

Hur står det till med uppreparheten?

Debatten om uppreparhet hos vetenskapliga resultat har fått en ny skjuts genom en rad systematiskt genomförda replikationsstudier inom psykologin. **Jesper Jerkert** berättar.

VETENSKAPLIGA RESULTAT är inte mycket värda om de inte är upprepbara, och detta är skälet till att man bör vara försiktig med att dra alltför stora slutsatser utifrån en enda undersökning. Först när andra studier som undersökt samma fråga ernått väsentligen samma resultat kan man börja känna sig säker.

Detta är skolboksuppfattningen, och den är förstås mycket rimlig. Somliga vill tona ned uppreparhetskravet, eftersom de egna favoritteorierna inte lyckats få stöd i upprepar forskning, men en sådan attityd är bara alltför

genomskinlig. Mer intressant är kanske att andra tvärtom intar en mycket strängare ståndpunkt: det kan finnas intressanta resultat som upprepats i många studier men som likväl är falska.

Kan det vara på det viset, och hur kan det i så fall komma sig?

JOHN IOANNIDIS' ANALYS

Saken har debatterats en del under senare år, en debatt där man ofta har hänvisat till en artikel från 2005 av den amerikanske epidemiologen John Ioannidis med titeln "Why most published research findings are false".¹ Ioannidis inleder med att referera några studier som tyder på dålig upprepbarhet inom framför allt medicinsk forskning. Han går sedan raskt vidare till sitt huvudärende, att modellera upprepbarheten hos en ny studie som funktion av följande faktorer:

1. Den valda statistiska signifikansnivån i studien (den högsta sannolikhet att resultatet skulle ha uppkommit av en slump som man är beredd att acceptera men ändå tro att resultatet inte har uppkommit av en slump).
2. Sannolikheten att man i studien påvisar ett statistiskt signifikant resultat, givet att det verkligen existerar.
3. Hur många forskargrupper som undersöker samma fråga.
4. Utrymmet för variationer i metodologi, dataanalys och presentation, som kan antas verka i sådan riktning att samband överdrivs.
5. Andelen sanna samband som finns bland de samband av ungefär samma slag som testas inom det aktuella

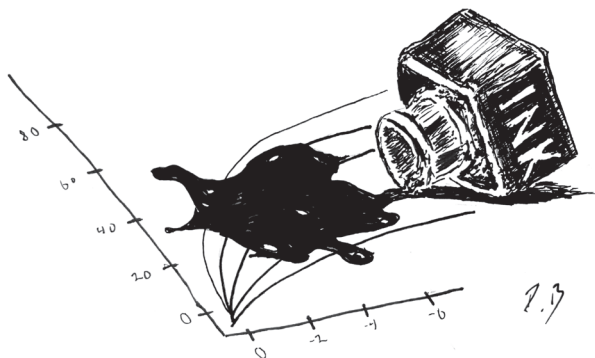


Bild: Daniel Bostedt

området. (Detta är liktydigt med en *a priori*-sannolikhet inom ett visst vetenskapligt område att ett testat samband är sant.)

Faktorena 1 och 2 kan i hög grad påverkas av forskaren som lägger upp studien. Faktor 3 beror på hur populärt ett visst område är inom forskarvärlden, vilket kan påverkas både av vetenskapliga analyser och av modenycker. Faktor 4 är omfattande, och inrymmer både sådant som den enskilda forskaren själv kan påverka (såsom hur strikt man hållit sig till i förväg fastställda metoder och analysverktyg) och sådant som har med attityder i forskarsamhället i stort att göra (såsom intresset att läsa om negativa resultat, de vetenskapliga tidskrifternas benägenhet att låta publicera rena replikationsstudier, med mera).

Även faktor 5, slutligen, verkar kunna uppdelas i flera

beståndsdelar. Den sökta *a priori*-sannolikheten beror på vilka samband som betraktas som potentiellt intressanta inom det givna vetenskapliga fältet. Man skulle kunna resonera på följande sätt. Inom ett fält med svag teoribildning är forskare benägna att testa alla möjliga samband, även när man saknar teoretiska skäl att tro att sambanden borde finnas. Om man vidare går med på att det finns svagare teoretisk grund inom, säg, psykologi än inom astronomi, borde man förvänta sig att fler osanna hypoteser testas inom psykologi än inom astronomi, och därmed borde man även få fler negativa resultat (inget samband styrkt) inom psykologi än inom astronomi. Är det så? Intressant nog finns forskning som säger att det är tvärtom: inom ämnen som psykologi och ekonomi verkar det vara mycket vanligare med positiva resultat än inom fysik och astronomi.² Så kanske testar inte forskare mer osannolika hypoteser inom psykologi än inom astronomi. Eller också gör de visst det, men andra faktorer som verkar i motsatt riktning är starkare, såsom att negativa resultat inom psykologi i högre grad förblir opublicerade, vilket i sin tur kan bero på attityder hos såväl forskarna själva som hos de vetenskapliga tidskrifternas redaktörer (och som i så fall hör hemma under faktor 4).

ALLTFÖR STARK SLUTSATS

Detta blir ju ganska rörigt, och man kan alltså invända mot Ioannidis' modell att den inte är särskilt transparent, att faktorerna kanske borde ha delats upp i flera, och att de definitivt inte är oberoende av varandra (vilket hade varit lättare att tillgodogöra sig). Utöver de exempel på beroende som redan givits kan följande nämnas. Hur många

forskargrupper som undersöker samma fråga (faktor 3) kan inte spela någon större roll i sammanhanget om det inte samtidigt är så att positiva resultat har lättare att bli publicerade och spridda – vilket i så fall ska fångas upp av faktor 4.

Icke desto mindre verkar många ha funnit Ioannidis' analys intressant, att döma av att hans artikel blivit flitigt citerad. Bland annat sätter Ioannidis in siffror, som han hävdar är rimliga, på de olika faktorerna och når slutsatsen att "de flesta forskningsfynd är falska för de flesta forskningsupplägg och inom de flesta forskningsfält". Man måste dock protestera mot hur slutsatsen formuleras och mot Ioannidis' artikelrubrik: han har knappast visat *varför* de flesta positiva vetenskapliga resultat *är* falska – snarare har han försökt visa *hur* de flesta positiva vetenskapliga resultat *skulle kunna vara* falska.

SYSTEMATISKA REPLIKATIONSSTUDIER

Vad som har saknats för den mer radikala slutsatsen är stora, systematiskt genomförda replikationsstudier av publicerade studier inom bestämda områden. Men en serie sådana replikationsstudier inom ett bestämt område har nu faktiskt genomförts av en sammanlutning av forskare

Ioannidis når slutsatsen att "de flesta forskningsfynd är falska för de flesta forskningsupplägg och inom de flesta forskningsfält".

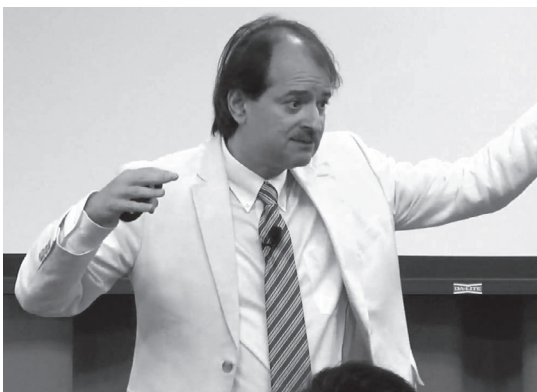
re som kallat sig "Open Science Collaboration". Denna grupp har låtit utföra upprepningsstudier av 100 studier som ursprungligen publicerades i de tre psykologiska tidskrifterna *Psychological Science*, *Journal of Personality and Social Psychology* och *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Utfallet av de 100 upprepningsstudierna rapporterades i *Science* i augusti 2015,³ vilket kom att uppmärksammas även utanför vetenskaps-samfundet.

Av de ursprungliga 100 studierna som Open Science Collaboration replikerade, hade 97% uppvisat statistiskt signifikant positiva resultat (dvs. förekomst av något samband). Bland de 100 replikationsstudierna var det däremot bara 36% som uppvisade signifikant positiva resultat. Totalt bedömdes 39% av de ursprungliga resultaten ha blivit framgångsrikt upprepade (bekräftade), oavsett om de från början varit positiva eller negativa. Replikationsstudierna uppges ha genomförts på så sätt att de i så stor utsträckning som möjligt efterliknade respektive ursprungsstudie. Man kan studera hur varje replikationsstudie genomfördes på en särskild webbplats.⁴

En upprepbarhetsgrad på 39% framstår som mycket låg. Det fanns också en allmän tendens att replikationsstudiens resultat blev svagare (mer negativt) än ursprungsstudiens, även när replikationsstudiens resultat i och för sig var positivt och statistiskt signifikant.

MÅNGA OBESVARADE FRÅGOR

Utfallet av de 100 upprepningsstudierna inom psykologi tyder alltså på att vi inte ska ha särskilt stort förtroende för resultaten från enstaka studier inom psykologi. Det-



John Ioannidis

ta visste vi egentligen redan – att man inte ska ha stort förtroende för enstaka studier är ju budskapet i den ”skolboksuppfattning” som jag redogjorde för i början. Resultatet väcker ändå många frågor. Bör vi alls lita mer på de 100 replikationsstudierna är på ursprungsstudierna? Det är inte givet, även om det finns ett viktigt argument som talar för det, nämligen att originalstudierna, men knappast replikationsstudierna, kan ha utsatts för publiceringsbias (eng. *publication bias*), med vilket menas en större benägenhet att publicera intressanta resultat än nollresultat.

Vidare undrar man: om uppreparhetsgraden inom psykologi bara är ca 39%, är det lika illa ställt inom andra fält? Det framstår som angeläget att försöka ta reda på detta, inte minst för att kunna resonera kring hur olika vetenskapliga områden skulle kunna lära av varandra i fråga om uppreparhet. För egen del tror jag att problemet kan vara större inom psykologi än inom de flesta andra veten-

skapsområden, och en indikation på detta tror jag just är den höga andelen positiva resultat som verkar publiceras inom psykologisk forskning. (Kanske är det symptomatiskt att det just är psykologiska forskare som känt sig kallade att starta tidskriften *Journal of Articles in Support of the Null Hypothesis*, för att uppmuntra kolleger att låta publicera även negativa resultat.⁵) Men som jag nämnt ovan kan man även tänka sig mekanismer i den motsatta riktningen: inom psykologi testar man ofta hypoteser som saknar starkt teoretiskt stöd, och därför borde man få många negativa resultat. Att fastställa vilka mekanismer som är starkast inom olika fält vore en grannliga uppgift.

PUBLICERINGSBIAS

Men om det ändå är ganska illa ställt med uppreparheten – åtminstone inom vissa områden, och åtminstone för positiva resultat – vad kan göras för att förbättra den? I den frågan är det just publiceringsbias som hittills mest kommit att diskuteras. Publiceringsbias uppkommer i två led, dels av att forskarna själva är mindre benägna att publicera negativa resultat, dels av att tidskrifternas redaktörer (och manusbedömare) är mindre benägna att vilja ta in artiklar med negativa resultat.

Forskare brukar ofta själva framhålla att negativa resultat (nollresultat) är lika viktiga som positiva resultat för vetenskapen. Varför skulle de ändå undanhålla negativa resultat? Två tänkbara förklaringar är lätta att komma på. För det första kan en forskare bli självkritisk och misstänka att något har gått snett när negativa resultat börjar trilla in, om andra forskare tidigare har publicerat positiva resultat inom samma område. Självttvivlen kan göra att man

avstår från att försöka publicera. För det andra blir positiva resultat mer uppmärksammade och citerade än negativa, och antal citeringar är en allmänt använd indikator på vetenskaplig vikt och framgång. Av denna anledning känns det kanske inte så angeläget att publicera negativa resultat. Samma förklaring skulle kunna tillämpas på tidskrifterna: även tidskrifternas redaktörer räknar gärna citeringar, och kan förmodas föredra att publicera artiklar som har en god chans att bli uppmärksammade, och som därmed ger tidskriften högre anseende.

Ett sätt att minska publiceringsbias, åtminstone i ledet som involverar forskarna själva, vore att införa ett system av förhandsregistrering av planerade studier. Då skulle man kunna kontrollera vilka registrerade studier som faktiskt till slut publiceras. Men naturligtvis skulle ett sådant system möta en del utmaningar, framför allt att se till att så många planerade studier som möjligt verkligen blir registrerade. Dyliga register finns redan på vissa områden, till exempel för kliniska prövningar inom medicinsk forskning.

Om registreringen även inkluderar uppgifter om hur forskarna exakt tänkt genomföra sin studie och sin dataanalys, skulle man också kunna minska problemet med att forskarna prövar sig fram och ändrar sig under studiens gång. Ett sådant "småfus", som

**Ett sätt
att minska
publiceringsbias
vore att införa
ett system av
förhandsregist-
ring av
planerade
studier.**

det kallas i ett välskrivet reportage i tidskriften *Modern psykologi*,⁶ torde kraftigt öka chanserna att finna positiva resultat. Forskarna behöver inte ens själva vara medvetna om att deras underhandsjusteringar är olämpliga och riskerar att snedvrیدا resultat.

Forskare måste i än högre grad göras uppmärksamma på dessa problem. Attitydförändringar behöver komma till stånd, men tyvärr är sådana nog inte så lätta att åstadkomma. Center for Open Science (COS) är en organisation som nyligen tagit initiativ i dessa frågor, och gruppen Open Science Collaboration, som genomförde de 100 uppföljningsstudierna inom psykologi, har verkat just under COS:s överinseende. COS har bland annat föreslagit riktlinjer för hur vetenskapliga tidskrifter ska agera för att öka forskningens upprepbarhet,⁷ och man har även skapat upprop för att få tidskrifterna att intyga att de tänker följa riktlinjerna.⁸

Det rör alltså på sig, och frågan lär debatteras vidare under lång tid. Det ska bli spännande att följa utvecklingen under de närmaste åren. ✍

Jesper Jerkert är doktorand i filosofi vid KTH.

REFERENSER

1. John P. A. Ioannidis: Why Most Published Research Findings Are False, *PLoS Medicine* 2(8), 2005, e124.
2. Daniele Fanelli: "Positive" Results Increase Down the Hierarchy of the Sciences, *PLoS One* 5(4), 2010, e10068.
3. Open Science Collaboration: Estimating the reproducibility of psychological science, *Science* 349(6251), 28 Aug 2015, aac4716.
4. <<https://osf.io/ezcuj/>>
5. <www.jasnh.com>

6. Nils Otto: Vem kan man lita på? *Modern psykologi* nr 4, 2015. Även tillgänglig på <modernpsykologi.se/2015/08/28/kan-man-lita-pa-psykologin/>.
7. B. A. Nosek m.fl.: Promoting an open research culture, *Science* 348(6242), 26 June 2015, 1422–1425.
8. <<https://cos.io/top/>>