



Hvordan vise likhet?

MEDISIN OG TALL

EVA SKOVLUND

E-post: eva.skovlund@ntnu.no

Eva Skovlund (f. 1959) er professor i medisinsk statistikk ved Institutt for samfunnsmedisin og sykepleie, Norges teknisk-naturvitenskapelige universitet, og seniorforsker ved Nasjonalt folkehelseinstitutt.

Forfatter har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

Hvordan kan man dokumentere at to behandlinger har like god effekt? Det kan man faktisk ikke, men man kan vise at forskjellen i effekt er så liten at den ikke antas å ha klinisk betydning.

I mange kontrollerte kliniske studier er forskningshypotesen at en ny behandling har bedre effekt enn standardbehandling, men slik er det ikke alltid. Noen ganger ønsker man i stedet å vise at en behandlingsstrategi ikke er dårligere enn en annen, for eksempel at en kortvarig antibiotikakur gir like god terapeutisk effekt som en lengre.

Hypoteser

I medisinsk forskning er det vanlig å teste hypoteser og beregne p-verdier. Den såkalte nullhypotesen er den hypotesen man ønsker å motbevise. Hvis målet er å vise forskjell mellom to behandlinger, er nullhypotesen at det ikke er forskjell. P-verdien leder enten til å forkaste eller ikke forkaste nullhypotesen.

En relativt vanlig misforståelse er at en høy p-verdi ($> 5\%$) kan tolkes som et uttrykk for at det ikke er forskjell i effekten av to behandlinger. Det er ikke en korrekt anvendelse av en p-verdi (1), og det er ikke slik man dokumenterer likhet (2).

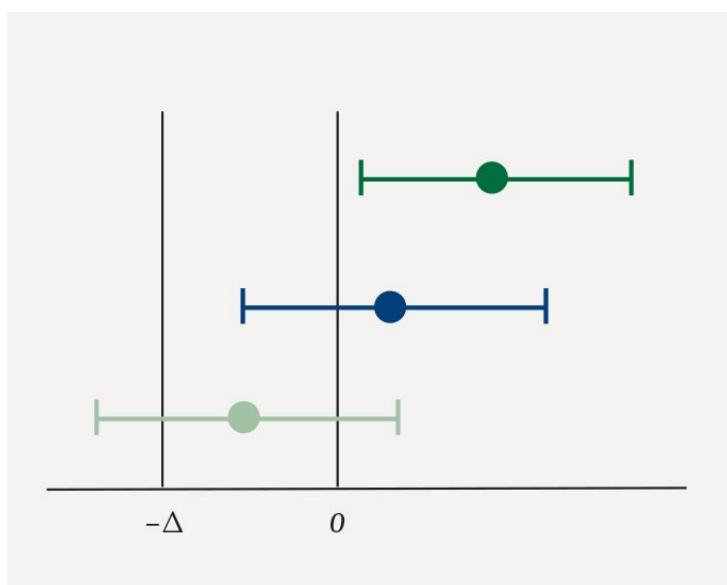
Dersom man kunne hevde at to behandlinger var like gode, kun basert på en høy p-verdi, ville det innebære at man kunne «dokumentere likhet» ved hjelp av små studier med så få pasienter at det ville være tilnærmet umulig å avdekke forskjell i effekt. At en nullhypotese ikke kan forkastes betyr på ingen måte at man har vist at den er sann.

Når man ønsker å vise at to behandlinger eller behandlingsstrategier har tilnærmet like god effekt, «snur» man i stedet testsituasjonen. Da skal nullhypotesen være at den eksperimentelle behandlingen er dårligere enn standardbehandlingen, mens forskningshypotesen, som ofte kalles alternativ hypotese, er at den eksperimentelle behandlingen er like god som, eller bedre enn standardbehandlingen. I praksis gjør man altså en ensidig test. På engelsk kalles slike studier «non-inferiority studies», og det er tatt til orde for å kalle dem «ikke-underlegenhetsstudier» på norsk (3). I mer tradisjonelle studier der man ønsker å vise effektforskjell, er tommelfingerregelen for øvrig å gjøre en tosidig test.

95 % konfidensintervall

Når man skal trekke slutninger om «ikke-underlegenhet», baserer man seg som regel på nedre grense i et 95 % konfidensintervall for differanse – heller enn å beregne en p-verdi.

Figur 1 viser resultater av tre forskjellige studier. Det øverste eksemplet viser et 95 % konfidensintervall som ikke inneholder verdien 0 og dermed dokumenterer statistisk signifikant forskjell i effekt på 5 %-nivå. Det midterste konfidensintervallet inneholder verdien 0, og det er altså ikke vist statistisk signifikant forskjell. Hvorvidt den eksperimentelle behandlingen kan sies å være «like god» som standardbehandlingen, avhenger av hvor stor forskjell man er villig til å anse at ikke har klinisk betydning (Δ). I eksemplet ligger nedre grense i konfidensintervallet til høyre for $-\Delta$, og man kan konkludere med at den eksperimentelle behandlingen ikke er underlegen. Det nederste eksemplet viser resultatet av en studie som ikke dokumenterer like god effekt – fordi den nedre enden av konfidensintervallet indikerer at forskjellen i effekt kan være større enn det man er villig til å akseptere.



Figur 1 Eksempler på 95 % konfidensintervaller for effektforskjell i ulike studier. Øverst: Eksperimentell behandling er statistisk signifikant bedre enn standardbehandling. I midten: Den eksperimentelle behandlingen er ikke underlegen («non-inferior») standardbehandlingen. Nederst: Eksperimentell behandling er potensielt underlegen. Δ = forskjellen man er villig til å anse at ikke har klinisk betydning

Akseptabel forskjell

En viktig vurdering når man planlegger å vise likhet eller «ikke-underlegenhet», er størrelsen av Δ , altså hvor mye dårligere man kan akseptere at den nye behandlingen kan være, men likevel akseptere den som likeverdig (4). Denne vurderingen må gjøres før man starter studien, slik at man unngår *post hoc*-definisjoner som er påvirket av innsamlede data.

La oss tenke oss at et nytt blodtrykkssenkende legemiddel er under utvikling og at man ønsker å vise at dette er likeverdig med et vanlig brukt legemiddel, i betydningen ikke klinisk underlegent. Hvor stor verdi av Δ er det rimelig å velge? Det finnes ikke noe fasitsvar, og man må først og fremst benytte klinisk skjønn og av praktiske årsaker kanskje også vurdere hvor mange individer det vil være nødvendig å inkludere i en studie. Det er vanlig å akseptere 5 mm Hg som en klinisk relevant forskjell i effekt på blodtryksreduksjon, så Δ må i alle fall være mindre enn dette. Jo lavere man setter Δ , desto flere pasienter må inkluderes. Ofte ser man at i hypertensjonsstudier at man benytter 2–3 mm Hg som grense for «ikke-underlegenhet». Det impliserer at man anser en slik forskjell i blodtryksreduksjon som så liten at den trolig ikke har betydning for risiko for

kardiovaskulære hendelser eller død.

Dersom man ikke vil akseptere at den eksperimentelle behandlingen kan ha litt dårligere effekt enn standardbehandlingen, må man i stedet dokumentere at den er (statistisk signifikant) bedre.

LITTERATUR:

1. Hellton KH, Røislien J. Verdens første p-verdi. Tidsskr Nor Legeforen 2017; 137: 897. [PubMed]
2. Altman DG, Bland JM. Absence of evidence is not evidence of absence. BMJ 1995; 311: 485. [PubMed][CrossRef]
3. Gjersvik P, Hem E, Jacobsen GW et al. Hva bør non-inferiority-studier kalles på norsk? Tidsskr Nor Legeforen 2014; 134: 852 - 3. [PubMed][CrossRef]
4. EMA. Guideline on the choice of non-inferiority margin. 2005.
http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003636.pdf (8.8.2017).

Publisert: 28. november 2017. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.17.0668
© Tidsskrift for Den norske legeforening 2020. Lastet ned fra tidsskriftet.no