

# Ei vaikutusta eikä eroa – ei oikeaa tilastollista päättelyä



Miten tulkitaan tilannetta, kun tilastollista merkitsevyyttä ei havaita? Hyvä tapa on hyödyntää luottamusvälejä.

**K**ahden ryhmän vertaaminen on keskeisimpiä menetelmiä hoitojen tai interventioiden tehon arvioinnissa lääketieteellisessä tutkimuksessa. Vertailtavat muuttujat voivat olla jatkuvia, kuten kipu, tai kategorisia, kuten tietyn komplikaation saaneiden potilaiden osuus. Jatkuvia muuttujia vertaillaan useimmiten Studentin t-testillä. Kategorisia muuttujia voidaan vertailla käyttämällä Fisherin tarkkaa testiä tai  $\chi^2$ -testiä. Lisäksi jatkuville ja kategorisille muuttujille voidaan laskea kerroinsuhde (odds ratio, OR) tai riskisuhde (risk ratio, RR). Aika-aineistoon pohjautuen voidaan laskea myös vaarasuhde (hazard ratio, HR).

Yleensä tilastollisen testin ja tunnusluvun, kuten t-testin tai kerroinsuhteen yhteydessä raportoidaan p-arvo, jonka perusteella tehdään tilastollinen päätely. Jos p-arvo on alle yleisesti käytetyn raja-arvon 0,05, voidaan todeta, että ryhmien välillä on tilastollisesti merkitsevä ero. Yleinen päätelmä tämän jälkeen on todeta, että tutkitulla menetelmällä on tilastollista vaikutusta päätemuuttujaan.

Jos p-arvo jää yli raja-arvon 0,05, voidaan todeta, että ryhmien välillä ei ole tilastollista eroa. Tämän jälkeen hyvin

usein päätellään, että ryhmien välillä ei ole eroa tai että tulosmuuttuja oli yhtä suuri molemmissa ryhmissä. Korrelaation osalta todetaan usein, että muuttujilla ei ole yhteyttä. Huolimatta päätelmien yleisyydestä mikään noista päätelmistä ei ole oikein.

Kaikki edellä mainitut testit pohjautuvat ns. nollahypoteesin merkitsevyyden testaamiseen. Studentin t-testissä nollahypoteesina on yleensä ryhmien keskiarvojen yhtäsuuruus. Riski- tai kerroinsuhteen p-arvon määrittäminen perustuu nollahypoteesiin, joka määrittää, että suhdeluku on 1. Kun p-arvo on alle 0,05, päätelmänä on nollahypoteesin hylkääminen. p-arvon ollessa yli 0,05 nollahypoteesia ei voida hylätä. Tämä ei tarkoita, että suurempi p-arvo kuin 0,05 vahvistaisi nollahypoteesin (1,2). Ainoa päätelmä, joka nollahypoteesin testämisessä voidaan tehdä, on nollahypoteesin hylkääminen, ei koskaan sen vahvistaminen tai todentaminen. Ryhmillä saattaa olla eroa tai riskisuhde voi olla eri suuri kuin 1, mutta otoskoko ei riittänyt sen toteamiseen.

”Ei eroa” tai ”ei vaikutusta” ei siis koskaan ole oikea tapa tilastollisessa päätelyssä (1–3). Miten sitten tulkitaan tilannetta, kun tilastollista merkitsevyyttä ei havaita? Hyvä tapa on hyödyntää luottamusvälejä.

Luottamusvälille on monta tulkintaa. Kätevinä lienee ajatella luottamusväliä arvojoukkona, jonka kanssa muuttuja, kuten ryhmien välinen ero, on parhaiten yhtenevä (compatible) tai jonka arvoille p-arvo olisi yli 0,05 (2,4). ”Ei eroa”-päätelmän sijaan olisi tärkeintä katsoa,

miten kliinisesti merkityksellinen ryhmien välinen ero tai riskisuhteen arvo sijoittuu luottamusvälille. Jos arvo sijaitsee luottamusvälillä, ryhmien väliltä ei voida pois sulkea kliinisesti merkitsevää eroa. Jos väli on hyvin kapea ja kliinisesti merkitsevä arvo ei sisälly siihen, on päätelmämahdollisuus jo erilainen. Ryhmien yhtäsuuruutta ei koskaan voida todeta, oli luottamusväli kuinka kapea tahansa.

Käytettäessä kerroin-, riski- tai vaarasuhdetta päätely on yleisesti hankalampaa, koska niille on harvemmin määritelty kliinisesti merkitseviä arvoja. Tästä huolimatta tilastollisessa päätelyssä pitää hyödyntää p-arvon ja itse piste-estimaatin lisäksi luottamusvälejä.

Tulkinta koskee myös korrelaatioker-toimia ja muita regressioker-toimia, koska niissä käytetään myös nollahypoteesin testausta. ●

**ALEKSI REITO**  
dosentti, ortopediaan ja  
traumatologiaan erikoistuva lääkäri  
TAYS, ortopedia

## SIDONNAISUUDET

Aleksi Reito: Apurahat (valtion tutkimusrahoitus 2018), luontopalkkiot (Orion).

## KIRJALLISUUTTA

- Greenland S ym. Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. *Eur J Epidemiol* 2016;31:337–50.
- Amrhein V ym. The earth is flat ( $p > 0.05$ ): significance thresholds and the crisis of unreplicable research. *PeerJ* 5:e3544.
- Marson Smith PR, Ware L, Adams C, Chalmers I. Claims of “no difference” or “no effect” in Cochrane and other systematic reviews. *BMJ evidence-based Med* 2020. <https://doi.org/10.1136/bmjebm-2019-111257>.
- Chow ZR, Greenland S. Semantic and Cognitive Tools to Aid Statistical Inference: Replace Confidence and Significance by Compatibility and Surprise. [arXiv:1909.08579 \[stat.ME\]](https://arxiv.org/abs/1909.08579).