

*Replikeringskrisen  
hotar vetenskapens  
trovärdighet*

**Gustav Nilsson**, docent vid Karolinska Institutet och forskare i neurovetenskap och metavetenskap, resonerar kring hur forskning som lärs ut på lärosäten och publiceras i vetenskapliga tidskrifter i hög grad kan vara vilseledande.

DEN SOM LÄNGE har läst *Folkvett* är van att se artiklar som kritiserar uppenbara exempel på pseudovetenskap: astrologi, homeopati, parapsykologi och andra irrläror. Ibland kontrasteras dessa mot riktig vetenskap, till exempel i form av referentgranskade artiklar publicerade av forskare vid etablerade universitet. Men även om den vetenskapliga metoden är den bästa vi har för att vinna kunskap,

finns det skäl att misstänka att mycket, kanske det mesta, av den vetenskapliga litteraturen i någon mening inte är korrekt vid den tid den publiceras. Här kommer jag att argumentera för att den forskning som bedrivs på våra lärosäten och som publiceras i traditionella tidskrifter faktiskt i ganska hög grad är infiltrerad av pseudovetenskap: den bär vetenskapens yttre former men saknar dess verkliga innehåll.

Under det senaste decenniet, ungefär, har det talats om en replikeringskris. Det har visat sig när man har försökt att systematiskt upprepa experiment, att det inte har gått att få samma resultat. Till exempel var jag själv delaktig i ett försök att upprepa 100 studier i experimentell psykologi. Replikerbarhet – att någon annan kan upprepa ett experiment och få samma resultat – har ansetts som en hörnsten i den vetenskapliga metoden sedan Galileis tid, men ändå var det ingen som innan vår studie 2015 hade försökt testa replikerbarheten i ett så stort urval av studier från ett specifikt forskningsfält. I vår studie samlade vi hundratals forskare för att så noga som möjligt upprepa de 100 experimenten på exakt samma sätt som de ursprungligen utförts, fast med nya försökspersoner. Vi fann att en tredjedel gick att upprepa med samma resultat.

I den debatt som följde var det många som tyckte att en tredjedel var en låg siffra. Men det fanns skäl att vänta sig att replikerbarheten skulle vara begränsad. En iögonfallande observation var att av de 100 experiment som replikerades hade 97 från början rapporterat positiva fynd, det vill säga resultat som bekräftade författarnas hypotes, och som var statistiskt signifikanta. Litteraturen som vi utgick ifrån verkade alltså ha en enormt stark tendens att publ-



icera positiva resultat. Men det är praktiskt taget omöjligt att få en så överväldigande dominans av positiva resultat om man öppet och ärligt rapporterar allt man undersöker. Det skulle bara vara möjligt om forskare hela tiden undersökte sanna hypoteser – om man bedrev forskning med extremt låg risk och aldrig gjorde något oväntat eller nytt. Och dessutom om forskarnas metoder hela tiden hade enormt hög kraft, så att de nästan aldrig på grund av slumpen skulle missa ett samband. Båda dessa antaganden är orimliga. Den höga andelen positiva resultat tyder därför på en stark publiceringssnedvridning (*publication bias*) där negativa resultat försvinner och positiva resultat lyfts fram.

Min kollega Michael Ingre och jag gjorde några ytterligare modelleringar av dessa resultat. En vetenskapsfilosof som heter Daniele Fanelli har uppskattat att publicerad

forskning i psykologi och psykiatri har mer än 90 % positiva fynd. Vi kombinerade siffrorna: en tredjedel resultat som kunde replikeras och 90 % positiva fynd, och gjorde en sannolikhetsmodell. Hur många resultat som faktiskt är sanna går inte att veta, så vi gjorde uppskattningar för en serie möjliga värden. Utifrån detta bedömde vi att stora mängder negativa resultat har observerats i forskning men aldrig publicerats: våra uppskattningar pekar på att av ca 50–100 negativa resultat som har observerats, är det i genomsnitt bara ett enda som publicerats. Vi bedömde att med så stark publiceringssnedvridning kommer praktiskt taget varje hypotes som undersöks att skenbarligen ha stöd i litteraturen, oavsett om den stämmer eller ej.

Vart försvinner då de negativa resultaten? Vi tror att det finns två möjliga vägar. Antingen hamnar resultaten i byrålådan och ser aldrig dagens ljus. Eller så analyserar man dem igen med en ny strategi, och prövar olika metodval, tills man får ett statistiskt signifikant resultat. Därefter konstruerar man ett narrativ kring det signifikanta resultatet. Man låtsas att man fått svar på en hypotes som man hade från början och då ser det ut som ett positivt resultat, fast man kanske bara hittat en variation i data som berodde på slumpen.

Ett exempel på hur dolda negativa resultat kan leda till skada är historien om depressionsläkemedlet reboxetin. Läkemedlet kom ut på marknaden 1997. Som stöd för dess effekt fanns en publicerad studie med 254 patienter som visade att reboxetin var bättre än placebo (sockerpiller) och tre publicerade studier med sammanlagt 507 patienter som visade att reboxetin inte var sämre än andra så kallade serotoninåterupptagshämmare (SSRI). Men



när den tyska myndigheten IQVIG år 2010 skulle utreda läkemedlets effekt lyckades de kräva ut fullständiga data från läkemedelföretagets kliniska studier. Då visade det sig att det hade gjorts ytterligare fem studier som hade jämfört reboxetin mot placebo (totalt 2002 patienter) och sex ytterligare studier som hade jämfört reboxetin mot andra SSRI (totalt 2 134 patienter). Den största delen av data hade alltså dolts. Och när IQVIG lade samman alla data visade det sig att reboxetin inte var bättre än placebo och att det var sämre än andra jämförda SSRI. Många tusentals patienter hade behandlats med reboxetin på ett felaktigt kunskapsunderlag. Sannolikt hade de flesta kunnat få ett annat mer effektivt läkemedel i stället om alla studiers resultat varit kända från början.

Forskare upplever en stark press att publicera positiva resultat. För att kunna söka nya anslag och tjänster är det

bäst att kunna göra troligt att man har ett konstruktivt forskningsprogram, där man hela tiden gör nya upptäckter. Men forskningslitteraturen i dess helhet tappar dramatiskt i kvalitet när enskilda forskare väljer att publicera endast sina positiva resultat. Teorier som egentligen saknar empiriskt stöd lever vidare i en slags zombietillvaro. Medicinska behandlingar som är verkningslösa eller direkt skadliga fortsätter att användas långt efter att mänsklighetens samlade kunskap borde varit tillräcklig för att mönstra ut dem.

Just inom klinisk medicinsk forskning har man varit särskilt uppmärksam på problemet att negativa resultat försvinner. Den starka kopplingen mellan forskning och praktik har lett till att man sett behovet av åtgärder för att få fram alla resultat – även negativa. Den som gör en klinisk studie behöver numera registrera studien i ett särskilt prövningsregister och sedan rapportera resultaten inom en bestämd tid. Meningen är dels att förhindra att resultat hamnar i byrålådan, dels att förhindra att man byter utfallsmått efter att ha sett resultaten. Antag till exempel att en studie av ett nytt cancerläkemedel har huvudfrågan: lever patienterna längre med den nya behandlingen? Då ska rapporten om studien också besvara den huvudfrågan, och inte bara någon sekundär fråga som till exempel om patienternas tumörer krympte på en röntgenbild, eller om ett visst blodprovsvärde förbättrades.

Men trots kraven på registrering och rapportering har vi långt ifrån löst problemet ens för de kliniska studierna. När man nyligen i Tyskland följde upp registrerade kliniska studier fann man, dels att färre än hälften hade rapporterat resultat inom två år från att studien var fär-

dig, dels att en fjärdede intel rapporterade något alls under hela uppföljningstiden. I Norden har vi nyligen startat ett projekt som ska göra en liknande uppföljning, ledd av Cathrine Axfors och undertecknad. Det är ett exempel på hur metavetenskap – forskning om forskningen själv – kan användas för att försöka förbättra kvaliteten på vår kunskapsförsörjning.

Sammanfattningsvis har vi en situation där forskares incitament på individnivå avviker kraftigt från vad som skulle behövas, för att den publicerade forskningen som helhet ska bli trovärdig och användbar. Jag tror att detta går att förändra. En väg framåt går genom att ta fram data som visar hur forskningens transparens och reproducerbarhet ser ut i dag. En annan väg handlar om att ta fram nya och bättre metoder för att bedöma vad som är bra forskning, utifrån hur väl den går att upprepa, till exempel om data är tillgängliga och en oberoende granskare kan få fram samma resultat som har rapporterats. I bästa fall är vi på väg mot en renässans för akademien som leverantör av verklig forskning i stället för pseudovetenskap. 🐣

*Gustav Nilsson*

## *Referenser*

- Axfors, Cathrine och Gustav Nilsson. *Systematic evaluation of clinical trial reporting at medical universities and university hospitals in the Nordic countries*. OSF, 2022. DOI: 10.17605/OSF.IO/BSA3D
- Eyding, Dirk, Monika Lelgemann, Ulrich Grouven, Martin Härter, Mandy Kromp, Thomas Kaiser, Michaela F Kerekes, Martin Gerken och Beate Wieseler. *Reboxetine for acute treatment of major depression: systematic review and meta-analysis of published and unpublished placebo and selective serotonin reup-*

- take inhibitor controlled trials. The BMJ*, 12 oktober 2010, 341. DOI: 0.1136/bmj.c4737
- Fanelli, Daniele. "Positive" results increase down the hierarchy of the sciences. *PLOS ONE*. 7 april 2010, 5(4). DOI:10.1371/journal.pone.0010068
- Ingre, Michael och Gustav Nilsson. *Estimating statistical power, posterior probability and publication bias of psychological research using the observed replication rate. Royal Society Open Science*, 12 september 2018, 5(9). DOI: 10.1098/rsos.181190
- Open Science Collaboration. *Estimating the reproducibility of psychological science. Science*. 28 augusti 2015, 349(6251): DOI: 10.1126/science.aac4716
- Riedel Nico, Susanne Wieschowski, Till Bruckner, Martin R Holst, Hannes Kahrass, Edris Nury, Joerg J Meerpohl, Maia Salholz-Hillel och Daniel Strech. Results dissemination from completed clinical trials conducted at German university medical centers remained delayed and incomplete. The 2014–2017 cohort. *Journal of Clinical Epidemiology*. 2022 Apr 1;144:1-7. DOI: 10.1016/j.jclinepi.2021.12.012