

Olika slutsatser från samma data

DEN ANALYTISKA FLEXIBILITETEN I MEDICINSK FORSKNING ÄR OFTA STOR

Gustav Nilssonne,
docent, Karolinska insti-
tutet, Stockholm
● gustav.nilssonne@ki.se

Magnus Johannesson,
professor, Handelshög-
skolan, Stockholm

Anna Dreber,
professor, Handelshög-
skolan, Stockholm

Den som prövat att analysera forskningsdata torde snabbt ha insett att det finns många olika val att göra. Forskaren behöver fatta en mängd beslut, exempelvis vilka variabler som ska undersökas, vilka observationer som ska inkluderas, hur en statistisk modell ska konstrueras och vilka kriterier som ska användas för att dra en viss slutsats. Friheten att använda olika analysstrategier gör att resultaten kan variera när olika forskare använder samma data för att testa samma hypotes [1].

Ett exempel på sådan variation är när tidskriften *Surgery* inom ett par månader publicerade två olika artiklar som undersökte samma fråga med samma data: huruvida en engångspåse (»retrieval bag«) minskade infektionsrisken vid laparoskopisk appendektomi. Båda artiklarna använde analysstrategier som föreföll fullt försvarbara, men de visade ändå olika resultat. Enligt den ena analysen minskade infektionsrisken [2], enligt den andra fanns en ökning som inte var statistiskt signifikant [3]. Analyserna i de båda artiklarna skilde sig med avseende på bland annat utfallsmått (endast intraabdominella abscesser respektive alla kirurgiska sårinfektioner), inklusionskriterier och variabler som användes som kovariat.

Det analytiska rummet

Inom sannolikhetsteori betecknar utfallsrummet mängden av alla möjliga utfall vid ett slumpmässigt försök. Forskarens valmöjligheter vid dataanalys kan i analogi med detta begrepp sägas ge upphov till ett analytiskt rum, bestående av de möjliga vägar som forskaren kan ta genom de val som står till buds (Figur 1). Det är välkänt att om man utforskar det analytiska rummet tills man får ett resultat som av någon anledning är önskvärt, till exempel att det är statistiskt signifikant (ofta definierat som $P < 0,05$), och sedan rapporterar endast detta resultat, så uppstår en snedvridning (bias) [1]. Det finns många namn för detta, till exempel »P-hackning« [4]. Denna typ av snedvridningar förekommer även i forskning som inte till sin natur är främst explorativ. Exempelvis har jämförelser mellan rapporterade kliniska prövningar och registreringar i prövningsregister visat att det är vanligt att utfallsmått bytts ut [5]. Utforskande av det analytiska rummet är delvis nödvändigt och behöver inte vara avsiktlig P-hackning. Forskare har dock starka incitament att visa positiva resultat för att kunna fortsätta sin verksamhet och göra karriär [6]. Även andra mekanismer kan ha betydelse, till exempel att sakkunniggranskare efterfrågar ytterligare analyser vid granskning av ett inskickat manuskript.

Ett sätt att bedöma det analytiska rummets omfattning och se hur resultaten kan variera är att göra ett stort antal analyser och rapportera alla. Detta har ibland kallats för »multiversell analys« [7] eller »effektvibration« [1]. Ett exempel är en metaanalys av randomiserade kontrollerade prövningar av naltrex-

on eller nalmefen jämfört med placebo vid alkoholberoende, som definierade 9 216 olika kombinationer av inklusionskriterier och modellspecifikationer och fann att 425 av dessa visade en statistiskt signifikant negativ effekt, 616 visade en statistiskt signifikant positiv effekt och resterande inte visade någon statistiskt signifikant skillnad mellan läkemedelsbehandling och placebo [8]. Här fanns alltså möjlighet att dra helt skilda slutsatser beroende på vilken väg som valdes genom det analytiska rummet. Det visar att utrymmet kan vara stort för att snedvrیدا resultaten genom P-hackning eller genom att analysstrategin anpassas i alltför hög grad till just de data som råkat observeras. I praktiken är det ofta svårt att utforska hela det analytiska rummet, eftersom det vanligen går att definiera fler alternativ än det är praktiskt möjligt att

»Metoden med multipla analytiker är ett sätt att placera forskaren under mikroskopet.«

beräkna, och dessutom eftersom inte alla experter är överens om vilka analysstrategier som faktiskt är försvarbara.

Multipla analytiker – en ny metavetenskaplig metod

Ett annat sätt att se hur mycket av det analytiska rummet som utnyttjas av forskare i verkligheten är att låta ett antal forskare oberoende av varandra analysera samma data med samma frågeställning. Denna metod har kallats för »multipla analytiker« och kan

HUVUDBUDSKAP

- Analys av forskningsdata medger vanligen stor flexibilitet i metodval, vilket ger upphov till ett »analytiskt rum« av möjliga strategier.
- Oberoende forskare kan nå skilda resultat med utgångspunkt från samma data och hypoteser.
- När forskare utforskar det analytiska rummet tills de får ett önskvärt resultat, och sedan rapporterar endast detta resultat, kan kraftiga snedvridningar uppstå.
- Att låta multipla analytiker undersöka samma data är en metavetenskaplig metod för att undersöka hur stor variation i resultat som kan följa av olika analysstrategier.
- Åtgärder som kan minska risken för snedvridningar inbegriper bland annat öppen delning av forskningsdata och kod samt förhandsregistrering.

användas för att uppskatta det analytiska rummets gränser, och variationen i resultat, avseende en viss fråga. Exempelvis har vi nyligen rapporterat en studie där 70 oberoende grupper analyserade samma data från ett experiment med funktionell magnetkamera-teknik [9]. Vi fann att alla 70 grupper hade en egen unik analysstrategi och att resultaten visade stor spridning när vi bad grupperna att rapportera huruvida 9 olika hypoteser kunde bekräftas. Ett relaterat projekt som vi för närvarande driver gör samma sak med data från experiment med elektroencefalografi (EEG) [10]. I denna typ av studier har de oberoende analytikerna vanligen inget incitament att dra en viss slutsats, utan metoden kan visa variationen på grund av att olika forskare fattar olika analytiska beslut.

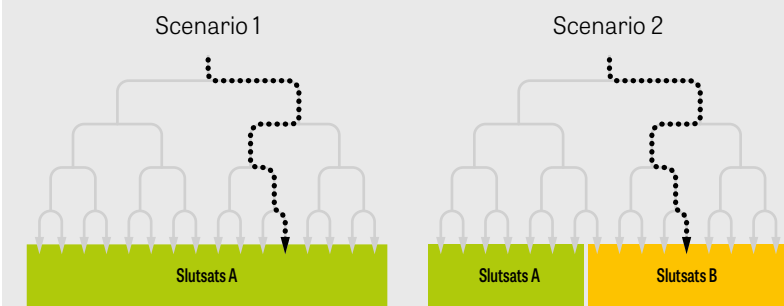
Metoden med multipla analytiker är ett sätt att placera forskaren under mikroskopet. Den tillåter oss att studera forskares beteende under relativt naturlistiska förhållanden. Det blir möjligt att studera hur forskarens beteende påverkas av olika typer av information och kontextfaktorer. Denna typ av forskning är en viktig del av det tvärvetenskapliga fältet som nu har börjat kallas för metavetenskap - kvantitativ forskning om forskningen själv.

Åtgärder som kan motverka snedvridningar

Öppen delning av data och kod för analys gör det möjligt för andra att återanalysera data med olika antaganden och kontrollera hur känsliga resultaten är för analysstrategin [9]. Förhandsregistrering med en analysplan för exakt vilka hypoteser som ska testas och hur, innan data samlas in, är ett sätt för forskare att begränsa friheten att utforska analysrummet och för läsare av vetenskapliga rapporter att kunna bedöma i vilken utsträckning forskarna följde sin ursprungliga plan [11]. Förhandsregistrering förhindrar inte explorativa analyser eller nya upptäckter, men gör det tydligare vilka rapporterade analyser som är explorativa.

Registrering av kliniska prövningar har förekommit länge inom medicinsk forskning, men för att detta effektivt ska motverka P-hackning behöver registreringen dels ange exakt hur analyser ska genomföras, dels sedan följas när studien rapporteras (så att exempelvis huvudutfall inte byts ut). Men även om dessa metoder används framgångsrikt, är det viktigt att notera att osäkerhetsintervall från statistiska test (såsom konfidensintervall) ändå underskattar den totala osäkerheten: de avspeglar hur slumpen kan in-

FIGUR 1. Schematisk illustration av det analytiska rummet



▶ Olika analysstrategier representeras av olika vägar. Scenario 1: Alla analysstrategier leder fram till samma slutsats (A). Scenario 2: Olika analysstrategier kan leda till olika slutsatser (A eller B). Figuren är anpassad från [13].

verka på ett statistiskt estimat, inte omfattningen av det analytiska rummet. Detta kan användas som argument för striktare statistiska signifikansskriterier [12], eller för att använda multipla analytiker för att uppskatta hur stor denna osäkerhet är [13].

Förhandsregistrering lämpar sig väl för randomiserade prövningar och andra prospektiva studier som kan registreras innan data samlas in. För retrospektiva studier är detta svårare, och det är också i denna typ av studier det analytiska rummet kan förväntas vara störst. Registerforskning, exempelvis, ger vittgående möjligheter att välja mellan olika inklusionskriterier, utfallsvariabler och kontrollvariabler. Förhandsregistrering kan vara möjlig vid analys av befintliga data om forskaren inte redan utforskat dessa, men försvåras bland annat av att det kan vara svårt att förutse alla val som behöver göras i analysen. Här kan andra lösningar behövas, såsom syntetiska (simulerade) dataset med samma struktur som det verkliga datasetet men utan något systematiskt samband mellan olika variabler [14]. Forskaren kan då bestämma exakt vilka analyser som ska genomföras baserat på det syntetiska datasetet, och sedan genomföra dessa analyser i det faktiska datasetet av en oberoende part. Genom att använda strategier som dessa kan vi förstärka trovärdigheten och användbarheten hos rapporterad forskning. ○

● Potentiella bindningar eller jävsförhållanden: Inga uppgivna.

Citera som: *Läkartidningen*. 2023;120:23015

REFERENSER

- Ioannidis JP. Why most discovered true associations are inflated. *Epidemiology*. 2008;19(5):640-8.
- Fields AC, Lu P, Palenzuela DL, et al. Does retrieval bag use during laparoscopic appendectomy reduce postoperative infection? *Surgery*. 2019;165(5):953-7.
- Turner SA, Jung HS, Scarborough JE. Utilization of a specimen retrieval bag during laparoscopic appendectomy for both uncomplicated and complicated appendicitis is not associated with a decrease in post-operative surgical site infection rates. *Surgery*. 2019;165(6):1199-202.
- Simmons JP, Nelson LD, Simonsohn U. False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol Sci*. 2011;22(11):1359-66.
- Dwan K, Altman DG, Cresswell L, et al. Comparison of protocols and registry entries to published reports for randomised controlled trials. *Cochrane Database Syst Rev*. 2011;2011(1):MR000031.
- Higginson AD, Munafò MR. Current incentives for scientists lead to underpowered studies with erroneous conclusions. *PLoS Biol*. 2016;14(11):e2000995.
- Steegen S, Tuerlinckx F, Gelman A, et al. Increasing transparency through a multiverse analysis: perspectives on psychological science. *Perspect Psychol Sci*. 2016;11(5):702-12.
- Palpacuer C, Hammam K, Duprez R, et al. Vibration of effects from diverse inclusion/exclusion criteria and analytical choices: 9216 different ways to perform an indirect comparison meta-analysis. *BMC Med*. 2019;17(1):174.
- Botvinik-Nezer R, Holzmeister F, Camerer CF, et al. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*. 2020;582(7810):84-8.
- Trübetschek D, Yang YF, Gianelli C, et al. EEGManyPipelines: a large-scale, grass-root multi-analyst study of EEG analysis practices in the wild. *MetaArxiv*. Epub 13 dec 2022. <https://osf.io/preprints/metaarxiv/jq342/>
- Nosek BA, Ebersole CR, DeHaven AC, et al. The preregistration revolution. *Proc Natl Acad Sci U S A*. 2018;115(11):2600-6.
- Benjamin DJ, Berger JO, Johannesson M, et al. Redefine statistical significance. *Nat Hum Behav*. 2018;2(1):6-10.
- Aczel B, Szasz B, Nilsson G, et al. Consensus-based guidance for conducting and reporting multi-analyst studies. *Elife*. 2021;10:e72185.
- Quintana DS. A synthetic dataset primer for the biobehavioural sciences to promote reproducibility and hypothesis generation. *Elife*. 2020;9:e53275.

SUMMARY

Same data – different conclusions: Analytical flexibility in medical research can be considerable

Analysis of research data entails many choices. As a result, a space of different analytical strategies is open to researchers. Different justifiable analyses may not give similar results. The method of multiple analysts is a way to study the analytical flexibility and behaviour of researchers under naturalistic conditions, as part of the field known as metascience. Analytical flexibility and risks of bias can be counteracted by open data sharing, pre-registration of analysis plans, and registration of clinical trials in trial registers. These measures are particularly important for retrospective studies where analytical flexibility can be greatest, although pre-registration is less useful in this context. Synthetic datasets can be an alternative to pre-registration when used to decide what analyses should be conducted on real datasets by independent parties. All these strategies help build trustworthiness in scientific reports, and improve the reliability of research findings.