

## Faraldsfræði í dag

# Rangflokkun



**María Heimsdóttir**

mariah@decode.is

Í ÖLLUM FARALDSFRÆÐILEGUM RANNSÓKNUM ER EINHVER HÆTTA Á AÐ UPPLÝSINGAR UM ÁREITI EÐA ÚTKOMU SÉU RANGAR. Slíkar villur geta orðið til við upphaflega mælingu eða söfnun gagna, við tölvufærslu upplýsinga eða við meðhöndlun þeirra síðar. Villurnar geta haft áhrif á flokkun þátttakenda með tilliti til áreitiss eða útkomu þannig að einstaklingur sem í raun tilheyrir tilteknum hópi er flokkaður með öðrum ólíkum hópi. Slíkt er kallað rangflokkun (misclassification). Telja verður nær ógerlegt að útrýma algerlega villum í gagnasöfnun eða -meðferð og er því sennilega í flestum rannsóknum um einhverja rangflokkun að ræða. Ef hún er aðeins til staðar í litlum mæli hefur hún mjög takmörkuð áhrif á niðurstöðurnar en ef um verulega rangflokkun er að ræða getur hún haft afgerandi áhrif og jafnvel snúið við niðurstöðum rannsóknarinnar.

Talað er um kerfisbundna rangflokkun ef hlutfall einstaklinga sem eru rangflokkaðir er mismunandi í hópunum sem bera á saman. Þannig er um kerfisbundna rangflokkun að ræða ef flokkun sjúklinga með tilliti til áreitiss er háð því hvort sjúkdómur er til staðar eða ef flokkun á sjúkdómsástandi (það er flokkun þátttakenda í sjúkdóms- eða samanburðartilfelli) er háð áreitinu. Til dæmis verður kerfisbundin misflokkun til ef meiri líkur eru á að tiltekna fæðuvenjur séu flokkaðar sem ofneysla meðal af þungra einstaklinga en meðal þeirra sem hafa kjörþyngd. Kerfisbundin rangflokkun leiðir almennt til kerfisbundinnar skekkju (bias) en umfang hennar og eðli fer eftir aðstæðum hverju sinni. Oft er mjög erfitt að átta sig á áhrifum kerfisbundinnar rangflokkunar á niðurstöðu rannsóknar, til dæmis hlutfallslega áhættu. Þetta stafar af því að kerfisbundin rangflokkun hinna ýmsu þátta getur haft mismunandi, jafnvel algerlega gagnstæð, áhrif á það sem verið er að mæla eða meta í rannsókninni. Samofin áhrif margra rangflokkaðra þátta geta leitt til brenglaðrar niðurstöðu en geta líka vegið hvern annan upp þannig að lokaniðurstaðan verði nokkurn veginn sú sama og ef rangflokkun hefði ekki verið til staðar.

Slembin rangflokkun (random misclassification) er hins vegar til staðar ef hlutfall einstaklinga sem eru rangflokkaðir er hið sama í hópunum sem bera á saman. Slembin rangflokkun vísar þannig til þess að villur í flokkun einstaklinga með tilliti til annars áss rannsóknar (til dæmis áreitiss) eru óháðar flokkun

með tilliti til hins ássins (til dæmis sjúkdóms). Oftast leiðir slík rangflokkun einfaldlega til „útpýningar“ á hinu raunverulega sambandi áreitiss og útkomu þannig að það sýnist veikara en það er í raun og veru. Þannig getur slembin rangflokkun til dæmis á áhættuþætti leitt til þess að reiknuð hlutfallsleg áhætta er lægri en ella.

Því hefur verið fleygt að slembin rangflokkun sé aldrei eins alvarleg og kerfisbundin systir hennar, þar sem sú slembna þynnir niðurstöðurnar út eða veikir þær (lækkar til dæmis hlutfallslega áhættu) en sú kerfisbundna getur breytt niðurstöðunum í hvora áttina sem er. Það er vissulega rétt að niðurstöður sem hafa orðið fyrir áhrifum af slembinni rangflokkun mega teljast varlega áætlaðar en sá böggull fylgir skammrifi að slík rangflokkun getur beinlínis falið áhugaverð en tölulega smá tengsl milli áreitiss og útkomu og haft þannig veruleg áhrif á túlkun og notkun niðurstæðna og frekari rannsóknir. Segjum sem svo að „sönn“ hlutfallsleg áhætta af tilteknum þætti sé 1,4, það er að einstaklingar sem hafa þennan áhættuþátt séu 40% líklegri til að fá sjúkdóminn en þeir sem ekki hafa áhættuþáttinn. Ef slembin rangflokkun verður til þess að helminga metna áhættu (í 1,2) og draga úr tölfræðilegum stöðugleika er ekki ólíklegt að litið verði fram hjá þessari niðurstöðu og henni ekki frekari gaumur gefinn, hvorki til forvarna né frekari rannsókna. Vegna þessa er sjálfsagt að hafa slembna rangflokkun alltaf í huga þegar rannsóknarniðurstöður benda til veiks eða einskis sambands milli áreitiss og útkomu.

Eins og áður sagði getur rangflokkun orðið til við söfnun gagna og meðhöndlun. Mestar líkur eru á kerfisbundinni rangflokkun þegar söfnun og flokkun gagna getur á einhvern hátt litast af upplifun eða reynslu þátttakenda af áreitinu eða útkomunni. Í samanburðarrannsóknum á sjúklingum (case-control studies) er þannig ákveðin hætta á að það hvort einstaklingur hefur sjúkdóminn eða ekki hafi áhrif á hve nákvæmlega hann man eftir áreitinu eða hve nákvæmlega hann skýrir frá því. Mun minni hætta er á kerfisbundinni rangflokkun ef upplýsingar um áreitid eru fengnar úr skrárm eða skýrslum sem haldnar voru áður en útkoman eða sjúkdómurinn kom í ljós. Vissulega geta slík gögn verið ófullkomin og innihaldið rangfærslur en mestar líkur eru á að slíkir gallar séu með slembnum hætti.