

TJUGO TIPS TILL DIG SOM TOLKAR VETEN- SKAPLIGA STUDIER

Olle Häggström uppmärksammar en artikel i Nature, se [5] i referenslistan. Det som följer här är baserat på den artikeln.

1 Såväl verkliga skillnader som slump orsakar variation. En stor del av all naturvetenskaplig forskning går ut på att ta reda på vad som orsakar de mönster och trender som vi kan se. Det finns vanligtvis många faktorer som spelar in, så forskningens största utmaning är att skilja de viktiga orsakerna från de mindre viktiga och från slumpens spel.

2 Mätvärden är sällan exakta. Nästan alla mätningar har något fel i sig. Om mätningen av något upprepas kan man få ett delvis annorlunda resultat. Alla mätvärden och skattningar bör publiceras tillsammans med någon kvantifiering eller kvalificerad uppskattning av precisionen.

3 Det finns systematiska fel på många håll och kanter ("bias is rife").

- ♦ Experimentdesigner och mätinstrument kan ge systematiska fel.
- ♦ Experiment kan ge systematiska fel om deltagarna vet om att de fått en behandling som förväntas vara effektiv, eller om forskaren vet vilka av deltagarna som fått den "effektiva" behandlingen. Därför bör experiment om möjligt vara dubbelblinda.
- ♦ Det är mer troligt att studier som redovisar "statistiskt signifikanta" resultat blir publicerade än studier som inte gör det. Därför kan vetenskaplig litteratur ge en överdriven bild av storleken på problem eller av effektiviteten hos föreslagna åtgärder.
- ♦ Även tränade forskare kan drabbas av "confirmation bias": fenomenet att man lägger särskilt märke till det som stödjer det man tror på, men bortser från det som talar emot.

4 Extrema resultat kan visa fel. Om en skola fått mycket bra resultat kan det bero på lärarnas skicklighet. Men slumpen kan också spela in; barnen som testades

kanske var ett urval av barnen i skolan, de lyckades kanske väldigt bra på provdagen, skolan kanske ligger i ett område med föräldrar med hög utbildning osv. Men ofta tolkas ett mycket bra resultat i någon mätning som evidens att det som mättes måste i grunden vara väldigt bra.

5 För stickprovsstorlekar är det sant att större är bättre. Det gäller särskilt studier som mäter förhållanden som har stor naturlig variation.

6 Korrelation betyder inte orsakssamband. Det kan ligga mycket nära till hands att anta att något vi ser orsakas av något annat i närheten. Till exempel trodde ekologer, helt naturligt, att orsaken till markant fiskdöd i deltat till floden Neuse i USA var tillväxt av alger. Utsläpp av kväve gynnar tillväxten av alger, som leder till syrebrist, som i sin tur tar död på fisk. Men orsakskedjan är lång och tunnas ut i varje steg. Minskade kväveutsläpp beräknades i det här fallet bara ge mycket måttlig förbättring av fiskens livsvillkor. [1]

7 Fenomenet "regression mot medelvärdet" kan vilseleda. Extrema händelser eller utfall är ofta till en del orsakade av slumpens spel. Om ett utfall varit extremt en gång, är det inte troligt att ett lika extremt utfall inträffar nästa gång. Om det till exempel på kort tid inträffat flera olyckor på en vägsträcka, kanske man sätter upp fartkameror där. Om olyckornas antal går ner kan man inte dra slutsatsen att det var fartkamerorna som orsakade nedgången. Om olyckorna var talrika under en period innan fartkameran, kan det bero på slumpen, och då hade antalet olyckor gått ner mot en mer normal frekvens ändå, även utan fartkameror.

8 Att extrapolera till förhållanden som man inte har data för är riskabelt. Det är mycket svårt att förutsäga till exempel »»»

Hela beståndet av torsk utanför Newfoundland kollapsade på grund av felräkning. På bilden fiskeriinspektörer som övervakar fisket i Nordsjön

FORTS: TJUGO TIPS TILL DIG SOM TOLKAR VETENSKAPLIGA STUDIER

Felräkning fick torskbestånd att kollapsa



FOTO: JAN VAN DE VELDE

» hur ekologiska system kommer att reagera på klimatförändringar när förändringarna är snabbare än vad vi har erfarenhet av.

9 **Fenomenet "base-rate fallacy" kan spöka.** Förmågan hos ett test att upptäcka ett fenomen beror på sannolikheten att fenomenet inträffar. Ett exempel: trots att ett test, som är 99% säkert, visar att du har en viss sjukdom, kan det vara osannolikt att du har sjukdomen! Om sjukdomen är ovanlig, kan det vara så att av 10 001 personer som testar sig har bara en av dem sjukdomen. Då kommer testet att nästan helt säkert att visa det (99 % säkert). Men ytterligare 100 personer kommer felaktigt att få positivt testresultat (99% säkert betyder 1% fel, och 1% av 10 000 utan sjukdomen ger 100 positiva testresultat – som är felaktiga). Så om du har fått ett positivt testresultat är det mer troligt att du tillhör gruppen av 100 personer, än att du är den enda som verkligen har sjukdomen.

10 **Det är viktigt att ha en kontrollgrupp.** När man testar om en behandling har någon effekt, måste man kunna jämföra med en kontrollgrupp, som inte har fått behandlingen. Effekten hade kunnat uppstå ändå, utan att vara orsakad av behandlingen.

11 **Slumpmässiga urval i experiment skyddar från systematiska fel.** Om möjligt ska individer fördelas slumpmässigt mellan den grupp som får en behandling och den grupp som utgör kontrollgrupp. Om man inte gör så, kan valet att ingå i den grupp som får behandlingen styras av något annat. Om man finner att barn till föräldrar som gått en föräldrakurs får bättre betyg än andra barn, kan det bero på att de föräldrar som valt att gå den kursen har högre utbildning än andra föräldrar.

12 **Forskare är människor.** Forskare kan övertolka eller fuska. Kollegial bedömning av kvaliteten av forskning är ett bra men inte 100% säkert system. Forskningsresultat som har studerats, diskuterats och upprepats av oberoende grupper av forskare är säkrare än resultat som kommer från bara en grupp av forskare.

13 **Statistisk signifikans är viktigt** ("significance is significant"). Att tolka icke-signifikanta resultat som "en tendens" är att övertolka.

14 **Att plocka ut russin.** Data och evidens kan vara plockade som russin ur kakan. Forskarna kan ha muddrat genom stora mängder data och lyft fram något som "intressant". Om man ska

kunna tro på att intag av lätt-yoghurt under graviditeten korrelerar med astma hos barnet behöver man veta om forskarna riktade in sig på att testa just denna hypotes, eller om de fann ett signifikant resultat genom att genomföra en stor mängd test. [4]

15 **Beroenden mellan händelser ändrar risken.** Man kan skatta risken för enskilda händelser som till exempel väldigt stor nederbörd, storm eller att det inte finns tillräckligt med personal i vården. Men om dessa händelser är beroende av varandra (storm och nederbörd tenderar att inträffa tillsammans och personalen kan inte komma till jobbet p.g.a. storm), då är risken att dessa tre saker inträffar samtidigt större än vad man har beräknat om man inte tog hänsyn till beroendet. Vissa amerikanska kreditinstituts försäkran att risken var försumbar att det skulle inträffa samtidigt att många inte skulle klara av att betala på sina lån visade sig vid finanskrisens start vara fel. Det berodde på att det fanns beroenden som man inte tog hänsyn till.

16 **Leta efter fullödiga upprepningar av studier och av insamling av data som inte i princip bara är replikat av varandra.**

Resultat som är samstämmiga över flera stu-

dier som upprepats på olika populationer är säkrare än studier som bara gjorts på en population.

Data som har samlats in och som inte beror av varandra ger starkare evidens än data som delvis innehåller samma information p.g.a av att de beror av varandra. I studier i början av 1990-talet av nedgången av torsk utanför Newfoundland drogs slutsatsen att det var naturlig mortalitet. Sedan kollapsade beståndet av torsk, som var det största i världen. Studierna hade inte tagit hänsyn till att de upprepade urval av torsk som man tog upp med trål delvis drogs ur samma kohort av torsk [3].

17 **Tro inte att ett icke-signifikant resultat måste betyda att fenomenet eller effekten inte finns.** Ett icke-signifikant resultat kan betyda det, men det kan också vara så att studien inte hade styrka att detektera fenomenet eller effekten.

18 **Hur stor effekten är har verklig betydelse.** Ett experiment med omfattande stickprovstorlek kan detektera effekter som är så små att de egentligen inte har någon betydelse. Under 1990-talet bad redaktören för tidskrif-

ten Epidemiology författarna att sluta skriva om statistisk signifikans eftersom tolkningen så ofta blev att statistiskt signifikanta effekter också måste vara signifikanta i verkligheten (i meningen praktiskt betydelsefulla), vilket i sin tur ledde till att mindre väsentliga åtgärder för förbättrad folkhälsa rekommenderades.

19 **En studies relevans begränsar dess möjlighet att generaliseras.** Ett exempel: det är besvärligt att generalisera resultat från djurförsök till människor.

20 **Riskbedömning är inte vår bästa gren.** Vår bedömning av risk influeras bland annat av om risken tas frivilligt och hur mycket egen kontroll vi tror att vi har. Till exempel underskattar folk i USA risken med att ha skjutvapen hemma med en faktor 100, och överskattar risken med bo nära ett kärnkraftverk med en faktor 10. [2]

DAN HEDLIN

Referenser

- [1] Borsuk, M.E., Stow, C.A. och Reckhow, K.H. (2003). Integrated Approach to Total Maximum Daily Load Development for Neuse River Estuary using Bayesian Probability Network Model. Journal of Water Resources, Planning and Management, 129, 271-282.
- [2] Fischhoff, B., Slovic, P. och Lichtenstein, S. (1982). Lay Foibles and Expert Fables in Judgments about Risk. The American Statistician, 36, 240-255.
- [3] Millar, R.B. och Anderson, M.J. (2004). Remedies for Pseudoreplication. Fisheries Research, 70, 397-407.
- [4] Maslova, E., Halldorsson, T.I., Ström, M. och Olsen, S.F. (2012). Low-Fat Yoghurt Intake in Pregnancy Associated with Increased Child Asthma and Allergic Rhinitis Risk: A Prospective Cohort Study. Journal of Nutritional Science, 1, 1-11.
- [5] Sutherland, W., Spiegelhalter, D. och Burgman, M. (2013). Policy – 20 Tips for Interpreting Scientific Claims, Nature, 503, 335-337.