



Jatkuvat muuttujat jatkuvina

Jatkuvien muuttujien pilkkomisesta koituu vain haittoja ilman hyötyjä.

Lääkietieteellisissä tutkimuksissa raportoidaan usein lähtömuuttujan vaikutusta tulosmuuttajaan ja tehdään ns. alaryhmä-analyyskejä kolmannen muuttujan näkökulmasta.

Yleisimpien suhteellisten vaikutuskokojen, kuten riskisuhteen, kerroinsuhteen tai vaarasuhteen, tulkinta puhtaasti kategorisille muuttujille ja luokkamuuuttujille on suoraviivaista. Tutkimuksessa voidaan esimerkiksi raportoida, että ”kuoleman riski on suurentunut miespotilailla (riskisuhde 1,3; 95 %:n LV 1,1–1,5)”. Koska sukupuoli tai tietyn sairauden olemassaolo on yksiselitteinen, myös tulosten tulkinta on selkeää.

Erittäin suuri ja edelleen yleinen ongelma on jatkuvien muuttujien pilkkominen kategorisiksi tai luokkamuuuttujiksi (1). Yleisin esimerkki tästä lienee iän pilkkominen kahteen tai useampaan luokkaan. 60 tai 65 vuotta on usein käytetty raja-arvo, jos käytetään kahta luokkaa; esimerkiksi että ”yli 65 vuoden ikäisillä oli suurentunut riski uusintaleikkaukselle (vaarasuhde RR 1,5; 95 %:n LV 1,1–1,9)”. Toisaalta usein raportoidaan hyvin ympärilyöreästi, että ”iällä oli vaikutusta uusintaleikkauksiriskiin”.

Jatkuvien muuttujien pilkkominen kategorisiksi ei ole koskaan suotavaa, eikä sillä saavuteta mitään hyötyjä. Muuttuja on tehokkainta analysoida siinä muodossa kuin se on mitattu (2).

Vain haittoja, ei hyötyjä

Pilkkominen vaatii jonkin raja-arvon, jonka perusteella jako tehdään. Se on

aina enemmän tai vähemmän mielivaltaisen päätös, ja sitä voidaan harvoin kovin vahvasti perustella. Keskiarvo- tai mediaaniperusteiset jaot ovat yleisiä, vaikka niille ei ole mitään oikeutusta. Koska raja-arvon valinta on satunnaista ja vaihtelee, jatkuvien muuttujien pilkkominen heikentää tutkimusten yleistä toistettavuutta ja keskinäistä vertailua.

Pilkkominen heikentää myös analyysien tilastollista voimaa, koska se vähentää aineiston informaatioarvoa. Mitä parempi aineiston informaatio on, sitä laadukkaampia analyysejä siitä voidaan tehdä. Pilkkominen mediaaniin kohdalla on sama kuin jättäisi aineistosta kolmasosan pois (3).

Lisäksi pilkkominen johtaa tuloksiin, jotka ovat epäloogisia ja epäintuitiivisia. Oletetaan, että tutkimus raportoi kaksinkertaisen uusintaleikkauksiriskin 65 ikävuoden jälkeen. Jos toinen potilaasi on 64-vuotias ja toinen 66-vuotias, onko loogista olettaa, että vanhemmalla potilaalla on kaksinkertainen uusintaleikkauksiriski? Entä jos toinen potilas on 66-vuotias ja toinen 92-vuotias? Tuloksen perusteella heillä riski olisi yhtä suuri, mikä tuntuu epäloogiselta.

Jatkuvien muuttujien pilkkominen on tehotonta aineiston hyödyntämistä. Vain muutaman kategoriaan pilkkominen ei edistä sitä, että aineistosta löytyisi uusia tai aiemmin kuvaamattomia yhteyksiä kahden muuttujan välillä, kuten erilaisia annos-vastesuhteita. Epälineaariset analyysimenetelmät saattavat antaa aivan uudenlaisia näkökulmia kerättyyn aineistoon.

Jatkuvat muuttujat pitää analysoida jatkuvina. Tuloksissa pitäisi raportoida esimerkiksi yhden vuoden tai yhden BMI-yksikön vaikutus tulostapahtuman riskisuhteeseen tai kerroinsuhteeseen. Tämän tyyppisen tuloksen hyödyntäminen edellä olevissa esimerkeissä on paljon hyödyllisempää, koska 2 vuoden tai 26 vuoden ero skaalautuu loogiseksi jatkumoksi pienestä riskierosta selkeästi suurempaan.

Jatkuvien muuttujien pilkkomisella eli luokittelulla ei siis ole saavuteta mitään etuja (2,3,4). Miksi raportoimme tulosmuuttujia satunnaisestiin luokkiin jakaen? Emmehän me vastaanotollakaan määrittä potilaan iäksi ”alle 65 vuotta” tai painoindeksiksi ”yli 30” vaan laskeimme kalenteriyän sellaisenaan ja mittaamme painoindeksin. ●

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

KIRJALLISUUTTA

- 1 Altman DG. The cost of dichotomising continuous variables. *BMJ* 2006;332:1080.
- 2 Altman DG. Categorizing continuous variables. Kirjassa: Armitage P, Coton T, toim. *Encyclopedia of Biostatistics*, 2. painos. John Wiley & Sons 2005.
- 3 Royston P ym. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med* 2006;25:127–41.
- 4 Osborne JW. Eleven: why dichotomizing continuous variables is rarely a good practice: debunking the myth of categorization. Kirjassa: Osborne JW. *Best practices in data cleaning: A complete guide to everything you need to do before and after collecting your data*. Sage Publications 2013:231–52.

SIDONNAISUDET

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