



Avoimen tieteen  
repositoriot osana  
toistettavuuskriisin  
ratkaisua

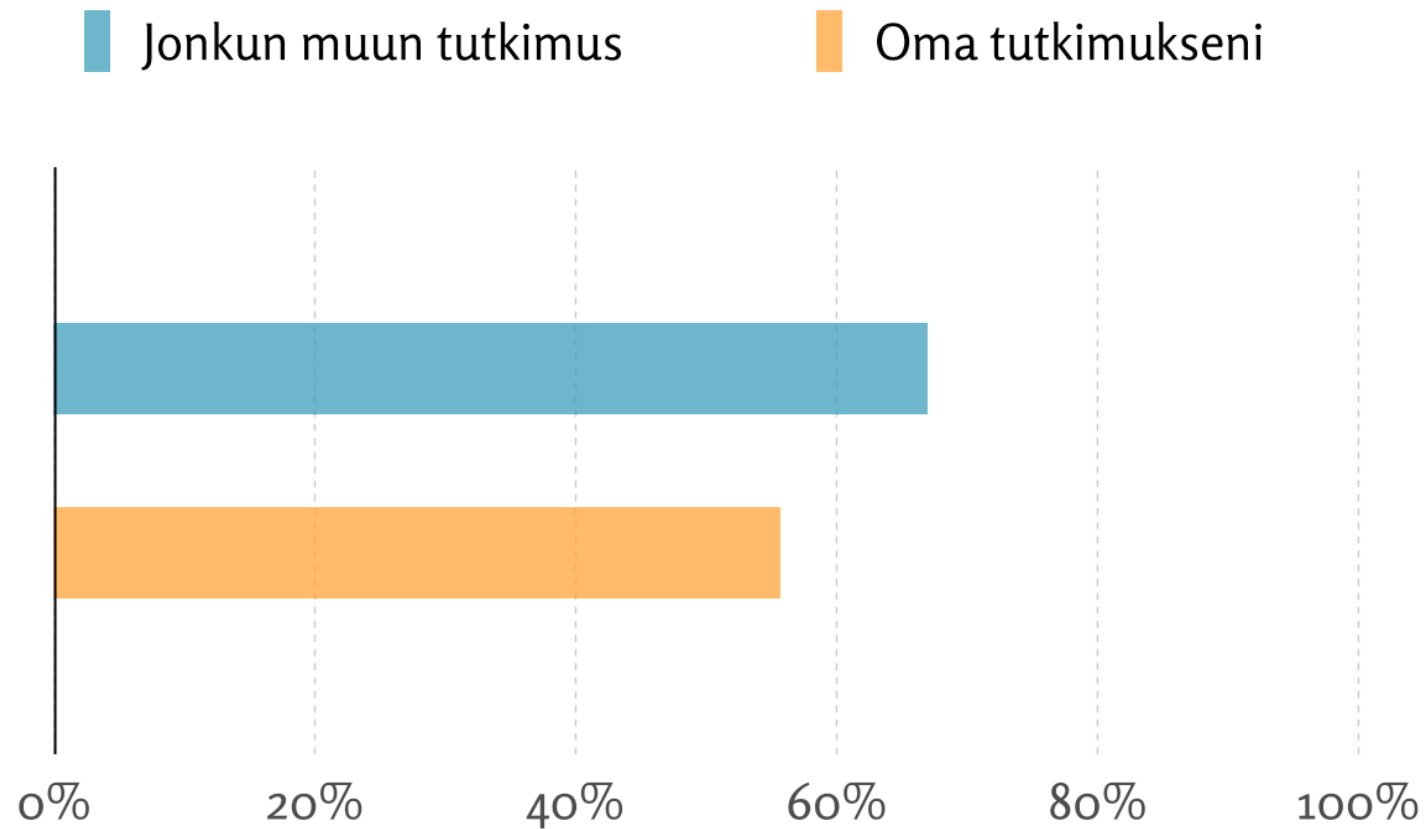
Ville Langén

“Yksittäiset havainnot,  
joita ei voida toistaa,  
eivät ole tieteelle merkityksellisiä.”

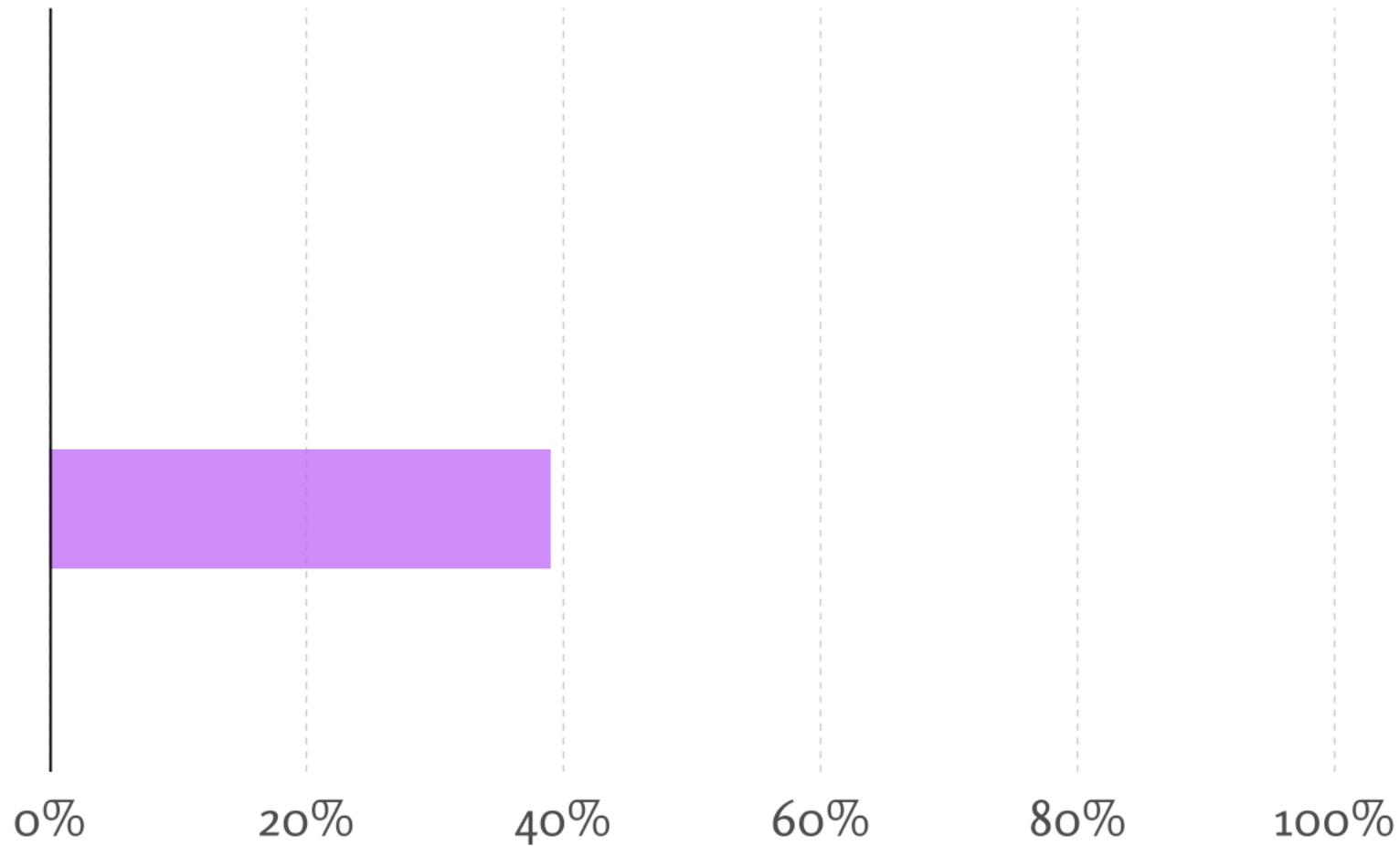
(Karl Popper, 1935)

# Oletko epäonnistunut tutkimustulosten toistamisessa?

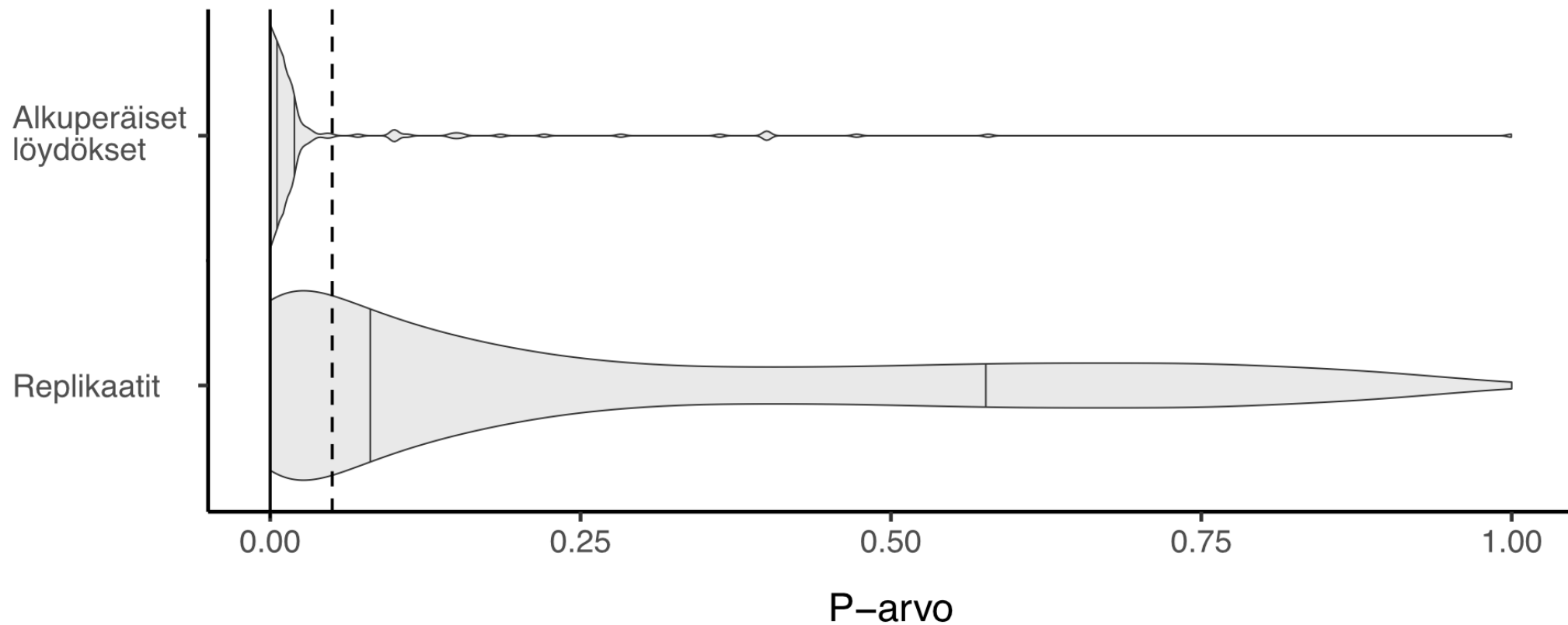
(N=1576 haastateltua tutkijaa)



Kuinka monta prosenttia tieteellisistä tutkimuksista on replikoitavissa?  
(Tieteenala: psykologia, N=100 replikoitua julkaisua)



# Miten P-arvot käyttäytyvät replikoiduissa töissä?



# Replikaatiokriisin (joitakin) syitä



Kyseenalaiset  
tutkimuskäytännöt



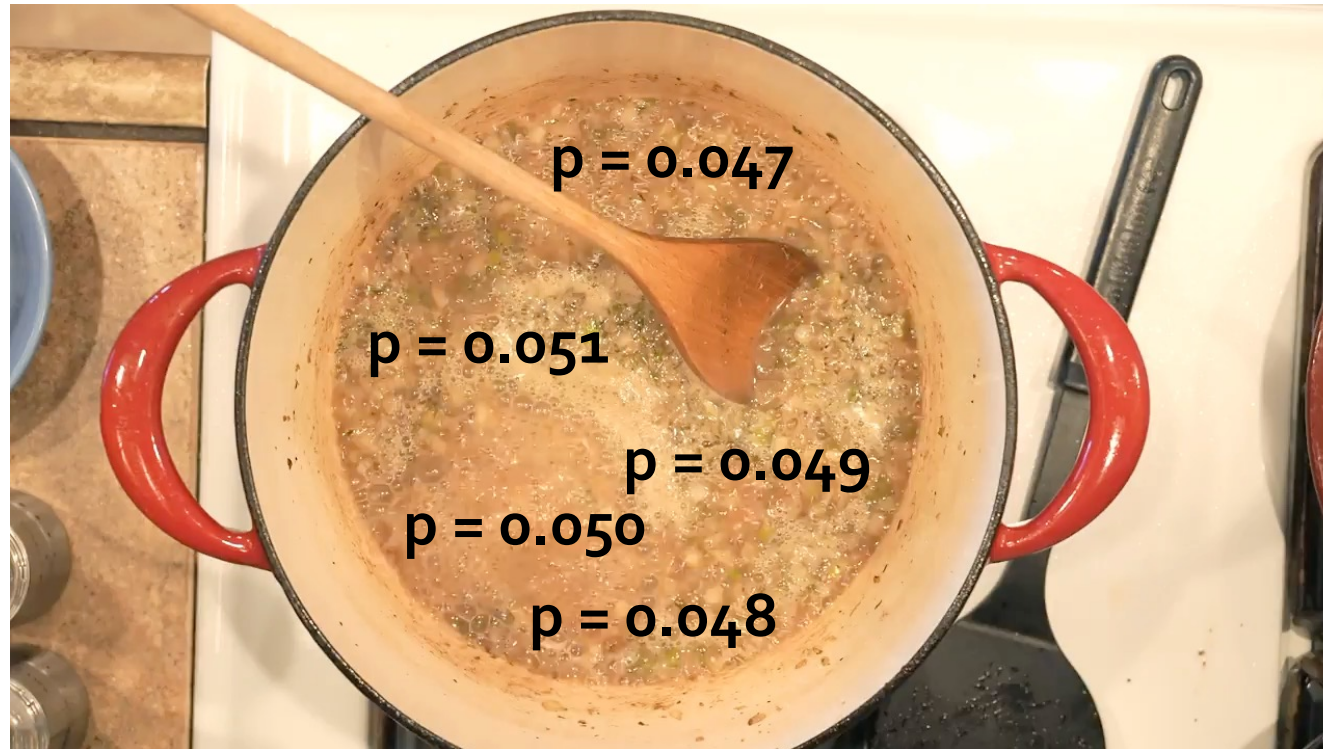
Analyysit  
virheellisiä



Läpinäkyvyys  
riittämätöntä

# Kyseenalaisten tutkimuskäytäntöjen klassikoita:

## 1. P-arvojen kalastus ("cooking")



Ilmiö tunnettu vähintään 1800-luvulta asti

A

B

C

D

E

A

B

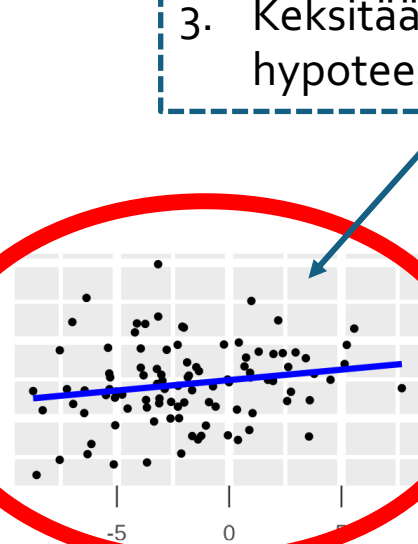
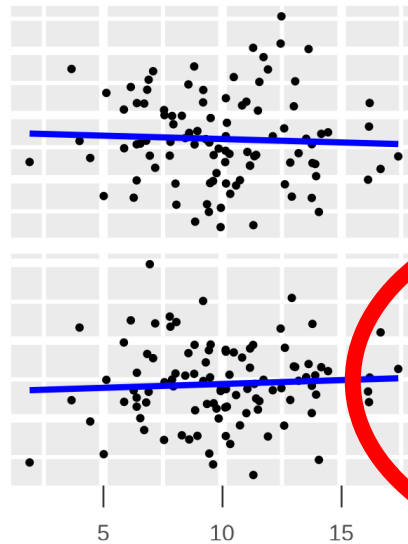
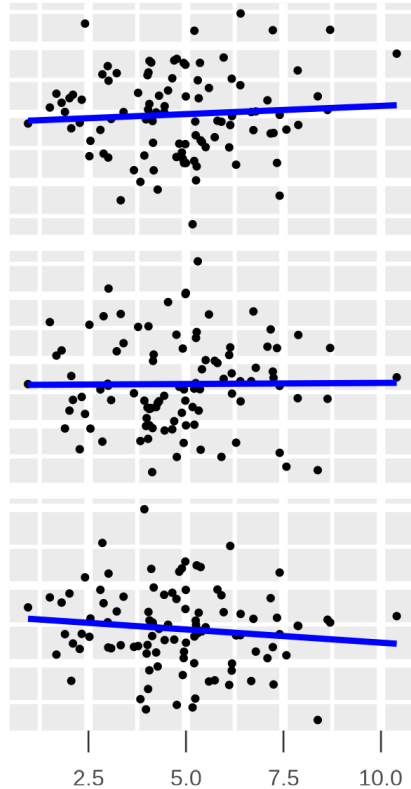
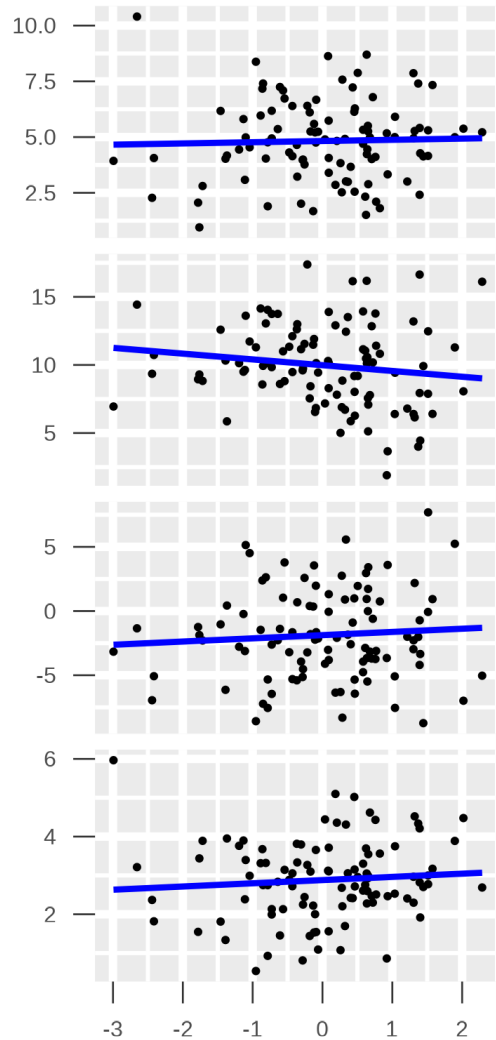
C

D

E

## 2. HARKing

("Hypothesizing After Results are Known")



1. Analysoidaan muuttujia summittaisesti
2. "Löydetään jotain"
3. Keksitään vasta sitten hypoteesi

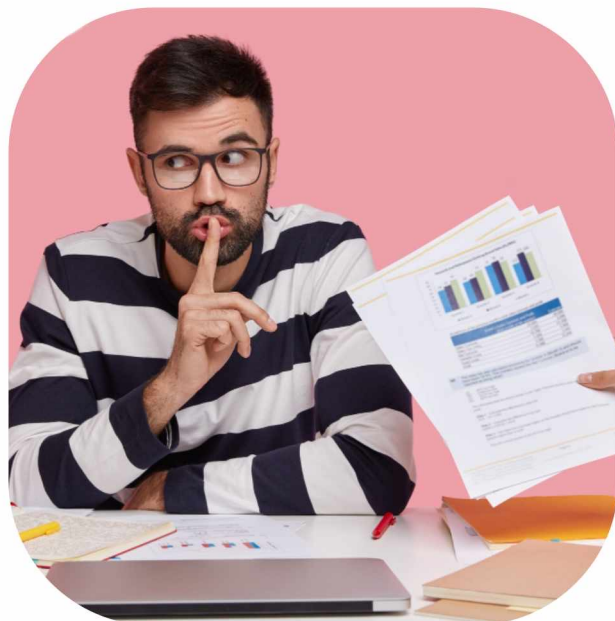


### 3. Julkaisuharha – “the file-drawer problem”



Ei julkaista tuloksia  
analyyyseistä, joissa P-  
arvot ovat ei-  
merkitseviä, mikä johtaa  
*julkaisuharhaan.*

# Näille kolmelle ratkaisuja...?



Kyseenalaiset  
tutkimuskäytännöt



Analyysit  
virheellisiä



Läpinäkyvyys  
riittämätöntä

# Näille kolmelle ratkaisuja...?



Kyseenalaiset  
tutkimuskäytännöt

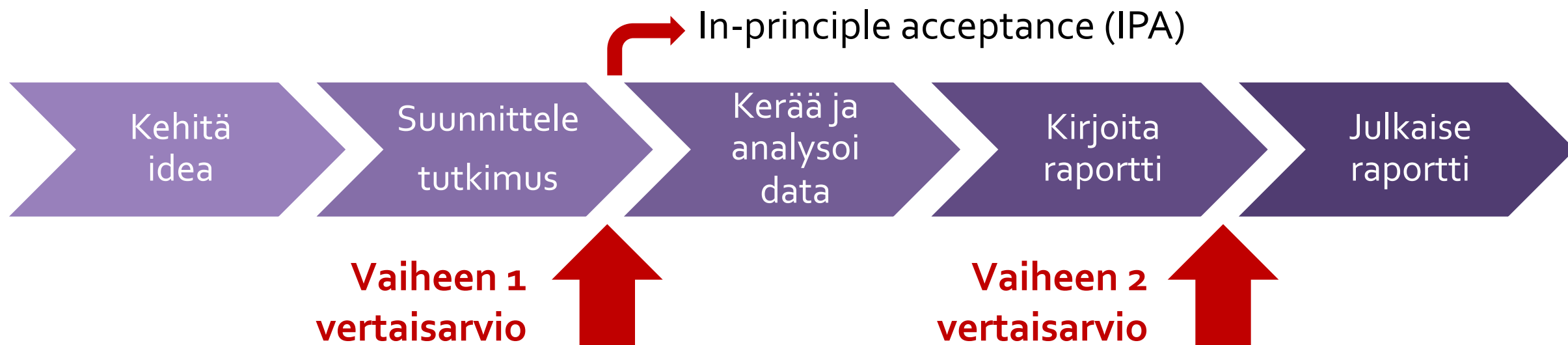


Analyysit  
virheellisiä



Läpinäkyvyys  
riittämätöntä

# Kyseenalaisten tutkimuskäytäntöjen ja virheellisten analyysien estämiseksi on ehdotettu tutkimusten esirekisteröintiä (= *Registered Reports*):



# Esirekisteröidyt tutkimukset

(Registered Reports)

> 300 lehteä jo mukana:  
Nature, BMC Medicine, Cochrane Reviews...

# Näille kolmelle ratkaisuja...?



Questionable  
research practices  
(QRPs)

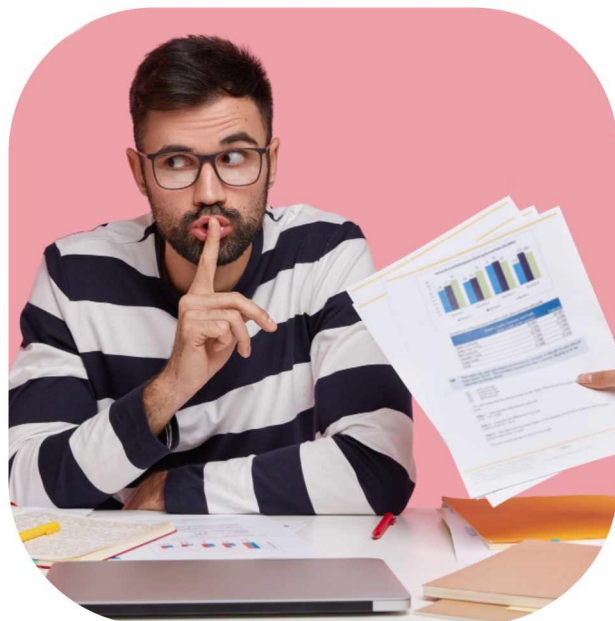


Analyysit  
virheellisiä



Läpinäkyvyys  
riittämätöntä

# Näille kolmelle ratkaisuja...?



Questionable  
research practices  
(QRPs)



Analyysit  
virheellisiä



Läpinäkyvyys  
riittämätöntä

# Läpinäkyvyyden maksimointi



Metodit tulisi kuvata  
käsikirjoituksessa  
riittäväällä tarkkuudella ja  
jakaa tutkimuskoodi (ja  
myös data silloin kun  
tämä on mahdollista)



# Tutkimusmateriaalin jakamisen fair data -prinsiipit

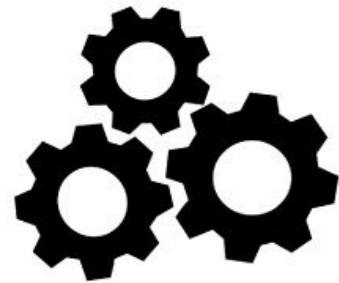
F<sub>indable</sub>



A<sub>ccessible</sub>



I<sub>nteroperable</sub>



R<sub>eusable</sub>

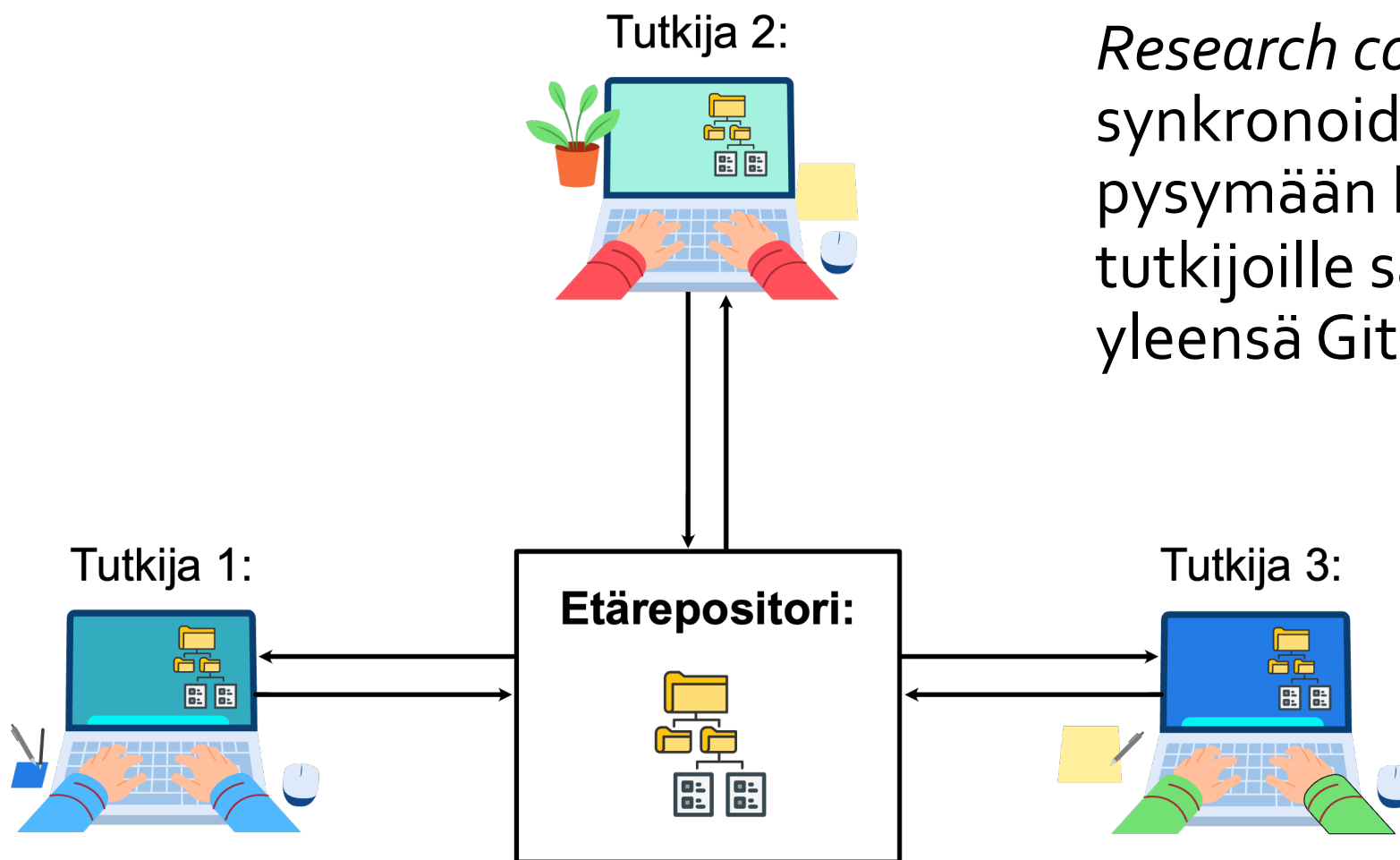


# Research compendium

Name	^	Date Modified	Size	Kind
> analysis		Today, 11:19	--	Folder
> data		Today, 11:19	--	Folder
DESCRIPTION.txt		Today, 11:20	11 bytes	Plain Text
README.md		Today, 11:20	6 bytes	md

Kaikki tutkimuksen tieto pitää olla siistissä järjestyksessä määrätyissä alihakemistoissa. "data"-alihakemiston sisältöä ei koskaan muuteta itsessään vaan ainoastaan "analysis"-alihakemiston koodiajoissa. Kyseiset alihakemistot sisältöineen sekä tarvittavat infotiedostot muodostavat *research compendiumin*.

# Research compendiumin työstäminen



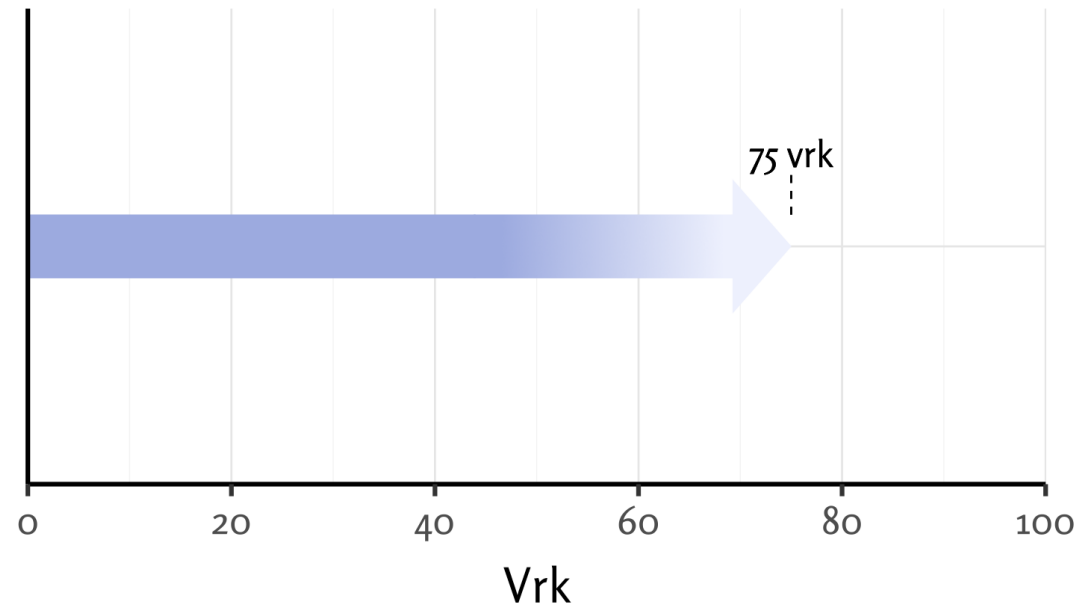
*Research compendium*  
synkronoidaan  
pysymään kaikille  
tutkijoille samana,  
yleensä Gitin avulla.

# Git-repositorioita



# Web-osoitteiden keskimääräinen elinikä

Pelkkä www-osoite  
johonkin repositorioon ei  
riitä tutkimusmateriaalin  
identifioimiseksi, sillä  
web-osoitteet eivät  
lähtökohtaisesti ole  
"pysyviä"



# Tutkimuksen materiaaleille tarvitaan DOI-tunnus (digital object identifier)

Tiederepositorioon  
tallentamalla  
tutkimusmateriaalille  
saadaan DOI-tunnus



<https://doi.org/10.1000/182>

# Yleiskäyttöisiä (tiede-)repositorioita

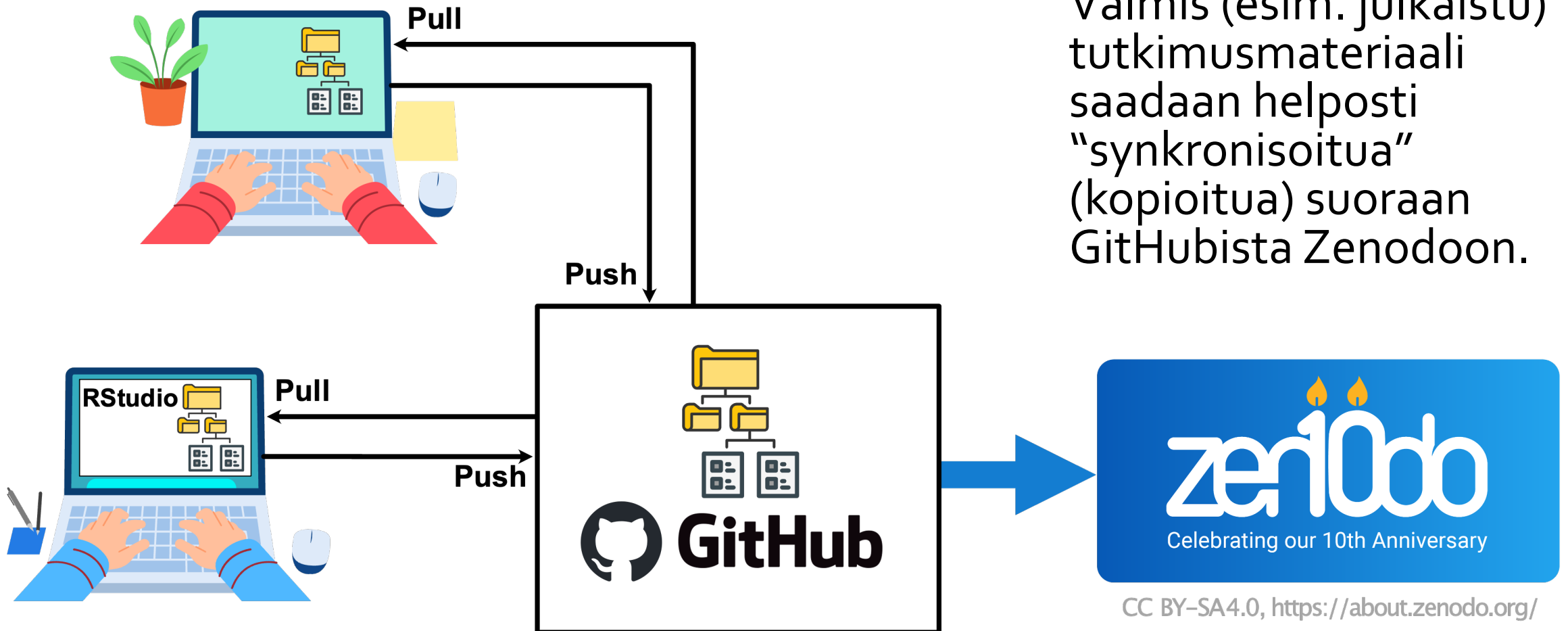


# Zenodo: CERNin OpenAIRE-projekti





# Workflow



Valmis (esim. julkaistu)  
tutkimusmateriaali  
saadaan helposti  
“synkronisoitua”  
(kopioitua) suoraan  
GitHubista Zenodoon.

Lisätietoa aiheesta saa osoitteesta  
[villelangen.com](http://villelangen.com)

Sivustolta löytyvästä YouTube-ikonista pääsee  
katsomaan luentovideon tästä aiheesta