



Utvärdering av statistiska

Nyligen kom en medelålders kvinna – ganska uppskakat – med en rapport från ett laboratorium i USA. Hon hade skickat dit lite blod och en ganska stor summa pengar varefter ett mycket stort antal genanalyser gjorts och en datoralgoritm rapporterat att den största risken för henne att få cancer förelåg för *pancreatic cancer*. Risken var angiven i procent och det var många nollor före den angivna siffran, men den var större än den för *breast cancer*, *colorectal cancer* och så vidare (glädjande nog hade hon ingen risk alls för prostatacancer, vilket angivits särskilt).

Om vi utgår från att de siffror det amerikanska laboratoriet hade var korrekta – vilket borde föranleda en diskussion i sig – kan man undra vad siffror som dessa kan och ska användas för. Rimligen grundas de i att någon stor databas med tusentals personers gener matchats mot resultat i någon annan stor databas där diagnoser, dödsorsaker med mera framgår. Har man tillräckligt stora databaser så kommer det med statistiskt signifikant säkerhet (sic!) att påvisas några korrelationer – kanske mycket svaga men ändå statistiskt signifikanta. Korrelationernas styrka eller svaghet kan man aldrig begära att en lekman ska förstå, oftast inte ens en rimligt välutbildad läkare. Dock kan man säga att en grupp av (friska, symtomfria) personer med en gen eller en kombination av gener har x gånger större sannolikhet att få en viss sjukdom än de som inte har denna gen eller grupper av gener. Detta kan vi kalla ”vetenskapliga” sanningar, men det kan ifrågasättas vad som i detta sammanhang menas med ordet *sanning*.

För det första har ordet *signifikant* fått en särskild betydelse i vetenskapliga sammanhang. Enligt ordböckerna kommer ordet signifikans med adjektivet signifikant av ett latinskt ord som betyder ”beteckna, betyda”. Adjektivet kan vardagligt översättas med betydelsefull. Statistiskt signifikant och statistiskt säkerställt (skillnad) kan användas synonymt. Arbiträrt har man emellertid satt gränsen för ”säkerställt skillnad”

till att risken att skillnaden inte är sann – utan slumpmässig – till en på 20 eller mindre (samtidigt som man inte ur filosofisk utgångspunkt inte säga att $p < 0,01$ är en mer säker skillnad än $p < 0,01$ – båda visar att risken att slumpen har avgjort resultatet är mindre än den gräns man testade, det vill säga en risk på 20 eller en risk på 100). Själva den statistiska grunden betyder emellertid att med användande av ordet *signifikant* har vi accepterat att det vi säger är sant har en stor risk att inte vara det i ett fall av 20 (om vi väljer $p < 0,05$). Mäter vi 40 gånger fördubblas risken att vi har fel och gör vi 100 mätningar borde statistiken lura oss i fem fall (statistiskt!). Märk särskilt att ”vi” valt intellektuellt förståeliga tal som 5, 100 och 1 000 istället för 7, 17 och 77, men valet är snarare gjort på psykologisk grund än på matematik eller logik.

Om man nu matar in tusentals data och ställer dem mot tusentals andra data är sannolikheten signifikant (!) att man får fram korrelationer som egentligen beror på slumpen. Ju mer statistiska beräkningar man gör, desto fler ”statistiska sanningar” är i verkligheten beroende på slumpen – samtidigt som man måste hålla i minnet att de flesta ”signifikanserna” (statistiskt sett 19 av 20) inte är beroende av slumpen. Det krävs alltså ett visst mått av intellektuell bedömning för att förstå de data som statistikerna och statistiken ger oss (vi måste ju inse att det mycket väl kan finnas verkliga, sanna, korrelationer trots att statistikerna inte kan visa det i ”signifikanta” siffror – om sannolikheten är 80 procent att en häst skulle vinna vore väl de flesta totalisatorbitna beredda att satsa stora pengar på den).

I nästa steg måste vi bestämma oss för vilken säkerhetsgrad vi behöver för att exempelvis sätta in en behandling. Det nuvarande teststickorna för urinprov lär ha en säkerhet i sina svar på 95 procent, men fortsätter vi med att göra en urinodling mer än halverar vi osäkerheten – men vi når inte 100 procent, inte ens om vi gör två separata odlingar. Men, en sä-



sanningar på individnivå

kerhet på 95 procent räcker för att sätta in urinvägsantibiotika om patienten har symtom på urinvägsinfektion på kvinnor enligt Socialstyrelsen (som samtidigt ifrågasätter om alla UVI på i övrigt friska kvinnor bör behandlas). Vill man utesluta urinvägsinfektion efter en ”negativ” sticka bör man ofta göra en urinodling också om man vill vara riktigt säker (fortfarande dock inte hundra procent säker, men säker med mycket stor sannolikhet). Exempelen ovan tagna i en situation där ett felaktigt behandlingsval i så gott som alla fall inte ger några allvarliga konsekvenser utan konsekvenser som kan rättas till i efterhand.

Inom onkologin är förutsättningarna nästan alltid helt annorlunda. Felaktiga beslut kan få mycket allvarliga konsekvenser och konsekvenser som kanske inte kan rättas till senare. Säkerheten måste således vara större – och innebörden mer precis. Samma test kan emellertid ha helt olika förutsättningar i olika situationer, medan p-värdena är lika. Tumörmarkörer kan exempelvis ha ett värde i en uppföljning av behandlad tumorsjukdom, men räcker nästan aldrig för att ställa diagnos. Man kan således inte ta ett CEA för att avgöra om en patient har koloncancer eller inte och ett högt CA 19-9 är inte tillräckligt för att bekräfta att patienten har pankreas-cancer. Om däremot en patient blivit radikalopererad för prostatacancer och har ett stigande PSA har det ett betydande (signifikant?) värde liksom för den unge mannen som har ett stigande AFP efter att testikelcancern opererats bort. Samma test, samma vetenskapliga exakthet, men helt olika förutsättningar och därmed krav på olika säkerhet.

Om man då återgår till frågeställningen till hur man ska se på statistiskt signifikanta risker för en individ att få en viss cancertyp måste man konstatera att det kan vara en sanning på gruppnivå, men att det inte är en sanning på individnivå. Om man dessutom lägger till att den genetiskt testade personen kan avlida av infektionssjukdomar, hjärt-kärlsjukdomar,

olyckor och annat inser man snabbt att den genetiska undersökningen i princip är näst intill värdelös trots att den kan generera många statistiska stjärnor i en tabell. Om man till detta lägger att det är svårt för läkaren att stå emot att inte göra en MR av pankreas när patienten kommer med sitt resultat på papper – (för patienten skulle det vara fullständigt orimligt om man inte gjorde det) även om det rimligen finns andra patienter (med symtom) som borde prioriteras före. Om det då visar sig att undersökningen utfaller negativt – det vill säga att man inte hittar några sjukdomstecken i pankreas – kommer ju patienten ändå att inse att risken finns kvar i minst lika hög grad som tidigare (minst eftersom cancerrisken ökar med åldern) och kommer då att önska nya undersökningar och nya undersökningar och nya undersökningar. Det innebär att man skapat en ”artificiell” rädsla och en källa till dyr resursåtgång som aldrig går att stoppa.

Den oetiska delen i berättelsen är naturligtvis att vetenskapliga rådgivare och kanske kliniker hjälper ekonomiska entreprenörer att profitera på människors rädsla för cancer. Det är omoraliskt och gör att också den kliniska trovärdigheten i andra sammanhang riskerar att bli skadad – och hur ska man argumentera för välunderbyggda cancerscreening-undersökningar när man förkastar genetiska studier som undersöker alla cancerformer? Det värsta är emellertid att man profiterar på oskyddade människor under den vetenskapliga manteln, en vetenskap där statistik ingår som en metod men som missbrukat kan ge upphov till skada för både individ och samhälle.

ÅKE ANDRÉN SANDBERG

