

Tilastotiede ja Etiikka

Jouko Miettunen, Kliinisen epidemiologian professori
Elinikäisen terveyden tutkimusyksikkö
Oulun yliopisto
jouko.miettunen@oulu.fi

Sisältö

- Eettiset säännöt tilastotieteessä
- Virheet tilastoanalyysissä
- Tilastotestien oletukset
- Monitestausta
- Tilastollinen voima ja kato
- Kliiniset kokeet
- Julkaisuharha
- Lähteitä

Tilastollisten menetelmien väärä käyttö voi rikkoa useita eettisiä sääntöjä tai velvollisuuksia, kuten velvollisuutta olla rehellinen ja objektiivinen, sekä velvollisuutta välttää virheitä ja olla avoin?

Huonot tilastometodit

→ huonoa tiedettä!

Miksi tilastometodeja käytetään väärin?

- Paineita julkaista, tuottaa tuloksia, tai saada avustuksia?
- Uratavoitteita tai toiveita?
- Eturistiriitoja ja taloudellisia syitä?
- Riittämätön ohjaus tai koulutus?

Tilastollisten käytäntöjen eettiset ohjeet

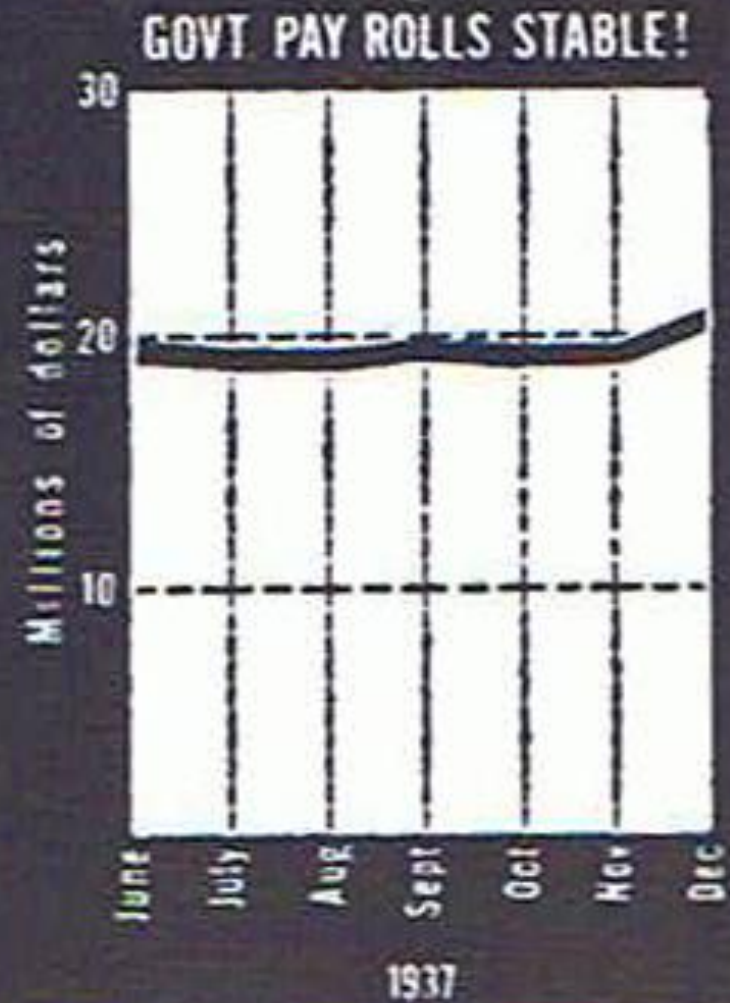
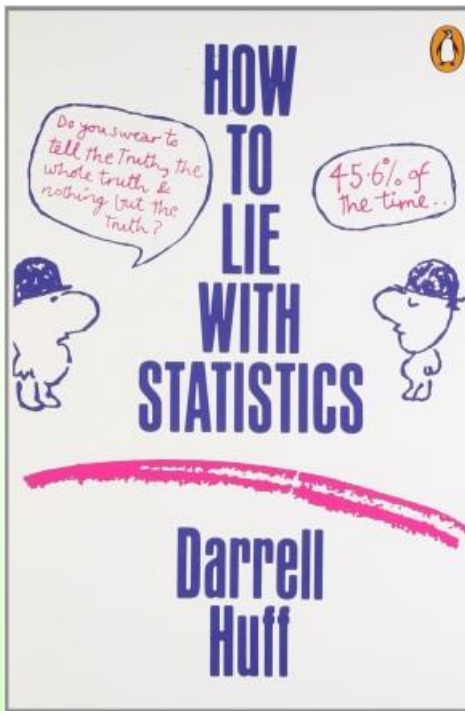
- Esitä havainnot ja tulkinnot rehellisesti ja puolueettomasti
- Vältä totuudenvastaisia, harhaanjohtavia tai epäselviä lausuntoja
- Selvitä mahdolliset taloudelliset tai muut eturistiriidat, jotka voivat vaikuttaa ammatilliseen lausuntoon (tai tutkimukseen)
- Kerää vain tarvittavat tiedot tutkimuksen tarkoitukseen
- Varmistettava, että aina kun tiedot siirretään muille henkilöille tai järjestöille siirto on sallittu ja luottamuksellisuus säilyy
 - Esim. filesender ohjelma, kirjallinen vakuutus (engagement)

Tilastollisten käytäntöjen eettiset ohjeet

- Tunnista että tilastollinen menettely saattaa jossain määrin olla tulkintakysymys
- Huomioida että asiakas (tutkija) tai työnantaja voi olla henkilö joka ei tunne tilastokäytäntöjä ja -metodeita
- Valitse tilastolliset menetelmät ilman huolta ”myönteisestä” tuloksesta (esim. $p < 0,05$ haluttuun suuntaan!)
- Esitä selvästi, tarkasti ja kattavasti vaihtoehtoiset (suositeltavat) tilastolliset menetelmät (*sensitivity analysis*) sekä kaikkien mahdollisten lähestymistapojen edut ja vaikutukset

Tilastollisten käytäntöjen eettiset ohjeet

- Tutkijoiden on käsiteltävä tilastollisia kysymyksiä kuten
 - Poikkeavat havainnot, tietojen muokkaus (imputointi) ja puhdistus, tiedon louhinta (*data mining*)
- Nämä toiminnot ovat usein käytännöllisiä, tai jopa välttämättömiä, ne on tärkeää kertoa rehellisesti ja avoimesti raportoidessa tutkimustuloksia.
- Asianmukaiset poistamiset (eksluusiot) tai puuttuvan tiedon korvaamiset (imputoinnit) voivat olla hyviä keinoja vähentää kohinaa muuttamatta signaalia, joka kuvaa tutkittavaa suhdetta tai vaikutusta.



Tilastolliset virheet

- Virheitä on helppo tehdä!
- Ei välttämättä helppo huomata!
- Tahallinen vai tahaton virhe?
- Kuka teki analyysit?
- Eroja tilasto-ohjelmien välillä?
- Kuinka usein?

Lang T. Twenty statistical errors even you can find in biomedical research articles. Croatian Med J 2004; 45:361-70.

Testien oletukset

- Normaalijakautuneisuus
 - Visuaalinen tarkastelu
 - Keskiarvo vs. mediaani
 - Regressioanalyysin oletus
 - Muunnokset (logaritmi tms.)
 - Voivat vaikeuttaa tulkintaa

Testien oletukset

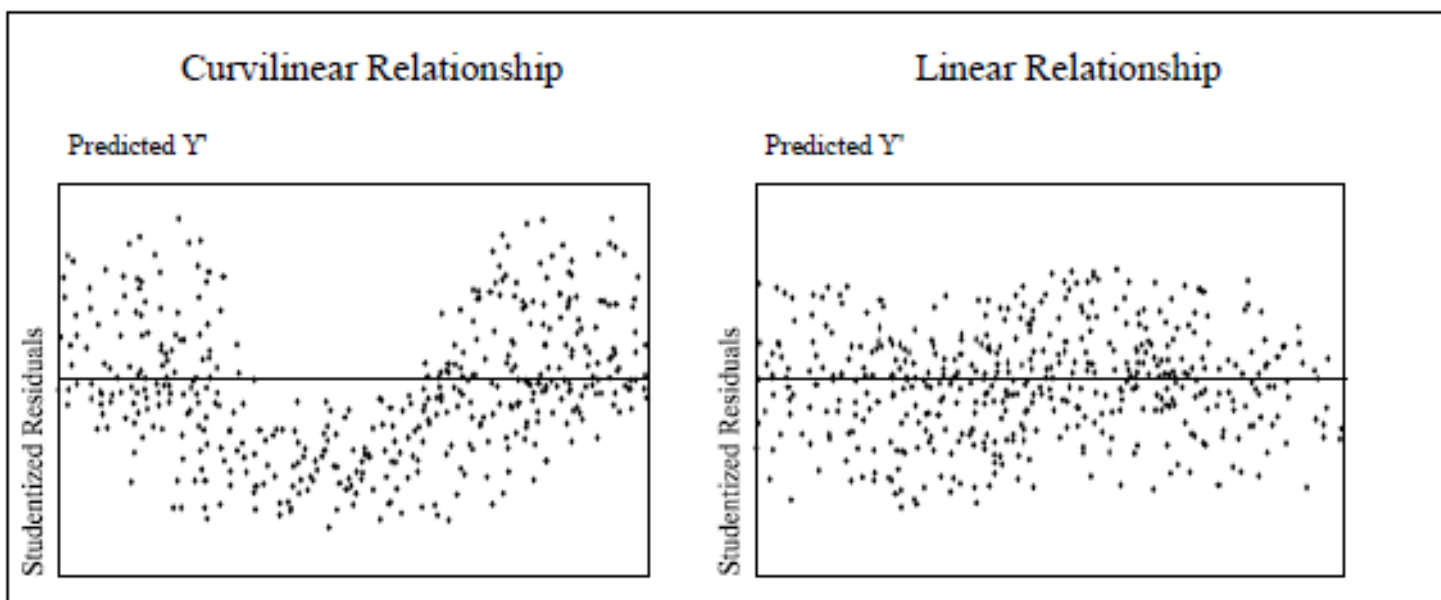
- Havaintojen riippumattomuus
 - Jos hyvin suunniteltu tutkimus ei ole yleensä ongelma
 - Isoissa tutkimuksissa ei ole ongelma
- Mittauksien luotettavuus
 - Huono luotettavuus vähentää tilastollista voimaa!

Testien oletukset

- Homoskedastisuus
 - Eli varianssin tulisi olla sama muuttujan eri tasoilla
 - Oletus regressioanalyysissä
 - Korkea heteroskedastisuus vähentää tilastollista voimaa!

Testien oletukset

- Ei-lineariset yhteydet vähentävät voimaa normaalissa regressioanalyysissa



Osborne and Waters 2002

Monitesta

If you torture the data long enough,
it will confess.

Ronald Coase



Monitestaus

- Hypoteesin asettaminen on tärkeää!
- Data louhinta tai kalastelu (*data fishing, data mining*)
- Monitestaus korjaukset
 - Bonferroni korjaus
 - Benjamini-Hochberg korjaus ym.
 - Bootstrap- menetelmä
 - Post-Hoc testaus varianssianalyysissa

Bonferroni korjaus

Dietary variable	P value
Total calories	<0.001
Olive oil	0.008
Whole milk	0.039
White meat	0.041
Proteins	0.042
Nuts	0.06
Cereals and pasta	0.074
White fish	0.205
Butter	0.212
Vegetables	0.216
Skimmed milk	0.222
Red meat	0.251
Fruit	0.269
Eggs	0.275
Blue fish	0.34
Legumes	0.341
Carbohydrates	0.384
Potatoes	0.569
Bread	0.594
Fats	0.696
Sweets	0.762
Dairy products	0.94
Semi-skimmed milk	0.942
Total meat	0.975
Processed meat	0.986

- Yksinkertainen, mutta konservatiivinen menetelmä
- Tilastollisen merkitsevyyden raja 0,05 muutetaan jakamalla se testien lukumäärällä
- Esimerkki:
 - 25 testiä
 - Ilman korjausta 5 merkitsevää ($p < 0,05$) muuttujaa
 - Korjauksen kanssa yksi merkitsevä ($p < 0,002$) muuttuja

Benjamini-Hochberg korjaus

Dietary variable	P value	Rank	Q=0.25 Q=0.05	
			$(i/m)Q$	
Total calories	<0.001	1	0.010	0.002
Olive oil	0.008	2	0.020	0.004
Whole milk	0.039	3	0.030	0.006
White meat	0.041	4	0.040	0.008
Proteins	0.042	5	0.050	
Nuts	0.060	6	0.060	
Cereals and pasta	0.074	7	0.070	
White fish	0.205	8	0.080	
Butter	0.212	9	0.090	
Vegetables	0.216	10	0.100	
Skimmed milk	0.222	11	0.110	
Red meat	0.251	12	0.120	
Fruit	0.269	13	0.130	
Eggs	0.275	14	0.140	
Blue fish	0.34	15	0.150	
Legumes	0.341	16	0.160	
Carbohydrates	0.384	17	0.170	
Potatoes	0.569	18	0.180	
Bread	0.594	19	0.190	
Fats	0.696	20	0.200	
Sweets	0.762	21	0.210	
Dairy products	0.94	22	0.220	
Semi-skimmed milk	0.942	23	0.230	
Total meat	0.975	24	0.240	
Processed meat	0.986	25	0.250	

- Merkitsevyydet järjestetään (Rank, i)
- Benjamini-Hochberg kriittinen arvo lasketaan kaavalla $(i/m)Q$, missä i = järjestys, m = testien lukumäärä ja Q = valittu *false discovery rate* (hylkäysvirheaste, kuinka monta väärää positiivista löydöstä hyväksytään)
- Esimerkki: $Q=0,25$
(usein $Q=0,05$ tai $0,10$)
- Esimerkissä ensimmäiset 5 ovat merkitseviä (myös "whole milk" ja "white meat" vaikkakin $P >$ kriittinen arvo, koska P on pienempi kuin P muuttujassa "proteins")

Tulkinta

- Tilastollinen merkitsevyys vs. vaikutus (*effect*) ?
- *“The difference between ‘significant’ and ‘not significant’ is not itself statistically significant”*
- *“Absence of evidence is not evidence of absence”*

American Statistical Association releases statement on statistical significance and p-values

Provides principles to improve the conduct and interpretation of quantitative science

1. P-values can indicate how incompatible the data are with a specified statistical model.
2. P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.
3. Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.
4. Proper inference requires full reporting and transparency.
5. A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.
6. By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis.

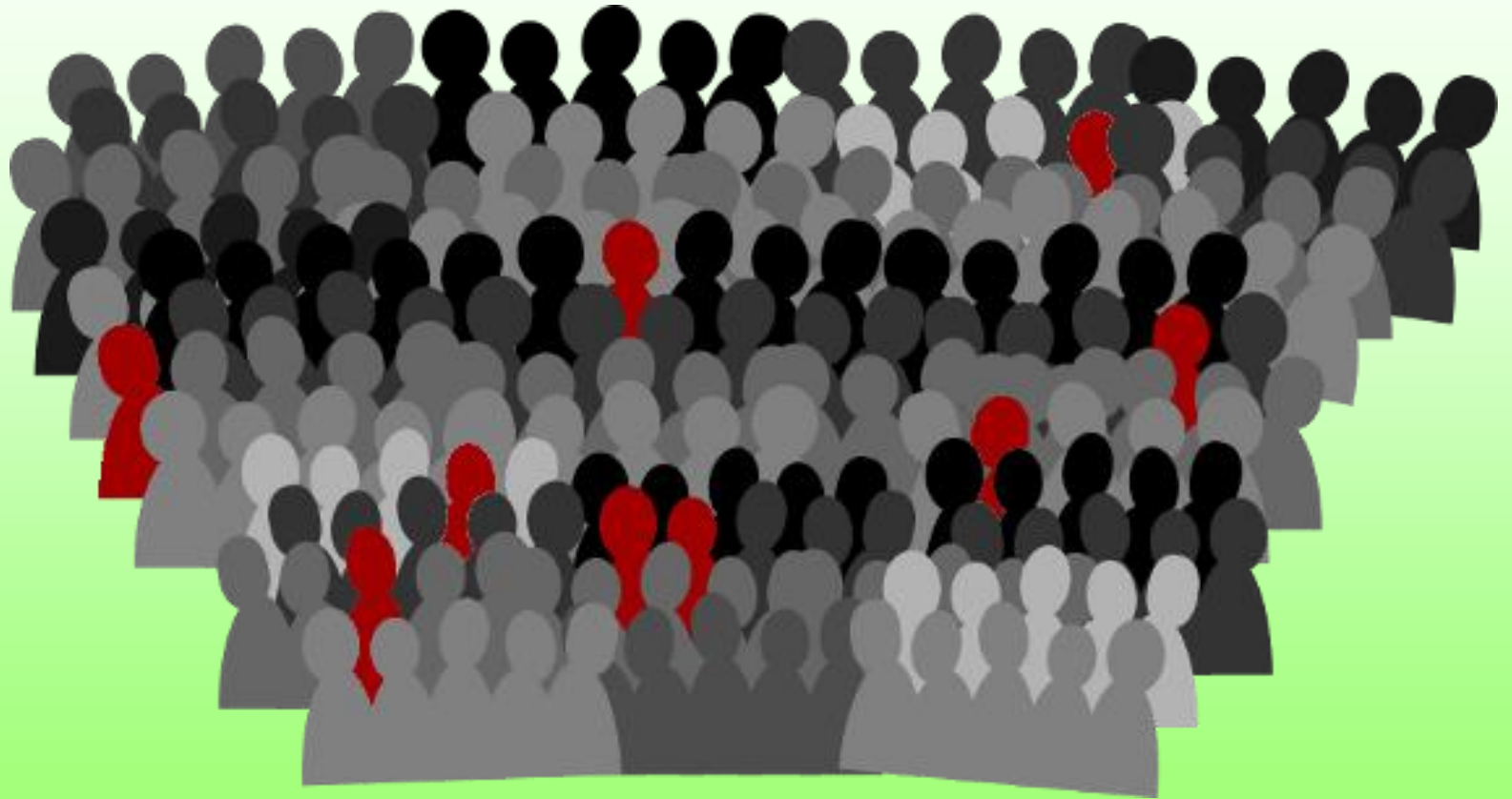
Mark DB, et al. Understanding the Role of P Values and Hypothesis Tests in Clinical Research. JAMA Cardiol. 2016; 1(9):1048-1054.

- Terapeuttisten kliinisten tutkimusten tulkinassa keskitytään ensisijaisesti hoitovaikutukseen. Tätä arvioidaan vaikutuksen suuruudella (*effect size*) ja sen tarkkuudella.
- Satunnaistetussa tutkimuksessa vaikutuksen koko on usein tiivistetty käyttämällä esim. vetosuhdetta (*odds ratio*), riskitiheyssuhdetta (*hazard ratio*) tai suhteellista riskiä (*relative risk*).
- Kuinka suuri hoitovaikutus tulisi olla arvioidaan kliinisen harkinnan mukaan.
- Vaikutuksen suuruus ja sen tarkkuus voivat olla molemmat hyödyllisiä, mutta tulee ottaa huomioon kliiniset vaikutukset.

Table 2. Common Misconceptions About P Value

Misconception	Comment
P value equals the probability that the null hypothesis is true.	P value is computed by assuming the null hypothesis is true.
P value equals the probability that the observed effect is due to "the play of chance."	P value is defined as the probability of a difference (effect) as large as that observed or larger if the null hypothesis is true. Even if the difference observed is consistent with a simple chance mechanism, other more complex explanations are also possible, and nothing in P value calculation allows one to conclude that this is the best or most likely explanation for the observed differences.
P value $\leq .05$ means the null hypothesis is false. P value $> .05$ means the null hypothesis is true.	P value is computed assuming the null hypothesis is true. It is not the probability that the null hypothesis is either true or false.
P value $\leq .05$ identifies a clinically or scientifically important difference (effect). P value $> .05$ rules out a clinically or scientifically important difference (effect).	Clinical or scientific importance of study results is a judgment integrating multiple elements, including effect size (expected and observed), precision of estimate of effect size, and knowledge of prior relevant research. At best, P value has a minor role in shaping this judgment.
A small P value indicates study results are reliable and likely to replicate.	P value provides no information about whether a given study result can be reproduced in a second, replication experiment. There are many other factors that must be considered in judging the reliability of study results. Understanding what works in medicine is a process and not the product of any single experiment.

Voima



Voima-analyysit

- *Power analyses*
- Hyvin tehty voima-analyysi tulisi olla osa kaikkia tutkimussuunnitelmia!
- Liian paljon tutkimusta liian pienillä aineistoilla → **eettinen ongelma!**

Voima-analyysit

- Otokoot kliinisissä kokeissa ovat yleensä pieniä, esim.
 - Nivelreuma: mediaani otoskoko 54 potilasta (196 kliinistä koetta)
 - Ihotaudit: 46 potilasta (73 koetta)
 - Skitsofrenia: 65 potilasta (2000 koetta)
- Otokoko ei yleensä perustu mihinkään!
- *Post hoc* voima-analyysit ovat turhia, luottamusvälit kertovat voimasta

Voima-analyysit

Tarvittavat tiedot

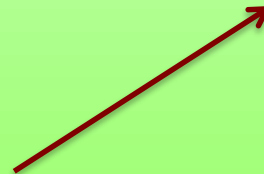
- Tutkittavien määrä
- Tärkeimmän vasteen esiintyvyys
(odotettu vastetapahtumien määrä)

Tehtävät oletukset

- Vaikutuksen koko
- Tilastollisen merkitsevyyden raja (α)
- Tilastollinen voima ($1-\beta$)

- Alfa eli merkitsevyystaso (esim. 0.05 tai 5%)
 - Todennäköisyys että ero löytyy vaikka sitä ei ole olemassa (väärä positiivinen löydös)
- Beta eli voima (esim 0.8 tai 80%)
 - Todennäköisyys että löydetty ero on todellinen
- Välianalyysit (*interim analyses*) ovat etukäteen (*a priori*) suunniteltuja analyyyseja meneillään olevaan kliniseen kokeeseen
 - Syyt eettisiä tai taloudellisia
 - α – virhe kasvaa
 - Voima voi jäädä riittämättömäksi?

- Erilaisia asetelmia
 - Keskiarvojen erot
 - Osuuksien erot
 - Monimuuttujamallit
- Erilaisia ohjelmia
 - Nettilaskurit
 - Spesifit ohjelmat
 - SPSS sample power, ...



The screenshot shows the 'Test of equality of two proportions' dialog box in SPSS. The window has a title bar with a standard icon, the text 'Test of equality of two pro...', and window control buttons. Below the title bar is a menu bar with 'Options' and 'Help'. The main area contains several sections with sliders and input fields:

- p1 = .325**: A slider on a scale from 0 to 1 with major ticks at 0, .2, .4, .6, .8, and 1. The slider is positioned at approximately 0.325.
- p2 = .52**: A slider on a scale from 0 to 1 with major ticks at 0, .2, .4, .6, .8, and 1. The slider is positioned at approximately 0.52.
- n1 = 100**: A slider on a scale from 0 to 140 with major ticks every 20 units. The slider is positioned at 100.
- n2 = 100**: A slider on a scale from 0 to 140 with major ticks every 20 units. The slider is positioned at 100.
- Equal ns**: A checkbox that is checked.
- Alpha = .05**: A slider on a scale from 0 to .07 with major ticks every .01 units. The slider is positioned at .05.
- Power = .7587**: A slider on a scale from 0 to 1 with major ticks at 0, .2, .4, .6, .8, and 1. The slider is positioned at approximately 0.7587.
- Continuity corr.**: A checkbox that is checked.
- Alternative**: A dropdown menu showing 'p1 != p2'.

<http://homepage.stat.uiowa.edu/~rlenth/Power/index.html>

- Tutkimusasetelma
 - Kliinisissä kokeissa pienempi otoskoko on riittävä
- Varianssi
 - Suurempi varianssi vaatii isompaa aineistoa ryhmäerojen löytämiseksi
- Seurantatutkimukset
 - Huomioi kato (muutot, kuolemat, kieltäytymiset)

Suresh KP & Chandrashekara S. J Hum Reprod Sci 2012; 5: 7–13.

Puuttuva tieto

- Henkilöt eivät osallistu?
- Henkilöt eivät tule seurantaan?
- Muuttujissa puuttuvaa tietoa?
- Tulisiko nämä asiat huomioida?

Kadon raportointi

- Onko tietoa osallistumattomista?
 - Perustutkimus, rekisterit, kliininen data?
- Onko aiemmin esitetty aineiston kadosta tuloksia?
- Voidaan raportoida osallistuman eri muuttujien avulla, esim. Sukupuoli, koulutus, sosioekonominen asema...
- Myös jatkuvia muuttujia (esim. hoitopäivät) voidaan vertailla osallistuneiden ja osallistumattomien kesken

Esimerkki raportoinnista

Miettunen J, Murray GK, Jones PB, Mäki P, Ebeling H, Taanila A, Joukamaa M, Savolainen J, Törmänen S, Järvelin MR, Veijola J, Moilanen I.

Longitudinal associations between childhood and adulthood externalizing and internalizing psychopathology and adolescent substance use.

Psychol Med. 2014 Jun; 44(8):1727-38.

Attrition analysis

In drop-out analyses for the 15–16-year follow-up we used register-based information. Of the adolescents who were alive at the time of the follow-up, 67.0% participated. Fewer males than females participated in the follow-up study (64% *v.* 71%; χ^2 test, $p < 0.001$), as did participants living in urban areas (66% *v.* 71%, $p < 0.001$). Adolescents with a parental history of psychiatric disorder (58% *v.* 69%, $p < 0.001$) participated less frequently than others. We weighted our adjusted analyses by these variables using inverse probability weighting (Haukoos & Newgard, 2007), that is on the proportions of these participants in the whole target population including non-participants. All the statistically significant ORs of unweighted analyses were

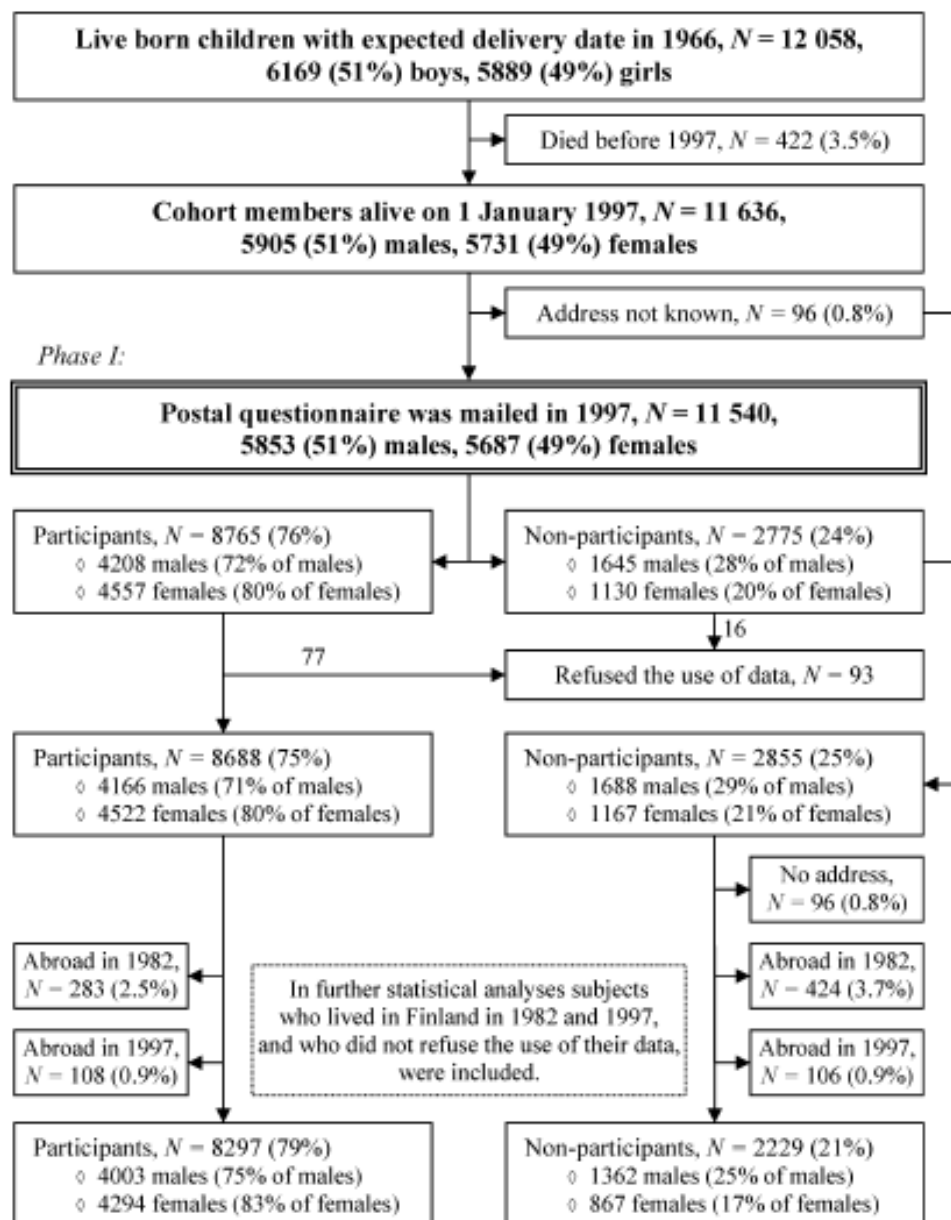


Figure 1. Data collection of phase I (postal questionnaire) in a health survey conducted in 1997 in the Northern Finland 1966 Birth Cohort.

Aineiston painottaminen

- Painotus sen mukaan miten ryhmässä on osallistuttu
- Tarvitaan tieto osallistumisesta esim. perustutkimuksesta tai rekistereistä
- Sama menetelmä suunnitellusti käytössä erilaisissa seulontatutkimuksissa kun arvioidaan esim. eri sairauksien yleisyyttä väestössä

Marianne Haapea, Non-response and information bias in population-based psychiatric research – The Northern Finland 1966 Birth Cohort study. Acta Universitatis Ouluensis, D 1049, University of Oulu, Oulu 2010.

Aineiston painottaminen

Esimerkki

- Painotus sen mukaan miten siviilisääty ja työssäolo vaikuttivat osallistumiseen

naimisissa	töissä	osallistuneet (%)	osallistuneet (n)	paino	painotettu n
ei	ei	44,1%	56	1,511	85
ei	kyllä	62,2%	350	1,072	375
kyllä	ei	53,4%	31	1,247	39
kyllä	kyllä	74,8%	566	0,891	504
aineisto		66,6%	1003	ka. 1	1003

- Painotus voidaan huomioida useissa erilaisissa tilastoanalyyyseissa

Moni-imputointi

Yksittäisen puuttuvan arvon korvaaminen

- Erilaisia yksittäisiä korvaamistapoja (keskiarvo, regressiomallin mukainen)
- Moni-imputoinnissa korvaaminen tehdään samanaikaisesti usealla arvolla (tuloksena useita dataa)

Moni-imputointi

- Muodostetaan useita dataja, esim. 5-20
- Analyysien tilastotuloksena tulee tulos kaikille datoille ja niistä yhteenveto
- Etuna että vähentää puuttuvasta tiedosta tulevaa harhaa, lisää luotettavuutta ja analyysien tarkkuutta

Moni-imputointi

- Kaikki puuttuva data korvataan määrättyllä tavalla käyttäen hyväksi tietoa muista määritellyistä muuttujista

Imputation_	X1	X2	kato43y
Original data	71	69	tutkittu
Original data	35	57	tutkittu
Original data	60	80	tutkittu
Original data	90	82	
Original data	81	×	.
Original data	40	29	
Original data	50	×	.
Original data	35	46	
Original data	50	60	
Original data	51	×	.

Imputation_	X1	X2
1	35	57
1	60	80
1	90	82
1	81	→ 44
1	40	29
1	50	→ 79
1	35	46
1	50	60
1	51	→ 69
2	51	58
3	51	71

Imputation_	X1	X2	kato43y
10	35	57	tutkittu
10	60	80	tutkittu
10	90	82	tutkittu
10	81	68	kato
10	40	29	tutkittu
10	50	31	kato
10	35	46	tutkittu
10	50	60	tutkittu
10	51	34	kato

Keskiarvojen vertailu kahden ryhmän välillä

			Group Statistics			
Imputation Number		Life_Satisfaction	N	Mean	Std. Deviation	Std. Error Mean
Original data	HSCL_Depression_New	>= 3	48	1,7444	,37165	,05364
		< 3	625	1,2789	,25299	,01012
1	HSCL_Depression_New	>= 3	81	1,7695	,47153	,05239
		< 3	844	1,2998	,28657	,00986
2	HSCL_Depression_New	>= 3	81	1,7770	,47131	,05237
		< 3	844	1,2992	,28753	,00990
3	HSCL_Depression_New	>= 3	81	1,7761	,49073	,05453
		< 3	844	1,2996	,28676	,00987
4	HSCL_Depression_New	>= 3	81	1,7671	,46597	,05177
		< 3	844	1,3007	,28883	,00994
5	HSCL_Depression_New	>= 3	81	1,7663	,46105	,05123
		< 3	844	1,2979	,28150	,00969
Pooled	HSCL_Depression_New	>= 3	81	1,7712		,05276
		< 3	844	1,2994		,00992

Elämäntyytyväisyys: hyvä <3 vs. huono >= 3

Logistinen regressioanalyysin imputoidut tulokset

Variables in the Equation									
Imputation Number			B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B) Lower Upper
Original data	Step 1 ^a	Gender(1)	,414	,200	4,268	1	,039	1,513	1,021 2,241
		Basic education(1)	,269	,207	1,689	1	,194	1,309	,872 1,965
		Constant	-1,834	,214	73,634	1	,000	,160	
1	Step 1 ^a	Gender(1)	,464	,165	7,889	1	,005	1,590	1,150 2,197
		Basic education(1)	,321	,171	3,519	1	,061	1,379	,986 1,930
		Constant	-1,777	,178	100,237	1	,000	,169	
2	Step 1 ^a	Gender(1)	,478	,166	8,321	1	,004	1,612	1,165 2,230
		Basic education(1)	,314	,172	3,351	1	,067	1,369	,978 1,916
		Constant	-1,787	,178	100,829	1	,000	,168	
3	Step 1 ^a	Gender(1)	,491	,166	8,766	1	,003	1,635	1,181 2,263
		Basic education(1)	,307	,172	3,186	1	,074	1,359	,970 1,903
		Constant	-1,796	,178	101,424	1	,000	,166	
4	Step 1 ^a	Gender(1)	,478	,166	8,321	1	,004	1,612	1,165 2,230
		Basic education(1)	,314	,172	3,351	1	,067	1,369	,978 1,916
		Constant	-1,787	,178	100,829	1	,000	,168	
5	Step 1 ^a	Gender(1)	,431	,165	6,856	1	,009	1,539	1,115 2,126
		Basic education(1)	,287	,171	2,826	1	,093	1,333	,953 1,863
		Constant	-1,737	,176	97,150	1	,000	,176	
Pooled	Step 1 ^a	Gender(1)	,468	,167			,005	1,597	1,151 2,217
		Basic education(1)	,309	,172			,073	1,362	,972 1,908
		Constant	-1,777	,179			,000	,169	,119 ,241

a. Variable(s) entered on step 1: Gender, Basic education.

Lieberman JA, et al. Antipsychotic drug effects on brain morphology in first-episode psychosis. Arch Gen Psychiatry. 2005 Apr;62(4):361-70.

OBJECTIVE:

To test a priori hypotheses that olanzapine-treated patients have less change over time in whole brain gray matter volumes and lateral ventricle volumes than haloperidol-treated patients.

DESIGN:

Longitudinal, randomized, controlled, multisite, double-blind study. Patients treated and followed up for up to 104 weeks. Neurocognitive and magnetic resonance imaging (MRI) assessments performed at weeks 0 (baseline), 12, 24, 52, and 104.

INTERVENTIONS:

Random allocation to a conventional antipsychotic, haloperidol (2-20 mg/d), or an atypical antipsychotic, olanzapine (5-20 mg/d).

RESULTS:

Of 263 randomized patients, 161 had baseline and at least 1 postbaseline MRI evaluation. Haloperidol-treated patients exhibited significant decreases in gray matter volume, whereas olanzapine-treated patients did not. A matched sample of healthy volunteers (n = 58) examined contemporaneously showed no change in gray matter volume.

CONCLUSIONS:

Patients with first-episode psychosis exhibited a significant between-treatment difference in MRI volume changes. Haloperidol was associated with significant reductions in gray matter volume, whereas olanzapine was not. Post hoc analyses suggested that treatment effects on brain volume and psychopathology of schizophrenia may be associated. The differential treatment effects on brain morphology could be due to haloperidol-associated toxicity or greater therapeutic effects of olanzapine.

Table 2. Changes in MRI Volumes for Primary Regions of Interest by Treatment Group (Baseline to Weeks 12, 24, 52, and 104)

Observed Case Mean Changes From Baseline and Mixed-Model <i>P</i> Values										
ROI	Therapy	Baseline		Week 12				Week 24		
		N	Mean (SE), cm ³	N	Mean (SE), cm ³	<i>P</i> Value*	<i>P</i> Value†	N	Mean (SE), cm ³	<i>P</i> Value*
CN	Olz	80	8.89 (0.19)	71	-0.36 (0.13)	.12	.18	64	-0.23 (0.15)	.03
	Hal	77	8.62 (0.15)	68	0.03 (0.13)		.92	46	-0.04 (0.14)	
	Con	52	9.09 (0.16)	52	-0.01 (0.15)					

ROI	Therapy	Week 52				Week 104		
		N	Mean (SE), cm ³	<i>P</i> Value*	<i>P</i> Value†	N	Mean (SE), cm ³	<i>P</i> Value*
CN	Olz	42	-0.52 (0.16)	.02	.003	25	-0.30 (0.28)	.04
	Hal	31	0.12 (0.18)		.58	10	0.39 (0.37)	
	Con	44	0.17 (0.13)					

Abbreviations: CN, caudate nucleus; Con, control; Hal, haloperidol; LV, lateral ventricle; MRI, magnetic resonance imaging; Olz, olanzapine; ROIs, regions of interest; TV, third ventricle; WB, whole brain; WBF, whole brain fluid; WBGm, whole brain gray matter; WBWM, whole brain white matter.

**P* values are from an *F* test at each time point, comparing olanzapine- and haloperidol-treated patients with respect to mean change for each ROI from baseline, within the context of a mixed model. Each model contained a random intercept and therapy, therapy by week, therapy by week squared, and duration of illness fixed effects. Each model was fit assuming a compound symmetric covariance structure with different variance components for each therapy.

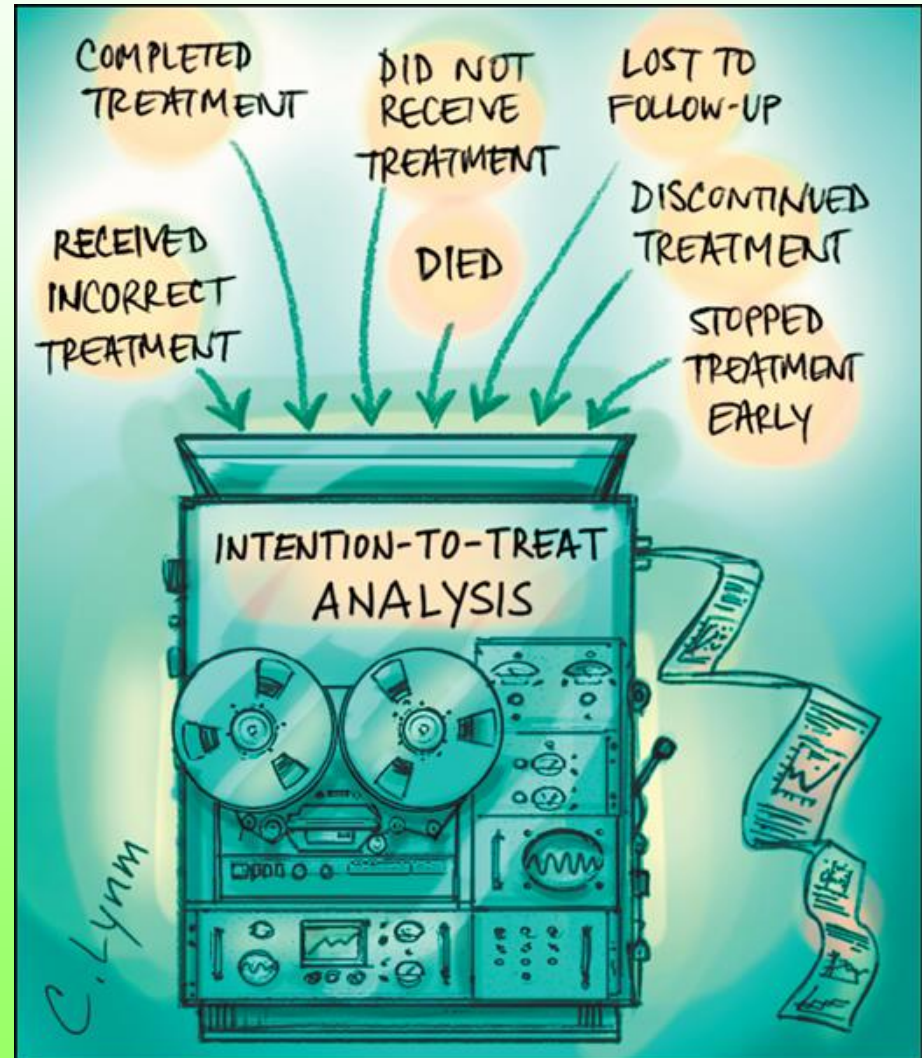
†*P* values are from an *F* test at each time point, comparing olanzapine-treated and control patients or haloperidol-treated and control patients with respect to mean change for each ROI from baseline, within the context of a mixed model. Each model contained a random intercept fixed effects for therapy and the therapy × week interaction. The model was fit assuming a compound symmetric covariance structure with different variance components for each of the 3 therapy groups.

Kliiniset kokeet

Intention-to-treat

- Aineisto analysoidaan satunnaistamisen mukaisesti

Satunnaistamisen
vaikutus säilyy!



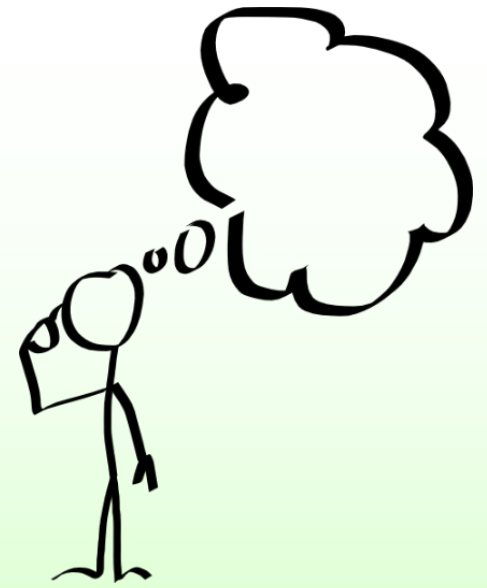
Menetelmät

- Intervention valinta
 - Perusteet?
 - Tutkimuksen kesto?
 - Yleistettävyyys?
- Ensisijainen vs. toissijainen vaste
- Alaryhmien analyysit

Tulokset

- Tilastolliset menetelmät tulisi esittää selkeästi
- Luottamusvälit tulisi esittää kun kuvataan vaikutuksen varmuutta
- Tarkat p-arvot (ei <0.05 tms.)

Pohdinta



- Heikkoudet?
- Vertailu aiempiin tutkimuksiin?
- Yleistettävyyys?
- Tulkinta?
- Johtopäätökset?

Julkaisuharha

- Tutkimuksen perusongelma
- Ongelma meta-analyyseissa
ja alkuperäistutkimuksissa
- Johtuu tutkijoista ja lehdistä!

Ei-merkitsevät tulokset



<http://mchankins.wordpress.com/2013/04/21/still-not-significant-2/>

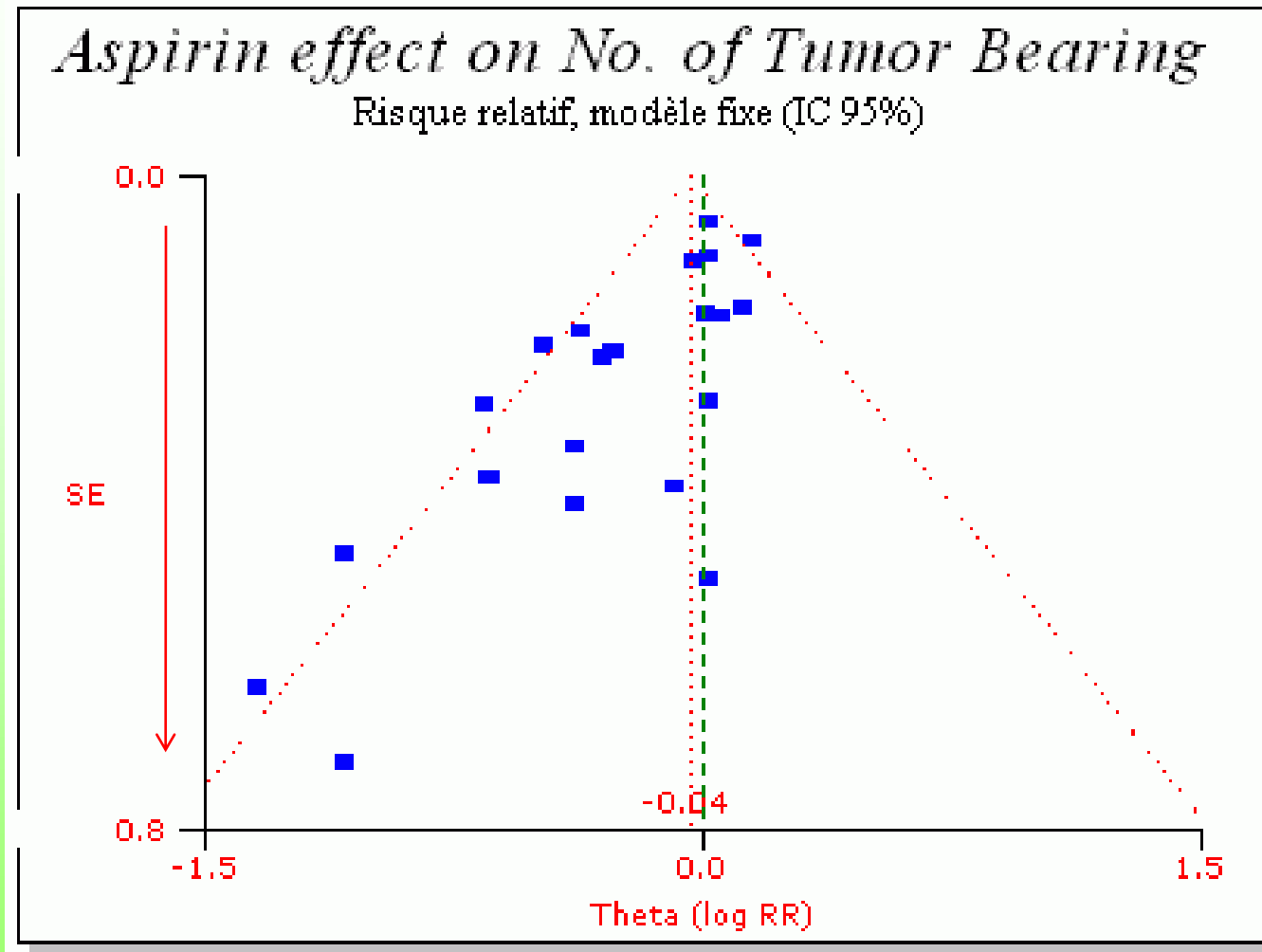
Miten ei tule raportoida ei-merkitseviä tuloksia!

- a clear, strong trend ($p=0.09$)
- an encouraging trend ($p<0.1$)
- an important trend ($p=0.066$)
- approached conventional levels of significance ($p<0.10$)
- below (but verging on) the statistical significant level ($p>0.05$)
- difference was apparent ($p=0.07$)
- essentially significant ($p=0.10$)
- failed to reach significance on this occasion ($p=0.09$)
- flirting with conventional levels of significance ($p>0.1$)
- leaning towards significance ($p=0.15$)
- narrowly escaped significance ($p=0.08$)
- not conventionally significant ($p=0.089$), but..
- not significant in the narrow sense of the word ($p=0.29$)
- on the very fringes of significance ($p=0.099$)

Meta-analyysit

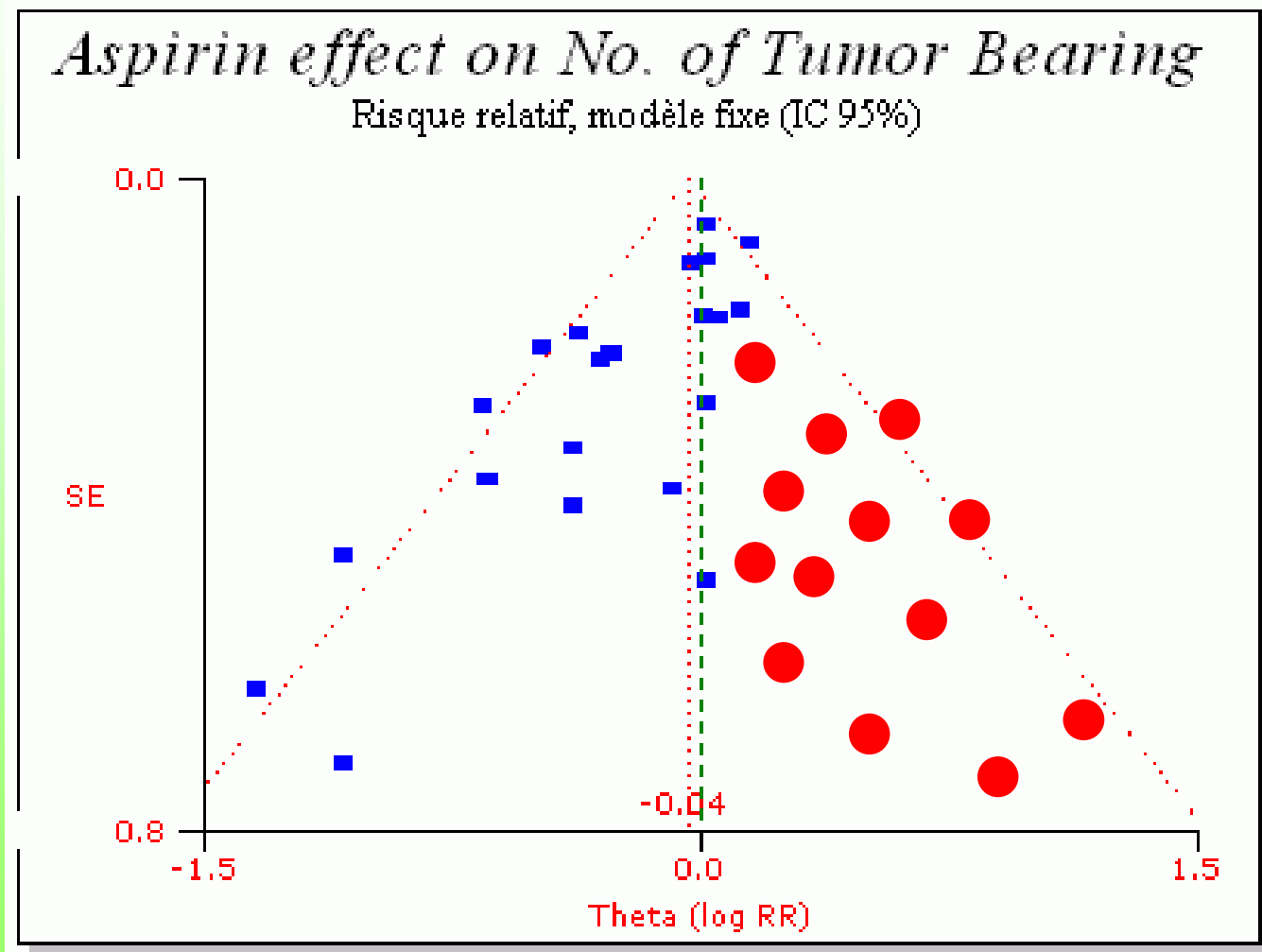
- Julkaisuharhaa voidaan arvioida funnel plotin avulla
 - Oletetaan että tarkimmat (usein suurimmat) tutkimukset antavat keskimääräiset tulokset, muut tutkimukset sijoittuvat molemmille puolille keskiarvoa
- "Trim and fill"

Funnel Plot



Trim and Fill

- Voidaan korjata julkaisuharhaa



Why most published research findings are false?

1. The smaller the studies conducted in a scientific field, the less likely the research findings are to be true.
2. The smaller the effect sizes in a scientific field, the less likely the research findings are to be true.
3. The greater the number and the lesser the selection of tested relationships in a scientific field, the less likely the research findings are to be true.
4. The greater the flexibility in designs, definitions, outcomes, and analytical modes in a scientific field, the less likely the research findings are to be true.
5. The greater the financial and other interests and prejudices in a scientific field, the less likely the research findings are to be true.
6. The hotter a scientific field (with more scientific teams involved), the less likely the research findings are to be true.

Ioannidis JPA. Why most published research findings are false. PLOS Medicine 2005;2:e124.

Ratkaisuja eettisiin ongelmiin?



- Enemmän tilastotieteen opetusta?
- Enemmän ohjeita?
- Ryhmätyö ja verkostoituminen?
- Tutkimusten rekisteröinti?
- Julkisesti saatavilla olevat aineistot?
- Sensitiivisyysanalyysit?

Kirjallisuutta

- Altman DG. Statistics and ethics in medical research. Misuse of statistics is unethical. Br Med J 1980; 281: 1182–4.
- DeMets DL. Statistics and ethics in medical research. Science and Engineering Ethics 1999; 5:97-117.
- Easterbrook PJ, et al. Publication bias in clinical research. Lancet 1991; 337:867–72.
- Gardenier J & Resnik D. The misuse of statistics: concepts, tools, and a research agenda. Accountability in Research: Policies and Quality Assurance 2002; 9:65-74.
- Greenland S, et al. Statistical tests, p-values, confidence intervals, and power: a guide to misinterpretations. Draft 1 Mar 2016.
- Ioannidis JPA. Why most published research findings are false. PLOS Medicine 2005;2:e124.
- Lang T. Twenty statistical errors even you can find in biomedical research articles. Croatian Med J 2004; 45:361-70.

Kirjallisuutta

- Mark DB, et al. Understanding the role of p values and hypothesis tests in clinical research. JAMA Cardiol. 2016; 1(9):1048-1054.
- Moher D, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010; 340:c869.
- Osborne JW & Waters E. Four assumptions of multiple regression that researchers should always test. Practical Assessment, Research, and Evaluation 2002: 8 (available online).
- Suresh KP & Chandrashekara S. Sample size estimation and power analysis for clinical research studies. J Hum Reprod Sci 2012; 5: 7–13.
- Wasserstein RL & Lazar NA. The ASA's statement on p-values: context, process and purpose. DOI: 10.1080/00031305.2016.1154108

Kiitos!

jouko.miettunen@oulu.fi
www.joukomiettunen.net

