

Gebruik van statistiek bij de detectie van wetenschapsfraude

Marcel A.L.M. van Assen

Meta-Research Center, TiU

Statistiek in Perspectief
2 juni, 2026, Utrecht

m.a.l.m.vanassen@uvt.nl

Meta-Research Center at Tilburg University (2012-)

Meta-research: onderzoek naar onderzoek, de studie van hoe onderzoek wordt gedaan en kan worden verbeterd

→ **Research misconduct**

Link Statistiek in Perspectief

- Werkgroep statistiek
- Formuleblad 'effectgrootte' statistiek havo
- Position paper: Statistiek in het Voortgezet Onderwijs

Boodschap

Mensen zijn slecht in het begrijpen en toepassen van toeval

Wetenschappers zijn mensen

Statistiek kan helpen bij het signaleren van
wetenschapsfraude

Een deel van de fraudeurs is slecht in het verzinnen van
data horende bij **nuleffecten** (en bestaande effecten)

Wetenschapsfraude

Casus “Diederik Stapel”

- University of Amsterdam ('94-'99), University of Groningen ('00-'06), Tilburg University ('07-'11)
- September 2011: fraude Stapel komt naar buiten
- Onderzoekers nodig die werk Stapel onder loep nemen
→ Marcel van Assen (TiU)
- Onderzoek fraude Stapel: Oct. 2011- Nov. 2012
- Nov. 28, 2012: publicatie rapport Levelt

Wetenschapsfraude

Voorbeeld Stapel

Tabel 2

Disgust	Fear	No emotion
2.87 _a (1.25)	3.07 _a (1.22)	1.60 _b (1.30)
2.60 _a (0.63)	1.66 _b (0.62)	1.60 _b (0.63)
1.87 _b (0.74)	3.13 _a (0.64)	2.13 _b (0.92)
3.00 _a (0.85)	2.73 _a (0.59)	1.73 _b (0.80)
1.93 _a (0.70)	1.87 _a (0.64)	1.60 _a (0.51)
2.00 _a (0.54)	1.87 _a (0.64)	1.93 _a (0.46)

Tabel 5 (ander experiment)

Prime		
Disgust	Fear	No emotion
6.27 _a (1.22)	6.07 _a (1.22)	4.93 _b (1.62)
2.67 _a (0.72)	1.73 _b (0.59)	1.60 _b (0.63)
1.93 _b (0.88)	3.60 _a (0.51)	2.13 _b (0.92)
6.00 _a (0.93)	6.00 _a (0.66)	4.87 _b (0.83)
1.93 _a (0.70)	1.87 _a (0.64)	1.60 _a (0.51)
2.00 _a (0.54)	1.87 _a (0.64)	1.93 _a (0.46)

Wetenschapsfraude

<https://www.youtube.com/watch?v=vJG698U2Mvo>

De onzichtbare gorilla (selective attention test)

Slechts ongeveer 50% van de proefpersonen heeft de gorilla gezien!

Je ziet wat je verwacht te zien, en niet het onverwachte

→ Onderzoekers zien wetenschapsfraude niet!

Wetenschapsfraude

Algemene lessen casussen wetenschapsfraude

- (1) Wetenschappers (mensen) zien het niet
- (2) Instituten stoppen wetenschapsfraude meestal in doofpot
- (3) Onderzoek naar wetenschapsfraude meestal amateuristisch (ontbreekt expertise)



Wetenschapsfraude

Voorbeeld Fuji (anesthesie)

- ‘Wereldkampioen’, met wetenschapsfraude ontdekt in 219 artikelen
- Fraude blootgelegd door Carlisle

Carlisle, J. B. (2012). The analysis of 168 randomised controlled trials to test data integrity. *Anaesthesia*, 67(5), 521-537.

- Principe statistische methode:

Als geen effect/verschil wordt verwacht, wordt te weinig ‘ruis’ (toevalsfluctuatie) toegevoegd aan verzonnen data

- Principe werkte ook bij andere casussen (o.a. Stapel)

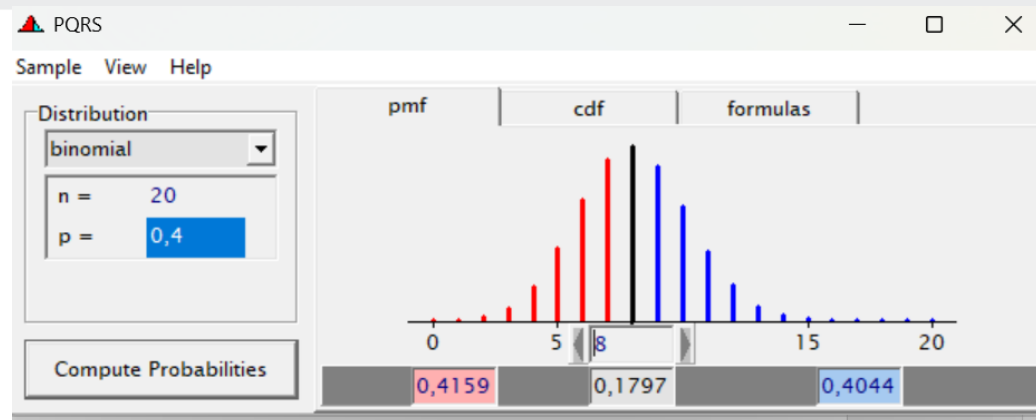
Wetenschapsfraude

Vervolg voorbeeld Fuji

Achtergrondvariabelen (gewicht, lengte, leeftijd, etc.) zouden geen effect moeten hebben

Tabel Carlisle, experiment Fuji, vijf conditions

	Rate	Placebo (n = 20)	10 (n = 20)	20 (n = 20)	40 (n = 20)	80 (n = 20)
Women	0.4	8	7	8	9	8
Motion sickness	0.09	2	1	2	2	2
Previous PONV	0.02	0	1	0	1	0
Operation two*	0.12	2	3	2	2	3
Operation three*	0.11	2	2	3	2	2
No analgesia	0.37	7	8	7	7	8



Wetenschapsfraude

Vervolg voorbeeld Fuji

Abstract Carlisle

databases, retrieving 168 RCTs published by this author between 1991 and July 2011. I extracted rates for categorical variables and means (SDs) for continuous variables, and compared these published distributions with distributions that would be expected by chance. The published distributions of 28/33 variables (85%) were inconsistent with the expected distributions, such that the likelihood of their occurring ranged from 1 in 25 to less than 1 in 1 000 000 000 000 000 000 000 000 000 000 000 (1 in 10^{33}), equivalent to p values of 0.04 to $< 1 \times 10^{-33}$, respectively. In 141 human studies, 13/13 published continuous variable distributions were inconsistent with



Wetenschapsfraude

Maar...

Dat principe werkt bij casussen wil niet zeggen dat:

- Alle/meeste fraudeurs dit doen (alleen 'idioten'?)
- Dit het beste principe is fraude te detecteren

→ EXPERIMENT

Hartgerink, C. H., Voelkel, J. G., Wicherts, J. M., & van Assen, M. A. L. M. (2019). Detection of data fabrication using statistical tools. PsyArXiv.

Wetenschapsfraude

Experimenten Hartgerink et al.

- Onderzoekers vragen data te fabriceren, EN echte data wordt verzameld
- Beste fraudeur krijgt beloning
- Vermeldenswaard: 'gezeik' met ethische commissie

Experiment 1: summary statistics

36 echte en 39 gefabriceerde datasets

Experiment 2: datapunten

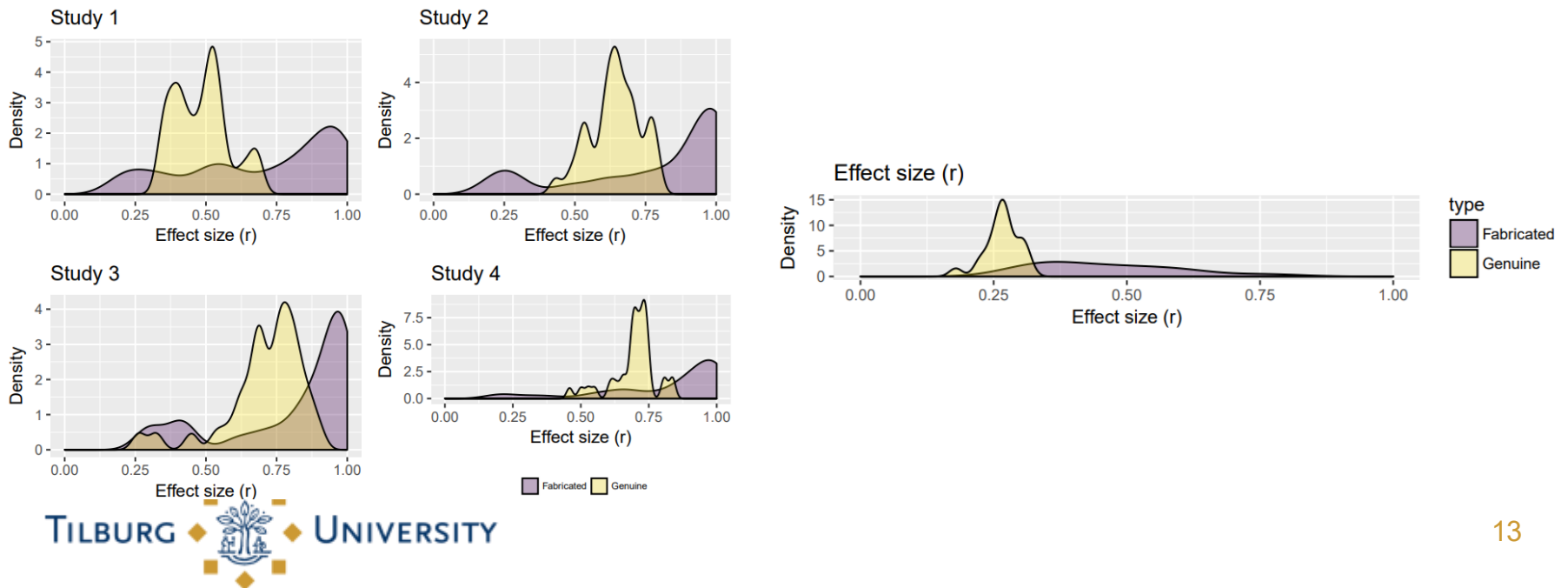
21 echte en 28 gefabriceerde datasets

→ Tools toegepast op alle datasets

Wetenschapsfraude

Experimenten Hartgerink et al. Resultaten

- Methode voor detecten nuleffecten pikt fraudeurs er niet goed uit
- Beter werkte de prevalentie van **extreme effecten**



Wetenschapsfraude

Experimenten Hartgerink et al.

Resultaten

- Beter werkte de prevalentie van **extreme effecten**
- Werkte ook bij andere casussen (incl Stapel)

Merk op – hetzelfde principe ligt hieraan ten grondslag

Als **wel een effect/verschil** wordt verwacht, wordt te weinig 'ruis' (toevalsfluctuatie) toegevoegd aan verzonnen data

Boodschap

Mensen zijn slecht in het begrijpen en toepassen van toeval

Wetenschappers zijn mensen

Statistiek kan helpen bij het signaleren van
wetenschapsfraude

Een deel van de fraudeurs is slecht in het verzinnen van
data horende bij **nuleffecten** (en bestaande effecten)

Thanks!

m.a.l.m.vanassen@uvt.nl

Meta-Research Center at Tilburg University