

Faraldsfræði í dag

# Réttmæti mælitækja

ÞESSUM FJÓRÐA pistli um hugtakið réttmæti (validity) verður varið til umræðu um þá þætti er geta haft áhrif á réttmæti mælitækja sem notuð eru í faraldsfræði og tölfræði. Jafnframt verður tæpt á þeim aðferðum er beita má til að meta hvort aðferðir eða mælitæki eru réttmæt í því þýði sem um ræðir hverju sinni. Tilgangurinn er auðvitað ekki að gefa tæmandi yfirlit um aðferðafræði við gerð og prófun mælitækja, heldur einungis að vekja athygli á helstu þáttum er huga þarf að.

Algengustu mælitæki sem notuð eru í faraldsfræði og sem grundvöllur tölfræðilegrar úrvinnslu eru sennilega ýmiss konar spurningalistar eða önnur sniðmát til gagnasöfnunar. Slíkum tækjum er beint að sjúklingum, aðstandendum, heilbrigðisstarfsfólki eða þau eru notuð til að vinna gögn úr fyrirleggjandi skrá. Réttmæti þeirra, eða hæfni til að mæla það sem á að mæla, getur takmarkast af ýmsum þáttum og var drepit á nokkra þeirra í síðasta pistli, svo sem óskýrt eða óviðeigandi orðalag, ónákvæmar eða óviðeigandi spurningar og svo framvegis. Segja má að eitt undirstöðuatriði þess að unnt sé að safna réttmætum gögnum með spurningalista sé að hver spurning hafi svarmöguleika, eða flokka, sem gera sérhverjum svaraanda kleift að velja einn en aðeins einn þeirra. Á ensku er talað um að svarmöguleikar verði að vera „exhaustive and mutually exclusive“, það er að svaraandi á að geta svarað sérhverri spurningu á viðeigandi hátt en aðeins einn kostur á að koma til greina. Algengasta dæmið um misbrestir á þessu er sennilega þegar spurt er um atriði eins og aldur eða þyngd þannig að bilin sem gefin eru upp sem svarmöguleikar skarast svo hluti svarenda getur tilheyrt tveimur bilum. Þetta gætu til dæmis verið aldursbilin 20-30 ár, 30-40 ár og svo framvegis. Hvort bilið, 20-30 eða 30-40, á þrítugur maður að velja? Afleiðingar þessa eru ónákvæmni í upplýsingum, eða rangflokkun (misclassification), sem almennt leiðir til þess að erfiðara er að átta sig á hugsanlegum tengslum milli aldurs í þessu tilviki og þeirra þátta sem taldir eru tengjast aldri. Annað atriði er spillt getur réttmæti spurningalista, og er í raun náskyllt hinu fyrra, er ónákvæmni í orðalagi þannig að erfitt er að svara spurningu afdráttarlaust. Þannig sá ég nýlega eftirfarandi spurningu í lista þar sem spurt var um reykingar: „Reykir þú eða hefur þú einhvern tíma reykt?“ Hvernig á maður að túlka neitandi svar við þessu? Þýðir það að

viðkomandi hafi aldrei snert sigarettur eða kannski að hann sé nýhættur eftir stórfelldar reykingar til margra ára? Orðalag spurningarinnar er slíkt að þessi gögn eru í besta falli lítils virði en í versta falli algerlega ónýt.

Ofantalin atriði snerta innra réttmæti mælitækja. Þegar þeim er hins vegar beitt utan þess þýðis sem þau voru upphaflega hönnuð fyrir og prófuð í geta ýmsir þættir orðið til þess að skaða ytra réttmæti þeirra eða gagnsemi þeirra til að mæla það sem átti að mæla upphaflega en nú undir öðrum kringumstæðum. Augljós dæmi eru hlutir eins og mismunandi málfar og málskilningur en aðrir þættir geta einnig komið til. Þar má nefna atriði eins og skilgreiningar svarmöguleika, viðhorf svarenda til þess sem spurt er um og samræmi milli svarmöguleika og raunverulegrar dreifingar þess sem spurt er um. Sem dæmi má nefna spurningalista um hreyfingu sem hannaður er í erlendri stórborg. Ef við höfum áhuga á að kanna hreyfingu almennt, hvort sem hún á sér stað sem hluti af daglegri vinnu, sem ferðamáti eða sem tómsunda- iða, er ekki víst að slíkur listi mæli nægilega vel hreyfingu meðal íbúa okkar samfélags þar sem þættir eins og almenningssamgöngur, veðurfar og atvinnu- hættir geta verið verulega ólíkir því sem upphaflega mælitækið miðaðist við.

Til að kanna réttmæti spurningalista og svipaðra mælitækja er nauðsynlegt að mæla á einhvern hátt fylgni eða tengsl svaraanna við þær útkomur sem mælitækið á að lýsa. Þá er nauðsynlegt að skilgreina einhvers konar „gull standard“ eða kvarða sem mælitækið er borið saman við. Innra réttmæti er þá metið með því að leggja spurningalistann fyrir ákveðið úrtak einstaklinga og safna jafnframt gögnum um þau atriði sem talin eru vera „gull standard“. Þetta tvennt er svo borið saman til að sjá hvort mælitækið gefur í raun réttar upplýsingar um það sem á að mæla. Ef til dæmis á að leggja fyrir almenning spurningalista um heilsufar og sjúkdómsgreiningar mætti bera svörin saman við gögn frá læknum og heilbrigðisstofnunum. Slíkur samanburður krefst auðvitað verulegrar vinnu en tilgangurinn er að staðfesta hvort mælitækið sé nægilega gott til að nota megi það til gagnaöflunar beint frá sjúklingum í stað þess að leita fanga inni í heilbrigðiskerfinu. Til að meta ytra réttmæti slíks tækis yrði á sama hátt að bera það saman við raun- gögn viðkomandi þýðis.



María  
Heimisdóttir

mariahei@landspitali.is