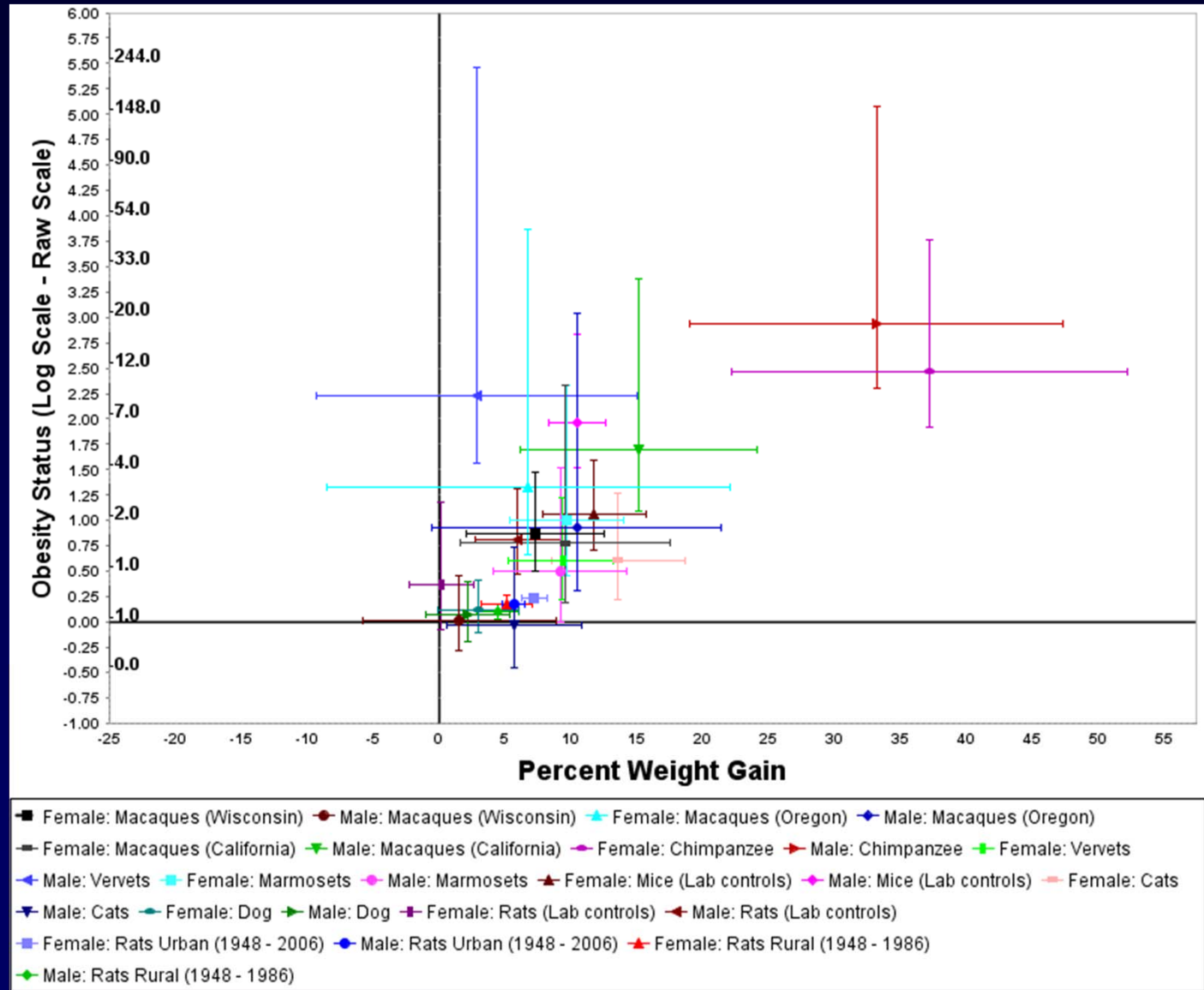
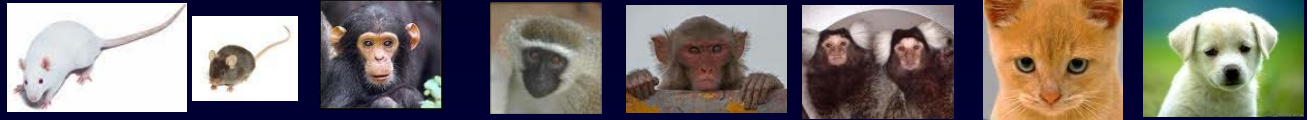
## Potential Benefit: Promoting New Findings

### Feral Rats in Baltimore are Getting Heavier



- Davis, D. E. The Weight of Wild Brown Rats at Sexual Maturity. *Journal of Mammalogy*, Vol. 30, No. 2. (May, 1949), pp. 125-130.
- Easterbrook et al. A survey of zoonotic pathogens carried by Norway rats in Baltimore, Maryland, USA. *Epidemiol Infect.* 2007 Jan 15;;1-8
- Glass, G.E. et al. (1989). *Occasional Papers of the Museum of Natural History*, 130, pp 1-33.

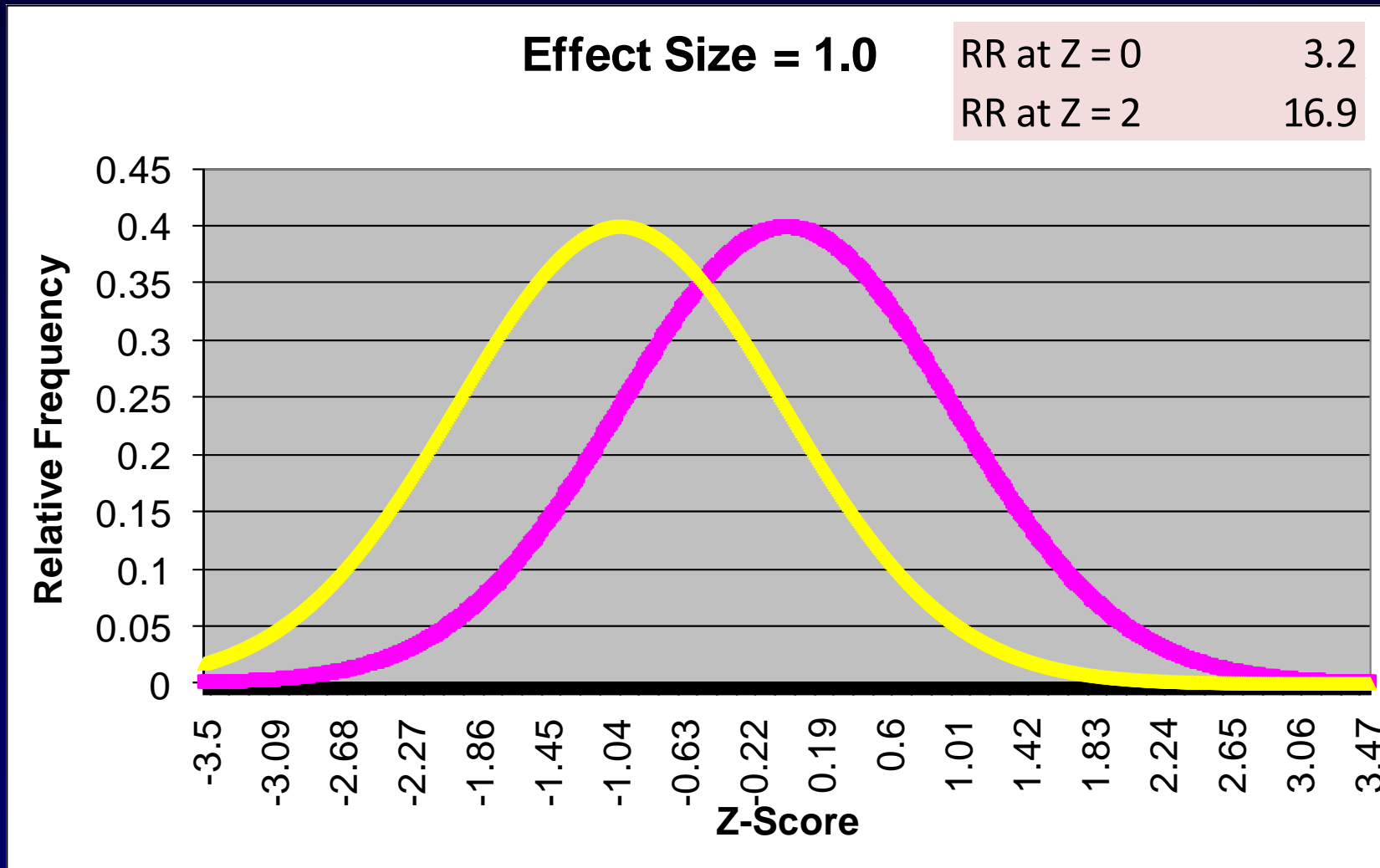
# The Plurality of Obesity Epidemics



Klimentidis Y.C., Beasley T.M., Lin H., Murati G., Glass G.E., Guyton M., Newton W., Jorgensen M., Heymsfield S.B., Kernitz J., Fairbanks L., Allison, D.B. (2010). *Canaries in the Coal Mine: A Cross-Species Analysis of the Plurality of Obesity Epidemics. Proceedings of the Royal Society B: Biological Sciences.*

## Potential Benefit: Ability to Reexpress Results

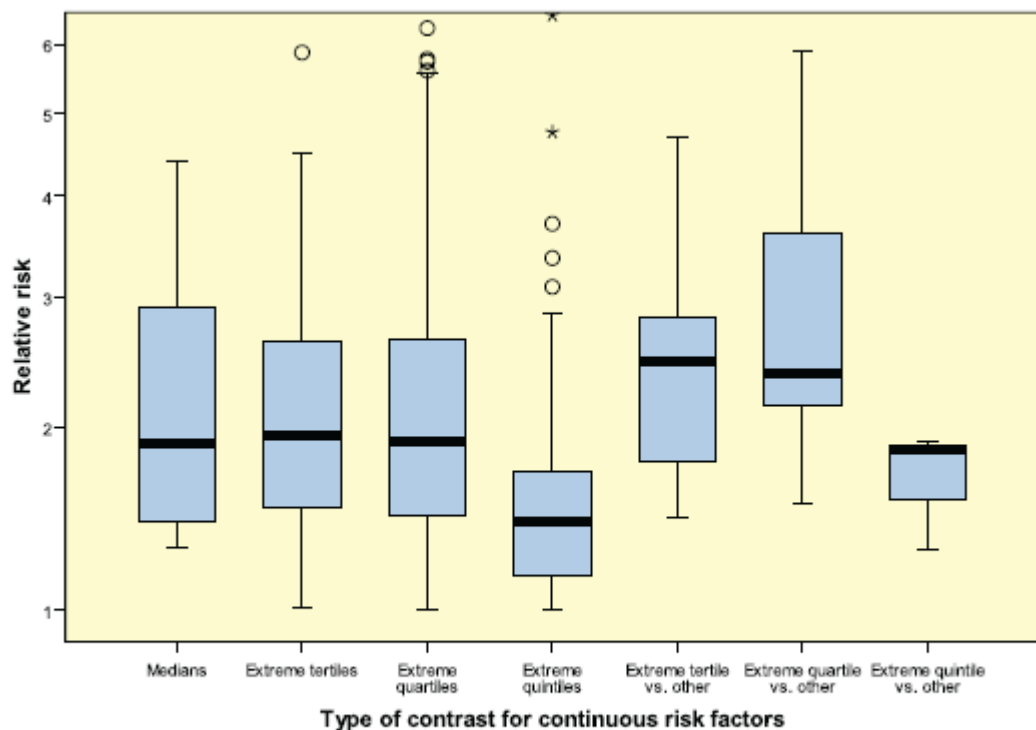
### Effect of Dependent Variable Threshold on Magnitude of Effect Estimate in Dichotomous Analysis



# Reporting Bias

Kavvoura FK, Liberopoulos G, Ioannidis JP. Selection in reported epidemiological risks: an empirical assessment. *PLoS Med.* 2007 Mar;4(3):e79.

“We evaluated 389 articles of epidemiological studies that reported, in their respective abstracts, at least one relative risk for a continuous risk factor in contrasts based on median, tertile, quartile, or quintile categorizations. ...Paradoxically, the smallest presented relative risks were based on the contrasts of extreme quintiles.”



**Figure 2.** Box Plots for Relative Risks for Different Contrasts of the Values of the Postulated Risk Factor  
All relative risks have been coined to be  $\geq 1.00$  for consistency.  
doi:10.1371/journal.pmed.0040079.g002

# Potential Benefit: Ability to Examine Fundamental Data

## ADVERSE EVENTS ASSOCIATED WITH DIETARY SUPPLEMENTS CONTAINING EPHEDRA ALKALOIDS

### ADVERSE CARDIOVASCULAR AND CENTRAL NERVOUS SYSTEM EVENTS ASSOCIATED WITH DIETARY SUPPLEMENTS CONTAINING EPHEDRA ALKALOIDS

(N Engl J Med 2000;343:1833-8.)

**TABLE 5.** OUTCOME IN 15 PATIENTS WITH ADVERSE EVENTS THAT WERE POSSIBLY RELATED TO THE USE OF SUPPLEMENTS CONTAINING EPHEDRA ALKALOIDS.

PATIENT No.	AGE (YR)/SEX	NAME OF SUPPLEMENT	ESTIMATED DAILY DOSE OF EPHEDRA ALKALOIDS mg	DURATION OF USE	ADVERSE EVENT	OUTCOME	PREEXISTING CONDITIONS OR CONCURRENT RISKS
1	46/M	Diect Fucl	Unknown	5-6 mo	Stroke	Death	None
2	22/M	Ripped Fucl	Unknown	Unknown	Hyperthermia, abnormal electrolyte levels, cardiac arrest	Death	None

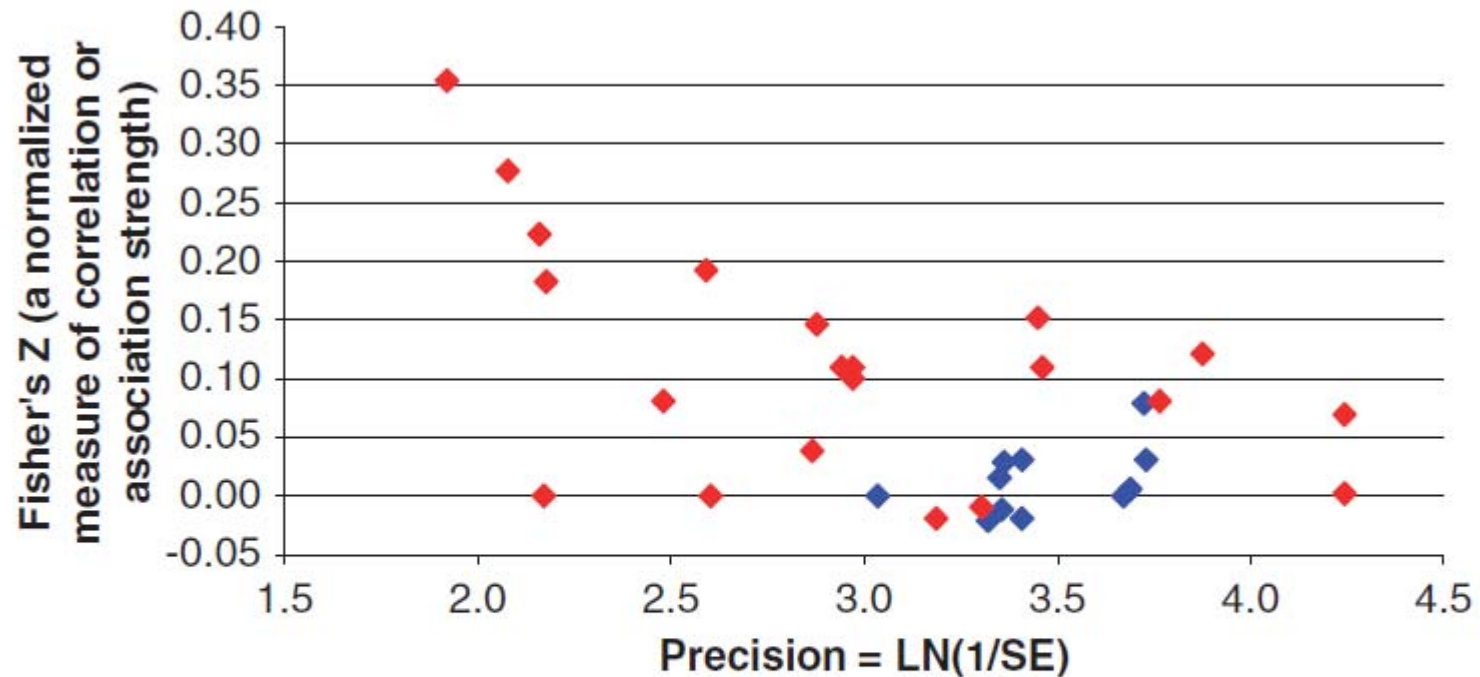
# No Concurrent Risks?

From the Investigating Officers' Reports (available on FDA website):

After the victim, Mr. [REDACTED], was removed, I talked to Mr. [REDACTED] and was told that Mr. [REDACTED] had been working out since 5:30 this morning to [REDACTED] for [REDACTED]. Just prior to Mr. [REDACTED] collapsing he had been riding a stationary bike in the shower room of locker room [REDACTED]. We proceeded to this area and upon entering noted that the room was quite warm and located in the showers were 4 stationary bikes. Lying on the floor, in a heap, was a sweatsuit and a rubberized body suit. Mr. [REDACTED] said that he

At 2:00 p.m., I visited the local Medical Examiner, Mr. [REDACTED]. Mr. [REDACTED] could not understand why the FDA and the press were so interested in the case. He stated that it was obvious the [REDACTED] died of hyperthermia. He stated that the individual was trying to lose weight by wearing a rubber suit under a sweat suit while riding a stationary bicycle in a steamy shower room. After being transported to the hospital and being pronounced dead, he had a temperature of 108 degrees Fahrenheit. He stated that one hour after death his temperature was still 107 degrees Fahrenheit. During this visit Mr. [REDACTED] provided me with the attached copies of:

# Potential Benefit: Alternative Insight



**Figure 1** Plot of sample effect sizes from cross-sectional studies of the association between sugar-sweetened beverage consumption and obesity indexes indicating publication bias among non-industry-funded studies (Blue diamonds = industry funded; Red diamonds = non-industry funded).

White hat bias: examples of its presence in obesity research and a call for renewed commitment to faithfulness in research reporting. Cope MB, Allison DB.. *Int J Obes (Lond)*. 2010 Jan;34(1):84-8;



# Potential Benefit: Try Alternative Analyses: Lesser et al, *PLoS Medicine*, 2007

**Table 2.** Relationship between Funding Source and Article Conclusions ( $n = 76$ )

Conclusion	Funding Source <sup>a</sup>			Exact $p$ for Trend
	All Industry Benefit <sup>b</sup>	No Industry	All Industry Antagonism <sup>c</sup>	
Favorable	14	24	0	0.037
Neutral	5	8	1	—
Unfavorable	3	20	1	—

## Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2- sided)
Pearson Chi-Square	6.774 <sup>a</sup>	4	.148	.131
Likelihood Ratio	7.894	4	.096	.101
Fisher's Exact Test	7.409			.069
N of Valid Cases	76			

a. 4 cells (44.4%) have expected count less than 5. The minimum expected count is .37.



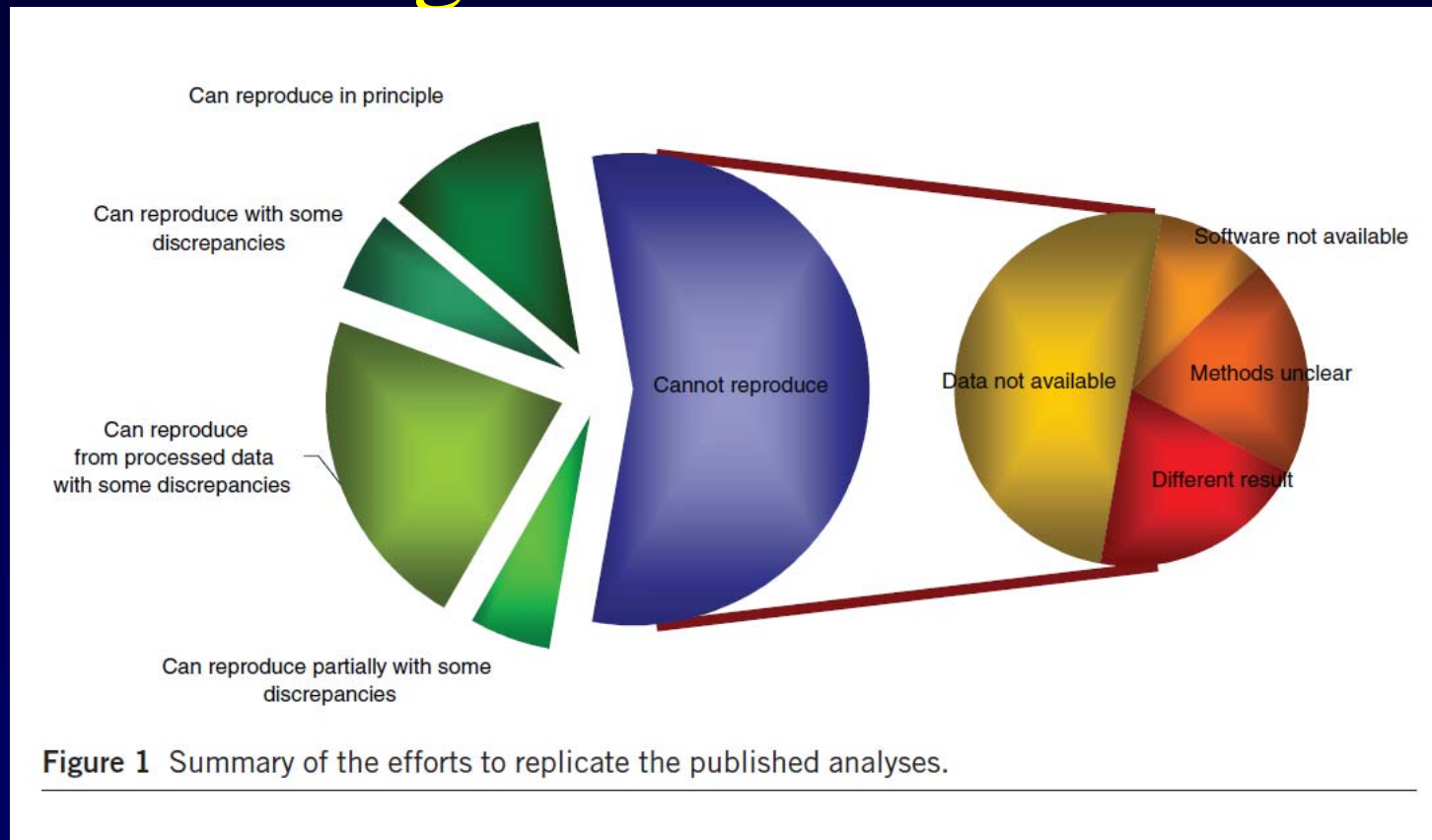
# Current State of Affairs

Allison, D. B. The antidote to bias in research. *Science*. 2009 Oct 23;326(5952):522-3.

“The movement toward open data has begun. NIH, Science, the Nature journals, and other journals (9) all have policies encouraging or mandating it. Still, compliance with data sharing is challenging. ... This indicates that effective implementation of data sharing policies requires resources to support implementation and monitoring.”

That is, largely a reactive and voluntary system with minimal compliance and little enforcement.

# An Example of the Challenges in Making Raw Data Available



Nat Genet. 2009 Feb;41(2):149-55. Repeatability of published microarray gene expression analyses. Ioannidis JP, Allison DB, Ball CA, Coulibaly I, Cui X, Culhane AC, Falchi M, Furlanello C, Game L, Jurman G, Mangion J, Mehta T, Nitzberg M, Page GP, Petretto E, van Noort V.

# Challenges to Implementing Public Data Sharing and Potential Responses

- Concerns about unqualified use – C'est La Vie.
- Concerns about use beyond consent – Contract unfettered use.
- Labor – including follow-up questions – See following slides.
- Fear of de-anonymization – Prosecute rather than gate-keep.
- Expense – Prioritize; triage proactive vs reactive.
- Fear of being seen as sloppy - ??
- Fear of honest errors being detected and portrayed as more than honest error – Set burden high for impugning intentions and do not accept defamation.
- Fear of 'bashing' by nefarious opponent camps – Special case of above?
- Loss of commercial IP – Different rules for different settings?
- Fear of losing priority of one's own data – Allow release (temporary?) of subsets of data based on only published variables.

# Some Potential Approaches

# LET JOURNALS TAKE THE LEAD?

PENG, R. D. (2009) REPRODUCIBLE RESEARCH AND  
*BIOSTATISTICS. BIOSTATISTICS VOLUME 10, ISSUE 3*  
*PP. 405-408. .*

## **2.1 Dimensions of reproducibility**

The Associate Editor for reproducibility (AER) will handle submissions of reproducible articles. Currently, the AER's involvement with a submission begins only when an article has been accepted for publication. The AER will consider three different criteria when evaluating the reproducibility of an article.

1. **Data:** The analytic data from which the principal results were derived are made available on the journal's Web site. The authors are responsible for ensuring that necessary permissions are obtained before the data are distributed.
2. **Code:** Any computer code, software, or other computer instructions that were used to compute published results are provided. For software that is widely available from central repositories (e.g. CRAN, Statlib), a reference to where they can be obtained will suffice.
3. **Reproducible:** An article is designated as reproducible if the AER succeeds in executing the code on the data provided and produces results matching those that the authors claim are reproducible. In reproducing these results, reasonable bounds for numerical tolerance will be considered.

# Provide Credit for Impact Via Data Provision

Table 2. Multivariate regression on citation count for 85 publications

	Percent increase in citation count (95% confidence interval)	p-value
Publish in a journal with twice the impact factor	84% (59 to 109%)	<0.001
Increase the publication date by a month	-3% (-5 to -2%)	<0.001
Include a US author	38% (1 to 89%)	0.049
<b>Make data publicly available</b>	<b>69% (18 to 143%)</b>	<b>0.006</b>

We calculated a multivariate linear regression over the citation counts, including covariates for journal impact factor, date of publication, US authorship, and data availability. The coefficients and p-values for each of the covariates are shown here, representing the contribution of each covariate to the citation count, independent of other covariates.

doi:10.1371/journal.pone.0000308.t002

Piowar HA, Day RS, Fridsma DB. Sharing detailed research data is associated with increased citation rate. PLoS One. 2007 Mar 21;2(3):e308.

# Proactive vs. Reactive Data Provision

Proactive:

*Globally Inefficient*

Reactive:

*Locally Inefficient*

*Socially Challenging*



# Centralize As Much of the Storage and Dissemination Burden as Possible

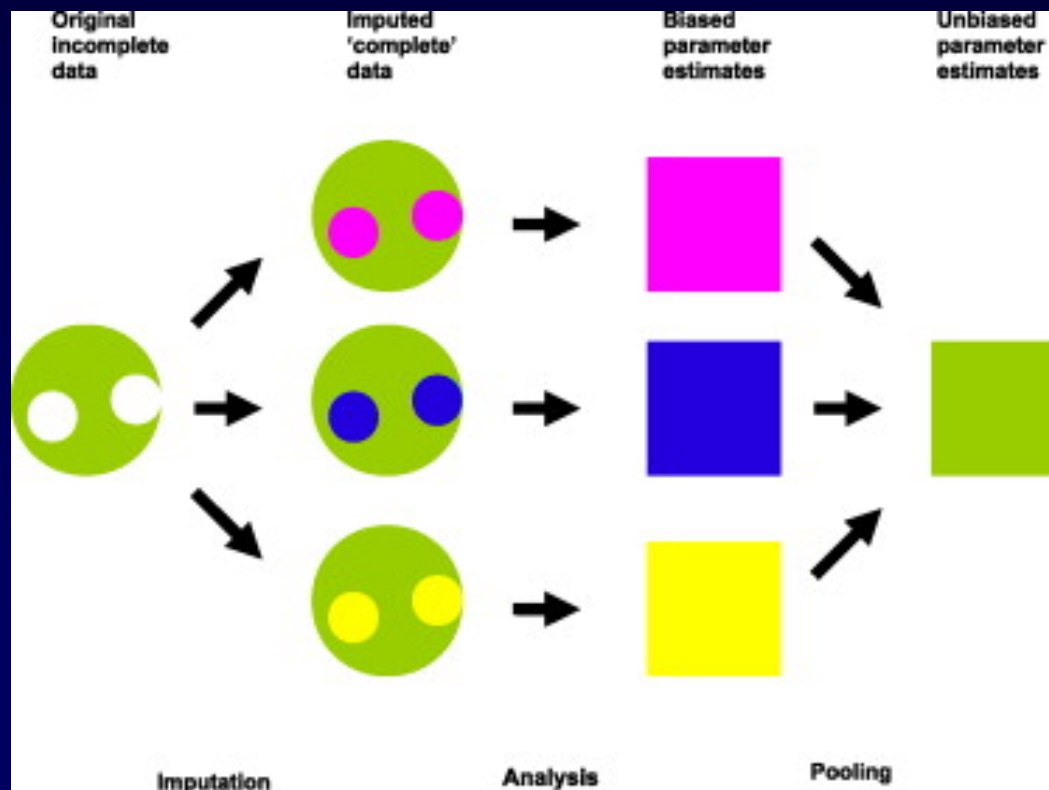
Public Repositories

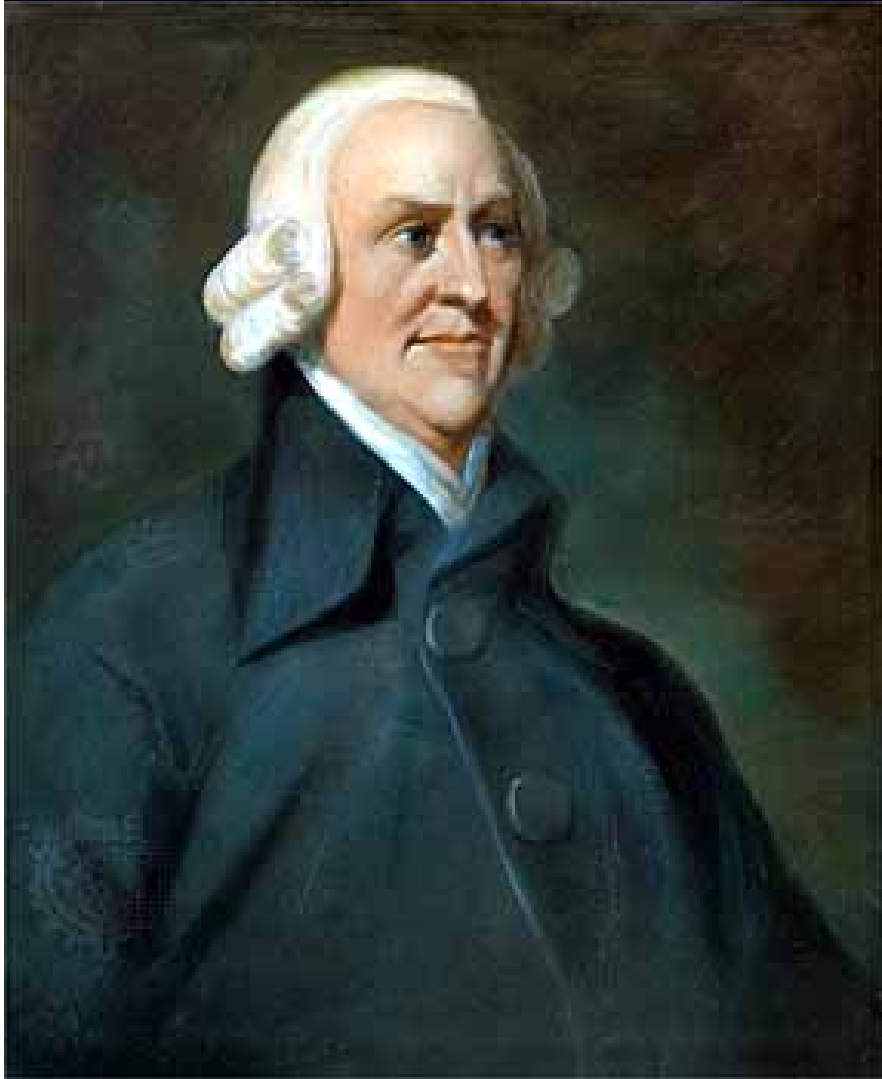
Institutional 'Tailors'

# Some Potential Levers to Mandate Open Data Sharing (and Publication)

- IRB – Ethical Obligation to Subjects
- IACUC – Ethical Obligation to Animals
- Ethical Obligation to students
- IRS – non-profit status
- Gov't funding – ethical obligation to taxpayers

# Statistical Approaches to Public Sharing of Data with Personal Identification Concerns: Multiple Imputation is One Example





“Science is the great  
antidote to the poison  
of enthusiasm and  
superstition.”

~ Adam Smith

# Acknowledgment

- Cathy Calloway
- Tapan Mehta
- Vinodh Srinivasasainagendra

“...let us take this  
path through the  
woods...”

~ Jean-Jacques  
Rousseau



**High Falls, Moss Rock Preserve, Al**

From: [http://farm3.static.flickr.com/2766/4226784012\\_7a1133a0e9.jpg](http://farm3.static.flickr.com/2766/4226784012_7a1133a0e9.jpg)