DETECTING DATA FABRICATION

CHRIS HARTGERINK
WITH
JAN VOELKEL
JELTE WICHERTS
MARCEL VAN ASSEN

HTTP://BIT.LY/WCRI20I7
WORKING MANUSCRIPT
HTTP://BIT.LY/WCRI20I7-MANUSCRIPT



General Article



Just Post It: The Lesson From Two Cases of Fabricated Data Detected by Statistics Alone

Populación (se la tenuar 2 de 18) 20 % 1 de 18 d

Uri Simonsohn Tre Whatton School, University of Perceybania

Abstract

I argue that requiring authors to post the raw data supporting their published results has the benefit, among many others, of making fraud much less likely to go undetected. I illustrate this point by describing two cross of suspected fraud I identified excisively through statistical analysis of reported means and standard deviations. Analyses of the raw data behind these published results provided invaluable confirmation of the fittial suspitions, ruling out benign explanations (e.g., reporting errors, unusual distributions), identifying additional signs of fabrication, and also ruling out one of the suspected fraud's explanations for his anomabus results. If journals, granting agencies, universities, or other entities overseeing research promoted or required data posting, it seems inevitable that fraud would be reduced.

Keywords

judgment, decision making, scientific communication, fake data, data sharing, data posting

Received7/20/12, Revision accepted1/90/13

Academic misconduct is a rare event just not rare emough. Its occurrence challenges the credibility of research, and the mission of science more generally. Although prevention is important, some misconduct is likely to occur no matter what steps are taken to preventit. Measures that facilitate identifying such cases can help mitigate their negative consequences. Furthermore, the risk of detection may constitute the ultimate determent.

To undetectably fabricate data is difficult. It requires both (a) a good understanding of the phenomenon being studied (e.g., what measures of a construct tend to bok like, which variables they conclute with and by how much) and (b) a good understanding of how sampling error is expected to influence the data (e.g., how much variation and the kind of variation the estimates of interest should exhibit given the observed sample size and design), in this article, if show hat although means and standard deviations can be analyzed in light of these two criteria to identify likely cases of fraud, the availability of raw data makes the task of detection caster and more diagnostic, and hence that of fabrication more difficult and intimidating.

Posting data has many advantages unrelated to, and possibly more variable than, prevention and date then of fauld. For example, as Wicherts and Bakker (2012) have noted, when raw data are posted, scientific evidence is preserved for longer periods of time, more researchers git to analyze and hence learn from a given amount of stentific evidence, and reporting errors become easier to prevent and detect.

In this article ,I illustrate how now data can be analyzed for identifying likely fraud through two case studies. Each began with the observation that summary statistics reported in a published article were too similar across conditions to have originated in random samples, an approach to identifying problematic data that has been employed before (Carlisle, 2012, Risher, 1936, Gaffan & Gaffan, 1992, Kalas, McKey, & Ber-Hillel, 1998, Roberts, 1987, Samberg & Roberts, 2006.13 These pre-liminary analyses of excessive similarity motivated me to omtact the authors and request the raw data behind their results. Only when the raw data were analyzed did these asspicions rise to a level of confidence that could trigger

Comment and Grant Comments of Persons Action Comments of Persons Action Action School, Christophy of Persons Action Action Fall, Philadelphia PA 19104
Final Action Comments of the Persons Action Comments of the Person

Flawed science:

The fraudulent research practices of social psychologist Diederik Stapel

Levelt Committee

Noort Committee

Drenth Committee

This document is an English translation of the Dutch report 'Falende wetenschap: De frauduleuce onderzoekspraktijken van social-psycholoog Diederik Stapel'. In the event of any differences between the Dutch report and the translation, the Dutch report will prevail.

28 november 2012





Just Post It: The Lesson From Two Cases of Fabricated Data Detected by Statistics Alone

Proplantigate i for terror
2 (4 (8) 2019-1980
C The Author (s) 2013
Expression and permissions
suppredictating from the formation and
100 to 117 (40)-1019 for 100 (40)
propagage infection
(SSEACE)

Uri Simonsohn Tre Whatton School, University of Pennsylvania Flawed science:

The fraudulent research practices of social psychologist Diederik Stapel

Abstract I argue that others, of a fraud I idea raw data by

explanation out one of other entitis

Keywords judgment, o METHODS DEVELOPED IN CASU

APPLICABLE AS GENERIC METHODS?

Academic n

Its occurrence challenges the credibility of research, and the mission of science more generally. Although prevention is important, some misconduct is likely to occur nomatter what steps are taken to preventit. Measures that facilitate identifying such cases can help mitigate their negative consequences. Furthermore, the risk of detection may constitute the ultimate determent.

To undetectably fabricate data is difficult. It requires both (a) a good understanding of the phenomenon being studied (e.g., what measures of a construct tend to bok like, which variables they conclute with and by how much) and (b) a good understanding of how sampling error is expected to influence the data (e.g., how much variation and the kind of variation the estimates of interest should exhibit given the observed sample size and design), in this article, if show hat although means and standard deviations can be analyzed in light of these two criteria to identify likely cases of fraud, the availability of raw data makes the task of detection caster and more diagnostic, and hence that of fabrication more difficult and intimidating.

Posting data has many advantages unrelated to, and possibly more variable than, prevention and date then of noted, when raw data are posted, scientific evidence is preserved for longer periods of time, more researchers git to analyse and hence learn from a given amount of scientific evidence, and reporting errors become caster to prevent and detect.

In this article ,I illustrate how new data can be analyzed for identifying likely fraud through two case studies. Each began with the observation that summary tatistics reported in a published article were too similar across conditions to have originated in random samples, an approach to dentifying problematic data that has been employed before (Carlisle, 2012, Fisher, 1936, Carlian & Goffan, 1992, Kalas, McKay, & Ber-Hild, 1998, Boberts, 1987, Sembarg & Roberts, 2006). These per-liminary analyses of excessive similarity motivated me to omtact the authors and request the raw data behind their resuls. Only when the raw data were analyzed did those suspicions rise to a level of coeffidence that could trigger.

Converp ending Authors
US Simonwohn, The Wharen School, University of Pennsylvania, 3730
Walnus Sr., 500 Hansenun Hall, Philadelphia PA 19104
Gravit unselbehanterungenn, edu

This document is an English translation of the Dutch report 'Falende wetenschap: De frauduleuce onderzoekspraktijken van social-psycholoog Diederik Stapel'. In the event of any differences between the Dutch report and the translation, the Dutch report will prevail.

28 november 2012



	GENUINE	FABRICATED
"GENUINE"		
"FABRICATED"		



	GENUINE	FABRICATED
"GENUINE"	??	
"FABRICATED"		



HOW CAN DATA BE FABRICATED?

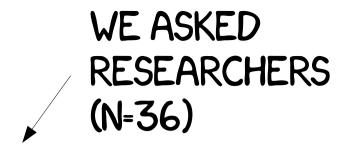


MANY LABS (N=36)

	GENUINE	FABRICATED
"GENUINE"		
"FABRICATED"	; ?	



MANY LABS (N=36)



	GENUINE	FABRICATED
"GENUINE"	??	
"FABRICATED"	; ?	



YES, WE ASKED RESEARCHERS TO FABRICATE DATA



YES, WE ASKED RESEARCHERS TO FABRICATE DATA

AND WE EVEN PAID THEM



YES, WE ASKED RESEARCHERS TO FABRICATE DATA

AND WE EVEN PAID THEM

AND WE PAID THEM EVEN MORE IF THEY WERE TOP 3



			Mean (true distance: 2,906.5 miles)	Standard Deviation
Low anchor	The distance from San Francisco to New York City is longer than 1,500 miles. How far do you think it is?	Female	2562.12	956.35
Low anchor	longer than 1,500 miles. How far do you think it is?		2540.36	942.14
High anghar	The distance from San Francisco to New York City is	Female	3421.25	845.21
High anchor	shorter than 6,000 miles. How far do you think it is?		3380.98	932.56



Anchoring study - distance from San Francisco to New York

Expectations			Current result	Supported
Main effect o	f condition	F(1, 96) = 21.33, p < .001		✓
No main effe	ct of gender		F(1, 96) = 0.03, p = 0.867	✓
No interaction effect of gender * condition			F(1, 96) = 0, p = 0.96	✓
			Mean (true distance: 2,906.5 miles)	Standard Deviation
Low anchor	The distance from San Francisco to New York City is longer than 1,500 miles. How far do you think it is?	Female	2562.12	956.35
LOW afficitor	longer than 1,500 miles. How far do you think it is?		2540.36	942.14
High anchor	The distance from San Francisco to New York City is shorter than 6,000 miles. How far do you think it is?	Female	3421.25	845.21

FABRICATION HYPOTHESES

- I. SIGNIFICANT CONDITION EFFECT
- 2. NONSIGNIFICANT GENDER EFFECT
- 3. NONSIGNIFICANT INTERACTION EFFECT



	Ancho	oring study - distar	nce from San Francisco to N	lew York				
	Expectations		Current r	esult	Supported			
Main effect o	f condition		F(1, 96) = 21.3	3, p < .001	/			
No main effe	ct of	Ancho	oring study - distance from	San Francisco to New Yo	rk		_	
No interactio	n effe							
	E	xpectations		Current result	Supp	orted	_	
	Main effect of condition			F(1, 96) = 21.33, p <	.001	/		
Low anchor	TheNo main effect of gender			F(1.96) = 0.03. n = 0		/		
LOVY GITCHOI	lon _i No interaction effect of gend	er * c		Anchoring study - dista	nce from San Francisco to N	ew York		
High anchor	The							
	shoı		Expectations	5	Current result		Supported	
	Low anchor The distance f	rom SMain effect of	of condition		F(1, 96) = 21.33, p < .001		- ' ' ' ' ' '	
	ionger than 1,	500 nNo main effe		Anchoring study - distance from San Francisco to New York				
	High anchor	rom SNo interactio	on effect of gend	From a state			Current result	Commanded
	shorter than 6	,000 r	Main offers	Expectations Main effect of condition				Supported
							F(1, 96) = 21.33, p < .001 F(1, 96) = 0.03, p = 0.867	v
		Low anchor	longer than 1.No interest	e fNo main effect of gender ¹ .No interaction effect of gender * condition			F(1, 96) = 0.03, p = 0.067	· · · · · · · · · · · · · · · · · · ·
			The distance f	lion enect of gender to	iluluoli		Γ(1, 70) - 0, μ - 0.70	•
		High anchor	shorter than 6				Mean (true distance: 2,906.5 miles)	Standard Deviation
		The distance from Can F		n Francisco to New York City	is Female		956.35	
			Low ancho	longer than 1,500 mi	les. How far do you think it i		2540.36	942.14
				The distance from Sa	n Francisco to New York City	is Female	3421.25	845.21
			High ancho	shorter than 6,000 m	iles. How far do you think it		3380.98	932.56

		Anchoring	study - distan	ce from San Francisco to Nev	w York		_			
	Expe	ectations		Current res	sult	Supported	-			
Main effect o	f condition			F(1, 96) = 21.33,	p < .001	1				
No main effe	ct of		Ancho	ring study - distance from Sa	an Francisco to New York	(
No interactio	n effe									
		Expec	tations		Current result		Supporte	ed		
	Main effect o	f condition			F(1, 96) = 21.33, p < .0	01	1			
	TheNo main effe	ct of gender			F(1.96) = 0.03. p = 0.8	67	/			
Low anchor	IoniNo interactio	n effect of gender *	С		Anchoring study - distan	ce from San Franci	sco to New \	⁄ork		ı
High angles	The									
High anchor	shor			Expectations		C	urrent resul	t	Supported	
		The distance from	_S Main effect o	f condition	F(1, 96) = 21.33, p) = 21.33, p	< .001	✓	l ,
	Low anchor	longer than 1,500 r	nNo main effe	t of gender		Anchoring study - distanc		from San Fra	ncisco to New York	
		The distance from	SNo interaction	n effect of gend						
	High anchor	shorter than 6,000	r		Expectations				Current result	Supported
				Main effect o	of condition			F(1	, 96) = 21.33, p < .001	✓
			Low anchor	The distance f No main effe	ain effect of gender			F(1, 96) = 0.03, p = 0.867		✓
			LOW arichor	longer than 1, No interactio	No interaction effect of gender * condition		F(1, 96) = 0, p = 0.96		✓	
			High anchor	The distance f						
shorter than 6		shorter than 6				Mea	n (true distance: 2,906.5 miles)	Standard Deviation		
			Low anchor	The distance from San	Francisco to New \	ork City is	Female	2562.12	956.35	
		Low anchor	The distance from San longer than 1,500 mile	s. How far do you	think it is?	Male	2540.36	942.14		
				High angles	The distance from San	Francisco to New \	ork City is	Female	3421.25	845.21
				High anchor	shorter than 6,000 miles. How far do you think it is?			Male	3380.98	932.56

WANT TO TRY IT YOURSELF? HTTP://BIT.LY/TRY-FABRICATION



DATA

EACH OF FOUR STUDIES PROVIDES

- I. FOUR VARIANCES
- 2. TWO NONSIGNIFICANT RESULTS
- 3. ONE SIGNIFICANT RESULT



DATA

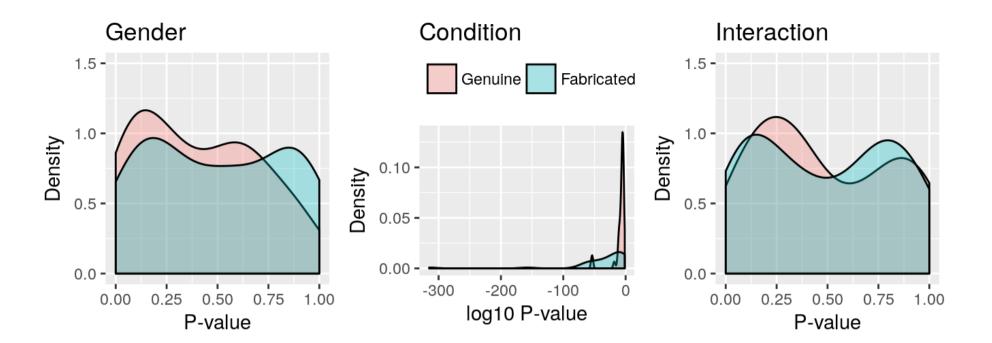
EACH OF FOUR STUDIES PROVIDES

- I. FOUR VARIANCES
- 2. TWO NONSIGNIFICANT RESULTS
- 3. ONE SIGNIFICANT RESULT

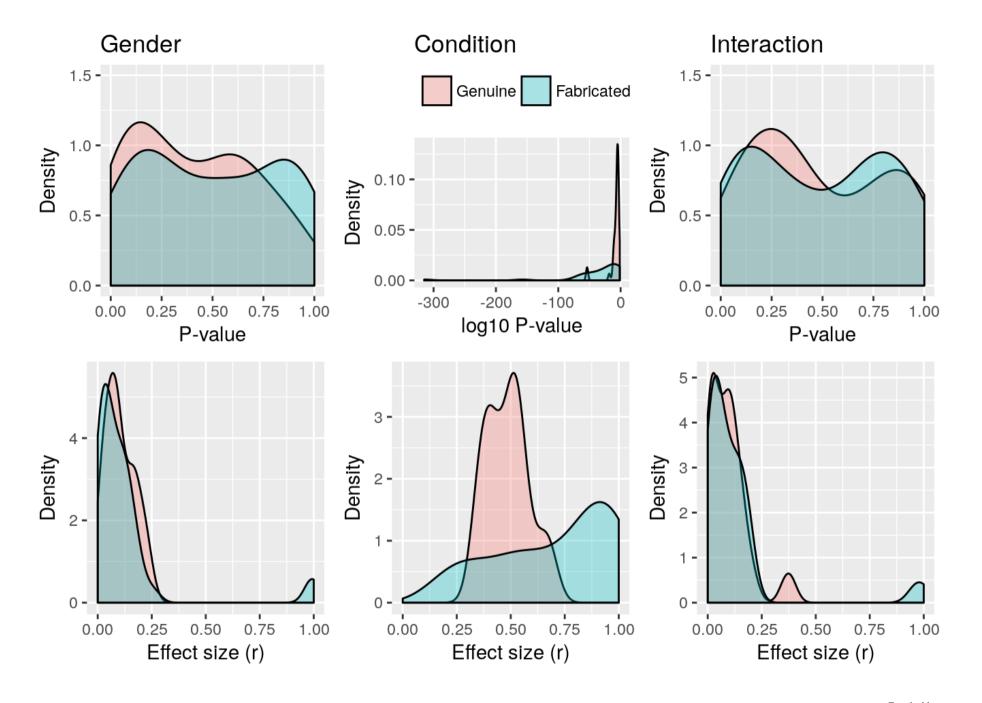
PER RESPONDENT:

- I. SIXTEEN VARIANCES
- 2. EIGHT NONSIGNIFICANT RESULTS
- 3. FOUR SIGNIFICANT RESULTS











METHOD	AUROC



METHOD	AUROC
VARIANCE ANALYSIS, ASSUMES = VARIANCE ACROSS CONDITIONS	.423
VARIANCE ANALYSIS, ASSUMES != VARIANCE ACROSS CONDITIONS	.770



METHOD	AUROC
VARIANCE ANALYSIS, ASSUMES = VARIANCE ACROSS CONDITIONS	.423
VARIANCE ANALYSIS, ASSUMES != VARIANCE ACROSS CONDITIONS	.770

$$z_j^2 \sim \left(rac{\chi_{N_j-1}^2}{N_j-1}
ight)/MS_w$$

$$MS_w = rac{\sum\limits_{j=1}^k (N_j-1) s_j^2}{\sum\limits_{j=1}^k (N_j-1)}$$



METHOD	AUROC
VARIANCE ANALYSIS, ASSUMES = VARIANCE ACROSS CONDITIONS	.423
VARIANCE ANALYSIS, ASSUMES != VARIANCE ACROSS CONDITIONS	.770
NONSIGNIFICANT P-VALUES GENDER EFFECT	.521
NONSIGNIFICANT P-VALUES INTERACTION EFFECT	.535



METHOD	AUROC
VARIANCE ANALYSIS, ASSUMES = VARIANCE ACROSS CONDITIONS	.423
VARIANCE ANALYSIS, ASSUMES != VARIANCE ACROSS CONDITIONS	.770
NONSIGNIFICANT P-VALUES GENDER EFFECT	.521
NONSIGNIFICANT P-VALUES INTERACTION EFFECT	.535

$$\chi^2_{2k} = -2\sum_{i=1}^k \ln(1-rac{p_i-t}{1-t})$$



METHOD	AUROC
VARIANCE ANALYSIS, ASSUMES = VARIANCE ACROSS CONDITIONS	.423
VARIANCE ANALYSIS, ASSUMES != VARIANCE ACROSS CONDITIONS	.770
NONSIGNIFICANT P-VALUES GENDER EFFECT	.521
NONSIGNIFICANT P-VALUES INTERACTION EFFECT	.535
EFFECT SIZE (I- CORRELATION)	.744



BUT ASSUMES 50/50 PREVALENCE



BUT ASSUMES 50/50 PREVALENCE

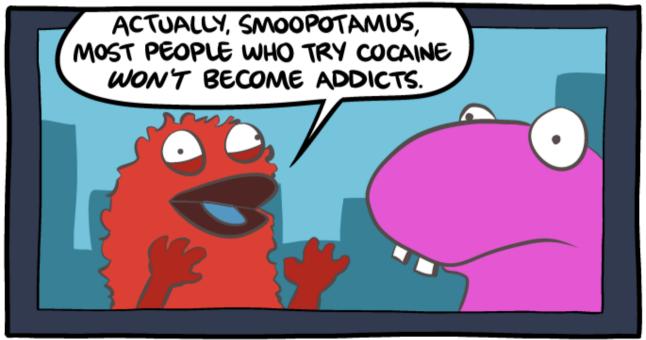
SO ONLY USE AS FLAG, NOT EVIDENCE OF



DUAL USE PROBLEM







smbc-comics.com



LESSONS

- I. FABRICATED NONSIGNIFICANT DATA LOOK RATHER GENUINE 2. VARIANCE ANALYSIS SENSITIVE TO POPULATION VARIANCES BEING HETERO- OR HOMOGENEOUS (IMPORTANT IN APPLICATION)
- 3. LARGE EFFECT SIZES EASY INDICATOR (R>.7)
- 4. USE AS FLAG, NOT EVIDENCE



THIS WAS JUST STUDY ONE



THIS WAS JUST STUDY ONE

STUDY TWO REPLICATES WITH RAW DATA



THIS WAS JUST STUDY ONE

STUDY TWO REPLICATES WITH RAW DATA

AND ADDS INTERVIEWS!



ONGOING, BUT YOU CAN FIND INCOMING TRANSCRIPTS @ HTTP://BIT.LY/WCRI2017-TRANSCRIPTS



ONGOING, BUT YOU CAN FIND INCOMING TRANSCRIPTS @ HTTP://BIT.LY/WCRI2017-TRANSCRIPTS

FEEL FREE TO USE FOR RESEARCH! (CC 0)

