

**Alexi Reito**

Dos, ortopedian ja traumatologian erikoislääkäri
Tays, TULES-keskus
Tutkimusjohtaja, Tekonivelsairaala Coxa

Tieteellisen tutkimuksen sudenkuoppia

– erityishuomio havainnoiviin tutkimuksiin

Sekä tieteen tekijöiden että käyttäjien on hallittava tutkimusmenetelmiä ja artikkelien kriittistä arviointia. Tässä artikkelissa käydään läpi erityisesti havainnoivien tutkimusten ongelmakohtia, kuten jäännössekoittuneisuutta, dikotomisaatiota ja riskitekijätutkimusten haasteita.

Erikoislääkärin työssä on tärkeää osata lukea ja tulkita tieteellisiä artikkeleita. Samoin tutkijan, joka näitä artikkeleita tuottaa, tulee hallita kliininen epidemiologia eli tutkimusmetodologia ja soveltava tilastotiede, jotta tulokset ja johtopäätökset olisivat päteviä ja luotettavia. Viime kädessä kyse on samoista taidoista: sekä tieteen käyttäjän että sen tuottajan on hallittava tutkimusmenetelmiä ja kyettävä kriittisesti arvioimaan artikkeleita.

Yhdessä artikkelissa ei voi kattavasti käsitellä kaikkia näihin teemoihin liittyviä asioita. Siksi suosittelen lukijaa tutustumaan tuoreeseen kotimaiseen kirjallisuuteen kliinisestä epidemiologiasta, joka sisältää sekä tutkimusmetodologian että soveltavan tilastotieteen näkökulmia. Aiemmissa kirjoituksissa on käsitelty muun muassa tutkimuslöydösten toistettavuutta, pätevyyttä, katsausartikkelien tyyppejä sekä sokkouttamisen merkitystä (1–3). Tässä artikkelissa keskitytään erityisesti tutkimuksen sudenkuoppiin, joita esiintyy usein takautuvissa eli retrospektiivisissä tutkimuksissa.

Jäännössekoittuneisuus ja sen hallinta (ja hallitsemattomuus)

Lääketieteellisen tutkimuksen keskeinen tehtävä on tuottaa tutkittua tietoa hoitopäätösten tueksi. Ihanteellisesti kaikki tutkimus olisi kokeellista: tarkastellaan yhtä päätöksenteon kannalta

keskeistä muuttujaa satunnaistetussa kokeessa ja selvitetään, miten sen muutos vaikuttaa potilaan terveyteen. Näin saadaan kausaalisia eli syy-seurauspäätelmiä tehtyä. Käytännössä tämä on kuitenkin harvinaista, sillä valtaosa lääketieteellisestä tutkimuksesta on havainnoivaa.

Havainnoiva tutkimus, nimensä mukaisesti, perustuu pelkkään havainnointiin. Siinä ei tehdä tutkimuksellisia interventioita, eikä keskeisiin altisteisiin pyritä manipuloimalla vaikuttamaan. Vaikka siinä voidaan tarkastella hoitojen vaikutuksia, nämä toimenpiteet tehdään osana potilaan normaalia hoitoa – eikä tutkimusta varten. Satunnaistetussa vertailukokeessa sen sijaan tehdään tietoinen tutkimusinterventio. Tiettyä altistetta manipuloidaan ja sen vaikutuksia arvioidaan, satunnaistamisen avulla. Satunnaistaminen tarkoittaa, että potilaalle annettava hoito tai lääke arvotaan vaihtoehtojen joukosta. Tyypillisesti kohteena on lääke tai hoitotoimenpide. Havainnoivien ja satunnaistettujen tutkimusten merkityksestä on keskusteltu jo vuosikymmeniä. Keskustelu on usein painottunut havainnoivan tutkimuksen rajoituksiin – juuri niihin ongelmiin, joita satunnaistetulla asetelmalla pyritään välttämään.

Satunnaistaminen ei ole aina mahdollista, ja joskus tutkimusnäyttö perustuu vain havainnoiviin tutkimuksiin. Potilaita ei voida esimerkiksi satunnaistaa tupakoiviksi ja tupakoimattomiksi tai kaupunki- ja maaseutuasukkaiksi. Sen sijaan lääkkeiden ja hoitotoimenpiteiden vaikuttavuutta

koskevan näytön tulisi aina nojata satunnaistettuihin vertailututkimuksiin. Tärkein syy tähän on jäännössekoittuneisuus (residual confounding) (4). Havainnoivassa tutkimuksessa esimerkiksi lääkettä saaneiden ja muiden potilaiden kuolleisuutta voidaan verrata, mutta tutkimus ei ole vaikuttanut siihen, kuka lääkettä saa. Tämä jättää mahdollisuuden perustavanlaatuisiin eroihin ryhmien välillä. Jos toisen ryhmän potilaat ovat keskimäärin 10 vuotta vanhempia, heidän kuolleisuutensa on todennäköisesti suurempi jo pelkästään iän vuoksi, jolloin lääkkeen vaikutusta ei voida luotettavasti arvioida. Tässä tapauksessa merkittävin osa jäännössekoittuneisuudesta muodostuu ryhmien ikäerosta, joka selittää osan kuolleisuuserosta. Yleisesti ottaen jäännössekoittuneisuus tarkoittaa sellaisten ryhmien epätasapainon aiheuttamien tekijöiden vaikutusta pääte-muuttujaan, mitä ei ole analyysissä huomioitu (5).

”Havainnoimalla ei voi tehdä kokeellisia päätelmiä”

Satunnaistetut kokeet ovat vaativia ja raskaita suorittaa. On siis varsin inhimillistä, että havainnoivasta aineistosta pyritään tekemään kausaalisia ja kokeellisia päätelmiä. Hyvin usein havainnoivassa tutkimuksessa tehdään vakiointia tai tasapainotuksia esimerkiksi propensity score -menetelmällä pyrkimyksenä vaikuttaa jäännössekoittuneisuuden määrään (6). PS-menetelmä on varsin suosittu havainnoivien aineistojen analysoinnissa, mutta siihenkin liittyy merkittäviä sudenkuoppia (7). Se ei poista tarvetta huolelliselle kausaalipäättelylle liittyen muuttujien välisiin riippuvuussuhteisiin, vaan edellyttää kuten muutkin kausaalimenetelmät, että tutkija tietää mitkä muuttujat on huomioitava ja mitä tehdä, jos tärkeitä muuttujia ei ole saatavilla. Lisäksi se heikentää tilastollisia ominaisuuksia, koska osa aineistosta jää väistämättä pois, kun hoitoryhmälle ei löydy riittävän samankaltaisia verrokkeja. Tämä johtaa havaintojen karsiutumiseen ja tilastollisen voiman pienemiseen. Se voi myös tuottaa harhaisia tuloksia, vaikka muuttujat olisi huolellisesti valittu ja se ei koskaan korvaa satunnaistettua koetta.

Muitakin vastaavia ”kausaalimenetelmiä” on olemassa, joissa pyritään tuottamaan satunnais-

tetun kokeen kaltaisia tuloksia. Yleensä ottaen kaikkien mahdollisten vaikuttavien tekijöiden mittaaminen ei ole kuitenkaan mahdollista, joten jäännössekoittuneisuuden määrä saattaa olla huomattava vakioinneista huolimatta. Tämä tarkoittaa, että jokin tekijä selittää edelleen, miksi potilas sai lääkettä tai toinen ei saanut. Jäännössekoittuneisuusharha saattaa siis jopa selittää tutkimuslöydöksen.

Riskialtis riskitekijätutkimus

– perinteisen riskitekijätutkimuksen rajat

Erittäin klassinen havainnoivan tutkimuksen muoto on ns. riskitekijätutkimus. Siinä kerätään aineisto tietyistä sairaudesta tai hoidosta ja pyritään tunnistamaan tekijöitä, jotka lisäävät komplikaatioiden, haittojen tai epäonnistumisen riskiä. Lääketieteellisessä koulutuksessa keskeinen osa onkin opetella tuntemaan eri sairauksien riskitekijät väestössä.

Epidemiologit Kayes ja Galea ovat todenneet, että ”riskitekijöiden pelto on jo melko hyvin kynnetty” (8). Perinteisesti riskitekijätutkimus pyrki löytämään potilas- ja toimenpidekohtaisia tekijöitä, jotka liittyvät sairastavuuteen ja kuolleisuuteen. Tyypillinen käytäntö on syöttää mallit täyteen muuttujia ja nimeä tilastollisesti merkitseviä tulokset riskitekijöiksi. Ongelmana on, että tällaiset tutkimukset harvoin kertovat vaikutusten todellisesta voimakkuudesta tai syy-seuraussuhteista, ja niiden kliininen arvo päätöksenteossa jää vähäiseksi. Viime kädessä riskitekijä on usein vain kahden muuttujan välinen tilastollinen assosiaatio ilman suurempaa merkitystä.

Riskitekijöitä kuvataan usein suhteellisinä vaikutuksina (esim. odds ratio, hazard ratio). Näillä mittareilla on yksinään vähän kliinistä merkitystä: esimerkiksi riskin kaksinkertaistuminen on kliinisesti vähäinen, jos lähtöriski on vain 0,3 % kymmenessä vuodessa. Erittäin usein tutkimuksessa raportoidaan vain taulukossa eri lähtömuuttujien regressiokertoimet monimuuttujamallin pohjalta. Muuttujat, joiden regressiokerroin on tilastollisesti merkitsevä, nimetään usein riskitekijäiksi. Erittäin harvoin, jos koskaan, kuitenkaan tutkimuksissa raportoidaan mikä absoluuttinen todennäköisyys oli

>>

Havainnoivassa tutkimuksessa ei tehdä tutkimuksellisia interventioita.

riskin lähtötaso tutkitulle päätemuuttujalle ilman yhtään riskitekijää. Näin ollen riskitekijätulosten soveltaminen on äärimäisen epäkäytännöllistä. Kliinisessä päätöksenteossa olennaisempia ovat absoluuttiset riskit, joita perinteiset riskitekijätutkimukset harvoin siis raportoivat. Tämän vuoksi ennustemallit, jotka tuottavat absoluuttisia riskiarvioita yksilötasolla, ovat käytännön kannalta hyödyllisempiä.

Edellä kuvattu ei kuitenkaan tarkoita, että riskitekijätutkimus pitäisi hylätä kokonaan. Hyvin suunnitellut ja metodologisesti laadukkaat tutkimukset voivat yhä olla hyödyllisiä – erityisesti uusien hypoteesien luomisessa, uusien riskitekijöiden tunnistamisessa uusissa hoitomuodoissa sekä ennustemallien kehittämisen perustana. On kuitenkin tärkeää, että riskitekijä määrittelyllään selkeästi: ilman tätä tutkimusten löydökset jäävät helposti kliinisesti merkityksettömiksi.

Haastan tutkijat pohtimaan, onko uuden tutkimusprojektin toteuttaminen tarpeen, jos sen keskeinen tarkoitus on tehdä riskitekijäanalyysi havainnoivasta aineistosta. Mikäli aiheesta on jo olemassa useita tutkimuksia, kannattaa kysyä, mitä lisäarvoa yksi uusi havainnoiva analyysi tuottaa kirjallisuuteen. Olisiko hyödyllisempää muotoilla tutkimuskysymys toisin tai jopa lähestyä ongelmaa kokonaan uudesta näkökulmasta? Vuoteen 2025 mennessä on yhä epätodennäköisempää, että perinteiset riskitekijätutkimukset tuottavat todella uusia tai kliinisesti merkittäviä löydöksiä. Niiden merkityshoidon kehittämisessä on rajallinen, ja tutkimuskenttä on siirtymässä kohti ennustemallintamista, koneoppimista, tekoälyä ja potilaskeskeisempiä lähestymistapoja.

Ennustemallinnus vs. selittävä mallinnus

Riskitekijätutkimus on klassisin tapa hyödyntää monimuuttujamalleja tutkimuksissa ja erityisesti havainnoivissa tutkimuksissa. Kuten sanottua, riskitekijän kliininen merkitys on erittäin ongelmallinen. Näissä tutkimuksissa ajaututaan erittäin helposti ns. taulukko 2 -harhaan. Termi viittaa yleiseen virhetulkintaan, joka syntyy, kun monimuuttujamallin kaikki tulokset esitetään samassa taulukossa ja tulkitaan itsenäisiksi vaikutuksiksi ("riskitekijän syy-seurausvaikutus

päätemuuttujaan"). Perinteisesti tällaiset mallit julkaistaan artikkelin toisessa taulukossa, josta nimitys on saanut alkunsa (9).

Ongelma liittyy siihen, että monimuuttujamallissa mukana olevat selittävät muuttujat eli lähtömuuttujat eivät ole keskenään samanarvoisia. Osa muuttujista on varsinaisia altisteita, joiden vaikutusta lopputulokseen halutaan arvioida. Toiset ovat sekoittavia tekijöitä, joiden avulla pyritään poistamaan harhaa. Lisäksi malliin voi huomaamatta päätyä muuttujia, jotka toimivat välittäjinä altisteen ja lopputuloksen välillä tai jopa kollidereina eli yhteisinä seurauksina. Jos kaikkien näiden muuttujien kertoimia tulkitaan yksiselitteisesti "itsenäisinä vaikutuksina", päädytään virheellisiin johtopäätöksiin.

Yksinkertainen esimerkki on tupakoinnin ja keuhkosyövän välinen yhteys. Jos malliin lisätään hengitystiesairaus, joka on tupakoinnin ja syövän välinen välittäjä (tupakointi aiheuttaa hengitystiesairauden mutta ei päinvastoin), tupakoinnin vaikutus näyttää pienemmältä kuin todellisuudessa. Väärä tulkinta voi olla, ettei tupakoinnilla olekaan merkitystä, vaan että sairaus on "itsenäinen riskitekijä". Tämä on klassinen taulukko 2 -harha.

Harhan välttämiseksi on tärkeää hahmottaa tutkimuskysymys ja kausaaliset suhteet ennen mallinnusta. Keskeistä on miettiä, onko monimuuttujamallin tarkoitus ennustaa tai selittää (10). Selittäminen ("miksi jokin tapahtuu") tarkoittaa kausaali- eli syy-seuraussuhteiden selvittämistä. Ylivoimaisesti paras keinohan tähän on satunnaistettu koe. Tietyin rajoituksen myös havainnoivasta aineistosta voidaan pyrkiä selvittämään kausaalisuhteita. Selittävässä mallinnuksessa kiinnostuksen kohteena on tietyn altisteen kausaalinen yhteys lopputulokseen. Siinä määritellään ensin altiste. Sitten selvitetään mikä on sekoittava tekijä ja mikä välitysmuuttuja (edellä tupakointiesimerkki). Vain sekoittavat tekijät otetaan mukaan analyysiin ja välitysmuuttujat jätetään pois, koska muuten analyysiin tulee harhaan. Oikeaoppinen muuttujavalinta vaatii perehtymistä ja edistyneitä menetelmiä, kuten kausaalikaavioita (directed acyclic graphs, DAG). Kausaalikaavioiden käyttö auttaa tunnistamaan, mitkä muuttujat on syytä sisällyttää malliin ja mitkä jättää pois (11). Lisäksi on syytä erottaa,

**Jäännössekoittuneisuusharha
saattaa jopa selittää
tutkimuslöydöksen.**

raportoidaanko altisteen kokonaisvaikutusta vai suoraa vaikutusta välittäjän ohii.

Ennustaminen tai ennustava mallinnus sen sijaan pyrkii vastaamaan kysymykseen ”mitä tulee tapahtumaan”. Siinä tutkitaan miten tietty lähtömuuttujien joukko ennustaa päätemuuttujaa, kuten hoidon epäonnistumista tai komplikaatiota. Ennustamisessa ei ensisijaisesti keskitytä lähtömuuttujien vaikutusten suuruuteen. Ennustemallinnuksessa olennaista on mallin suorituskyky eli miten hyvin ennustemalli ennustaa päätemuuttujaa. Tämä kuvataan eri tavoin, joista AUC-luku on kaikista tavallisin. Lisäksi suorituskykyä kuvaa pseudo- R^2 -luku, c-indeksi sekä kalibraatiokuvaaja. Jos mallin ennuste- tai erottelukyky on heikko, yksittäisten muuttujien regressiokertoimilla on silloin myös vähäinen merkitys.

Ennustaminen on noussut lääketieteellisen tutkimuksen megatrendiksi 2010-luvun loppupuolella, kun koneoppimismenetelmät alkoivat yleistyä erityisesti suurten rekisteriaineistojen analyysissä. Tekoälystä tuli vieläkin merkittävämpi megatrendi 2020-luvun alussa. Koneoppiminen on toki tekoälyn alatyyppejä, mutta yleensä tekoälyllä viitataan erityisesti syväoppimista hyödyntävään analyysiin. Yksi tekoälyn sovellusalue on luokittelevat mallit, jotka keskittyvät päätemuuttujien ennustamiseen. Näiden uudempien menetelmien tulkinta on usein samankaltainen kuin yksinkertaisessa logistisessa regressiossa. Mielenkiintoinen havainto on ollut, että edistyneemmät koneoppimismenetelmät ennustamisessa eivät ole keskimäärin parempia kuin perinteinen logistinen regressio (12).

Tilastollisen merkitsevyyden dikotomisaatio

Varsinkin havainnoivassa tutkimuksessa syyllistytään usein tilastollisen merkitsevyyden dikotomisaatioon. Tärkeää on toki tiedostaa, että vastaavaa tapahtuu erittäin usein myös kokeellisissa, satunnaistetuissa tutkimuksissa, kuten myöhempi esimerkki kertoo. Havainnoivassa tutkimuksessa on kuitenkin usein paljon enemmän laskettavia tilastollisia vaikutussuureita, kuten edellä kuvattua monimuuttujamallien tulokset (”riskitekijät”) tai yksinkertaisemmat korrelaatiokertoimet sekä kahden muuttujan vertailut.

Frekventistisessä tilastotieteessä päätös siitä, tuleeko altisteen ja lopputuloksen välistä havaittua yhteyttä pitää merkityksellisenä eli tilastollisesti merkitsevä, perustuu yksinomaan p-arvon tulkintaan. Yleinen tilastollisen merkitsevyyden raja-arvo on 0,05. Jos p-arvo alittaa tämän rajan, eli todennäköisyys havaita yhtä suuri vaikutus kuin tutkimusaineistossa, mikäli nollahypoteesi on tosi, on alle 5 %, tutkijat voivat hylätä nollahypoteesin. Valinta 0,05:n käytöstä merkitsevyyden rajana on mielivaltaisen, vaikka sen tueksi on esitetty näennäisen tieteellisiä perusteluja. Esimerkiksi on viitattu normaalijakaumaan, jossa yli kahden keskihajonnan päässä keskiarvosta sijaitsevat havainnot – noin 5 % populaatiosta – voidaan katsoa poikkeaviksi tai muuten keskiarvosta merkittävästi eroaviksi.

Tutkimustulosten tulkitseminen p-arvon perusteella binäärisesti niin, että havaittu vaikutus luokitellaan joko ”merkitseväksi” tai ”ei-merkitseväksi”, on yleistä mutta erittäin ongelmallista – sekä kliinisessä käytännössä että tutkimuksen suunnitteluvaiheessa (13). Vaikutus ei muutu olemassa olevaksi vain siksi, että p-arvo muuttuu 0,06:sta 0,04:ään, tai päinvastoin. P-arvon määritelmän mukaan muutos 0,06:sta 0,04:ään merkitsee vain kahden prosenttiyksikön eroa siinä todennäköisyydessä, että näin suuri vaikutus havaitaan aineistossa, jos vaikutusta ei todellisuudessa ole populaatiossa. Biologiassa ja lääketieteessä kaikki vaikutukset ja yhteydet ovat jatkumoa. Kliinikon näkökulmasta jotkin vaikutukset voivat olla erityisen tärkeitä (esim. insuliinihoito tyypin 1 diabeteksessa), kun taas toiset voidaan jättää huomiotta niiden vähäisen merkityksen vuoksi (esim. tulehduskipulääkegeelien vaikutus murtumien paranemiseen). Tutkimustulosten dikotomisaatio kaventaa monimutkaista, jatkuvaa dataa liian yksinkertaistetuiksi johtopäätöksiksi, ja voi siten johtaa vakaviin virhetulkintoihin.

Dikotomisaatioon liittyy kaksi keskeistä virhetyyppiä: tyypin I virhe (väärä positiivinen) ja tyypin II virhe (väärä negatiivinen). Tyypin I virhe syntyy, kun nollahypoteesi hylätään virheellisesti, mikä voi johtaa tehottomien tai haitallisten hoitojen käyttöönottoon. Esimerkiksi kirurgiasa tämä voi tarkoittaa näennäisesti tehokkaan leikkaustekniikan yleistymistä sattumanvaraisen

On yhä epätodennäköisempää, että riskitekijätutkimukset tuottavat uusia löydöksiä.

>>

löydöksen perusteella. Tyypin II virhe taas ilmenee, kun todellista vaikutusta ei havaita, usein riittämättömän otoskoon vuoksi, jolloin hyödyllinen hoito jää huomaamatta. Molemmat virheet vääristävät kliinistä päätöksentekoa ja vahvistuvat julkaisuharhan kautta: positiiviset löydökset julkaistaan todennäköisemmin kuin ”negatiiviset” tulokset, mikä vinouttaa tietopohjaa ja johtaa virheellisiin johtopäätöksiin.

ANDROMEDA-SHOCK-tutkimus septisen shokin hoidosta tarjoaa havainnollisen esimerkin tulosten dikotomisaation ongelmista (14). Tässä tutkimuksessa verrattiin hiusverisuonten täyttymisajan (capillary refill time) perusteella ohjattua nestehoitoa ja seerumin laktaatin perusteella ohjattua nestehoitoa septistä shokkia sairastavilla tehohoitopotilailla. Tutkimuksessa kuolleisuus 28 päivän kohdalla oli 34,9 % hiusverisuonten täyttymisajan ryhmässä ja 43,4 % laktaattiohjatussa ryhmässä. Arvioitu kuolleisuuden riskiero oli -8,5 % (95 %:n luottamusväli -18,2 %–1,2 %). Hazard ratio oli 0,75 (95 %:n luottamusväli 0,55–1,02), ja p-arvo 0,06. Johtopäätökseksi annettiin, että ”eroa ei ole”, koska p-arvo ylitti 0,05. Perinteisen, p-arvo-keskeisen ja dikotomisen tulokinnan mukaan tämä olisi ollut perusteltu.

On kuitenkin epäintuitiivista, että yhden prosenttiyksikön ero p-arvossa (0,06 vs. 0,05) olisi johtanut täysin päinvastaiseen johtopäätökseen, jossa hiusverisuonten täyttymisaikaan perustuva hoito olisi tulkittu laktaattiohjattua paremmaksi. Vaikka p-arvo viittasi hieman suurempaan epävarmuuteen, tulokset olivat kuitenkin pikemminkin suuntaa antavia hiusverisuonten täyttymisajan hyväksi. Keskittyminen pelkästään tilastolliseen merkitsevyyteen osoitti, miten jäykkä dikotomisaatio voi estää potilashoitoa kehittymästä ja optimaalisten hoitostrategioiden käyttöönottoa.

Kuten Douglas Altman totesi jo vuonna 1995, ”näytön puute ei ole todiste vaikutuksen puuttumisesta” (16). Tämä pätee niin ”riskitekijöille” kuin kliinisten kokeiden tuloksille. Viimeaikaiset suositukset kannustavat muuttamaan tapaa, jolla p-arvoista keskustellaan ja niitä tulkitaan tutkimuksessa. NHST:lle (null hypothesis significance testing) on olemassa useita vaihtoehtoisia tai täydentäviä lähestymistapoja, kuten luottamusvälien käyttö tai bayesilainen analyysi. Nämä menetel-

mät tuottavat arvioita mahdollisesta vaikutuksesta, ja niitä on helpompi tulkita, mikä tekee niistä kliiniseen päätöksentekoon soveltuvampia.

Jatkuvien muuttujien dikotomisaation sudenkuopat

Varsinkin havainnoivassa tutkimuksessa tehdään erittäin usein jatkuvien muuttujien dikotomisatiota. Erittäin tavallinen esimerkki on esimerkiksi jakaa potilaat kahteen tai kolmeen ryhmään iän perusteella ja käyttää tätä kolmiluokkaista muuttujaa esimerkiksi riskitekijäanalyysissä. Tämä toimintatapa ei ole koskaan missään olosuhteissa kannatettavaa, ja tilastoeksperit ovatkin tästä kliinikkotutkijoita varoittaneet jo vuosikymmenten ajan (17).

Tätä lähestymistapaa pidetään usein houkuttelevana, koska sen ajatellaan yksinkertaisen analyysin, helpottavan kliinistä tulkintaa ja vastaavan käytännössä vakiintuneisiin raja-

arvoihin, kuten diagnostisiin cut-off-pisteisiin. Kuitenkin, vaikka se on näennäisen kätevä, dikotomisaatio johtaa tiedon menetykseen ja voi tuoda mukanaan merkittäviä harhoja. Siitä huolimatta menetelmää on käytetty ja käytetään yhä kliinisessä tut-

kimuksessa, ja jotkin tilastolliset mallit – kuten päätöspuut – jopa vaativat muuttujien dikotomisoinnin ennen tai analyysin aikana.

Suurin ongelma jatkuvien muuttujien kategorisoinnissa on merkittävä tiedon menetys. Jatkuvat muuttujat sisältävät runsaasti tietoa, joka kuvaa väestön hienovaraisia eroja ja vaihtelua. Kun nämä arvot pakotetaan kahteen luokkaan, tämä tarkkuus menetetään: vaikutusten koko aliarvioidaan, trendit peittyvät ja näennäisesti samankaltaisiin ryhmiin yhdistetään hyvin erilaisia havaintoja. Lisäksi otoskoko kussakin kategoriasa pienenee, mikä vähentää analyysin tilastollista voimaa. Kliinisessä tutkimuksessa tämä tarkkuuden menetys on erityisen ongelmallinen, koska se voi peittää annos-vastesuhteet ja estää kliinisesti merkittävien vaikutusten havaitsemisen. On arvioitu, että muuttujan jakaminen kahteen ryhmään mediaanin kohdalla vähentää analyysin tilastollista voimaa yhtä paljon kuin jos kolmannes aineistosta olisi jätetty kokonaan pois (17).

Lisäksi jatkuvan riskitekijän ja lopputuloksen välinen yhteys voi olla epälineaarinen (esim. eks-

Jäykkä dikotomisaatio voi estää potilashoitoa kehittymästä.

ponentiaalinen, sigmoidinen tai U-muotoinen). Dikotomisaatio johtaa tällöin helposti virheelisiin johtopäätöksiin, ellei epälineaarisuuden mahdollisuutta huomioida asianmukaisilla menetelmillä (18).

Toinen merkittävä ongelma on, että dikotomisaatio edellyttää aina kynnyksarvon valintaa. Tämä tarkoittaa, että kaikki kynnyksen yläpuoliset arvot käsitellään samanlaisina, vaikka erot voivat olla kliinisesti merkittäviä. Esimerkiksi iän jakaminen kahteen ryhmään, < 65 vuotta ja ≥ 65 vuotta, johtaa siihen, että 66-vuotias ja 92-vuotias tulkitaan riskiprofililtaan samanlaisiksi. Samoin painoindeksiin ja komplikaatorisikin välillä on usein U-muotoinen yhteys, jossa sekä alipainoisilla että sairaalloisen lihavilla potilailla riski on kohonnut verrattuna normaali-painoisiin. Raja-arvojen käyttö voi olla joskus perusteltua, mutta usein se perustuu mielivaltaisiin valintoihin tai vakiintuneisiin käytäntöihin ilman vahvaa tieteellistä perustaa. Lisäksi pienet muutokset valitussa rajassa voivat johtaa huomattaviin eroihin tuloksissa, mikä tekee löydöksistä epävakaita ja herkkiä valitulle kynnykselle. Lisäksi emme kliinisessä työssä totea ja kirjaa, että potilaan ikä on yli 65 ja BMI alle 30, vaan käytämme näiden muuttujien tarkkoja arvoja. Miksi siis toimisimme eri lailla ja epäintuitiivisesti tutkimuksen osalta?

Lopuksi

Tässä katsauksessa käytiin vain pintapuolisesti läpi havainnoivaan tutkimukseen liittyviä sudenkuoppia. Niin kirjoittajan kuin lukijankin kannattaa syventää tietopohjaansa jatkolukemisen avulla. Kliinisestä epidemiologiasta on nykyisin saatavilla yhä enemmän myös selkokielistä suomenkielistä materiaalia. Kliininen epidemiologia ei ole vain tutkijoiden osaamisalue, vaan myös klinikon tärkeä työkalu. John R. Paul (1938) määritteli sen epidemiologisten määrällisten menetelmien ja kliinisessä lääketieteessä tehtävän potilaskohtaisen päätöksenteon yhdistelmäksi. Tämä määritelmä havainnollistaa hyvin, miksi aiheet ovat myös klinikoille olennaisia. ■

Viitteet

- Reito A, Raittio L, Helminen O. Tutkimustulokset eivät toistu: missä syy? *Duodecim*. 2020 ; 136 (10) : 1155–63.
- Kuitunen I, Ponkilainen V, Uimonen M. Mikä on sokkoutuksen merkitys lääketieteellisessä tutkimuksessa? *Duodecim*. 2025 (12): 988–92.
- Reito A. Miten tutkimusartikkelia kannattaa lukea? *Duodecim*. 2023 ; 139 (2) : 143–50.
- Ranstrom J, Cook JA. Causal relationship and confounding in statistical models. *Br J Surg*. 2016 ; 103 (11) : 1445–6.
- Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med*. 2018 ; 210 : 2–21.
- Desai RJ, Franklin JM. Alternative approaches for confounding adjustment in observational studies using weighting based on the propensity score: a primer for practitioners. *BMJ*. 2019 ; 367 : 15657.
- Elze MC, Gregson J, Baber U, ym. Comparison of Propensity Score Methods and Covariate Adjustment: Evaluation in 4 Cardiovascular Studies. *J Am Coll Cardiol*. 2017 ; 69 (3) : 345–57.
- Keyes K, Galea S. The limits of risk factors revisited: is it time for a causal architecture approach? *Epidemiol Camb Mass*. 2017 ; 28 (1) : 1–5.
- Westreich D, Greenland S. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. *Am J Epidemiol*. 2013 ; 177 (4) : 292–8.
- Schooling CM, Jones HE. Clarifying questions about "risk factors": predictors versus explanation. *Emerg Themes Epidemiol*. 2018 ; 15 (1) : 10.
- Shrier I, Platt RW. Reducing bias through directed acyclic graphs. *BMC Med Res Methodol*. 2008 ; 8 : 70.
- Christodoulou E, Ma J, Collins GS, ym. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019 ; 110 : 12–22.
- Reito A. Past, Present and Future With p-Values, Confidence Intervals and Effect Sizes. *J Foot Ankle Surg*. 2021 ; 60 (3) : 642–3.
- Hernández G, Ospina-Tascón GA, Damiani LP, ym. Effect of a Resuscitation Strategy Targeting Peripheral Perfusion Status vs Serum Lactate Levels on 28-Day Mortality Among Patients With Septic Shock: The ANDROMEDA-SHOCK Randomized Clinical Trial. *JAMA*. 2019 ; 321 (7) : 654–64.
- Effect of a Resuscitation Strategy Targeting Peripheral Perfusion Status vs Serum Lactate Levels on 28-Day Mortality Among Patients With Septic Shock: The ANDROMEDA-SHOCK Randomized Clinical Trial | Emergency Medicine | JAMA | JAMA Network. Saatavilla: <https://jamanetwork.com/journals/jama/fullarticle/2724361>
- Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ*. 1995 ; 311 (7003) : 485.
- Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ*. 2006 ; 332 (7549) : 1080.
- Lopez-Ayala P, Riley RD, Collins GS, Zimmermann T. Dealing with continuous variables and modelling non-linear associations in healthcare data: practical guide. *BMJ*. 2025 ; 390 : e082440.