Tutkimustulokset eivät toistu – Missä syy?

LT, dos Aleksi Reito
TULE-keskus, PSHP

Luennon rakenne

- Taustaa
- Syyt ja syylliset
- Entäs sitten ja mitä huomenna?

→ Reito A, Raittio L, Helminen O. Tutkimustulokset eivät toistu - missä syy? Duodecim. 2020;136(10):1155-63

Tieteellisen tutkimuksen toisteisuus on empiirisen tutkimuksen merkittävin lähtökohta

”Scientific theories are built on replicable phenomena”

- **1600-luvulta alkaen tieteellisen kokeiden menetelmien kuvaus**
- **Satunnaisvaihtelun vuoksi epätarkkuutta identtisissäkin tutkimuksissa**
- **Yksi poikkeava tulos ei tarkoita että teoria on väärässä**

“non-reproducible single occurrences are of no significance to science”

Miller 2009, Popper 1934

Tieteen toistettavuus lisääntyvästi esillä

- Ovatko tutkimustulokset toistettavissa = luotettavuus
- Tieteessä on toistettavuuskriisi/-ongelma
- Hukkatutkimus, research waste
- Laineet eivät ole lyöneet kovasti vielä suomenkieliseen lääketieteelliseen diskurssiin

LÄÄKETIETEELLINEN TIEDEKUNTA / AJANKOHTAISTA / UUTIS- JA TIEDOTEARKISTO /

Voiko tieteeseen vielä luottaa?

22.12.2017 | KORKEAKOULU- & TIEDEPOLITIIKKA | KOULUTUS, KASVATUS JA OPPIMINEN | OPETUS JA OPISKELU YLIOPISTOSSA

HENKILÖ

TEKSTI: PÄIVI LEHTINEN

Tieteen uskottavuus ei kaadu trumpismiin, vaihtoehtoisiin totuuksiin eikä huuhaa-uskoon. Todellinen uhka tieteen luotettavuudelle voi nousta vain tiedeyhteisön sisältä, sanoo professori Kimmo Kontula.

– Julkisuudessa on viime aikoina puhuttu paljon totuuden jälkeisestä ajasta, asiantuntijoiden vähättelystä, false balancesta ja vaihtoehtoisista totuuksista. Ne ovat huolestuttavia ilmiöitä, mutta eivät pysty vakavasti horjuttamaan tieteen luotettavuutta. Todelliset uhat tieteen luotettavuudelle nousevat tiedeyhteisön sisältä.

Näin toteaa sisätautiopin professori **Kimmo Kontula**, joka otsikoi 21.12. pitämänsä jäähyväisluennon haastavasti "Voiko tieteeseen vielä luottaa?"

Voiko julkaistuun tutkimustietoon vielä luottaa?

LÄÄKETIETEELLINEN AIKAKAUSKIRJA DUODECIM

2020; 136(24):2750-7

Kimmo Kontula

Miksi toistettavuus(kriisi) on nyt kuuminta hottia?

RESEARCH

RESEARCH ARTICLE

PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*†

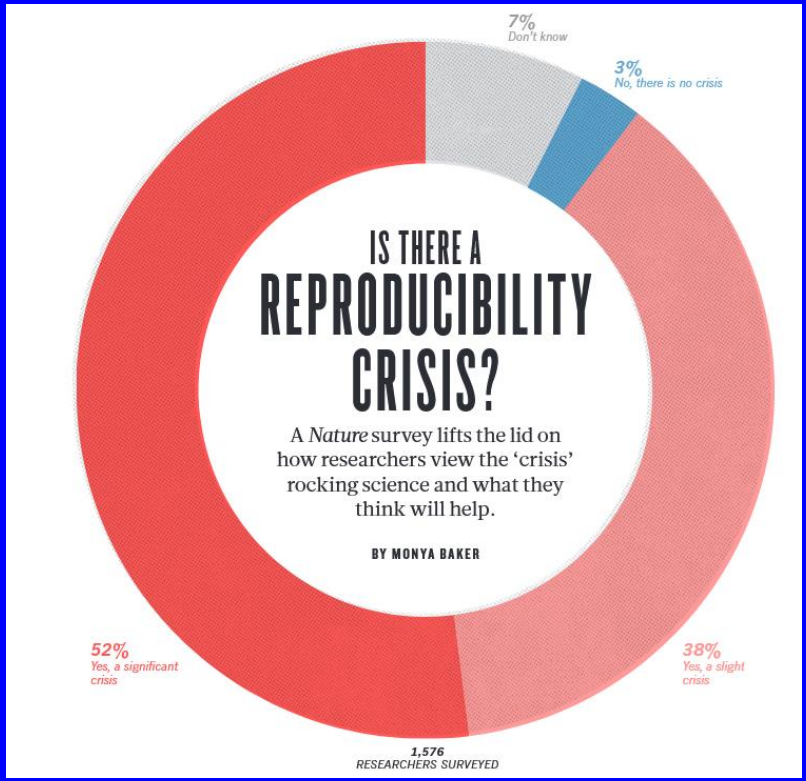
Seurauksena

THE AMERICAN STATISTICIAN
2016, VOL. 70, NO. 2, 129–133
http://dx.doi.org/10.1080/00031305.2016.1154108

Taylor & Francis
Taylor & Francis Group

EDITORIAL

The ASA's Statement on *p*-Values: Context, Process, and Purpose



eLIFE
elifesciences.org

EDITORIAL

REPRODUCIBILITY IN CANCER BIOLOGY

The challenges of replication

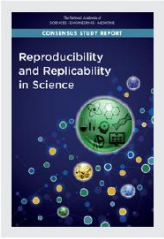
REPRODUCIBILITY PROJECT CANCER BIOLOGY

Interpreting the first results from the Reproducibility Project: Cancer Biology requires a highly nuanced approach.

The National Academies of SCIENCES ENGINEERING MEDICINE THE NATIONAL ACADEMIES PRESS

This PDF is available at http://nap.edu/25303

SHARE



Reproducibility and Replicability in Science (2019)

DETAILS

218 pages | 6 x 9 | PAPERBACK
ISBN 978-0-309-48616-3 | DOI 10.17226/25303

Toisteisuusongelma ei ole uusi ilmiö

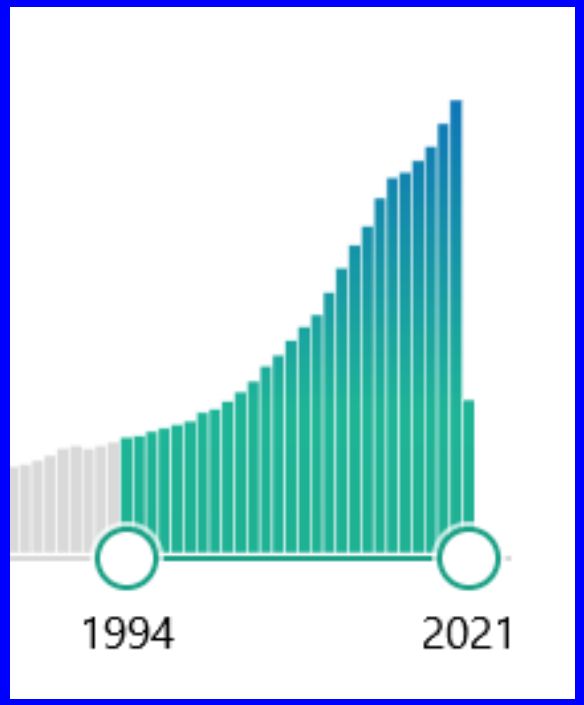
BMJ

LONDON, SATURDAY 29 JANUARY 1994

The scandal of poor medical research

We need less research, better research, and research done for the right reasons

”fracture OR cancer”



Mistä on siis kyse?

- Tutkimus 1: A aiheuttaa B, A on parempi kuin B
- Tutkimus 2: A ei aiheuta B, A ei ole parempi/on yhtä hyvä kuin B
  → TOISTETTAVUUSONGELMA (!/?)
- Mutta
  - Taustalla sama biologinen ilmiö
  - Todellista heterogeenisyyttä tottakai esiintyy
  - Onko degeneratiivinen meniskiruptuura Suomessa erilainen kuin Saksassa?
  - Onko luun (solu)biologinen erilaista Amerikassa?

Onko kaikki pielessä?

- Menetelmätoistettavuus
- Tulostoistettavuus
- Päätelmätoistettavuus

Midshaft Clavicle Fractures: Surgery Provides Better Results as Compared With Nonoperative Treatment

A Meta-analysis

Enrico Guerra,* MD, Davide Previtali,† MD, Simone Tamborini,†‡ MD, Giuseppe Filardo,*† MD, PhD, Stefano Zaffagnini,* MD, Prof, and Christian Candrian,† MD
Investigation performed at Unità di Traumatologia e Ortopedia, Ospedale Regionale di Lugano, Lugano, Switzerland

Surgical treatment of MCFs [...], leads to better shoulder functional scores at short- and long-term follow-up.

Goodman et al. 2016

COPYRIGHT © 2017 BY THE JOURNAL OF BONE AND JOINT SURGERY, INCORPORATED

Plate Fixation Versus Nonoperative Treatment for Displaced Midshaft Clavicular Fractures

A Meta-Analysis of Randomized Controlled Trials

Sarah Woltz, MD, Pieta Krijnen, PhD, and Inger B. Schipper, MD, PhD
Investigation performed at the Leiden University Medical Center, Leiden, the Netherlands

plate fixation significantly reduces the risk of nonunion, but does not have a clinically relevant advantage regarding final functional outcome.

Erittäin suuri syyllinen ja kaiken pahan alku ja juuri on…

- **Nollahypoteeesin (merkitsevyyden) testaus (*null hypothesis significance testing, NHST*)**
  - H_0: $u_1=u_2$
  - H_1: $u_1 \neq u_2$
- **Fisher ja Neyman-Pearson hypoteesin testaus**
- **<u>Kukaan ei ole keksinyt NHST, se vain syntyi</u>**

p-arvon rajasta 0.05 on muodostunut tieteen pohja

- Kalastelu
- p-hakkerointi (p-hacking)
- Julkaisupaine

Although our results show that there is good evidential value for orthopaedic literature published in our top journals, there is some evidence of p-hacking of which authors and readers should be wary.

Abd Razak et al. 2016

HARKing

- *Hypothetizing after the results are known*
- Jos nollahypoteesi on totta, 65% todennäköisyydellä löytyy tilastollisesti merkitsevä löydös 20 testissä (satunnaisvaihtelu!)
- Hypoteesin tuottava vs hypoteesin vahvistava tutkimus
- Avoimuus ja läpinäkyvyys

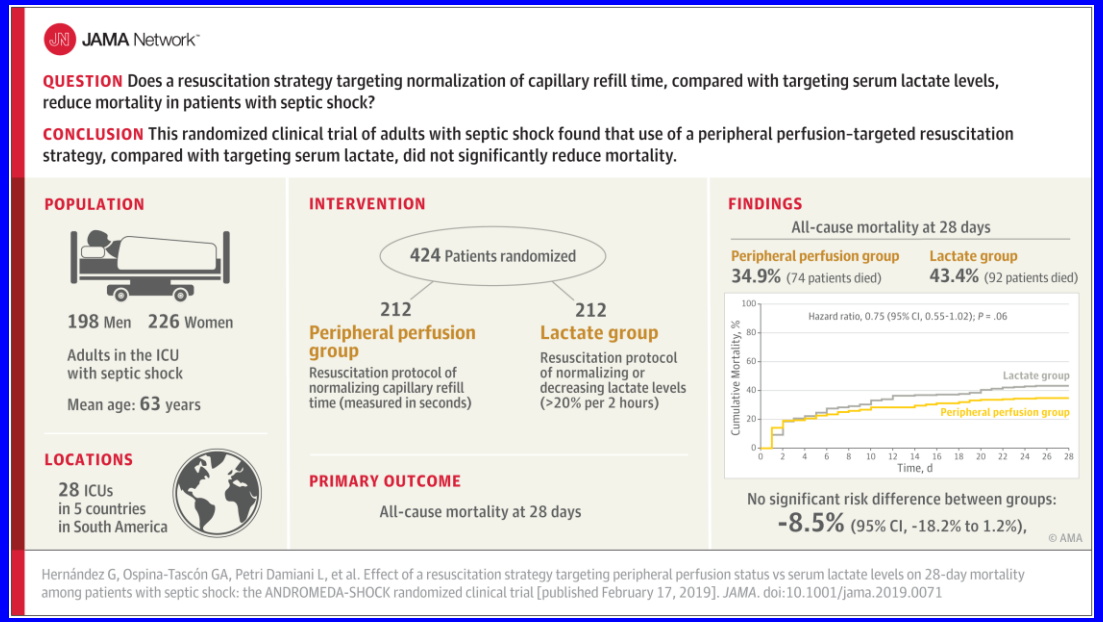
Tiede

Tutkimus: Kakkostyypin diabetes on yleisempi miehillä, joilla on koira

Naisten diabeteksella ja koiranomistajuudella ei ilmennyt minkäänlaista yhteyttä.

Tutkimuslöydösten dikotomisaatio

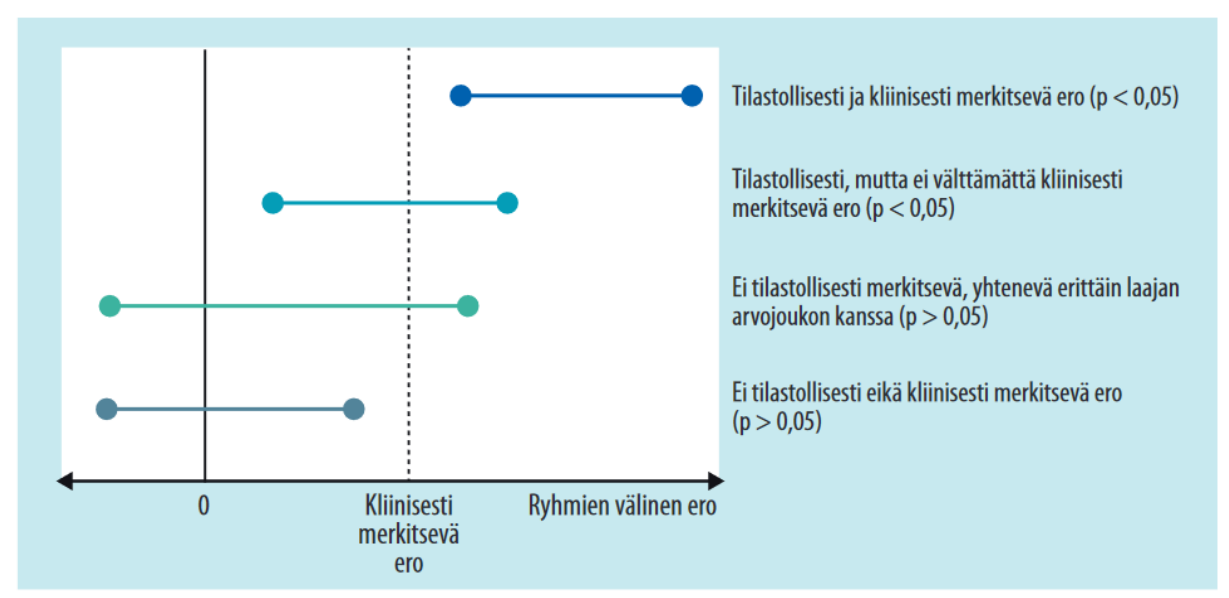
- Lääkkeen A interventiotutkimus, p=0.053 → Ei hyötyä
- Lääkkeen A interventiotutkimus, p=0.048 → Hyötyä on
- Mitään maagista ei tapahdu, kun p ylittää/alittaa 0.05
  - Vaikutukset ovat jatkumoita
- Kaikki vaikuttaa kaikkeen, kyse on suuruudesta



Negatiiviset tulokset

- p-arvo >0.05 → ei eroa/vaikutusta/yhteyttä
- Ei osoitettavaa eroa ≠ ei eroa (yhtä hyvä)

Zhou et al.	"The risk of re-rupture after surgical or non-surgical treatment was equivalent if a functional rehabilitation protocol with early range of motion was performed, [...]"
Van der Eng et al.	"In conclusion, we found no difference in the re-rupture rate between the surgically and nonsurgically treated patients followed by early weightbearing. Weightbearing after 4 weeks also resulted in no differences in the rupture rate in the surgical vaersus conservatively treated patients."
Soroceanu et al.	"[Conservative treatment] resulted in re-rupture rates similar to those for surgical treatment while offering the advantage of a decrease in other complications."
Ochen et al.	"No significant difference in re-rupture rate was seen between operative and nonoperative treatment in studies that used accelerated functional rehabilitation with early range of motion"



Kausaalisuus vs ennustaminen

- **Kausaali- eli syy-seurauspäättelyn periaatteet laajalti hukassa**
- **Oikein tehty havainnoiva tutkimus usein vaikeampaa kuin RCT**

Aleksi Reito @AleksiReito · 24. huhtik. 2020

Vastauksena käyttäjälle @TeemuVKarjalain

Mielenkiintoista. Sekoittavan tekijän pitää vaikuttaa sekä altisteeseen ("kahvi") että outcomiin ("kuolleisuus"). Miksi verenpaine vaikuttaisi kahviin? Ettei vain olisi toisinpäin (kuva)? Tällöin ovat adjustoineet mediaattorilla mikä ei ole oikein. Mitä sanoo @pronutritionist?

3 1 2

Jouni Helske @jouni_helske · 24. huhtik. 2020

Samaa katsoin että menty klassisella lähestymistavalla että isketään kaikenlaista malliin mukaan ja oletetaan että tämä hoitaa sekoittavien tekijöiden vaikutuksen pois. Lisäksi mielivaltaisia kategorisointeja esim. iän suhteen, p-arvoihin tuijottelua (95% lv:ksi naamioituna).

Jatkuvien muuttujien pilkkominen

- Heikentää tilastollista voimaa noin 30%
- Mikään raja-arvo ei ole koskaan perusteltu
- “Potilaan BMI on yli 30 ja ikä alle 65 vuotta”

And the list goes on…

- Tutkijan vapausasteet
- Tutkimuslöydöksen ennakkotodennäköisyys
- Signaali-kohina –suhde
- Tilastollisen voima ja otoskoko
- Julkaisuharha

		Päätetapahtuma	
		Ei	Kyllä
Riskitekijä	Kyllä	71	4
	Ei	37	8

Kaksisuuntainen Fisherin tarkan todennäköisyyden testi:
p-arvo = 0,055

Khiin neliö -testi Yatesin korjauksella:
p-arvo = 0,059

Khiin neliö -testi ilman Yatesin korjausta:
p-arvo = 0,028

Logistinen regressio: kerroinsuhde = 3,83,
95 %:n luottamusväli: 1,13–15,2
p-arvo = 0,037

Ortopedinen Grand slam

■ ARTHROPLASTY

The effect of triclosan-coated sutures on the rate of surgical site infection after hip and knee arthroplasty: a double-blind randomized controlled trial of 2546 patients

Conclusion

This trial provided no evidence that the use of triclosan-coated sutures in THA and TKA leads to a reduction in the rate of SSI.

Table III. The rate of surgical site infection in the two groups, defined according to the Health Protection Agency criteria

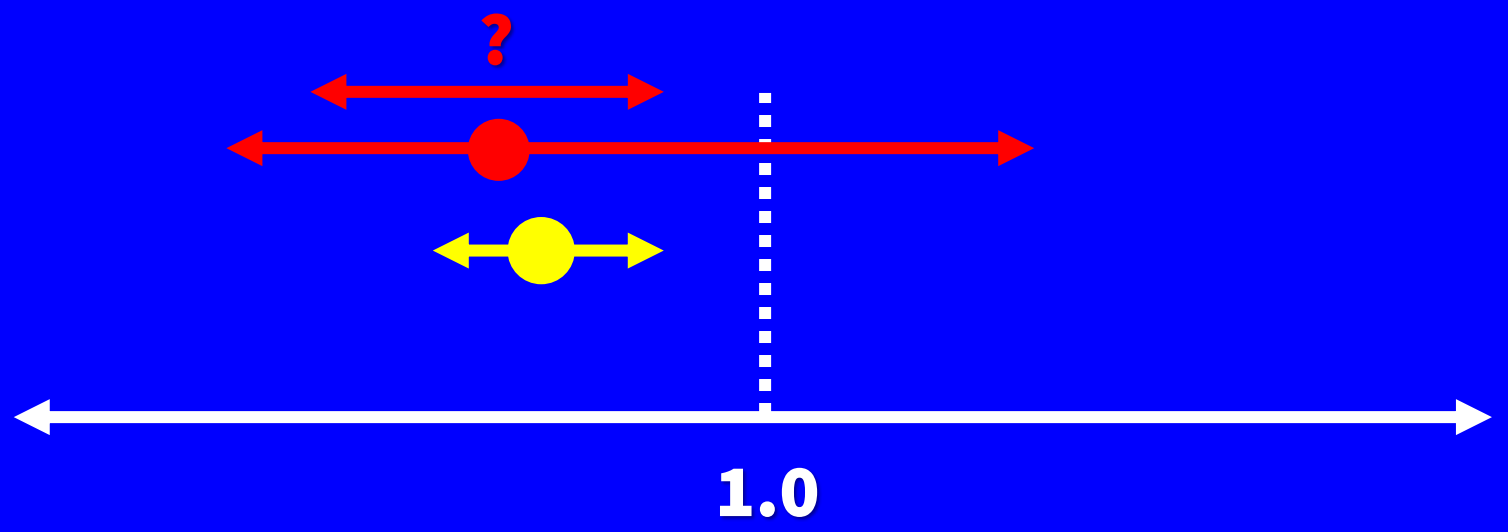
Outcome	Vicryl	Vicryl+	Total	p-value*
Superficial SSI, n (%)	11/1273 (*0.8*)	8/1164 (*0.7*)	19/2437 (*0.7*)	0.651
Deep SSI, n (%)	21/1273 (*1.6*)	13/1164 (*1.1*)	34/2437 (*1.3*)	0.300
Deep and superficial SSI, n (%)	32/1273 (*2.5*)	21/1164 (*1.8*)	53/2437 (*2.1*)	0.266

*Fisher's exact test
†SSI, surgical site infection

Ortopedinen Grand slam

Syvät infektiot 1.6% -> 1.1% (”*No evidence*”)

RR = 0.68 (95% LV: 0.34-1.35)



RR 0.72 (95% LV: 0.60-0.82)

De Jonge et al. Br J Surg

Miksi sitten niin?

- **Tilastotiede ja metodologia on oikeasti vaikeaa!**
- **Väärät ja huonot menetelmät kulkevat isältä pojalle**

Tilastollinen merkitsevyystaso (p-arvo):

- Todennäköisyys, että hylätään nollahypoteesi, vaikka se on totta (tyypin I virhe)

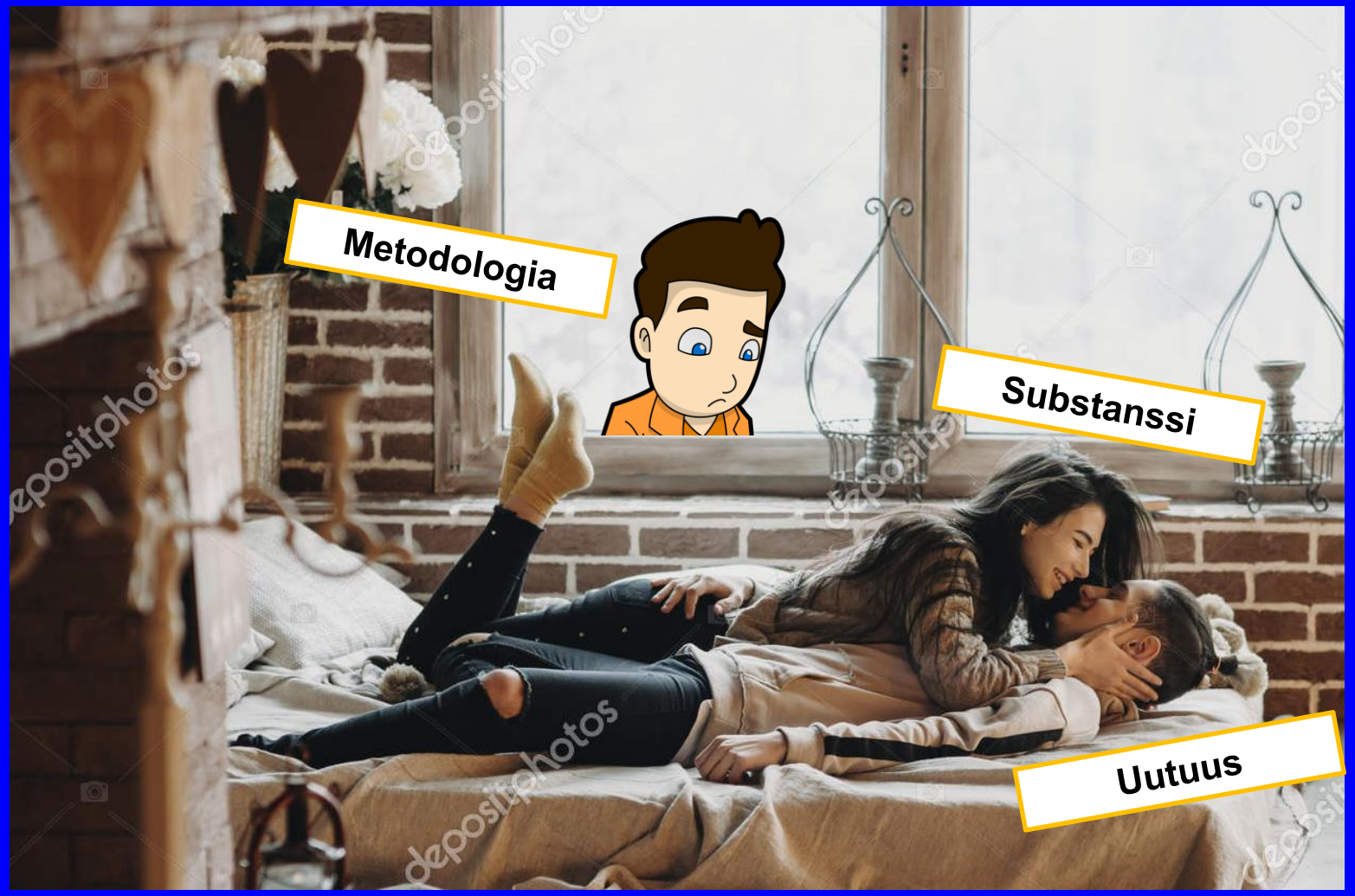
Käytännössä p-arvo tarkoittaa todennäköisyyttä, että nollahypoteesi on totta. Jos on

Miksi sitten niin?

- LL/MD tutkinto antaa kaikkivoipat oikeudet suorittaa mitä tahansa tilastollisia suoritteita

“science undervalues its own backbone, methodology”

- **Metodologia on yhteistä kaikielle tieteelle, substanssi ei**
- **Ensin metodologia, sitten substanssi**
- **Eminenssin voima**



No mitä jatkossa?

- **Down-to-top vai top-to-down?**

Q: Why do so many colleges and grad schools teach $p = 0.05$?
A: Because that's still what the scientific community and journal editors use.
Q: Why do so many people still use $p = 0.05$?
A: Because that's what they were taught in college or grad school.

Wasserstein, Lazar 2016

Ihan vain ajatus lopuksi

Tutkimuksen perusteella miessukupuoli ja leikkausaika yli 70 minuuttia lisäävät syvän infektion riskiä ACL-leikkauksen jälkeen noin puolitoistakertaisesti (~50%)

Miten tämän tieto vaikuttaa potilaan hoitoon = päätöksenteko hoitolinjasta?

Yhteenveto

- Vähäinen koulutus ja heikko osaaminen sekä tiedostamaton että tiedostettu kyseenalaisten tutkimusmenetelmien käyttö yhdessä julkaisupaineen ja epäterveen kilpailun kanssa on johtanut siihen, että merkittävä osa lääketieteellisestä tutkimuksesta on laadutonta ja lähes käyttökelvotonta = toistettavuusongelma

Lähteitä

Bin Abd Razak HR, Ang JE, Attal H, Howe TS, Allen JC. *P-Hacking in Orthopaedic Literature: A Twist to the Tail.* J Bone Joint Surg Am. 2016 Oct 19;98(20):e91. doi: 10.2106/JBJS.16.00479.

Goodman SN, Fanelli D, Ioannidis JP. *What does research reproducibility mean?* Sci Transl Med. 2016 Jun 1;8(341):341ps12. doi: 10.1126/scitranslmed.aaf5027

Miller J. *What is the probability of replicating a statistically significant effect?* Psychon Bull Rev. 2009 Aug;16(4):617-40. doi: 10.3758/PBR.16.4.617.

Altman DG. *The scandal of poor medical research.* BMJ. 1994 Jan 29;308(6924):283-4. doi: 10.1136/bmj.308.6924.283.

Wasserstein RL, & Lazar NA. *The ASA Statement on p-Values: Context, Process, and Purpose*, The American Statistician, 70:2, 129-133, DOI: 10.1080/00031305.2016.1154108

Kiitos!

Email: ***aleksi@reito.fi***
Twitter: @AleksiReito
Luentoslidet löytyvät: ***reito.fi/luennot***

Tilasto-
nurkka

TilastoKUNTO