A Spreadsheet Program for Use in the Detection of Anomalous Numerical Data of the Type Frequently Encountered in Cell and Radiation Biology Colony Survivals

Helene Z Hill

Rutgers NJ Medical School, Newark, NJ

And

Joel Pitt

Renaissance Associates, Princeton, NJ

American Society for Photobiology Biennial Meeting, San Diego, CA

June 19, 2014

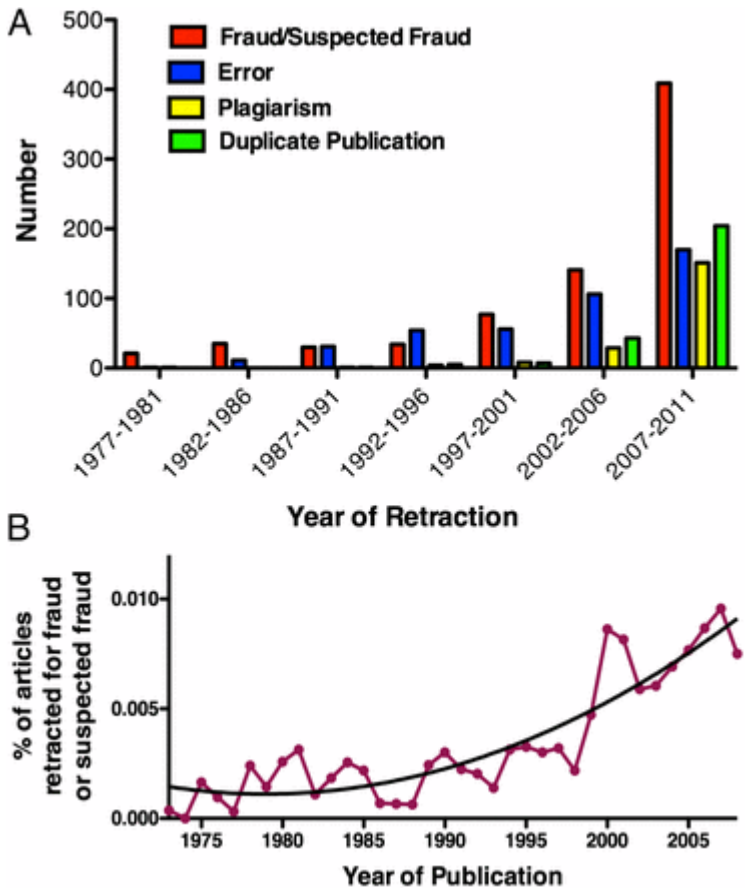
Scientific Misconduct

Falsification, Fabrication, Plagiarism

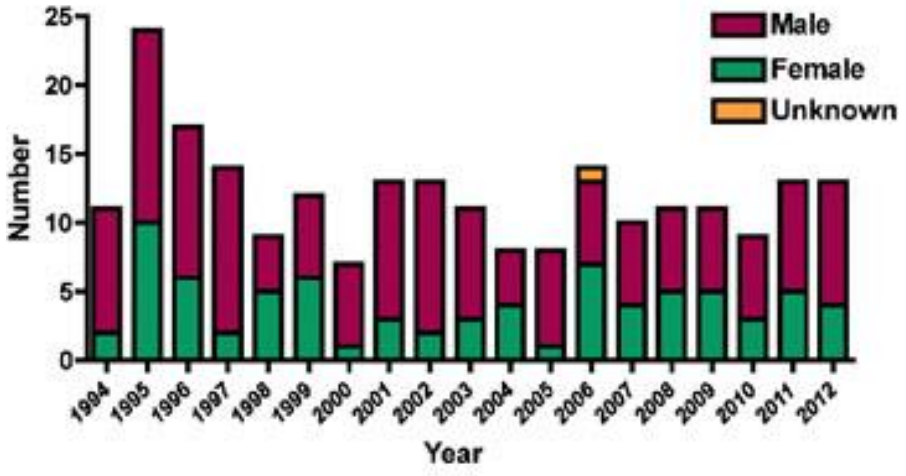
- How much is there?
- Who does it?
- How much does it cost?
- What to do about it?

Misconduct accounts for the majority of retracted scientific publications PNAS 109: 17028 (2012)

1. Ferric C. Fang[a],[b],[1],
2. R. Grant Steen[c],[1], and
3. Arturo Casadevall[d],[1],[2]

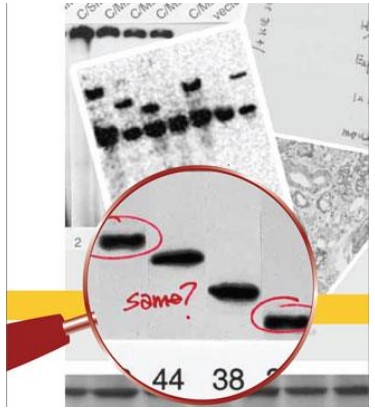


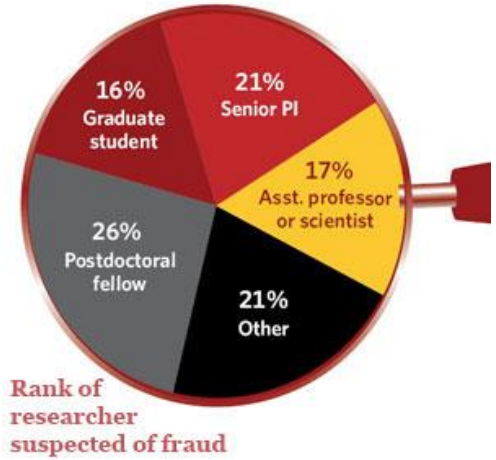
Men commit more misconduct than women
Williams, SCP *Biotechniques 1/23/2013*

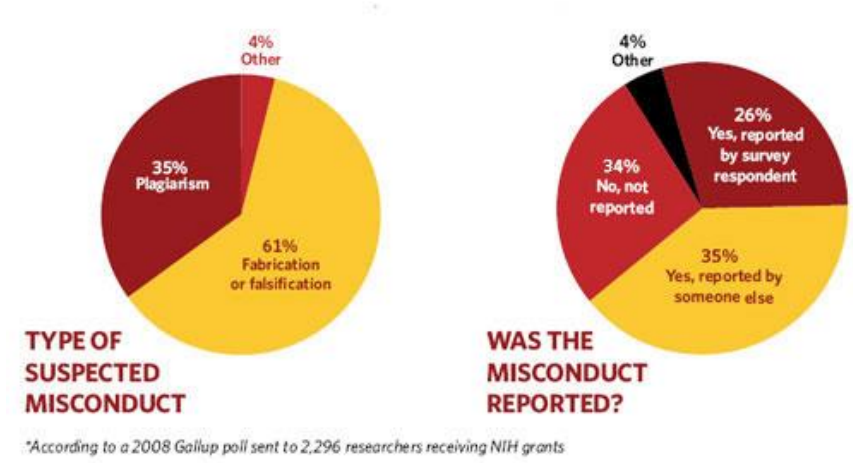


A Gawrylewski Fixing Fraud *The Scientist* **23: 67** (2009)

Images are the easiest to spot







Fanelli D (2009) How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data. **PLoS ONE 4(5):** e5738. doi:10.1371/journal.pone.0005738

"A pooled weighted average of 1.97% (N = 7, 95%CI: 0.86–4.45) of scientists admitted to have fabricated, falsified or modified data or results at least once –a serious form of misconduct by any standard– and up to 33.7% admitted other questionable research practices. In surveys asking about the behaviour of colleagues, admission rates were 14.12% (N = 12, 95% CI: 9.91–19.72) for falsification, and up to 72% for other questionable research practices. "

"…misconduct was reported more frequently by medical/pharmacological researchers than others."

The Costs of Research Misconduct

From the Ithenticate[R] website

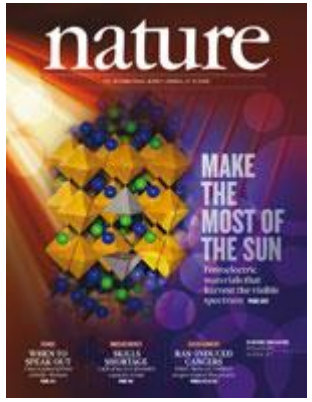
- 2002: 1.09m journal articles published annually
  2010: 1.94m journal articles published annually
- 7,000,000 researchers/ca 32,000 scholarly journals
- 23% of submissions to one leading scholarly journal rejected for plagiarism
- Types of damage
  - job losses, revoked PhDs and awards, damaged reputations, retractions
  - Est cost of single investigation in US $525,000
  - ca 71,000 patients treated in ca 900 retracted studies
  - $110,000,000 Total cost of investigations into research misconduct in US in 2010

Research ethics: 3 ways to blow the whistle

Reporting suspicions of scientific fraud is rarely easy, but some paths are more effective than others.

•Ed Yong,
•Heidi Ledford
•& Richard Van Noorden

27 November 2013



The Analytical

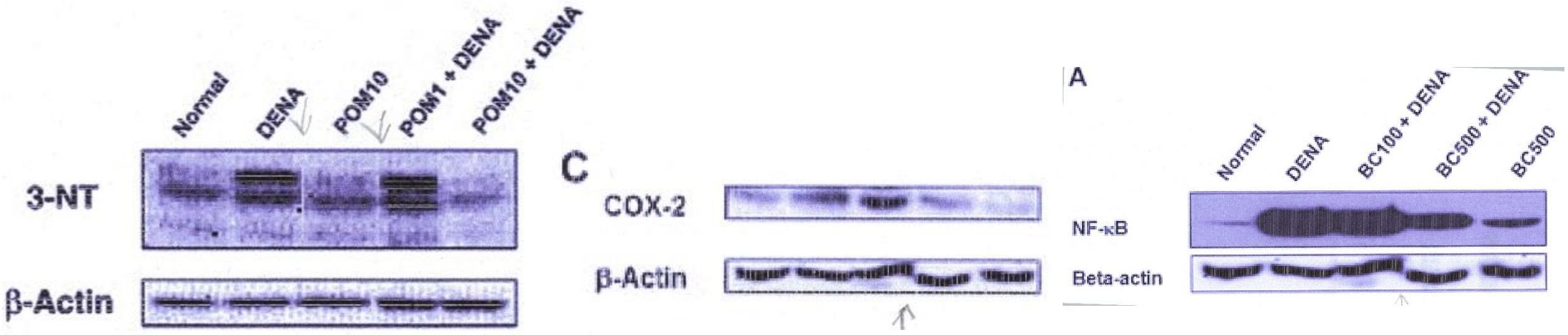
The Quixotic

The Anonymous

What to do about it?

Clare Francis: the mysterious anonymous whistleblower

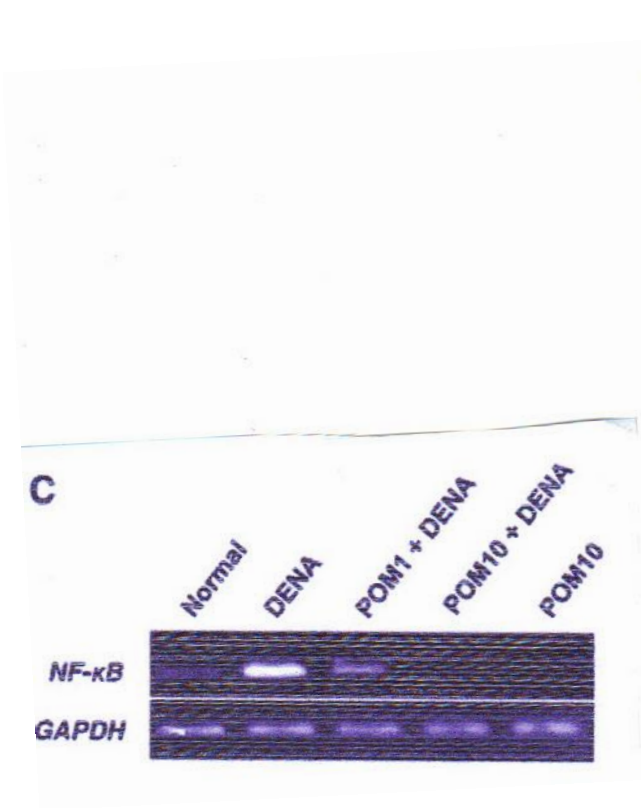
Image Manipulations



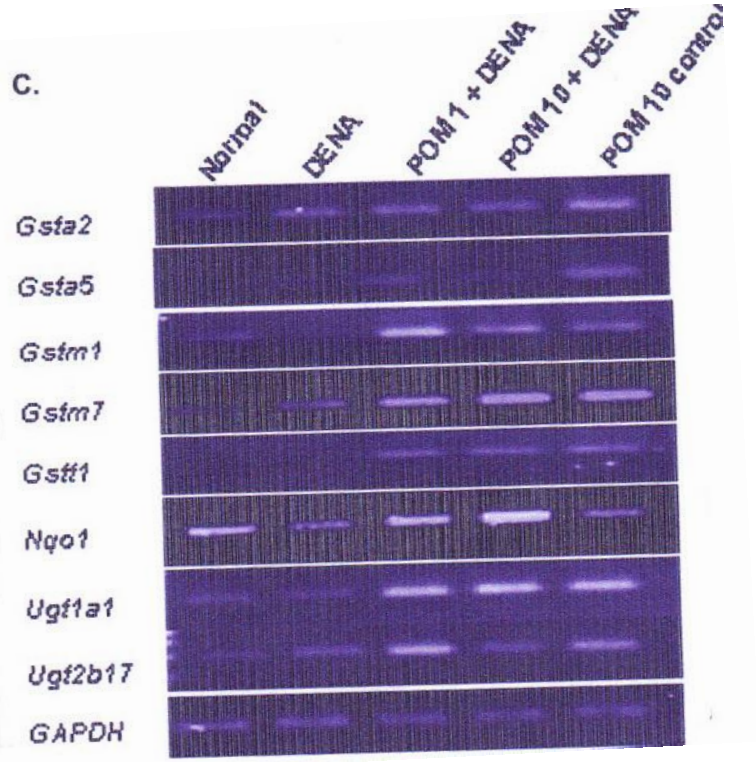
3NT: Sharp vertical lines between lanes 2/3 and 3/4, background change lane 4 versus lanes 3 and 5. Possible figure manipulation

Beta-actin: large vertical steps between bands in lanes 3 and 4 versus cox-2 and NF-κB: no vertical step between bands 3 and 4: unlikely these are from the same blot

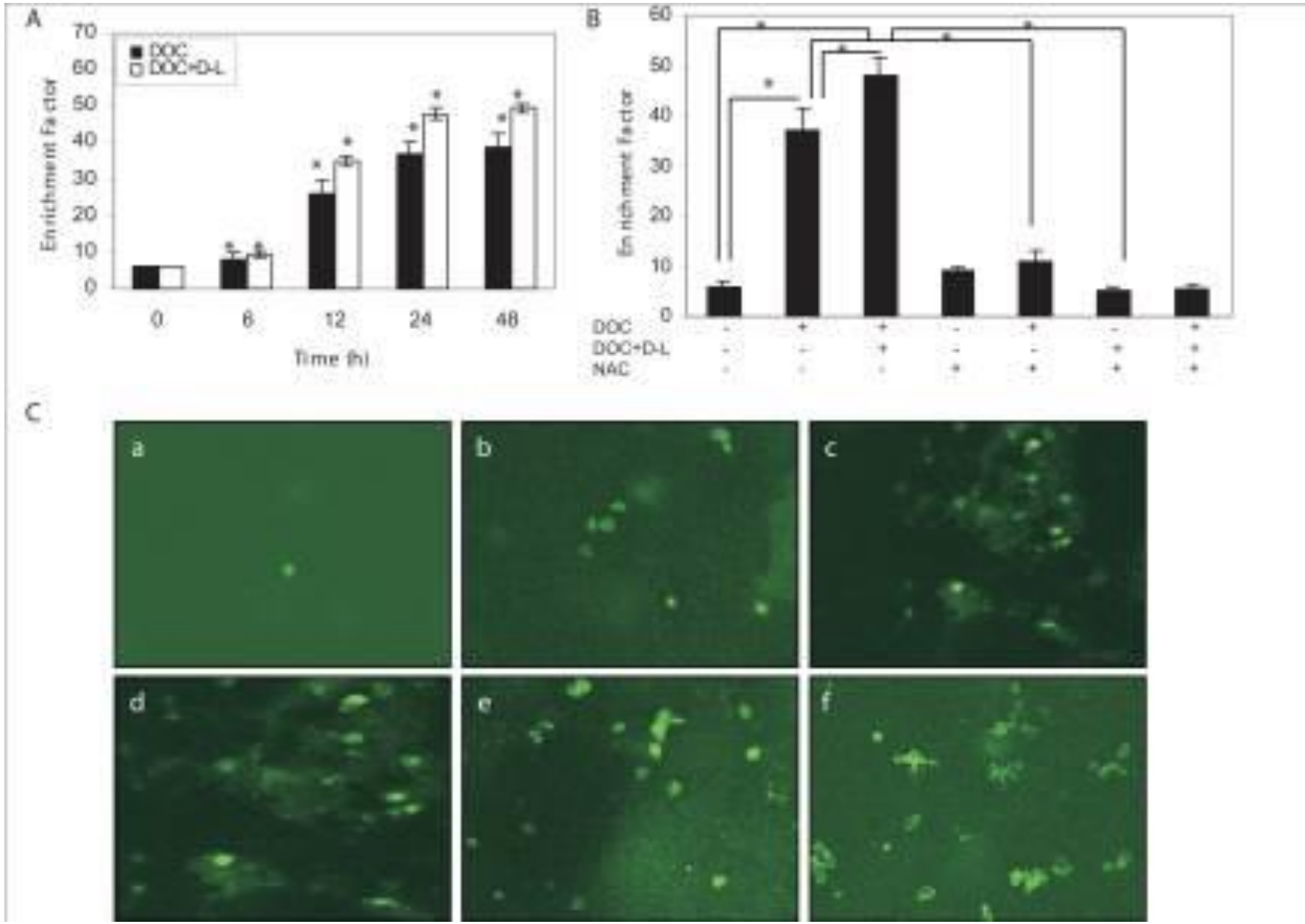
Data Reuse: same GAPDH in 2 different studies



J. Nutr Biochem
(2013) **24:** 178-187



Carcinogenesis (2011)
32: 888-896

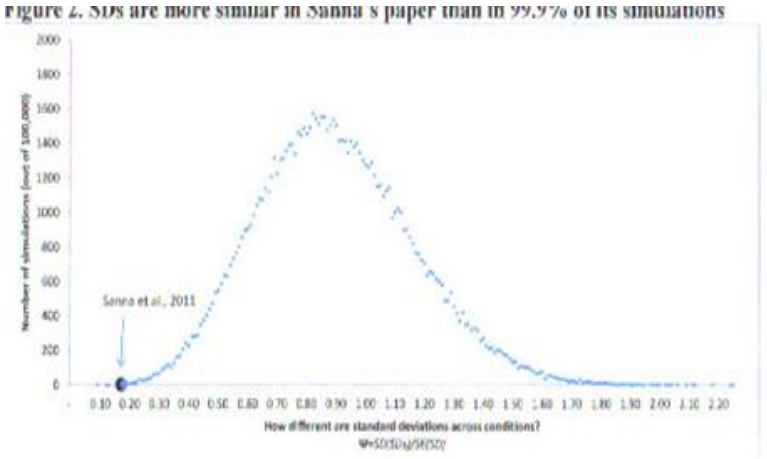

A
Enrichment Factor
DOC
DOC+D-L
Time (h)
0
6
12
24
48
B
Enrichment Factor
DOC
DOC+D-L
NAC
C
a
b
c
d
e
f

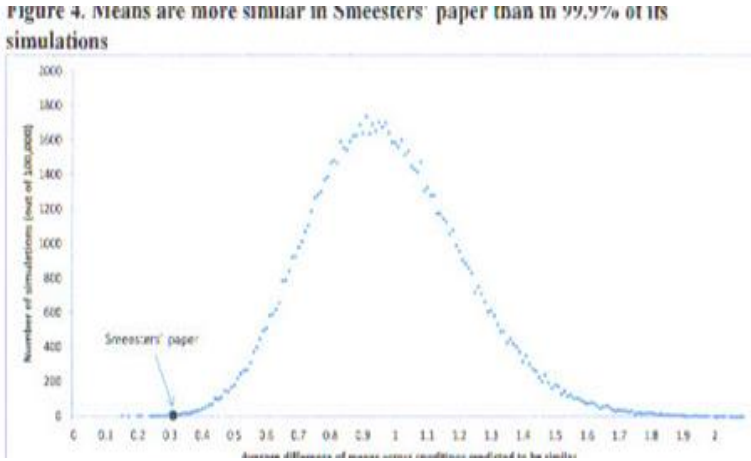
Statistical Sleuthing

Uri Simonsohn: the analytical whistleblower

Just post it: The lesson from two cases of fabricated data detected by statistics alone. Uri Simonsohn The Wharton School University of Pennsylvania uws@wharton.upenn.edu

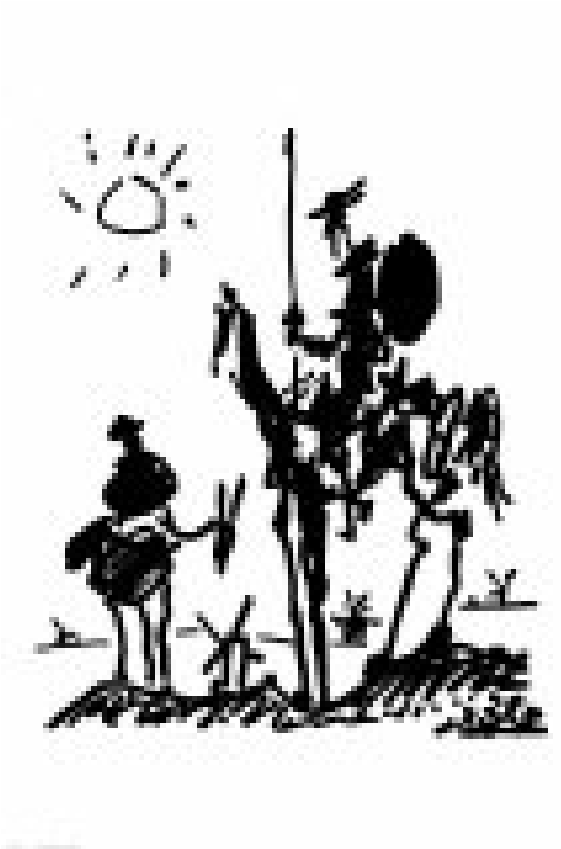
I argue **that requiring authors to post the raw data supporting their published results** has, among many other benefits, that of making fraud much less likely to go undetected. I illustrate this point by describing **two cases of fraud I identified exclusively through statistical analysis** of reported means and standard deviations. … If journals, granting agencies, universities or other entities overseeing research promoted or **required data posting, it seems inevitable that fraud would be reduced.**





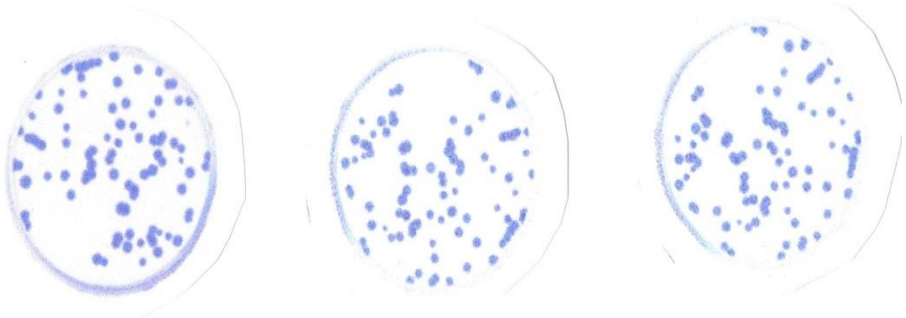
Statistical Sleuthing:

Helene Z Hill: the quixotic whistleblower and Joel Pitt = Sancho Panza



Data Sets:

Colony Counts in triplicate



Cell Counts (not necessarily in triplicate)

Coulter ZM

The Mid-Ratio

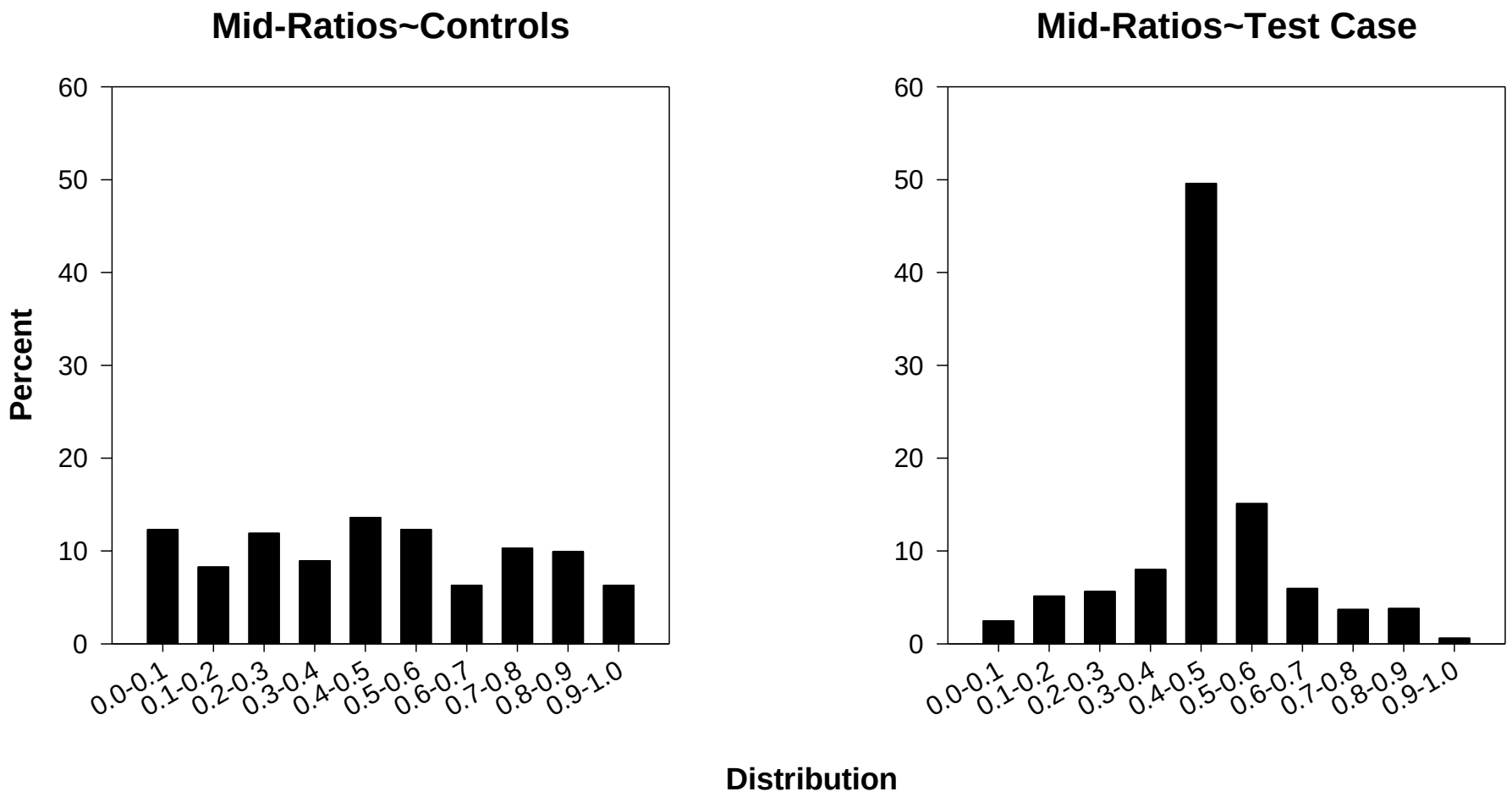
Mid-Ratio: Unusually high frequency of the rounded average as one of the triplicate sample counts

Expt. #: 2 Date: 02/22/99

Colony Counts and Survival Fraction

Tube.dilution	Colony 1	Colony 2	Colony 3	Avg Colony for x.2	SF	
1·2	130	149	142	140·33	—	
2·2	131	137	143	137·0	0·9762	
3·2	123	131	138	130·66	0·9311	
4·2	128	134	140	134	0·9548	
5·2	125	130	136	130·33	0·9287	uH
6·3	115	126	137	12·6	0·089	0
7·2	17	20	24	20·33	0·1484	20
8·2	29	35	41	35	0·2678	40
9·2	62	70	54	62	0·4626	80
10·2	70	79	62	70·33	0·5396	100

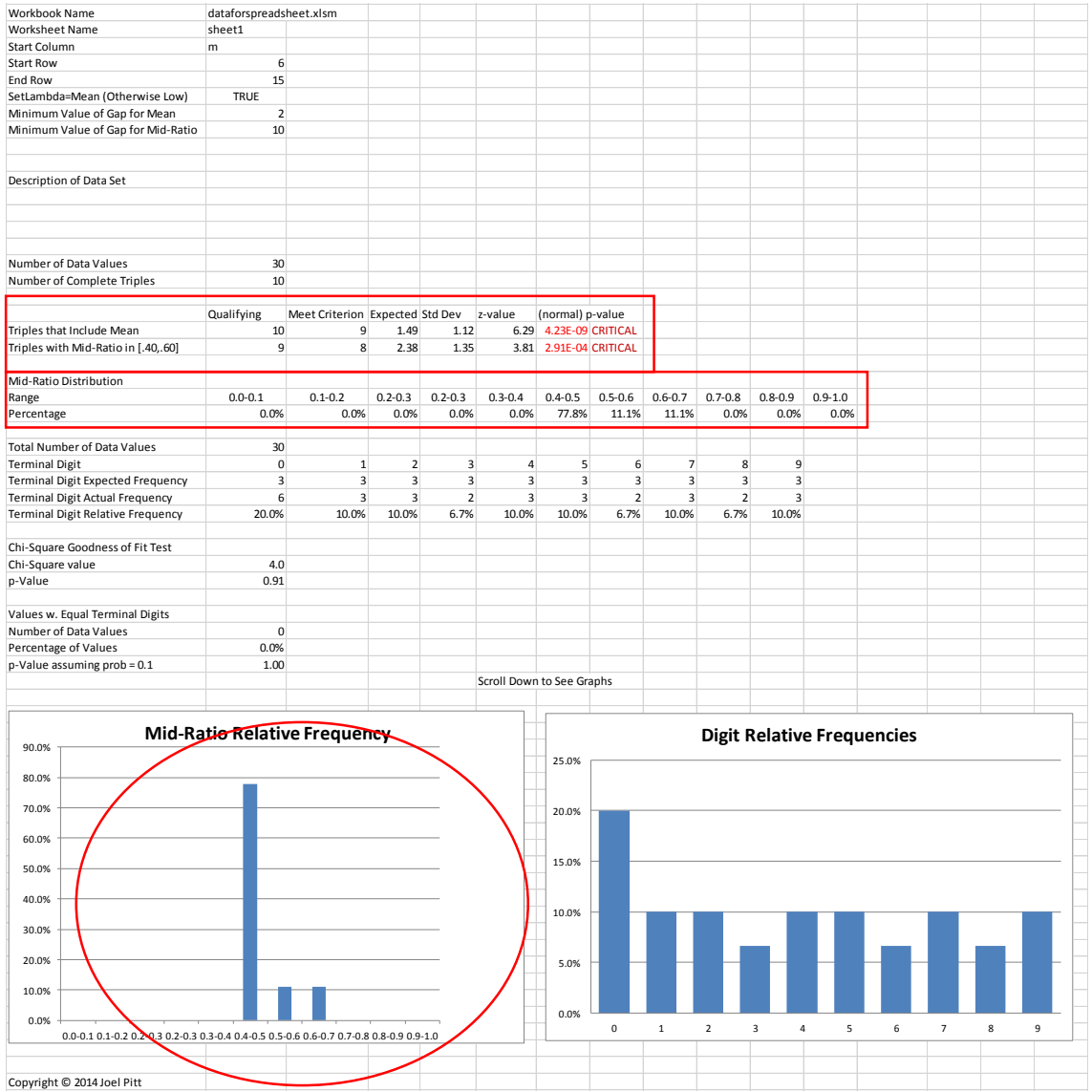
Mid-Ratio Distributions



Test Case Colony Counts

Mid Ratio: (mid-lo)/(hi-lo)

Sample #	T Triplicate Counts			Average	Mid-ratio (b-a)/(c-a)	C Triplicate Counts			Average	Mid-ratio (b-a)/(c-a)
1	130	149	142	140.3	**0.63**	92	111	119	107.3	**0.7**
2	131	137	143	137	**0.5**	78	85	74	79	**0.36**
3	123	131	138	130.6	**0.53**	142	126	120	129.3	**0.27**
4	128	134	140	134	**0.5**	120	129	121	123.3	**0.11**
5	125	130	136	130.3	**0.45**	64	68	79	70.3	**0.27**
6	115	126	137	126	**0.5**	92	101	78	90.3	**0.61**
7	17	20	24	20.3	**0.43**	74	62	94	76.7	**0.38**
8	29	35	41	35	**0.5**	89	69	67	75	**0.091**
9	62	70	54	62	**0.5**	85	87	97	89.7	**0.17**
10	70	79	62	70.3	**0.47**	71	58	55	61.3	**0.19**



Workbook Name	dataforspreadsheet.xlsm										
Worksheet Name	sheet1										
Start Column	m										
Start Row	6										
End Row	15										
SetLambda=Mean (Otherwise Low)	TRUE										
Minimum Value of Gap for Mean	2										
Minimum Value of Gap for Mid-Ratio	10										
Description of Data Set											
Number of Data Values	30										
Number of Complete Triples	10										
	Qualifying	Meet Criterion	Expected	Std Dev	z-value	(normal) p-value					
Triples that Include Mean	10	9	1.49	1.12	6.29	4.23E-09	CRITICAL				
Triples with Mid-Ratio in [.40,.60]	9	8	2.38	1.35	3.81	2.91E-04	CRITICAL				
Mid-Ratio Distribution											
Range	0.0-0.1	0.1-0.2	0.2-0.3	0.2-0.3	0.3-0.4	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1.0
Percentage	0.0%	0.0%	0.0%	0.0%	0.0%	77.8%	11.1%	11.1%	0.0%	0.0%	0.0%
Total Number of Data Values	30										
Terminal Digit	0	1	2	3	4	5	6	7	8	9	
Terminal Digit Expected Frequency	3	3	3	3	3	3	3	3	3	3	
Terminal Digit Actual Frequency	6	3	3	2	3	3	2	3	2	3	
Terminal Digit Relative Frequency	20.0%	10.0%	10.0%	6.7%	10.0%	10.0%	6.7%	10.0%	6.7%	10.0%	
Chi-Square Goodness of Fit Test											
Chi-Square value	4.0										
p-Value	0.91										
Values w. Equal Terminal Digits											
Number of Data Values	0										
Percentage of Values	0.0%										
p-Value assuming prob = 0.1	1.00										
Scroll Down to See Graphs											

Coulter Counts: Terminal Digits and Doubles

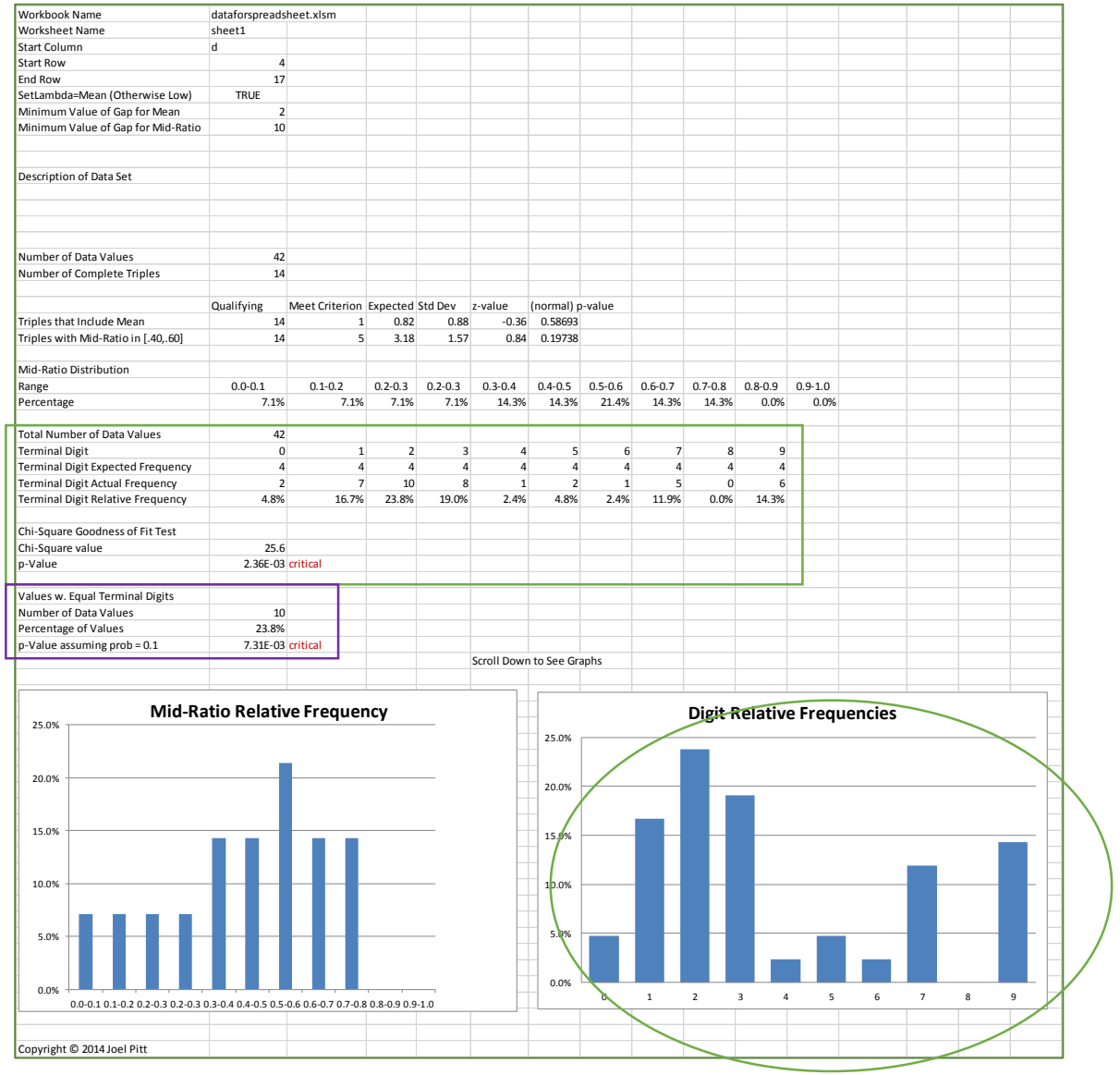
Sample #	T Triplicate Counts			C Triplicate Counts		
1	577	592	563	89	97	86
2	611	607	653	331	316	329
3	581	593	617	378	330	375
4	633	645	619	333	404	367
5	511	537	549	396	382	408
6	544	562	573	342	331	344
7	666	672	693	340	349	344
8	601	572	633	325	347	304
9	511	529	541	315	291	283
10	532	555	562	307	339	323
11	513	549	562	285	314	323
12	562	539	547	260	262	284
13	560	542	522	361	315	298
14	680	669	671	355	324	356

10 doubles p = 7.31×10^{-3} **4 doubles p= 0.616**

Term Digit	0	1	2	3	4	5	6	7	8	9	Total	Chi Sq	Chi sq p for uniform
T Freq	2	7	10	8	1	2	1	5	0	6	42	21.8	2.4×10^{-3}
C Freq	3	4	3	4	7	6	4	4	3	4	42	3.7	0.93
Uniform Freq	4.2	4.2	4.2	4.2	4.2	4.2	4.2	4.2	4.2	4.2	42		

Test Case Coulter Counts

577	592	563	89	97	86
611	607	653	331	316	329
581	593	617	378	330	375
633	645	619	333	404	367
511	537	549	396	382	408
544	562	573	342	331	344
666	672	693	340	349	344
601	572	633	325	347	304
511	529	541	315	291	283
532	555	562	307	339	323
513	549	562	285	314	323
562	539	547	260	262	284
560	542	522	361	315	298
680	669	671	355	324	356



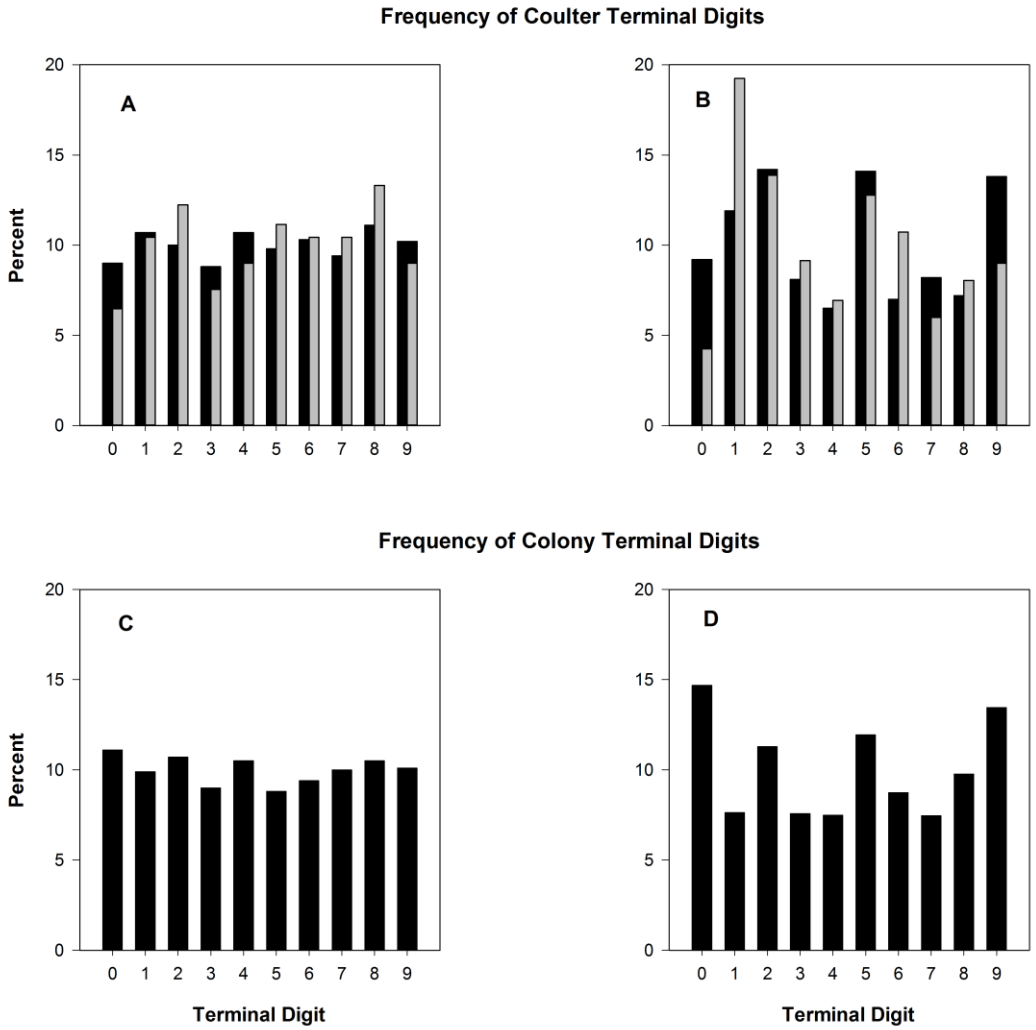
Workbook Name	dataforspreadsheet.xlsm										
Worksheet Name	sheet1										
Start Column	d										
Start Row	4										
End Row	17										
SetLambda=Mean (Otherwise Low)	TRUE										
Minimum Value of Gap for Mean	2										
Minimum Value of Gap for Mid-Ratio	10										
Description of Data Set											
Number of Data Values	42										
Number of Complete Triples	14										
	Qualifying	Meet Criterion	Expected	Std Dev	z-value	(normal) p-value					
Triples that Include Mean	14	1	0.82	0.88	-0.36	0.58693					
Triples with Mid-Ratio in [.40,.60]	14	5	3.18	1.57	0.84	0.19738					
Mid-Ratio Distribution											
Range	0.0-0.1	0.1-0.2	0.2-0.3	0.2-0.3	0.3-0.4	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1.0
Percentage	7.1%	7.1%	7.1%	7.1%	14.3%	14.3%	21.4%	14.3%	14.3%	0.0%	0.0%
Total Number of Data Values	42										
Terminal Digit	0	1	2	3	4	5	6	7	8	9	
Terminal Digit Expected Frequency	4	4	4	4	4	4	4	4	4	4	
Terminal Digit Actual Frequency	2	7	10	8	1	2	1	5	0	6	
Terminal Digit Relative Frequency	4.8%	16.7%	23.8%	19.0%	2.4%	4.8%	2.4%	11.9%	0.0%	14.3%	
Chi-Square Goodness of Fit Test											
Chi-Square value	25.6										
p-Value	2.36E-03	critical									
Values w. Equal Terminal Digits											
Number of Data Values	10										
Percentage of Values	23.8%										
p-Value assuming prob = 0.1	7.31E-03	critical									

Scroll Down to See Graphs

Terminal Digits and Doubles

Others

Test Case



What's To Do:
Retraction Watch

[New post] Science retracts two papers for image manipulation mailbox:///C:/Users/Lanie/AppData/Roaming/Thunderbird/P...

Subject: [New post] Science retracts two papers for image manipulation
From: Retraction Watch <comment-reply@wordpress.com>
Date: 5/29/2014 6:22 PM
To: hzhill@verizon.net

Respond to this post by replying above this line

New post on **Retraction Watch**

Science retracts two papers for image manipulation
by ivanoransky

Science has retracted two papers by Frank Sauer, of the University of California, Riverside, after the university found evidence of serious image manipulation. Here's the notice, signed by Science editor-in-chief Marcia McNutt:

Read more of this post

ivanoransky | May 29, 2014 at 6:22 pm | Categories: freely available, germany retractions, image manipulation, science (journal) retractions, united states | URL: http://wp.me/pYKlt-5nN

Comment See all comments Like

Unsubscribe to no longer receive posts from Retraction Watch.
Change your email settings at Manage Subscriptions.

Trouble clicking? Copy and paste this URL into your browser:
http://retractionwatch.com/2014/05/29/science-retracts-two-papers-for-image-manipulation/

Thanks for flying with WordPress.com

1 of 1 6/11/2014 2:41 PM

Uri Simonsohn's Blog

Data Colada

Thinking about evidence and vice versa

One size won't fit all: Each case will require its own set of analyses

Science Fraud

Highlighting Misconduct in Life Sciences Research

Main menu

- About This Site

Nothing Found

Apologies, but no results were found for the requested archive. Perhaps searching will help find a related post.

Search

Search

Tag Cloud

Search

Recent Comments

RSS

- RSS - Posts

Proudly powered by WordPress

Paul Brookes

497 papers for which data integrity had been questioned. … 70 (14%) were subjected to some type of corrective action.

Replacement Blog (under construction)

coming soon

Integritywatchforscienceandmedicine.com

Goal: to reduce scientific misconduct by providing a site where suspicious findings can be gathered, aired and discussed

- A replacement for Science Fraud where people can post questions regarding specific image manipulations and data anomalies
- A site to direct whistleblowers to resources for data analysis
- A site for posting methods of data analysis
- A place to talk about the latest news relative to scientific integrity and to report the latest from *PubPeer*

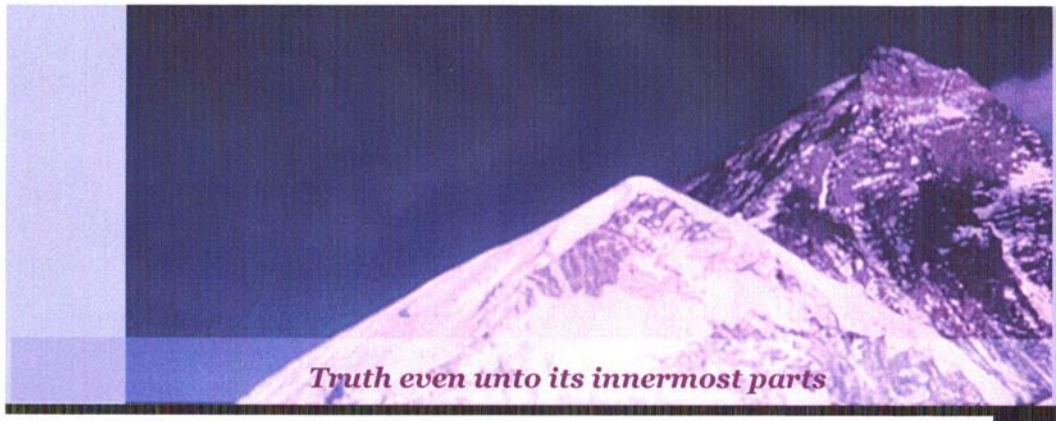
The Obligations for Journals

- Run every submission through plagiarism testing
- Require that complete images for gels be submitted for review
- All raw data must be posted and publically accessible
- Don’t be afraid of lawsuits ~ the truth is the best defense

www.helenezhill.com

Have Comments?
Email Me

Helene Z. Hill, Ph.D.



Home | My Family | My Career | Committees | Qui Tam | Publishing | Raw Data | Envoi | Blog

This Website Exposes a Scientific and Medical Cover Up

This website represents my own personal views as a scientist, not as a Rutgers/NJMS faculty member. Please visit our most recent preprint: Pitt, JH and Hill, HZ. Statistical Detection of Potentially Fabricated Data. Posted on arXiv: 1311.5517, and on Figshare: http://dx.doi.org/10.6084/m9.figshare.858921

I made it in *Nature*!! http://www.nature.com/news/research-ethics-3-ways-to-blow-the-whistle-1.14226!! 11/27/13.
And I published an Opinion in *The Scientist:* http://www.the-scientist.com/?articles.view/articleNo/39139/title/Opinion--Reducing-Whistleblower-Risk/

Welcome to my WebPage. Here you will find the story of a woman's struggle to make her mark in the modern scientific world in the face of uncounted obstacles. Read on (if you want to know more, click on the links -- underlined text in red):

In October of 2003, I filed a case for *qui tam* in the Federal District Court of Newark charging my