

# Naturliga antal förbättrar förmågan att tolka testresultat

**Sebastian von Schreeb**, läkar- och psykologstudent, Lunds universitet  
 ● sebastianvonschreeb@gmail.com

**Juan Merlo**, professor i samhällsmedicin, överläkare i socialmedicin, Lunds universitet; Region Skåne

**Elia Psouni**, leg psykolog, docent i utvecklingspsykologi, Lunds universitet

**Redan på 1980-talet** konstaterade David M Eddy att flertalet läkare har svårt att hantera villkorade sannolikheter [1]. Han granskade litteraturen kring bröstcancerscreening och visade hur sensitivitet, specificitet och prediktiva värden ofta blandades ihop, vilket ledde till felaktiga rekommendationer. Detta har inspirerat en mängd forskare att studera läkares förmåga att dra slutsatser utifrån sannolikheter, bland annat genom följande hypotetiska scenario [2], här översatt till svenska:

»För en 40-årig kvinna som deltar i rutinmässig screening är sannolikheten för bröstcancer 1 procent. Om en kvinna har bröstcancer är sannolikheten 80 procent att hon kommer att testa positivt vid mammografi. Om en kvinna inte har bröstcancer är sannolikheten 10 procent att mammografin kommer att vara positiv. Hos en kvinna i denna åldersgrupp var mammografin positiv i en rutinmässig screening. Vad är sannolikheten att hon faktiskt har bröstcancer?»

I scenariot ges screeningens prevalens, sensitivitet och falska positiva värde. Det som efterfrågas är det positiva prediktiva värdet, vilket är 7 procent. Dessa mått är grundläggande statistiska termer, vilka ingår i en läkares utbildning. Trots detta har det upprepade gånger visats att en majoritet av både läkare och läkarstudenter misslyckas med uppgiften, vilket indikerar en oförmåga att hantera villkorade sannolikheter. Sådana kunskapsluckor kan ha konsekvenser för diagnostik och tolkning av screeningresultat [3]. I förlängningen motverkar det möjligheten att på ett effektivt sätt kommunicera med patienten kring dennes individuella sjukdomsrisik [1, 4, 5].

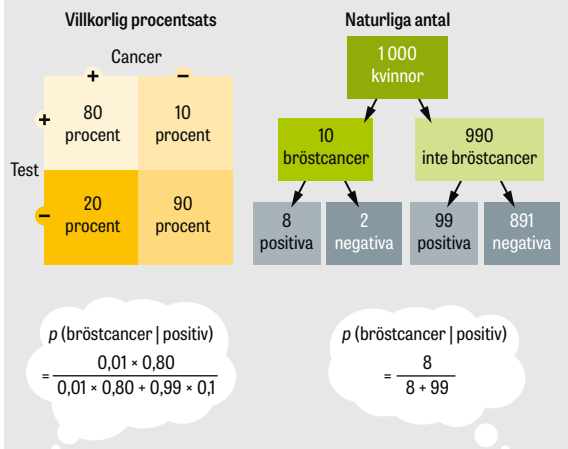
## Teorin om ekologisk rationalitet

Flera förklaringar har föreslagits till denna oförmå-

### HUVUDBUDSKAP

- Läkares förmåga att tolka screeningresultat har visats vara bristande.
- Naturliga antal är ett format att presentera statistik på som kan underlätta den förmågan.
- Den här studien är först i Sverige att undersöka om naturliga antal förbättrar läkarstudenters förmåga att finna ett testresultats prediktiva värde.
- 74 procent lyckades med uppgiften när naturliga antal användes, jämfört med 54 procent då gängse format användes (oddskvot 3,2;  $P = 0,002$ ; 95 procents konfidensintervall 1,52–6,81).
- Resultaten indikerar att effekten av att använda naturliga antal kvarstår i en svensk kontext.

**FIGUR 1. Alternativa statistiska format**



► Naturliga antal är ett alternativt sätt att presentera statistisk information, vilket kan förbättra förmågan att dra slutsatser utifrån sannolikheter. Figuren ger en illustration av den tankegång som krävs för att lösa uppgiften om den presenteras med en procentsats (till exempel 1 procent) eller med naturliga antal (till exempel 1 av 100).

»Som jägare och samlare hanterade människan inte procentsatser, utan skapade i stället en uppfattning om risker och chanser utifrån händelser och hela antal ...«

ga att dra statistiska slutsatser. Den mest inflytelserika teorin fokuserar på så kallad ekologisk rationalitet [6]. Där andra förklaringsmodeller har utropat människan till misslyckad statistiker [7] flyttar teorin om ekologisk rationalitet fokus till samspelet mellan hjärna och omgivning. Den tar fasta på att människans hjärna har utvecklats och anpassats i samklang med våra tidiga förfäders yttre förhållanden. Som jägare och samlare hanterade människan inte procentsatser, utan skapade i stället en uppfattning om risker och chanser utifrån händelser och hela antal [8].

## Naturliga antal

Mot denna bakgrund har en grupp psykologer föreslagit ett nytt format för att presentera statistik, vilket

kallas naturliga antal (natural frequencies). Exemplet med bröstcancer som presenterades tidigare ser ut på följande sätt om det i stället framställs med naturliga antal.

»10 av 1000 kvinnor vid 40 års ålder som deltar i rutinmässig screening har bröstcancer. Hos 8 av 10 kvinnor med bröstcancer kommer mammografin att vara positiv. Hos 99 av 990 kvinnor utan bröstcancer kommer också mammografin att vara positiv. Föreställ dig ett representativt urval av 40-åriga kvinnor med en positiv mammografi vid rutinmässig screening. Hur många av dessa kvinnor förväntar du faktiskt har bröstcancer?«

I motsats till villkorliga procentsatser, vilka illustrerades i förra exemplet, inkluderar naturliga antal både del och helhet. Viktigt att notera är att naturliga antal inte uppstår genom att flytta decimaltecken. I stället är alla värden relaterade till samma gruppstorlek, så att de med lätthet kan jämföras med varandra. Det blir därmed möjligt att föreställa sig en kohort på 1 000 kvinnor där 8 har fått ett korrekt positivt screeningresultat medan 99 har fått ett felaktigt positivt resultat. Genom att på så sätt visualisera uppgiften blir den uträkning som följer betydligt enklare [6], vilket illustreras av Figur 1.

I figuren visas hur villkorliga procentsatser tvingar uträknaren att hålla abstrakta siffror i huvudet. Naturliga antal, å andra sidan, underlättar visualisering av de värden som behövs för att utföra beräkningen [6]. Det bör understrykas att effekten av naturliga antal inte avser kommunikation av enkla sannolikheter (till exempel att prevalensen av bröstcancer är 1 procent). Inte heller anger naturliga antal de diagnostiska måttens konfidensintervall, vilka behövs för att utvärdera det empiriska stödet bakom siffrorna. Formatet är snarare att betrakta som ett pedagogiskt verktyg som gör tolkningen av medicinska testresultat mer begriplig, både för läkare och för patienter [5]. En Cochraneöversikt konstaterade att naturliga antal underlättar förståelsen av villkorade sannolikheter inom ramen för diagnostik och screeningstest [3].

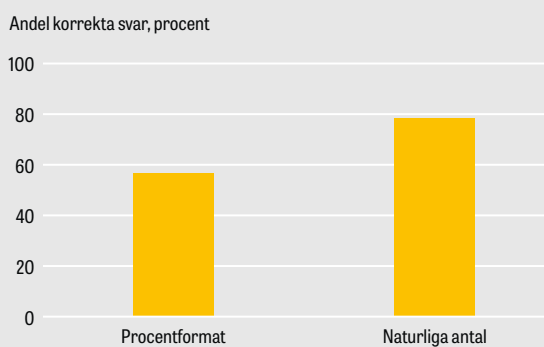
Trots att formatet använts i läkarutbildningar internationellt har, så vitt vi vet, ingen studie hittills undersökt effekterna av att använda naturliga antal bland svenska läkarstudenter. Vi ville därför undersöka om även blivande läkare i Sverige brister i förmågan att hantera villkorade sannolikheter samt om denna förmåga kan stärkas genom att använda naturliga antal. Det skulle ge en fingervisning om huruvida formatet är gångbart även i en svensk kontext.

## METOD

En enkätundersökning genomfördes på 97 läkarstudenter vid Lunds universitet. Inbjudna att medverka var studenter på termin 9, 10 och 11, sammanlagt 300 personer. Svarsfrekvensen var 46 procent från termin 11, 12 procent från termin 10 och 37 procent från termin 9. Studenterna blev inbjudna att delta via sina studentmejladresser och fick via en bifogad länk tillgång till den webbaserade enkäten.

Studien använde en intraindividuell crossover-design för att utvärdera effekten av presentationsformat. Deltagarna blev slumpmässigt indelade i två grupper. Den första gruppen tilldelades den tidigare beskrivna

FIGUR 2. Andel som klarar uppgiften



► Staplarna visar andel av deltagarna som svarade rätt på uppgiften när den presenterades med procent respektive naturliga antal.

uppgiften om bröstcancerscreening med procentsatser, medan den andra tilldelades samma uppgift där siffrorna presenterades med naturliga antal. Därefter fick den första gruppen utföra en likartad uppgift med ett nytt scenario, där siffrorna presenterades med naturliga antal. Den andra gruppen utförde samma nya uppgift, med enda skillnad att siffrorna stod i procentformat. Deltagarnas svar kodades som korrekta om de låg inom 2,5 procentenheter från svaret. Som sensitivitetsanalys gjordes en striktare kodning där korrekta svar inte fick avvika mer än 1 procentenhet från svaret och måste vara korrekt konverterade till procent. Data analyserades sedan genom en betingad logistisk regression som tog hänsyn till den intraindividella studiedesignen.

## RESULTAT

Av de som deltog uppgav sig 44 som kvinnor (12, 5 och 27 från termin 9, 10 respektive 11) och 51 som män (16, 5 och 30 från respektive termin), medan 2 personer (en från termin 9 och en från termin 10) kryssade för alternativet »vet ej/vill ej svara«.

När uppgifterna presenterades med naturliga antal svarade 74 procent (72/97) av deltagarna korrekt, jämfört med 54 procent (52/97) då procent användes som presentationsformat, se Figur 2. På den första uppgiften svarade 44 procent (21/48) rätt med procent och 61 procent (30/49) rätt med naturliga antal. På den andra uppgiften svarade 63 procent (31/49) rätt med procent och 88 procent (42/48) rätt med naturliga antal. Den betingade logistiska regressionen visade att oddset att svara rätt var högre då naturliga antal användes som presentationsformat, med oddskvoten 3,2 ( $P = 0,002$ ; 95 procents konfidensintervall 1,52-6,81) jämfört med procent som presentationsformat. Resultaten kvarstod när det striktare kodningssystemet användes.

## DISKUSSION

En dator som utför en statistisk uträkning kommer inte att påverkas av vilket format informationen presenteras i. Detsamma verkar inte gälla för läkarstudenter. För denna kohort av svenska läkarstudenter ökade förmågan att dra slutsatser utifrån sannolikheter betydligt då uppgiften presenterades med naturliga antal. Det ligger i linje med den internationella

forskningen på området [2, 9, 10]. Utifrån vår känedom är denna studie den första att undersöka effekten av att använda naturliga antal i en svensk kontext. Eftersom effekten bestod över geografiska och språkliga gränser kan dessa resultat anses stärka dess externa validitet. Effekten kvarstod även vid en sensitivitetanalys gällande rättningsystemet, vilket stärker fyndens interna validitet.

Andelen rätta svar var relativt hög jämfört med tidigare studier, men resultaten bör inte föranleda några slutsatser om svenska läkarstudenters kompetens. Svartsfrekvensen var låg och en svartsbias kan ha förelagat, där studenter som fann uppgifterna särskilt besvärliga kan ha avbrutit enkäten och avstått från att skicka in ett svar. Dessutom användes en intraindividuell crossover-design som gör det troligt att en övningseffekt spelat in, vilket påpekats i Cochraneöversikten [3]. Det är därför rimligt att misstänka att den totala andelen rätta svar är en överskattning. Detta resonemang är dock inte i samma mån applicerbart för att kritisera den funna skillnaden mellan presentationsformaten. Det är naturligtvis möjligt att en indirekt effekt av urvalet påverkat resultatet, men svartsfrekvensen var jämnt distribuerad mellan grupperna och presentationsformaten, vilket talar emot att en direkt svartsbias var orsaken till den funna resultat-skillnaden mellan formaten.

### Teoretiska förklaringsmodeller

Det är anmärkningsvärt att en betydande andel av läkarstudenter misslyckades med att finna det prediktiva värdet av ett medicinskt screeningresultat. En grupp forskare som försökt förklara detta fenomen, däribland nobelpristagare Daniel Kahneman, har föreslagit en prevalensbias (base rate fallacy) [11]. Det innebär att deltagarna misslyckas för att de negligerar prevalensen. Den förklaringen har mött kritik eftersom empirin inte bekräftar de förutsägelser som teorin implicerar [6]. Även i vår studie låg de felaktiga svaren utanför det intervall som förutsågs av prevalensbias, varför inte heller våra resultat stödjer teorin. Ett annat förslag är att deltagarna misslyckades då de konverterade sina svar till procent. För att undersöka det inkluderades i vår studie en separat fråga som ombad deltagarna omvandla sina svar. 9 av 10 lyckades, vilket talar emot att misslyckandena berodde på bristande förmåga att konvertera till procent. Den förklaring som bäst stöds av våra resultat är teorin om ekologisk rationalitet. Teorin grundar sig i antagandet att människans förmåga att dra slutsatser utifrån sannolikheter är beroende av presentationsformat [4]. Jäm-

fört med det konventionella sättet att presentera statistik var det märkbart fler som svarade rätt då uppgiften presenterades med naturliga antal.

### Begränsningar och vidare forskning

Den mest uppenbara felkällan i den aktuella studien var den låga svartsfrekvensen. Det kan möjligtvis ha berott på att studien var internetbaserad, tidsmässigt krävande och saknade materiell ersättning. Den här studien är dessutom en tvärsnittsstudie på en avgränsad population, vilket ger begränsade möjligheter att uttala sig generellt om läkarstudenters förmåga. Vidare jämförs ett presentationsformat som ingår i studenternas utbildning med ett som de aldrig tidigare stött på. Det är därför rimligt att misstänka ytterligare skillnad mellan formaten om naturliga antal varit en del av den statistiska utbildningen. Tidigare forskning har rekommenderat att naturliga antal ska ingå i kursplanen för medicinsk statistisk, vilket har prövats med positiva resultat [12]. Vinsterna bör ställas mot eventuella nackdelar av att undervisa svenska läkarstudenter i naturliga antal. För att göra en mer trovärdig utvärdering föreslår vi därför en longitudinell studie där en grupp av studenter utbildas i naturliga antal, vilket sedan kan jämföras med en grupp som getts traditionell undervisning.

Läkares förmåga att hantera villkorade sannolikheter är viktig vid diagnostik och tolkning av screeningresultat [3]. I förlängningen möjliggör det effektiv riskkommunikation med patienten, så att denne ges en riktig bild av sin individuella sjukdomsrisk [5]. Internationell forskning inom området har visat att förmågan kan förbättras genom att använda naturliga antal som presentationsformat. Innevarande studie ger en fingervisning om att formatet kan vara gångbart även i en svensk kontext. Vidare forskning behövs för att utvärdera om naturliga antal kan vara användbara för att förbättra svensk läkarutbildning. ○

● Potentiella bindningar eller jävsförhållanden: Inga uppgivna.

Citera som: *Läkartidningen*. 2018;115:EY3D

## Läs mer!

Engelsk sammanfattning på [Läkartidningen.se](http://Lakartidningen.se)

### REFERENSER

- Eddy DM. Probabilistic reasoning in clinical medicine: problems and opportunities. In: Kahneman D, Slovic P, Tversky A (editors). *Judgment under uncertainty: heuristics and biases*. Cambridge: Cambridge University Press; 1982. p. 249-67.
- Hoffrage U, Gigerenzer G. Using natural frequencies to improve diagnostic inferences. *Acad Med*. 1998;73(5):538-40.
- Akl EA, Oxman AD, Herrin J, et al. Using alternative statistical formats for presenting risks and risk reductions. *Cochrane Database Syst Rev*. 2011;(3):CD006776.
- Gigerenzer G, Gaissmaier W, Kurz-Milcke E, et al. Helping doctors and patients make sense of health statistics. *Psychol Sci Public Interest*. 2007;8(2):53-96.
- Gigerenzer G. What are natural frequencies? *BMJ*. 2011;343:d6386.
- Gigerenzer G. Calculated risks. How to know when numbers deceive you. New York: Simon and Schuster; 2015.
- Kahneman D. *Thinking, fast and slow*. New York: Farrar, Straus and Giroux; 2011.
- Cosmides L, Tooby J. Better than rational: evolutionary psychology and the invisible hand. *Am Econ Rev*. 1994;84:327-32.
- Hoffrage U, Krauss S, Martignon L, et al. Natural frequencies improve Bayesian reasoning in simple and complex inference tasks. *Front Psychol*. 2015;6:1473.
- Kurzenhäuser S, Hoffrage U. Teaching Bayesian reasoning: an evaluation of a classroom tutorial for medical students. *Med Teach*. 2002;24(5):516-21.
- Tversky A, Kahneman D. Evidential impact of base rates. Technical Report No 4. Stanford, CA/Arlington, VA: Stanford University/Office of Naval Research; 1981. <http://www.dtic.mil/dtic/tr/fulltext/u2/a099501.pdf>
- Fuks A, Boudreau JD, Cassell EJ. Teaching clinical thinking to first-year medical students. *Med Teach*. 2009;31(2):105-11.

## SUMMARY

### **Natural frequencies improved diagnostic inference among medical students in Sweden**

The ability to draw statistical inferences from test results may be generally limited among physicians. A new way of presenting statistics, called natural frequencies, has been shown to improve this ability. The current study is the first to investigate this effect in Sweden, involving senior medical students. An intra-individual crossover design was used in which participants answered a statistical inference task presented with natural frequencies and another with percentages, which has been the customary format. A total of 74% made a correct inference on the natural frequencies task, as compared to 54% on the percentages task, with an odds ratio of 3.2,  $p = 0.002$ , 95% CI 1.52-6.81. The findings indicate that the effect of using natural frequencies arises also in a Swedish context.