



Selittää vai ennustaa?

Lääketieteellisissä tutkimuksissa on erittäin yleinen ongelma, että selitysmalleja ei juuri pohdita tai esitetä.

Tilastollinen päättely on keskeinen vaihe tieteellistä tutkimusta. Yksinkertaisin tilastollisen päättelyn muoto on hyödyntää aineistoa kuvaavia tunnuslukuja, kuten keskiarvoja ja keskihajontalukuja. Niistä voidaan kuitenkin tehdä vain melko rajoittuneita päätelmiä. Varsinkin havainnoivassa tutkimuksessa sen selvittämiseksi, miten selittävät eli lähtömuuttujat ovat yhteydessä selitettävään eli päätemuuttujaan, käytetään usein lineaarista tai logistista regressiota tai aika-analyysiä. Näiden pohjalta voidaan arvioida muuttujien yhteyksiä ja riippuvuussuhteita.

Tutkimuksessa tulisi aina erottaa selittäminen ja ennustaminen (1). Selittäminen tarkoittaa kahden muuttujan välisen yhteyden tutkimista. Tutkijaa saattaa kiinnostaa esimerkiksi tupakoinnin vaikutus sepelvaltimotautikuolleisuuden tai ylipainon vaikutus leikkauksen jälkeisiin infektioihin. Tällöin pitäisi aina rakentaa yksiselitteinen selitysmalli, jossa on perustellut syy-seuraussuhteet lähtömuuttujasta päätemuuttujaan. Tarvittaessa mukaan otetaan sekoittavia tekijöitä, joilla tiedetään olevan syy-seuraus-suhde kumpaankin.

Erittäin yleinen ongelma lääketieteellisissä tutkimuksissa on se, että selitysmalleja ei juuri pohdita tai esitetä. Tämä lisää mahdollisuutta, että tutkimuksessa on ongelmallisia selityssuhteita, jotka vääristävät tuloksia (2).

Esimerkiksi tuoreessa tutkimuksessa, jossa raportoitiin suodatinkahvin juonin pienentävän kuolleisuutta, tutkijat ottivat verenpaineen analyysiin sekoittavaksi tekijäksi (3). Se ei kuitenkaan ole todellinen sekoittava tekijä, vaan toden-

näköisemmin mediaattori, eli se välittää kahvin juonin vaikutusta kuolleisuuteen (kuvio).

Todellinen sekoittava tekijä on ns. yhteinen syy eli sekä altisteeseen että päätemuuttujaan vaikuttava tekijä. Verenpaineella tuskin on vaikutusta kahvin juontiin.

Mahdollisuus virhepäätelmiin on erityisen suuri, jos tutkimuksessa on otettu kaikki mitatut lähtömuuttujat lopulliseen monimuuttujamalliin suoraan, pohtimatta niiden keskinäisiä suhteita. Tästä menetelmästä käytetään myös termiä ”taulukko 2 -harha” (4).

Ennustekyky tutkittava erikseen

Ennustaminen tarkoittaa, että tietyillä lähtömuuttujien yhdistelmillä tai arvoilla voidaan ennustaa päätemuuttujan arvoa (5). Ennustemalleja tutkittaessa varsinaisilla selityssuhteilla ei välttämättä ole merkitystä. Tärkeintä on mallin ennustekyky, ei niinkään yksittäisen regressiokertoimen arvo.

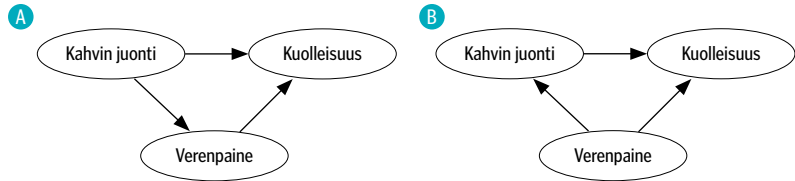
Jos tutkimuksessa raportoidaan ennustekykyä, tuloksissa pitäisi tulla esille

esimerkiksi mallin selityssaste, ns. käyrän alla oleva pinta-ala, c-indeksi tai kalibraatiokäyrä. Kaikki nämä kuvaavat, miten malli pystyy ennustamaan uusia havaintoarvoja.

Hyvin usein tutkimuksissa raportoidaan, että tietty muuttuja ”ennustaa” (predicts) toista. Usein kuitenkin taustalla on monimuuttujamalli, jonka regressiokertoimien tilastollista merkittävyyttä on tulkittu. Yleensä on kyse selittämisestä. Tilastollisesti merkittävä regressiokerroin ei tarkoita, että muuttujalla olisi välttämättä ennustekykyä. Ennustekyky pitää tutkia aina erikseen. ●

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

Mediaattori vai sekoittava tekijä?



A) Mediaattori välittää vaikutusta yhdestä muuttujasta toiseen. Mediaattoria ei saa koskaan ottaa mukaan monimuuttujanalyysiin, jos tarkoitus on selvittää altisteen yhteyttä päätemuuttujaan.
B) Todellinen sekoittava tekijä on ns. yhteinen syy eli tekijä, joka vaikuttaa sekä altisteeseen että päätemuuttujaan. Sekoittava(t) tekijä(t) on tärkeä ottaa mukaan monimuuttujanalyysiin.

KIRJALLISUUTTA

- 1 Shmueli G. To Explain or to Predict? *Statist Sci* 2010;25;289–310.
- 2 Ranstam J, Cook JA. Causal relationship and confounding in statistical models. *Br J Surg* 2016;103;1445–1446.
- 3 Tverdal A, Selmer R, Cohen JM, Thelle DS. Coffee consumption and mortality from cardiovascular diseases and total mortality: Does the brewing method matter? *Eur J Prev Cardiol* 22.4.2020. doi: 10.1177/2047487320914443
- 4 Hernan MA, Hsu J, Healy B. A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks. *CHANCE* 2019;32;42–49.
- 5 Westreich D, Greenland S. The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients. 2013;177;292–298.

SIDONNAISUUDET

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