Túlkun marktækni þegar öryggisbil skarast



Sigrún Helga Lund
prófessor í tölfræði við Háskóla Íslands

Í bæði rannsóknarsamstarfi og kennslu hef ég oft orðið vör við þann misskilning að þegar öryggisbil fyrir meðaltöl tveggja eða fleiri hópa skarast, þýði það sjálfkrafa að ekki sé marktækur munur milli hópanna. Þetta er skiljanlegt, enda höfum við vanist því að ef einungis eitt meðaltal er metið og borið saman við fast viðmið, þá nægi að skoða hvort viðmiðið liggi utan öryggisbilsins til að álykta um marktækni.

Vissulega gildir að ef öryggisbil tveggja hópa eru sundurlæg þá gefur það örugglega til kynna marktækan mun milli þeirra, en það gagnstæða gildir ekki. Það er nefnilega vel mögulegt að tvö 95% öryggisbil fyrir meðaltöl skarist þó marktækur munur sé á milli meðaltalanna. Þegar tveir hópar eru bornir saman, eins og með t-prófi, er réttara að skoða eitt öryggisbil fyrir muninn á meðaltölunum tveimur, fremur en bera saman tvö öryggisbil fyrir hvort meðaltal fyrir sig.

Þetta hefur verið sýnt með stærðfræðilegum rökum, meðal annars í grein eftir Peter Austin og Janet Hux sem birt var í *Journal of Vascular Surgery* árið 2002. Þar kemur fram að ef við gefum okkur jafnstór úrtök og sömu dreifni í báðum hópum, þá má búast við marktækum mun á 5% stigi ef skörun 95% öryggisbilanna er minni en um það bil 29%. Þau sýna einnig að ef 83% öryggisbil rétt snertast, þá samsvarar það marktækni á 5% stigi.

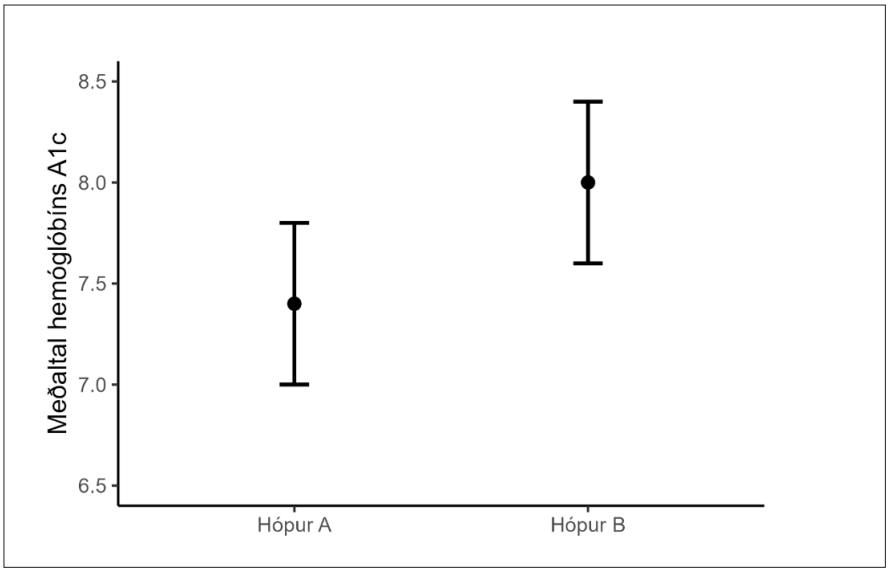
Skýringu á þessu má hugsa sem svo: þegar við reiknum tvö aðskilin öryggisbil, erum við í raun að meta tvö mismunandi punktmöt (meðaltölin sjálf) og óvissuna sem fylgir þeim hvoru um sig. Þegar við reiknum eitt öryggisbil fyrir mismun meðaltala er okkur hins vegar sama um staðsetningu einstakra meðaltala, því við metum eingöngu eina stærð: fjarlægðina milli meðaltalanna tveggja. Þar sem nú aðeins ein stærð er til skoðunar verður óvissan minni og öryggisbilið þar af leiðandi þrengra. Þetta skýrir hvers vegna slíkt öryggisbil getur sýnt marktækan mun, jafnvel þó aðskildu öryggisbilin fyrir hvort meðaltal fyrir sig skarist.

Það er því ekki örugg leið að meta marktækni með því einu að athuga hvort öryggisbil fyrir meðaltölin skarist eða ekki. Sú leið getur reynst of íhaldssöm og leitt til þess að læknar og aðrir vísindamenn líti fram hjá raunverulegum mun sem gæti haft klíníska þýðingu, sérstaklega í fámennum rannsóknum með lítið afl eins og við glímum oft við á Íslandi.

Til að meta hvort marktækur munur sé milli hópa þarf annað hvort að nota p-gildi úr samanburðarprófi eða reikna öryggisbil fyrir muninn sjálfan. Það á ekki bara við um einföld t-próf heldur einnig um fervikagreiningu (ANOVA), aðhvarfsgreiningar eða hverja aðra tölfræðiaðferð þar sem verið er að bera saman meðaltöl hópa.

Heimild

Austin PC, Hux JE. A brief note on overlapping confidence intervals. Journal of vascular surgery 2002;36(1): 194-195.



Mynd 1 *sýnir niðurstöðu úr grein Austin og Hux myndrænt: Meðaltal hemóglóbíns A1c í hópi A var 7,4 með 95% öryggisbili 7,0 til 7,8 og í hópi B var það 8,0 með 95% öryggisbili 7,6 til 8,4. Þessi öryggisbil skarast en munurinn reyndist samt marktækur (p = 0,0376). Þetta dæmi undirstrikar að það getur verið villandi að meta marktækni eingöngu út frá skörun öryggisbila.*