

Mystiikkaa vai vääriä tulkintoja: p-arvo



Tieteellinen päättely, joka perustuu p-arvoon, on ollut runsaan kritiikin kohteena. Virhetulkinnat ovat erittäin yleisiä.

Modernissa lääketieteellisessä tutkimuksessa tilastollinen päättely perustuu lähes yksinomaan ns. frekventistiseen tilastotieteeseen. Sen keskeisiä työkaluja ovat p-arvo, luottamusväli ja vaikutuskoko (kuten vakioitu keskiarvoerotus ja riskisuhde).

Viime vuosikymmenten aikana p-arvosta on tullut selvästi käytetyin tilastollinen päättelyn työkalu, kun taas vaikutuskoko ja luottamusväli ovat jääneet vähemmälle huomiolle. Tällä on ollut monia epäsuotuisia vaikutuksia nykytieteeseen, osin siksi, että p-arvon tulkintaan liittyy runsaasti virhekäsityksiä (1,2).

Yleisimmin p-arvoa käytetään nollahypoteesin testaamisen yhteydessä. Nollahypoteesi olettaa esimerkiksi, että ryhmien keskiarvot ovat yhtä suuret. Tällöin vastahypoteesi olettaa, että ryhmien keskiarvot eivät ole yhtä suuret. Valitun tilastollisen testin, esimerkiksi Studentin t-testin, perusteella määritetään tilastollinen testisuure ja sitä vastaava p-arvo.

Määritetty p-arvo suhteutetaan ennalta määrättyyn tilastollisen merkitsevyyden rajaan, joka on lähes aina 0,05. Kun p-arvo on pienempi kuin 0,05, tutkimuksen tulos on tilastollisesti merkitsevä. Tämän seurauksena hylätään nollahypoteesi (eli esimerkitapauksessa oletetaan, että keskiarvot eivät ole yhtä suuret). Kun p-arvo on suurempi kuin 0,05, tutkimustulos on tilastollisesti ei-merkitsevä, ja tällöin nollahypoteesia ei voida hylätä. Esimerkitapauksessa ei siis voida sulkea pois sitä, että ryhmien keskiarvot ovat todellisuudessa yhtä suuret.

p-arvo vastaa todennäköisyyttä saada vielä enemmän poikkeava tulos kuin tutkimuksessa todettu, olettaen, että kaikkien laskemisessa käytetyt taustaoletukset ovat voimassa. Toisen tulkinnan mukaan p-arvon voi ajatella kuvaavan, kuinka yhtenevä aineisto on taustaoletusten kanssa, joihin myös nollahypoteesi kuuluu.

Esimerkiksi tutkimuksessa, jossa on verrattu kahden verenpainelääkkeen tehoa, ryhmien 1 ja 2 verenpainemittausten keskiarvojen erotus oli 5 mmHg ja p-arvo 0,20. Tämä tarkoittaa, että ryhmien keskiarvojen ero olisi 20 %:n todennäköisyydellä yli 5 mmHg, jos tutkimus toistettaisiin samankaltaisena ja kaikki taustaoletukset, nollahypoteesi mukaan lukien, olisivat totta.

Ei kerro kliinisestä merkitsevyydestä

p-arvo ei siis ole todennäköisyys, että nollahypoteesi olisi totta. Se ei kuvaa myöskään sattuman tai satunnaisvaihtelun osuutta tutkimustuloksessa: p-arvo 0,02 ei tarkoita, että tutkimustulos oli 2 %:n todennäköisyydellä vain sattuman aikaansaama. Sen sijaan p-arvo laskeaan aina olettamuksella, että vain ja ainoastaan sattuma on saanut aikaan tuloksen. Laskuteknisesti siis sattuman osuus on aina 100 %. Vastaavasti 1 – p ei tarkoita todennäköisyyttä, että tutkimustulos olisi ”oikea”.

Myöskään tyyppin 1 virheen eli väärän positiivisen tuloksen todennäköisyyttä p-arvo ei kuvaa. Tyyppin 1 virheellä on merkitystä vain, kun samaa hypoteesia testataan useita kertoja peräkkäin pyrki- myksenä minimoida sekä väärin negatiivisten (hypoteesin hylkääminen) että väärin positiivisten (hylkäämättä jättäminen) päätösten vaikutus.

Tutkittavan ilmiön kliinisestä merkitsevyydestä p-arvo ei kerro mitään. Nyrkisääntö on, että otoskoon kasvaessa yhä pienemmät vaikutuskoot ja erot saattavat tilastollisen merkitsevyyden.

Varsinaista mystiikkaa p-arvoon ei siis liity, mutta virhetulkinnat ja -käsitteykset ovat erittäin yleisiä. Tieteellinen päättely, joka perustuu p-arvoon, on ollut runsaan kritiikin kohteena erityisesti viime vuosikymmenen lopulta lähtien (2). Ilman vaikutuskokoa ja sen luottamusväliä p-arvo onkin hyvin epäkäytännöllinen päättelyn työkalu. Merkittävimmät keskustelunavaukset ovat koskeneet tilastollisen merkitsevyyden laske- misen raja-arvoa 0,05 sekä luopumista kokonaan tilastollisen merkitsevyyden käsitteestä (3,4). Keskustelu jatkuu myös tulevaisuudessa. ●

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

SIDONNAISUUDET

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

KIRJALLISUUTTA

- Gigerenzer G. Mindless statistics. *J Soc Econ* 2004;33:587–606.
- Wasserstein RL, Lazar NA. The ASA statement on p-values: Context, process, and purpose. *The American Statistician* 2016;70:129–133.
- Benjamin DJ, Berger JO, Johannesson M ym. Redefine statistical significance. *Nat Hum Behav* 2018;2:6–10.
- Amrhein V ym. Scientists rise up against statistical significance. *Nature* 2019;567:305–307.