

I replikationskrisens spår agerar länder mot utbredda metodbrister

REPRODUCERBARHETSNETVERK DRIVER PÅ EN NY FORSKNINGSKULTUR

Cathrine Axfors,
med dr, postdok, Stan-
forduniversitetet; Upp-
sala universitet
● caxfors@stanford.edu

Robert Thibault,
med dr, postdok,
Stanforduniversitetet

Steven N Goodman,
professor i epidemio-
logi och folkhälsa
samt medicin, biträ-
dande dekanus för kli-
nisk och translationell
forskning, specialist-
läkare i pediatrik,
Stanforduniversitetet

Våren 2011 syntes notiser i Nature Medicine, Lancet Oncology och New England Journal of Medicine: en serie studier om tumörcellers genaktivitet som medel för att skraddarsy cellgiftsbehandling, skrivna av forskare vid välansedda Duke University i USA, återkallades [1-3]. Det hade inte gått att »återskapa vissa viktiga experiment«, och man erkände »korruption« av flera uppsättningar forskningsdata. Skandalen kom att hota hela universitetets anseende. Snarare än fusk rörde det sig om bristande forskningsrutiner av förbluffande omfattning [4, 5]. Databasens rubrikrad hade felaktigt kodats som data, vilket förskjutit ordningen. Gener vars uttryck aldrig hade mätts fanns med i data, och uppgifter om resistens och känslighet för kemoterapi hade förväxlats. Dessa och en rad andra misstag i datahanteringen och analyserna förvanskade fynden, som initialt hade firats som stor vetenskaplig framgång och föranlett flera kliniska prövningar. Efterföljande långdragna rättsfall har kostat Duke University tiotals miljoner dollar i skadestånd.

Felaktigheterna uppdagades av bioinformatiker vid ett annat universitet efter en förnyad analys av offentligt tillgängliga originaldata [6]. Det var inte deras första erfarenhet av en vetenskaplig återvändsgränd. Samma forskare försökte några år tidigare validera resultaten bakom Ovacheck, ett screeningtest för ovariecancer [7] som författare från bland annat USA:s motsvarighet till Läkemedelsverket rapporterat om i topptidskrifter. Testets enastående precision visade sig förklaras av bland annat artefakter från provhanteringen, som inte standardiserats mellan kontroller och cancerfall [7]. Gemensamt för dessa högprofilerade fall är en orsakskedja mestadels bestående av misstag. Som granskarna noterade var de vanligaste fe-len enkla, såsom rad- eller kolumnförskjutningar [6]. Fallen rön-te uppmärksamhet eftersom de indikerade djupgående problem med hur medicinsk forskning bedrivs. Om enkla fel passerar obemärkta genom hela forskningsprocessen på toppnivå, vad pågår då inom forskningen i stort, och hur kan situationen förbättras?

Replikationskrisen satte ljuset på tvivelaktig praxis

»Replikationskrisen« är ett namn på den trovärdighetskris som drabbat vetenskapen efter att en omfattande metaforskningslitteratur dokumenterat brister. En grundbult i forskningen är att fynd bör kunna återupprepas (Fakta 1). Det var därför alarmerande när läkemedelsföretaget Amgen år 2012 angav att enbart 6 av 53 banbrytande prekliniska studier kunde replikeras av deras forskare [8]. Försöken hade utförts i letan-

det efter molekylära mål för utvecklingen av cancerläkemedel, under konsultation av och tidvis samarbete med originalförfattarna. En författare ska ha medgett att man enbart rapporterat 1 av 6 genomförda experiment - det enda som var lyckosamt - eftersom det gav en »bättre story« [9]. Även liknande försök vid företaget Bayer visade att en minoritet av fynden var reproducerbara (14 av 67 studier) [10].

Frånvaron av bekräftande resultat verkar inte förklaras av triviala metodskillnader [11, 12]. Systematis-

»När ett uppmärksammat projekt syftade till att återupprepa 193 inflytelserika experiment i cancerbiologi var inte ett enda av dem beskrivet i tillräcklig detalj för att upprepas.«

ka litteraturoversikter har kartlagt hur experiment ofta utförs med suboptimala metoder (till exempel utan randomisering och maskering) och tenderar att rapportera överdrivna effekter [13, 14]. Den medicinska litteraturen har ett »överskott« på statistiskt signifikanta P-värden, särskilt strax under det vanliga

HUVUDBUDSKAP

- »Replikationskrisen« hänvisar till den förtroendekris som utlösts av att inflytelserika fynd i flera discipliner inte kunnat reproduceras. Medicinsk forskning är i hög grad drabbad.
- Forskningsmetoder i gränslandet mellan uppsåtlig vilseledning och oavsiktliga misstag, s k tvivelaktig forskningspraxis, är mycket utbredda.
- Högprofilerade skandaler och en omfattande metaforskningslitteratur har fått välrenommerade forskningsaktörer att agera för att stärka vetenskaplig reproducerbarhet och stringens.
- Reproducerbarhetsnätverk som involverar universitet, finansärer, lagstiftare och medicinska forskare ger möjlighet att påverka forskningskulturen på bred front.

gränsvärdet 0,05 [15-17], vilket tyder på en omfattande framselektion av slumpfynd hos både tidskrifter och forskare, men även manipulering. Problematiskt är även att bristfälliga metodbeskrivningar hindrar utomstående replikeringsförsök. När ett uppmärksammat projekt syftade till att återupprepa 193 inflytelserika experiment i cancerbiologi var inte ett enda av dem beskrivet i tillräcklig detalj för att upprepas [14]. En tredjedel av originalförfattarna svarade inte trots upprepade kontaktförsök eller avböjde att bistå. Under skandalen vid Duke University hade granskarna initialt tillgång till analysdata, men originalteamets datadelning och samarbetsvilja minskade när problemen upptäcktes. Även individuella forskare har vittnat om stora svårigheter att återupprepa andras studier, trots aktiva kontaktförsök, och att få kontrasterande fynd publicerade [18]. Den sammantagna bilden står i kontrast till devisen att forskningen är självkorrigering.

Tvivelaktig forskningspraxis kräver andra åtgärder

Replikationskrisen har lett till tesen att utbredd tvivelaktig forskningspraxis (Fakta 2), som i de flesta fall inte är avsiktligt vilseledande, åsamkar vetenskapen som helhet värre skada än oredlighet. I USA är definitionen av forskningsintegritet under omdaning [19] genom att fokus flyttas allt mer från forskarens intentioner till resultatens trovärdighet. Vad är skillnaden i praktiken? Trots de slutgiltiga bevis som tidigt lades fram under Duke University-skandalen [20] var det först efter upptäckten av en oegentlighet i författarens meritförteckning som effektiva åtgärder vidtogs, såsom avbrott av kliniska prövningar [5]. När skandalen briserat kommenterade universitetsledningen att inget granskningssystem kan ersätta tilliten till forskaren [4]. En modell där forskarens rykte och garantier tillåts ersätta oberoende utredningar är illa anpassad att hantera tvivelaktiga forskningsrutiner som framodlas i en komplex vetenskapskultur.

Anonyma enkäter bland forskare har visat att beteenden i gränslandet mellan oavsiktliga misstag och uppsåtlig vilseledning är mycket vanliga, exempelvis genom selektiv rapportering av resultat baserat på en »magkänsla« som bekräftar den egna hypotesen [21]. En stor andel medicinska forskare verkar känna osäkerhet inför vad som utgör sunda metoder. Duke University reviderade senare sin preventiva plan för

att skapa en miljö där oliktankande uppmuntras, lämpliga datahanteringssystem används och biostatistiker involveras i metodbeslut [4].

Hur uppstår tvivelaktig forskningspraxis? Förklaringarna omfattar bland annat forskarutbildningar som ger bristande kunskap om design och statistik, publiceringskulturer som lägger större vikt på nyhetsvärde än felkontroll, incitamentsystem som premierar hög produktivitet utan ansvar för om resultaten är sanna eller inte och hierarkiska strukturer med intressekonflikter [22].

Reproducerbarhetsnätverk en lovande strategi

Efter skandalen har Duke University utvecklat ett av USA:s mest ambitiösa program för att stärka vetenskaplig reproducerbarhet och stringens, med institutionsomfattande kurser, diskussionsforum för forskargrupper och ett nytt visseblåsarsystem riktat till juniora forskare [5]. Offentliga forskningsfinansiärer i USA, däribland National Institutes of Health, har infört ansökningskrav relaterade till vetenskaplig reproducerbarhet och stringens (Fakta 1). Tidskrifter som Nature har publicerat över 100 artiklar om replikationskrisen och infört checklistor för fullständig rapportering samt statistisk expertis i referentgranskningen. Även i Sverige har reportage och debattinlägg synts, om än sporadiskt [23, 24].

Det EU-sponsrade projektet »Standard operating procedures for research integrity« har sammanställt material som kan underlätta för universitet och finansierare att strukturera sin reformverksamhet (www.sops4ri.eu) [25]. Under arbetet identifierades flera exempel från europeiska universitet på innovativ verksamhet för att stärka forskningens integritet. Dessa var dock spridda skurar utan en sammanhållen struktur [25]. Oss veterligen följer även det lilla antalet svenska initiativ samma mönster [26-28]. Ett mer troligt framgångskoncept är koordinerade insatser med utbildningar såsom workshoppar i öppen vetenskap, resurser såsom datahanteringsstöd, användarvänliga verktyg såsom »Experimental design assistant« [29], anpassade regelverk åtföljt av granskning och återkoppling, karriärstegar med mindre fokus på publika-

FAKTA 1. Vetenskaplig reproducerbarhet och stringens [31, 32]

- Reproducerbarhet av metoder innebär möjligheten att exakt återupprepa en studie eller ett experiment baserat på tillgänglig metodbeskrivning.
- Reproducerbarhet av resultat innebär att upprepning av en studie eller ett experiment leder till resultat samstämmiga med originalet. Detta kallas även replikerbarhet.
- Vetenskaplig stringens (rigor) innebär att experimentell design, metodik, analys, tolkning och resultatrapportering utförs på ett sätt som minimerar snedvridning och bias.

FAKTA 2. Tvivelaktig forskningspraxis [33, 19]

Tvivelaktig forskningspraxis (questionable research practices) är beteenden som faller utanför oredlighet, men som är problematiska eftersom de kan snedvridera forskningsresultat. Följande är några exempel:

- Att fatta beslut om ytterligare insamling av data efter att ha testat om resultaten är statistiskt signifikanta eller inte.
- Att utesluta datapunkter baserat på en »magkänsla« om att de är felaktiga eller efter att ha undersökt hur resultatet påverkas.
- Att avrunda ett P-värde så att resultatet blir statistiskt signifikant (t ex rapportera att ett P-värde på 0,053 är <0,05).
- Att inte rapportera samtliga utfallsmått.
- Att enbart rapportera de experiment som »fungerade« och inte andra.
- Att rapportera ett explorativt fynd som om det funnits en hypotes om det från början.
- Att använda en databas för att testa många olika samband och enbart rapportera dem som visade statistiskt signifikanta resultat.

tionssiffror och en gemensam förståelse kring vikten av vetenskaplig reproducerbarhet och stringens. Här leder Storbritannien ligan, med åtskilliga landsomfattande utredningar om forskningskulturen och ett

och forskningsförslag. Reproducerbarhetsnätverkets idé är bland annat att stödja en förändring inifrån själva forskarskrået genom att samordna gräsrotsrörelser som i sin tur kan driva på förnyade arbetsätt hos exempelvis finansiärer och förläggare. Liknande nätverk i Sverige (www.swern.org), Danmark, Norge och Finland har tagit sina första steg på initiativ av engagerade forskare, medan andra parter ännu visar ett ljumt intresse. »Självrannsakan ligger i luften«, kommenterar den brittiske metaforskaren Marcus Munafò [30]. De närmaste åren kommer att utvisa om svenska universitet, finansiärer, lagstiftare och medicinska forskare tar chansen att lära sig av andras kostsamma misstag. ○

»En stor andel medicinska forskare verkar känna osäkerhet inför vad som utgör sunda metoder.«

aktivt »reproducerbarhetsnätverk« (reproducibility network). Sedan 2019 har nätverket skapat noder i över 50 universitet samt samarbeten med finansiärer

- Potentiella bindningar eller jävsförhållanden: Inga uppgivna.
- Cathrine Axfors postdokstipendium från Knut och Alice Wallenbergs stiftelse finansierade detta arbete.

Citera som: *Läkartidningen*. 2023;120:23011

REFERENSER

- Potti A, Dressman HK, Bild A, et al. Retraction: Genomic signatures to guide the use of chemotherapeutics. *Nat Med*. 2011;17(1):135.
- Bonnefoi H, Potti A, DeLorenzi M, et al. Retraction - Validation of gene signatures that predict the response of breast cancer to neoadjuvant chemotherapy: a substudy of the EORTC 10994/BIG 00-01 clinical trial. *Lancet Oncol*. 2011;12(2):116.
- Potti A, Mukherjee S, Petersen R, et al. Retraction: A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. *N Engl J Med*. 2006;355:570-80. *N Engl J Med*. 2011;364(12):1176.
- Appendix B: Gene expression-based tests developed at Duke University and used in clinical trials. In: Micheel CM, Nass SJ, Omenn GS (editors). *Evolution of translational omics. Lessons learned and the path forward*. Washington, DC: National Academies Press; 2012 p. 239-80.
- Appendix D: Detailed case histories. In: National Academies of Sciences, Engineering, and Medicine. *Fostering integrity in research*. Washington, DC: National Academies Press; 2017. p. 265-85.
- Baggerly KA, Coombes KR. Deriving chemosensitivity from cell lines: forensic bioinformatics and reproducible research in high-throughput biology. *Ann Appl Stat*. 2009;3(4):1309-34.
- Baggerly KA, Morris JS, Coombes KR. Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. *Bioinformatics*. 2004;20(5):777-85.
- Begley CG, Ellis LM. Drug development: raise standards for preclinical cancer research. *Nature*. 2012;483(7391):531-3.
- Begley S. In cancer science, many »discoveries« don't hold up. *Reuters*. 28 mar 2012.
- Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov*. 2011;10(9):712.
- Ioannidis JPA. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124.
- Lazic SE. Experimental design for laboratory biologists: maximising information and improving reproducibility. Cambridge: Cambridge University Press; 2016.
- Dirnagl U. Bench to bedside: the quest for quality in experimental stroke research. *J Cereb Blood Flow Metab*. 2006;26(12):1465-78.
- Errington TM, Denis A, Perfito N, et al. Challenges for assessing replicability in preclinical cancer biology. *Elife*. 2021;10:e67995.
- Ridley J, Kolm N, Freckelton RP, et al. An unexpected influence of widely used significance thresholds on the distribution of reported P-values. *J Evol Biol*. 2007;20(3):1082-9.
- Chavalaris D, Wallach JD, Li AHT, et al. Evolution of reporting P values in the biomedical literature, 1990-2015. *JAMA*. 2016;315(11):1141-8.
- van Zwet EW, Cator EA. The significance filter, the winner's curse and the need to shrink. *Stat Neerl*. 2021;75(4):437-52.
- Mobley A, Linder SK, Brauer R, et al. A survey on data reproducibility in cancer research provides insights into our limited ability to translate findings from the laboratory to the clinic. *PLoS One*. 2013;8(5):e63221.
- Ch 4. Context and definitions. In: National Academies of Science, Engineering, and Medicine. *Fostering integrity in research*. Washington, DC: National Academies Press; 2017. p. 57-76.
- Coombes KR, Wang J, Baggerly KA. Microarrays: retracing steps. *Nat Med*. 2007;13(11):1276-7; author reply 1277-8.
- Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS One*. 2009;4(5):e5738.
- Begley CG, Buchan AM, Dirnagl U. Robust research: institutions must do their part for reproducibility. *Nature*. 2015;525(7567):25-7.
- Ennart H. 53 synade cancerstudier gav »chockerande« insikt. *Svenska Dagbladet*. 9 nov 2021.
- Luleå tekniska universitet; Bergvall-Kåreborn B. Forskningskvalitet och reproducerbarhet [blogg]. 28 okt 2019 [citerat 10 mar 2023]. <https://www.ltu.se/ltu/Organisation/Rektor/blogg/Forskningskvalitet-och-reproducerbarhet-1.192305?l=sv>
- Mejlgaard N, Bouter LM, Gaskell G, et al. Research integrity: nine ways to move from talk to walk. *Nature*. 2020;586(7829):358-60.
- Göteborgs universitet. Reproducerbarhet inom medicinsk forskning, 3 hp [citerat 10 mar 2023]. <https://kompetensutveckling.gu.se/seminar/detail/3904>
- Karolinska institutet. Kurskatalog forskarutbildning - VT 23. Open science and reproducible research [citerat 10 mar 2023]. <https://kiwas.ki.se/katalog/katalog/kurs/5390?sessionid=ceaaa1423fe8b8e43e6d-8658cf5?lang=sv>
- Uppsala universitet; Gerdin B. Medfarm Play. Icke reproducerbar forskning [föreläsning]. 19 mar 2020 [citerat 10 mar 2023]. <https://media.medfarm.uu.se/play/video/10026>
- Percie du Sert N, Bamsey I, Bate ST, et al. The experimental design assistant. *PLoS Biol*. 2017;15(9):e2003779.
- Munafò M. Raising research quality will require collective action. *Nature*. 2019;576(7786):183.
- Goodman SN, Fanelli D, Ioannidis JP. What does research reproducibility mean? *Sci Transl Med*. 2016;8(341):341pi12.
- National Institutes of Health (NIH). Rigor and reproducibility [citerat 10 jan 2023]. <https://www.nih.gov/research-training/rigor-reproducibility>
- John LK, Loewenstein G, Prelec D. Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychol Sci*. 2012;23(5):524-32.

SUMMARY

In the wake of the replication crisis, reproducibility networks take action against widespread methodological shortcomings

The »replication crisis« refers to the crisis of confidence triggered by the failure to reproduce a large portion of important findings in several scientific disciplines, including medicine. Failed replication occurred in high-profiled scandals such as the »omics« case at Duke University, as well as in systematic attempts to reproduce influential preclinical studies. An extensive meta-research literature attests to problems with suboptimal methods choices and indicates that behaviors bordering between deliberate misleading and well-intentioned mistakes (questionable research practices) are very common (e.g. selective reporting of particular results »based on a gut feeling«). As a consequence, influential international institutions have been prompted to take action to strengthen research rigor and reproducibility. So-called reproducibility networks, pioneered in the UK, seem particularly promising to organize necessary coordinated efforts among a wide range of stakeholders.