How do researchers fabricate data and how to detect fabrication?

Chris Hartgerink
@chartgerink
www.chjh.nl
May 29, 2016

CENTER FOR OPEN SCIENCE
SHUTTLEWORTH FUNDED
TILBURG UNIVERSITY

"I had a student who fabricated the results of his MA study, but he simply copy-pasted subjects"

-Anonymous

CENTER FOR OPEN SCIENCE
SHUTTLEWORTH FUNDED
TILBURG UNIVERSITY

“I had a student who fabricated the results of his MA study, but he simply copy-pasted subjects”

“this did not require a genius to find out.”

-Anonymous

CENTER FOR OPEN SCIENCE
SHUTTLEWORTH FUNDED
TILBURG UNIVERSITY

TABLE 1

Mean Number of Fragments Completed as Category-Related Words, as a Function of Prime Emotion and Exposure Duration: Study 1

Exposure duration and fragment type	Prime emotion			
	Disgust	Fear	Anger	Neutral
Quick (120 ms)				
Disgust fragments	2.33_a (0.62)	1.20_b (0.94)	1.20_b (0.68)	1.53_b (0.74)
Fear fragments	0.80_b (0.78)	1.87_a (0.92)	1.13_b (0.92)	1.00_b (0.93)
Anger fragments	0.93_b (0.70)	0.93_b (0.70)	1.80_a (0.86)	0.80_b (0.78)
Negative fragments	2.27_a (0.46)	2.33_a (0.82)	2.20_a (0.41)	1.33_b (0.98)
Super-quick (40 ms)				
Disgust fragments	1.27_a (0.96)	1.07_a (0.80)	1.27_a (0.96)	1.33_a (0.72)
Fear fragments	1.07_a (0.59)	0.87_a (0.74)	1.07_a (0.59)	1.00_a (0.66)
Anger fragments	0.87_a (0.74)	1.07_a (0.80)	0.87_a (0.74)	0.87_a (0.83)
Negative fragments	1.80_a (0.56)	2.07_a (0.80)	2.27_a (0.46)	0.93_b (0.88)

Note. Standard deviations are given in parentheses. For each fragment type, the score could range from 0 to 3. Within each row, means with different subscripts differ significantly from each other ($p < .05$).

Conclusions

- Applied statistical tools to 36 genuine datasets, 36 fabricated datasets
- How do they perform in detecting fabrication?
- PPV = 91%, NPV = 59%, detected 10/36

CENTER FOR OPEN SCIENCE
SHUTTLEWORTH FUNDED
TILBURG UNIVERSITY

How can we detect data fabrication?*

CENTER FOR OPEN SCIENCE

SHUTTLEWORTH FUNDED

TILBURG UNIVERSITY

How can we detect data fabrication?*

**today: summary results*

CENTER FOR OPEN SCIENCE
SHUTTLEWORTH FUNDED
TILBURG UNIVERSITY

1. Variance of variance analysis

CENTER FOR OPEN SCIENCE
SHUTTLEWORTH FUNDED
TILBURG UNIVERSITY

1. Variance of variance analysis

$$\tilde{s}_j^2 \sim \left[\frac{\chi^2_{N_j-1}}{N_j-1} \right] / MS_w$$

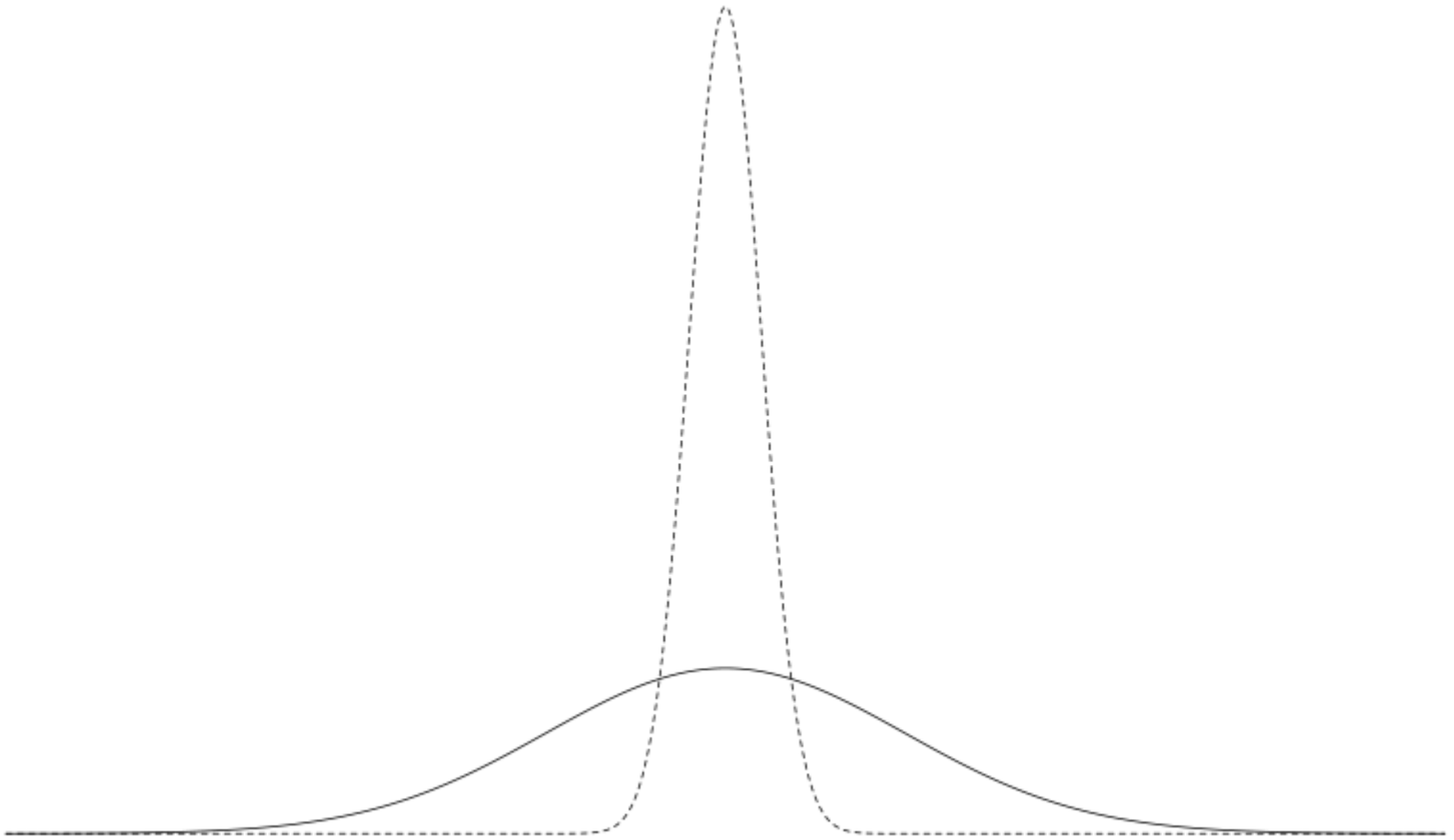
CENTER FOR OPEN SCIENCE
SHUTTLEWORTH FUNDED
TILBURG UNIVERSITY

What do (potentially) fabricated variances look like?

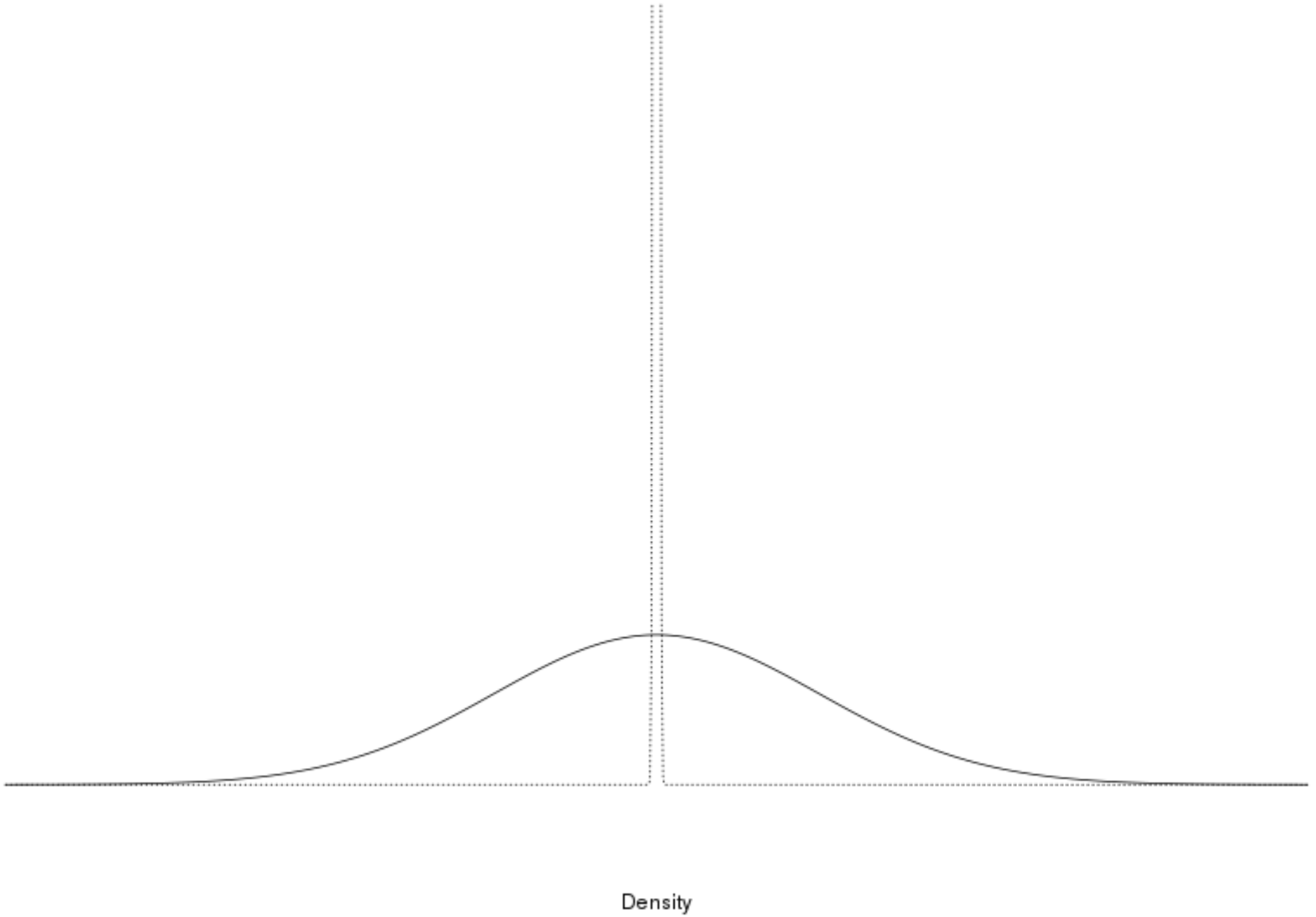
Density

CENTER FOR OPEN SCIENCE
SHUTTLEWORTH FUNDED
TILBURG UNIVERSITY

What do (potentially) fabricated variances look like?

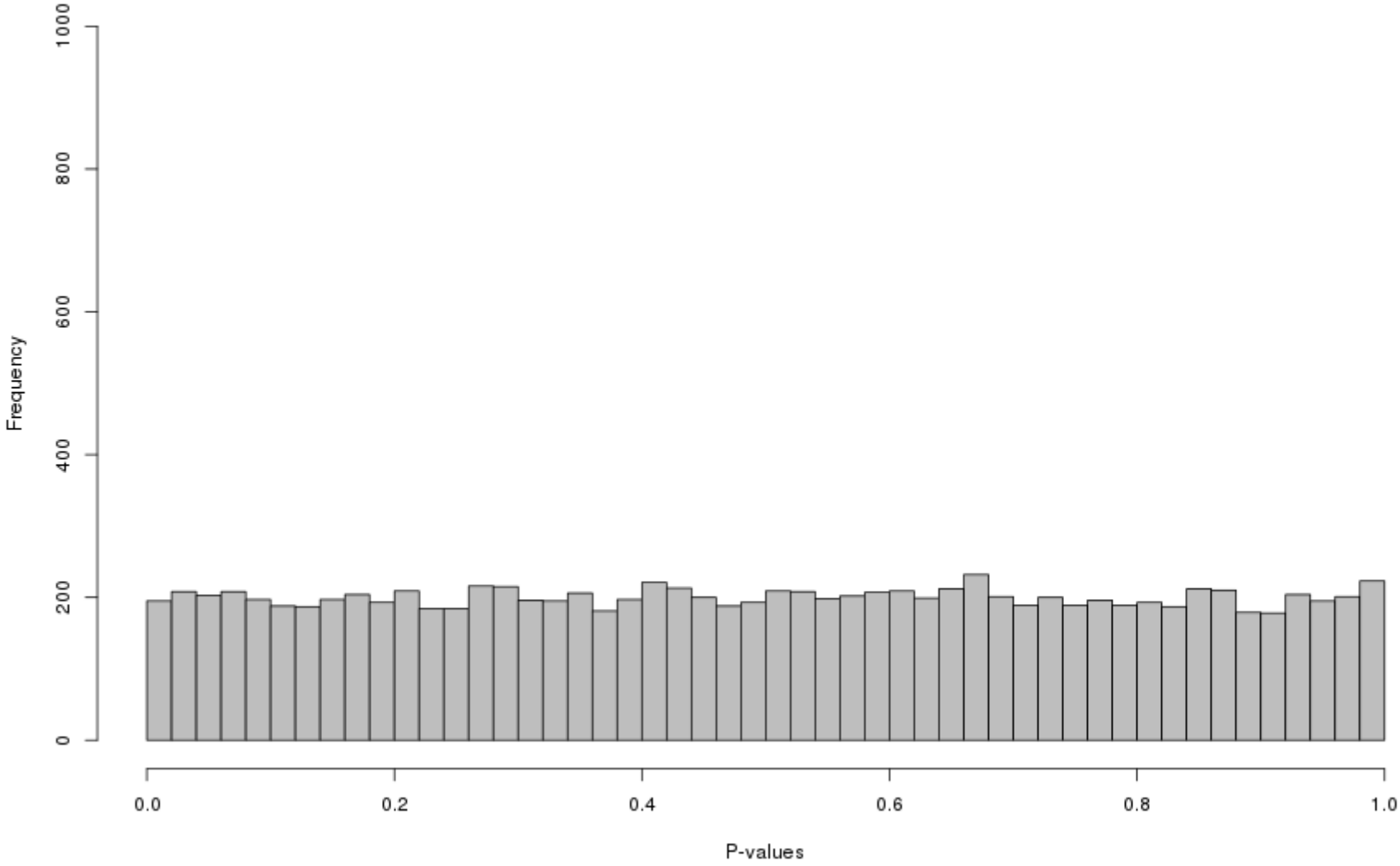


What do (potentially) fabricated variances look like?



2. Nonsignificant *p*-values

What do (potentially) fabricated p-values look like?

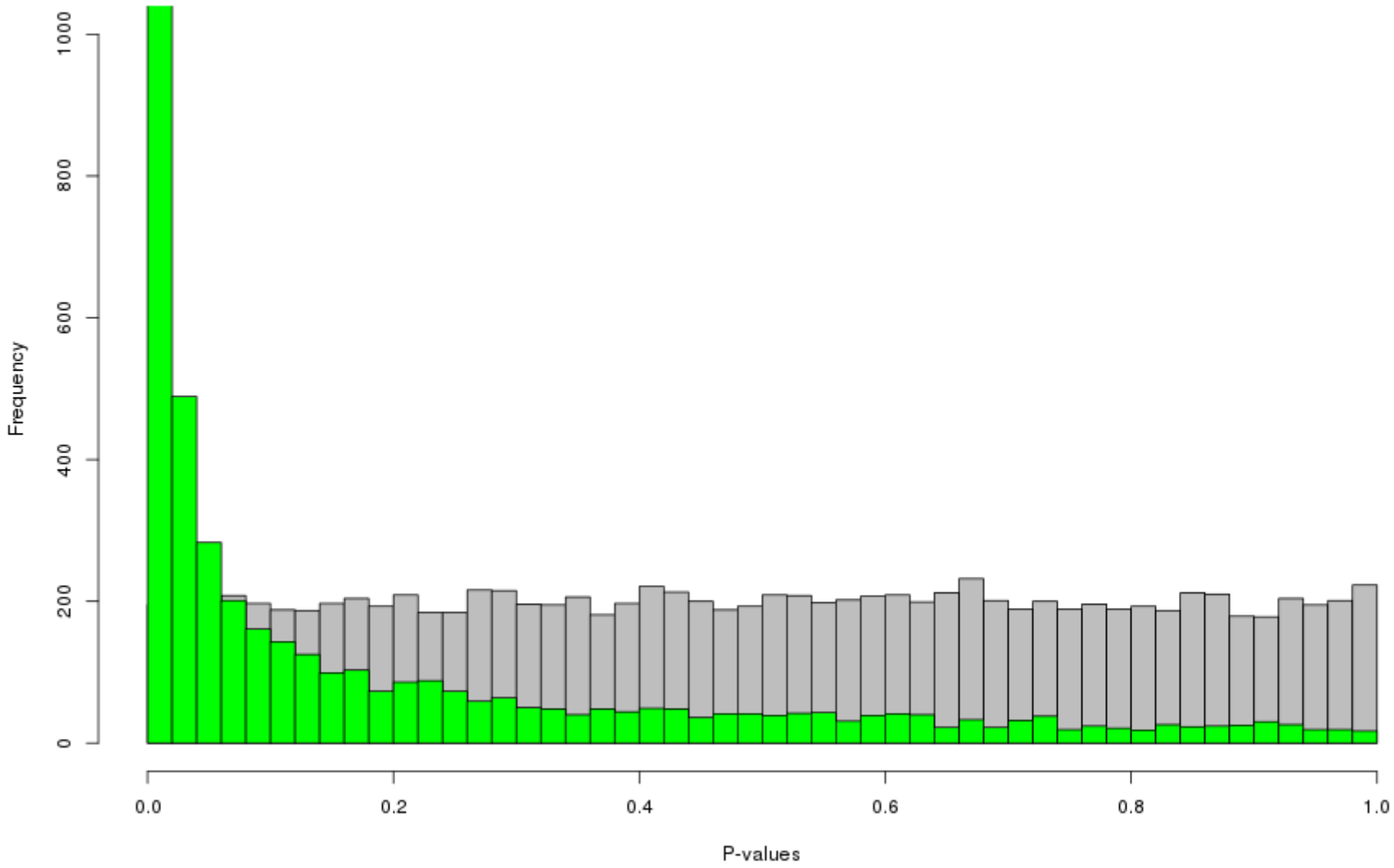


CENTER FOR OPEN SCIENCE

SHUTTLEWORTH FUNDED

TILBURG UNIVERSITY

What do (potentially) fabricated p-values look like?

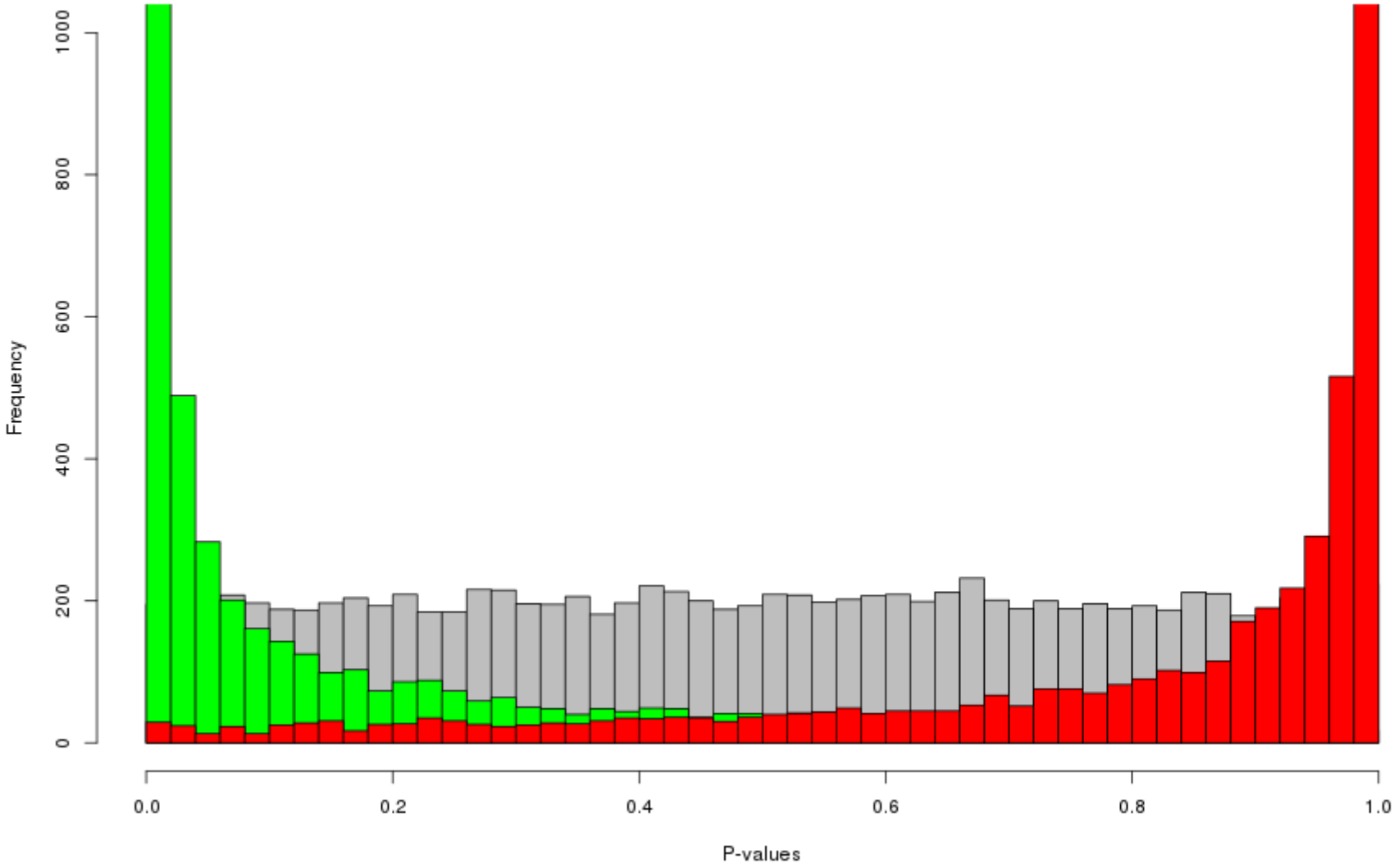


CENTER FOR OPEN SCIENCE

SHUTTLEWORTH FUNDED

TILBURG UNIVERSITY

What do (potentially) fabricated p-values look like?



CENTER FOR OPEN SCIENCE

SHUTTLEWORTH FUNDED

TILBURG UNIVERSITY

3. Simple meta-analysis of 1 + 2*

3. Simple meta-analysis of 1 + 2*

**Fisher method*

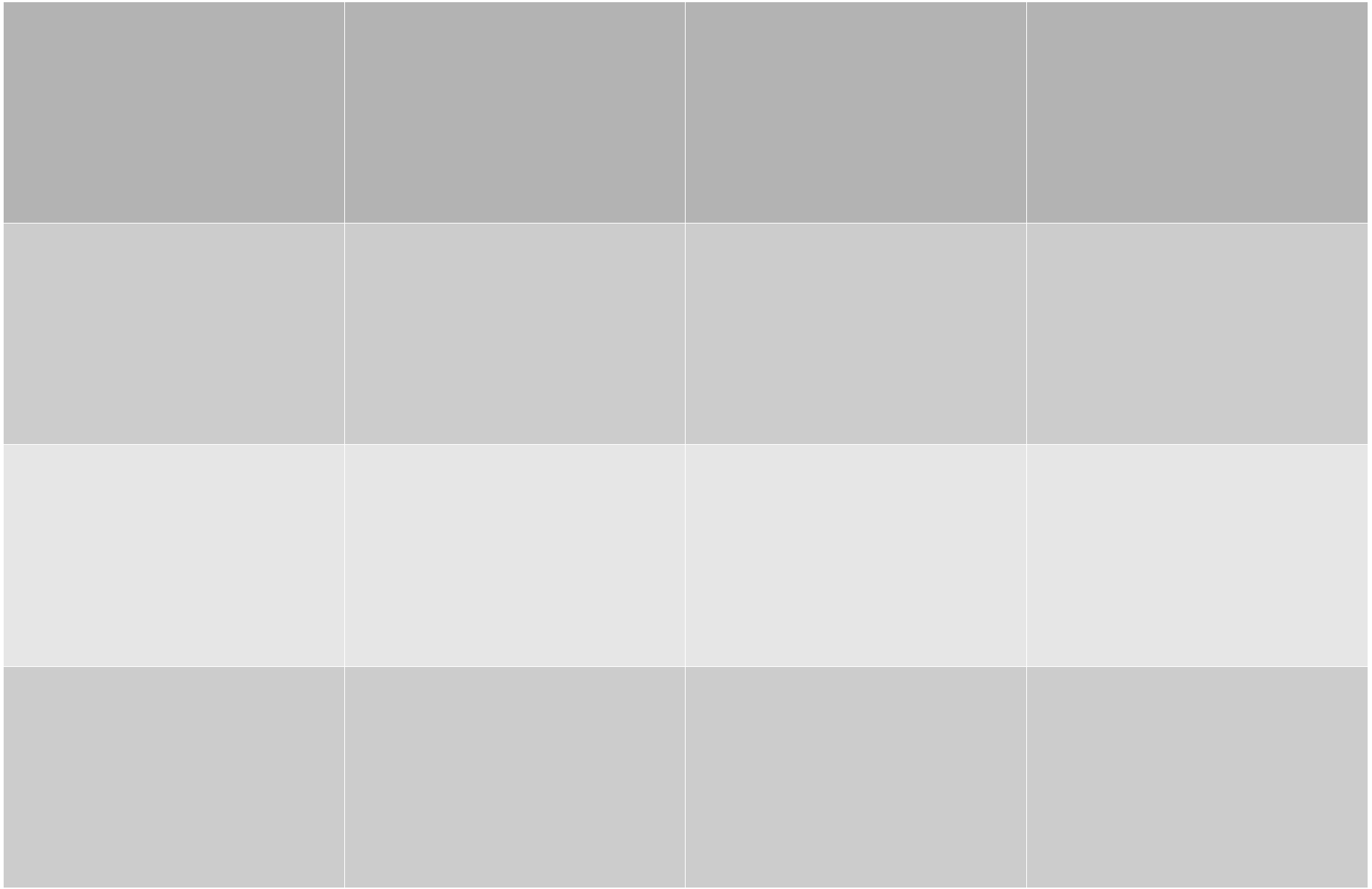
3. Simple meta-analysis of 1 + 2

$$\chi^2_{2k} = -2\sum \ln(p_i)$$

CENTER FOR OPEN SCIENCE
SHUTTLEWORTH FUNDED
TILBURG UNIVERSITY

Test this on genuine and fabricated data

CENTER FOR OPEN SCIENCE
SHUTTLEWORTH FUNDED
TILBURG UNIVERSITY



CENTER FOR OPEN SCIENCE

SHUTTLEWORTH FUNDED

TILBURG UNIVERSITY

	Genuine data (Many Labs 1)		

	Genuine data (Many Labs 1)		
	36		

	Genuine data (Many Labs 1)	Fabricated data (asked psychology researchers)	
	36		

	Genuine data (Many Labs 1)	Fabricated data (asked psychology researchers)	
	36	36	

	Genuine data (Many Labs 1)	Fabricated data (asked psychology researchers)	
“Genuine”			
“Fabricated”			
	36	36	

	Genuine data (Many Labs 1)	Fabricated data (asked psychology researchers)	
“Genuine”			
“Fabricated”			
	36	~~36~~ 34	

CENTER FOR OPEN SCIENCE
SHUTTLEWORTH FUNDED
TILBURG UNIVERSITY

For the best performing method

*Fisher method, at $\alpha = .05$

CENTER FOR OPEN SCIENCE
SHUTTLEWORTH FUNDED
TILBURG UNIVERSITY

	Genuine data (Many Labs 1)	Fabricated data (asked psychology researchers)	
“Genuine”			
“Fabricated”	1		
	36	34	

	Genuine data (Many Labs 1)	Fabricated data (asked psychology researchers)	
“Genuine”			
“Fabricated”	1	10	
	36	34	

	Genuine data (Many Labs 1)	Fabricated data (asked psychology researchers)	
“Genuine”			
“Fabricated”	1	10	11
	36	34	

	Genuine data (Many Labs 1)	Fabricated data (asked psychology researchers)	
“Genuine”	35		
“Fabricated”	1	10	11
	36	34	

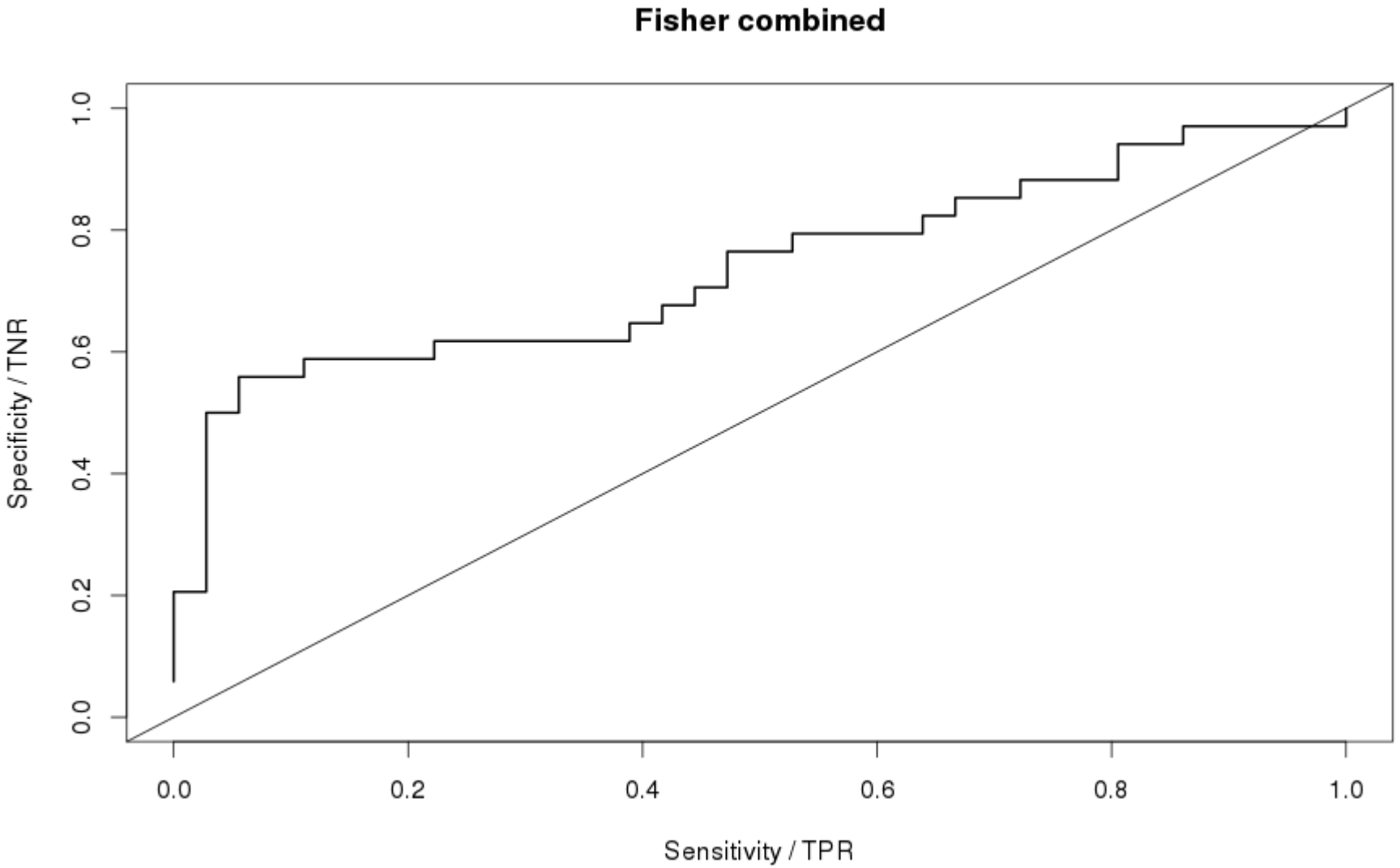
CENTER FOR OPEN SCIENCE
SHUTTLEWORTH FUNDED
TILBURG UNIVERSITY

	Genuine data (Many Labs 1)	Fabricated data (asked psychology researchers)	
“Genuine”	35	24	
“Fabricated”	1	10	11
	36	34	

CENTER FOR OPEN SCIENCE
SHUTTLEWORTH FUNDED
TILBURG UNIVERSITY

	Genuine data (Many Labs 1)	Fabricated data (asked psychology researchers)	
“Genuine”	35	24	59
“Fabricated”	1	10	11
	36	34	

CENTER FOR OPEN SCIENCE
SHUTTLEWORTH FUNDED
TILBURG UNIVERSITY



CENTER FOR OPEN SCIENCE

SHUTTLEWORTH FUNDED

TILBURG UNIVERSITY

PPV = 91%

CENTER FOR OPEN SCIENCE
SHUTTLEWORTH FUNDED
TILBURG UNIVERSITY

NPV = 59%

CENTER FOR OPEN SCIENCE
SHUTTLEWORTH FUNDED
TILBURG UNIVERSITY

Conclusions

- Applied statistical tools to 36 genuine datasets, 36 fabricated datasets
- How do they perform in detecting fabrication?
- PPV = 91%, NPV = 59%, detected 10/36
- Slides at osf.io/ucfpv
- Data, materials etc. osf.io/b24pq

CENTER FOR OPEN SCIENCE
SHUTTLEWORTH FUNDED
TILBURG UNIVERSITY