



Herkkyys ja tarkkuus

– klinikon epäkäytännölliset työkalut

Herkkyys ja tarkkuus perustuvat ns. käänteisiin todennäköisyyksiin eivätkä ne tarjoa selkeää vastausta kliiniseen käytännön kysymyksen.

Sairauksien diagnostiikka on yksi keskeisiä ammattimme autonomisia oikeuksia. Diagnostiikassa hyödynnämme erilaisia mittauksia ja määrittämenetelmiä, kuten verikoikeita, mikrobiologisia näytteitä ja kuvantamisia. Jokainen lääkäri oppii peruskoulutuksen hyvin varhaisessa vaiheessa herkkyyden ja tarkkuuden käsitteet: Herkkyys tarkoittaa ns. oikeita positiivisia eli niiden potilaiden osuutta tutkittavista, joiden testitulokset on positiivinen ja joilla on tutkittava sairaus. Tarkkuus tarkoittaa ns. oikeita negatiivisia eli niiden potilaiden osuutta, joiden testitulokset on negatiivinen ja joilla ei ole tutkittavaa sairautta. Oppimamme nyrkisääntöjen mukaan, mitä lähempänä nämä luvut ovat sataa prosenttia, sitä parempi diagnostinen menetelmä on.

Herkkyys ja tarkkuus ovat edelleen keskeisiä raportoitavia tuloksia, kun uusien diagnostisten menetelmien hyötyjä arvioidaan, vaikka niitä on kritisoitu vuosikymmeniä (1). Ne ovat ns. käänteisiä todennäköisyyksiä ja siksi oikeastaan epäkäytännöllisiä käytännön työssä.

Herkkyyden määritelmä on todennäköisyys, että testi on positiivinen, jos potilaalla on sairaus: $P(\text{Testi+}|\text{Sairaus+})$. Tarkkuuden määritelmä taas vastaa todennäköisyyttä, että testi on negatiivinen, jos potilaalla ei ole sairautta: $P(\text{Testi-}|\text{Sairaus-})$. Nämä ns. ehdolliset todennäköisyydet ovat siis ajallisesti käänteisiä: todennäköisyys lasketaan, kun sekä sairauden tila ja testitulokset on tiedossa (2). Tämä on kliinisen päätök-

senteon kannalta hyvin epäkäytännöllistä (3).

Hyvin usein näitä todennäköisyyksiä sovelletaan Bayesin teoreeman perusteella, kun halutaan korostaa prognoosia ja ennustearvoa. Silloin ennakkotodennäköisyys on taudin esiintyvyys. Herkkyyttä ja tarkkuutta hyödyntämällä voidaan laskea taudin kokonaistodennäköisyys. Ei kuitenkaan ole perusteltua käyttää kiinteää ennakkotodennäköisyyttä kliinisessä työssä. On selvää, että esimerkiksi koronataudin kokonaistodennäköisyys on erilainen potilailla, joista toisella on kuumetta ja puuttuva hajuaihi ja toisella on vain lievä yskä.

Kliinisessä työssä ongelma onkin yleensä päinvastainen: meitä kiinnostaa todennäköisyys, että potilaalla on sairaus, jos testi on positiivinen eli $P(\text{Sairaus+}|\text{Testi+})$, tai todennäköisyys, että potilaalla ei ole sairautta, jos testi on negatiivinen eli $P(\text{Sairaus-}|\text{Testi-})$. Nämä ehdolliset todennäköisyydet ovat huomattavasti käytännöllisempiä.

Näillä ns. ennustetodennäköisyyksillä on keskeinen rooli tutkimuksessa, kun tavoitteena on laatia ennuste- ja riskimalleja. Silloin pyritään arvioimaan mahdollisia todennäköisyyksiä suhteessa ennustekyvyltään parhaisiin lähtömuuttujiin.

Hyvä esimerkki tästä on suomalainen Acute Appendicitis Score (4). Malli antaa ehdollisen todennäköisyyden akuutille umpilisäkkeen tulehdukselle, kun tietyt lähtötiedot on kuvattu. Lukuisissa tutkimuksissa on selvitetty esimerkiksi CRP:n herkkyyttä ja tarkkuutta

akuutissa appendisiitissa, mutta näiden takaperoisten todennäköisyyksien soveltaminen käytännön työssä on epäkäytännöllistä ("CRP-arvon 77 herkkyys ja tarkkuus"). Paljon hyödyllisempää on arvioida taudin todennäköisyyttä ehdollista lähtömuuttujien arvoilla. Perustuen lähtöarvoihin ennustemalli voi antaa sairauden todennäköisyydeksi laadullisen tuloksen (pieni/suuri) tai suoran prosenttimäärän.

Herkkyys ja tarkkuus ovat muodostuneet rutiininomaisiksi työkaluiksi tautien ja sairauksien diagnostiikassa. Ne esitetään yleensä ikään kuin ne olisivat universaaleja vakioita. Todellisuudessa ne vaihtelevat erittäin paljon sen mukaan, millainen on kysymyksenasettelu. Niiden käyttöä ei edistä taudin ennakkotodennäköisyyden käyttö.

Paljon hyödyllisempää kliinisessä tutkimuksessa on pyrkiä raportoimaan ennuste- ja riskimalleja, jotka antavat taudin tai sairauden mahdollisia todennäköisyyksiä tärkeimpien lähtöarvojen perusteella. ●

ALEKSI REITO
aleksi@reito.fi

KIRJALLISUUTTA

- Guggenmoos-Holzmann I, van Houwelingen HC. The (in)validity of sensitivity and specificity. *Stat Med* 2000;19:1783–92.
- Harrell FE, Slaughter JC. 19.1 Problems' with Traditional Indexes of Diagnostic Utility. *Biostatistics for Biomedical Research* 2020.
- Sensitivity, specificity, and ROC curves are not needed for good medical decision making - topical areas / diagnosis - *Datamethods Discussion Forum*. <https://discourse.datamethods.org/t/sensitivity-specificity-and-roc-curves-are-not-needed-for-good-medical-decision-making/1152>
- Adult Appendicitis Score Calculator. <https://www.appendicitisscore.com/>