

Luottoa luottamusväleihin



Luottamusvälit ovat ainoa keino arvioida tulosten merkityksellisyyttä, varsinkin jos halutaan ottaa huomioon tulokseen sisältyvä epävarmuus.

Frekventistisessä tilastotieteessä p-arvo ja luottamusvälit ovat keskeisiä tilastollisen päättelyn työkaluja. p-arvo voidaan jopa jättää raportoimatta tulosten yhteydessä, mutta kaikessa raportoinnissa tulisi hyödyntää luottamusväliä, sillä se ilmaisee tulokseen liittyvää epävarmuutta.

Luottamusvälille raportoidaan yleensä jokin leveys prosenttiyksiköissä ilmaistuna. 95 %:n luottamusväli on yleisin, samaan tapaan kuin p-arvosta 0,05 on muodostunut lähes universaali tilastollisen merkitsevyyden raja. Mitään matemaattista perustetta sille ei ole. Luottamusväli voidaan raportoida muillekin leveyksille, esimerkiksi 99 %:lle.

Luottamusvälin tulkintaan liittyy erittäin yleinen virhekesitys. 95 %:n luottamusväli ei tarkoita sitä, että muuttujan todellinen arvo sijaitsisi 95 %:n todennäköisyydellä raportoidulla luottamusvälillä (1). Määritelmän mukaan muuttujan todellinen arvo, kuten ryhmien välinen keskiarvojen erotus, sijaitsee silloin 95 kertaa 100:sta raportoidulla luottamusvälillä, jos sama tutkimus toistetaan 100 kertaa.

Luottamusvälin arvot tutkimuksissa vaihtelevat, eivätkä ne ole koskaan samat edes identtisissä kokeissa. Yksittäisen tutkimuksen kohdalla todellinen arvo siis joko sijaitsee tai ei sijaitse luottamusvälillä, eikä sille ei voida antaa todennäköisyysarvoa. Kuten koko frekventistinen tilastotiede, luottamusvälin

määritelmä perustuu otantaan useita identtisiä kokeita.

p-arvon tulkinta tilastollisessa päätelyssä on perinteisesti ollut kaksijakoista; vaikutuskoko on julistettu joko tilastollisesti merkitseviksi tai ei-merkitseviksi vaikka p-arvo olisi kuinka lähellä raja-arvoa 0,05. Biolääketieteessä ilmiöt ovat kuitenkin harvoin jyrkän kaksijakoisia – ja vielä vähemmän täysin keinotekoinen p-arvo 0,05 toimii keskeisenä raja-arvona. p-arvoa tulisi sen sijaan aina tulkita jatkumona, miten hyvin tutkimushypoteesi sopii aineistoon. Tämä on kuitenkin haasteellista, koska yksittäinen p-arvo ei anna mitään viitettä tuloksen kliinisestä asiayhteydestä tai epävarmuudesta. Tässä korostuukin luottamusvälien merkitys. Ne antavat käsityksen siitä, millainen on löydöksen epävarmuus ja miten löydös suhtautuu potilaalle tai väestölle merkityksellisiin lukuarvoihin.

Olipa kyseessä sitten keskiarvoerotus, regressiokerroin tai riskisuhde, niiden yhteydessä tulisi aina raportoida luottamusvälit. Luottamusvälin voidaan ajatella kuvaavan niitä arvoja, joiden kanssa aineisto on yhtenevä (compatible) tietyllä virhetasolla (2). Toisaalta voidaan ajatella, että luottamusväli kuvaa niitä arvoja, joille p-arvo olisi suurempi kuin 0,05, eli arvoja, jotka voidaan hylätä 5 %:n virhetasolla. Tämä tulkinta on askel parempaan suuntaan kaksijakoisesta tulkinnasta. Keskeistä tässä tulkinnassa on miettiä, millaiset arvot ovat oleellisia

ja potilaan tai väestön kannalta merkityksellisiä. Tämä on luonnollisesti haastavaa ja joskus varsin epämääräistä, mutta toisaalta siinä on silti vähemmän tulkinnanvaraa kuin p-arvon käytössä yhden kiinteän raja-arvon suhteen. Lisäksi on tärkeä muistaa, että viime kädessä kaikkia tutkimustuloksia tulkitaan jonkinlaisen arvomaailman kannalta.

Luottamusvälien merkitys perustuu tulkintaan suhteessa kliinisesti merkitseviin tai potilaalle merkityksellisiin arvoihin. Oletetaan, että tutkimuksessa on selvitetty lääkkeen vaikutusta potilaiden sydäntautikuolleisuuteen 5 vuoden aikana. Tutkimuslääkkeen riskitehysuuhde on 0,8 suhteessa verrokkiryhmään. Sen p-arvo on 0,6 ja luottamusväli 0,4–2,5. Olisi virheellistä todeta, että ”lääkkeellä ei ole vaikutusta kuolleisuuteen”. Luottamusvälin perusteella voidaan sulkea pois vain yli 60 %:n suhteellinen vähenemä tai 2,5-kertainen kuolleisuus lääketä saaneilla potilailla. Jos potilaalle tai väestölle merkityksellinen muutos olisi ollut kuolleisuuden puolittuminen tai kaksinkertaistuminen, tutkimuksen ainoa rehellinen päätelmä on, että siitä ei voida päätellä mitään (inconclusive). Jos sen sijaan luottamusväli samalle riskitehysuhteelle olisi 0,67–1,05, voisi päätellä jo eri lailla lääkkeen tehosta, joskin 5 %:n kasvua kuolleisuudessa ei voitaisi sulkea pois. ●

ALEKSI REITO
aleksi@reito.fi

KIRJALLISUUTTA

- Greenland S, Senn SJ, Rothman KJ ym. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol* 2016;31:337–50.
- Rafi Z, Greenland S. Semantic and cognitive tools to aid statistical science: replace confidence and significance by compatibility and surprise. *BMC Med Res Methodol* 2020;20:244.