

Faraldsfræði í dag

$p < 0,05$?



**María
Heimisdóttir**

Netfang:
mariah@decode.is

Lýst er eftir hugmyndum að góðum íslenskum þýðingum á orðaforða faraldsfræðinnar. Íðorðasafn lækna inniheldur nokkuð af þeim orðum sem nauðsynleg eru en betur má ef duga skal. Ég leitast við að nota íslensk orð yfir þau hugtök sem rædd hafa verið en læt ensku orðin yfirleitt fylgja í sviga til að forðast misskilning. Enska hugtakið **point estimate** var mér erfitt og væri gott að fá hugmyndir lesenda um þjált og skýrt íslenskt orð sem mætti nota. Ég notaði **metið tölugildi** en hugsanlega er gott orð þegar í notkun og væri vel þegið að heyra af því. Hugtakið point estimate er einfalt, það merkir einfaldlega niðurstöður tiltekinnar rannsóknar varðandi þau atriði sem leitast er við að meta á tölulegan hátt í hvert skipti. Slíkar niðurstöður geta verið á margvíslegu formi, til dæmis nýgengi, aldur eða hlutfallsleg áhætta (risk ratio). Nýleg rannsókn á kynþroska íslenskra drengja sýndi til dæmis að meðalaldur við upphaf kynþroska var 11,89 ár í úrtakinu (1). Þetta gildi er þá point estimate fyrir meðalaldur allra íslenskra drengja við upphaf kynþroska.

Heimild

1. Þórssón ÁV, Dagbjartsson A, Pálsson GI, Arnórsson VH. Kynþroski íslenskra drengja. Læknablaðið 2000; 86: 655-9.

NIÐURSTÖÐUR VÍSINDAGREINA BIRTAST oft sem umfangsmiklar tölur þar sem fara gjarnan saman ýmiss konar metin tölugildi (point estimate), vikmörk og p-gildi. Ekki er alltaf augljóst hvernig best er að ráða í allar þessar upplýsingar og það er óneitanlega freistandi að hvíma hratt yfir og leita að lyklinum: $p < 0,05$! Því miður gengur hann ekki alltaf að lásnum og nauðsynlegt er að gefa náinn gaum að notkun og túlkun p-gilda til að átta sig á þýðingu niðurstaðnanna.

Best er að útskýra p-gildi með því að nota dæmi. Setjum sem svo að gerð sé ferilrannsókn (cohort study) þar sem tengsl áhættuþáttar X við sjúkdóm Y eru metin með því að reikna hlutfallslega áhættu (risk ratio), sem reynist vera 2,0 (95% CI 1,8-2,5) með p-gildi 0,04 ($p=4\%$). Metið tölugildi þessarar rannsóknar, hvað varðar hlutfallslega áhættu, bendir til að einstaklingar með áhættuþáttinn X séu tvisvar sinnum líklegri til að fá sjúkdóminn Y en einstaklingar án þessa áhættuþáttar. P-gildið er túlkað sem svo: Ef í raun eru engin tengsl milli áhættuþáttarinnar og sjúkdómsins þá eru líkurnar á því að finna hlutfallslega áhættu (eða öllu heldur, metið tölugildi hlutfallslegrar áhættu) af þessari stærð eða stærri, aðeins fjórir af hundrað.

Aragrúi tölfræðilegra prófa er notaður til að kanna hvort tilteknar niðurstöður eru líklegar til að stafa af tilviljun, það er hvort þær eru tölfræðilega marktækar. Þessi próf eiga það sameiginlegt að birta niðurstöðuna á formi p-gildis og sú hefð hefur skapast að miða tölfræðilegan marktækileika við $p < 0,05$. Hefðir eru ágætar en varasamt er að binda sig algerlega við ákveðið p-gildi. Stærð þess er ekki eingöngu háð raunverulegum mun á milli hópanna sem bornir eru saman (til dæmis stærð hlutfallslegrar áhættu) heldur einnig fjölda einstaklinga í úrtakinu. Þannig getur of lítið úrtak valdið því að niðurstöður virðast ekki tölfræðilega marktækar jafnvel þó um raunverulegan mun á milli hópa sé að ræða. Hið gagnstæða getur einnig gerst, það er mjög stórt úrtak getur leitt til þess að óverulegur munur milli hópa reynist

tölfræðilega marktækar. Til að varast slíkt má nota vikmörk í staðinn fyrir eða samhliða p-gildum til að meta hlut tilviljana í niðurstöðunum. Vikmörk bera í sér þrenns konar upplýsingar. Í fyrsta lagi sýna þau umfang mismunar á milli hópanna sem bornir eru saman, það er hinn raunverulegi mismunur er sagður liggja innan vikmarkanna. Í öðru lagi sýna þau, eins og p-gildið, hvort niðurstaðan er tölfræðilega marktæk (ef vikmörkin innihalda ekki gildið 1,0). Í þriðja lagi gefur vídd vikmarka vísbendingu um stærð úrtaksins og þar með um stöðugleika niðurstöðunnar. Vídd vikmarka er einmitt sérlega mikilvæg þegar niðurstöður eru ekki tölfræðilega marktækar. Þröng vikmörk, sem innihalda 1,0, benda til að það sé í raun enginn munur milli hópanna. Víð vikmörk, sem innihalda 1,0, gefa hins vegar í skyn að niðurstöðurnar geti samræmst hvort sem er jákvæðum og neikvæðum áhrifum (til dæmis að tiltekinn umhverfisþáttur gæti verið, hvort sem er, verndandi eða aukið áhættu) og að smæð úrtaksins hafi takmarkað tölfræðilegt afl (power) rannsóknarinnar til að útiloka tilviljun sem skýringu niðurstöðunnar.

Túlkun tölfræðilega marktækra niðurstaðna krefst nokkurs sjálfsgætu. Í fyrsta lagi, og í ljósi þess sem rætt var hér að ofan, ber ekki að líta á tölfræðilegan marktækileika sem afdráttarlaust svar heldur aðeins sem eina vísbendingu um hvernig túlka megi niðurstöðurnar. Niðurstaða sem er tölfræðilega marktæk getur að sjálfsgæðu verið tilviljun, en líkurnar á því eru takmarkaðar. Hið gagnstæða getur einnig gerst, það er niðurstöður sem ekki eru tölfræðilega marktækar þurfa ekki að stafa af tilviljunum. Því er mikilvægt að sýna sjálft p-gildið en ekki aðeins upplýsa að það sé stærra en til dæmis 0,05. Túlkun slíkra niðurstaðna verður að byggjast á bæði nákvæmu p-gildi og vídd vikmarkanna, eins og lýst er að framan. Í öðru lagi verður að gera skýran greinarmun á tölfræðilega marktækum niðurstöðum annars vegar og orsakatengslum hins vegar. Smátt p-gildi bendir til að tilviljun sé ekki líkleg skýring á niðurstöðunum en gefur engar upplýsingar um gæði rannsóknarinnar eða hlutverk kerfisbundinnar skekkju (bias) eða truflandi þátta (confounding). Á hinn bóginn geta niðurstöður, sem ekki eru tölfræðilega marktækar, endurspeglad raunverulegt orsakasamband. Í þriðja lagi segir p-gildi ekkert um hagnýti eða klíniska þýðingu niðurstaðnanna. Örsnátt p-gildi getur fylgt óverulegum mun milli hópa sem hefur enga klíniska eða lífræðilega þýðingu.