

Imidlertid har systemet en uheldig bieffekt: *publikasjonsbias*. Det er i tidsskrifters interesse å publisere interessante og nye funn. Og studier som for eksempel viser at medikament X virker mot tarmkreft, kan oppfattes som mer interessante for tidsskriftene enn studier som viser at X *ikke* fungerer, selv om de siste kan være like viktige for *kreftforskningen*. Situasjonen fører for det første til at negative studier overses, noe som fører til ressurssløsing når forskere undersøker noe som allerede har vært undersøkt, men som de ikke vet om fordi resultatene aldri ble publisert. Men for det andre: Dersom *mange* undersøkelser utføres om hvorvidt X fungerer (kanskje mot forskjellige ting), er det statistisk sett sannsynlig at *noen* får positive resultater ved ren tilfeldighet, og dersom disse har større sannsynlighet for å bli publisert enn de negative resultatene, kan det etter hvert se ut som om X er et rent vidundermiddel, selv om det i realiteten ikke har noen effekt overhodet.

Et konkret eksempel er såkalt *p-hacking*: I enhver statistisk undersøkelse er det en sjanse for at et interessant – eller *signifikant* – resultat skyldes ren tilfeldighet. Tradisjonelt angir man at et resultat regnes som signifikant dersom sannsynligheten for ren tilfeldighet er tilstrekkelig lav, f.eks. under 5 % – at « $p \leq 0,05$ ». Det betyr at opptil én av 20 tilsynelatende signifikante resultater fremdeles bare er støy. Det kan vi kanskje leve med: Ingen forventer at vitenskap gir *sikkerhet*. Problemet oppstår når forskere gjør mange analyser på samme datamateriale: Man kan undersøke ikke bare om sjokolade er assosiert med kreft, men om det er assosiert med kreft blant østfoldinger, blant vestlendinger, blant dem over 30, blant fyrbøtere osv. Dersom hver analyse har 1/20 sjanse for et falskt positivt resultat, vil man fort ende opp med noen falske positive resultater om

man gjør mange nok analyser. Og dersom man er motivert til å publisere *positive* resultater, vil en slik måte å arbeide på fort føre galt av sted.

De siste årene har det vært mye snakk om *replikasjonskrisen*, særlig i psykologi, dvs. at et stort antall publiserte studier ikke har latt seg replikere (Nosek et al., 2015). P-hacking har nok noe av skylden.

Så må det tilføyes at mange av studiene faktisk også har latt seg replikere, og ikke minst: At vi *oppdager* problemet, kan være et sunnhetstegn. Pseudovitenskaper, i hvert fall definert som Popper gjør det, vil aldri oppleve replikasjonskriser.

Verdier påvirker hvordan vi tolker og sorterer data

Data er viktig, men målet for vitenskapelige undersøkelser er ikke bare å samle masse data. Vi ønsker å samle data som er *relevante og interessante* for oss, for eksempel fordi de bidrar til å gi oss forklaringer som tilfredsstiller våre behov. Enhver observasjon av verden gir oss egentlig utallige data, og interesser og verdier påvirker i) hvilke vi oppfatter som relevante og legger merke til, og ii) hvordan vi sorterer dem.

I biologien sorterer vi individer i *arter*, men de fleste biologer vil nok si både at det ikke finnes noen enhetlig definisjon av «art», og at kategoriene vi kaller «arter», ikke nødvendigvis markerer entydige skillelinjer i naturen selv. Snarere bruker vi kategorien fordi den egner seg til å sortere populasjoner av individer på måter som belyser detaljer som er viktige for oss, og som hjelper oss å formulere forklaringene vi trenger. Når historievitenskapen deler historiske