Tidsskriftet Tidsskrift for Den norske legeforening

Negativ studie – et misvisende begrep

MEDISIN OG TALL

EVA SKOVLUND

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.

Mange kliniske studier lykkes ikke i å påvise statistisk signifikante forskjeller i effekt, og disse blir ofte kalt «negative studier». Dette begrepet er etter min oppfatning uheldig fordi det kan tolkes som evidens for mangel på effekt. I virkeligheten kan det dreie seg om studier som har inkludert for få pasienter til å påvise en reell forskjell som statistisk signifikant.

Det er velkjent at et statistisk signifikant resultat ikke nødvendigvis innebærer en klinisk relevant forskjell i effekt (1). En stor studie kan lede til en lav p-verdi selv om den estimerte forskjellen er så liten at den neppe har klinisk betydning. Et større problem er trolig feiltolkning av resultater som ikke er statistisk signifikante.

I medisinsk forskning anses typisk en p-verdi som er større enn 5 % ($p > 0,05$) som ikke statistisk signifikant. Hvis vi antar at det ikke er svakheter knyttet til forsøksplan eller gjennomføring av studien, kan mangel på statistisk signifikans enten skyldes at den sanne effektforskjellen faktisk er lik eller nær null eller at studien har inkludert for få pasienter til at en klinisk relevant forskjell avdekkes som statistisk signifikant (type 2-feil).

Det er blitt vanlig å benytte begrepet «negativ studie» om randomiserte studier som ikke har vist statistisk signifikant forskjell. En negativ studie burde vel strengt tatt bety at den eksperimentelle behandlingen har vist dårligere effekt enn standardbehandlingen eller at den ikke er bedre enn placebo (2). Men selv om man aksepterer at uttrykket oftest betyr «ingen statistisk signifikant

forskjell», er det uheldig fordi det kan gi inntrykk av at studien har vist at det ikke er forskjell i effekten av to behandlinger. Det korrekte er at det ikke er *vist* at det er forskjell. Disse to påstandene er ikke identiske. Såkalte «negative funn» blir likevel ofte fortolket som dokumentasjon på likhet mellom to behandlinger uten diskusjon om hvorvidt studien inkluderer et tilstrekkelig antall pasienter til at man kan trekke denne konklusjonen (3).

Utvalgsstørrelse

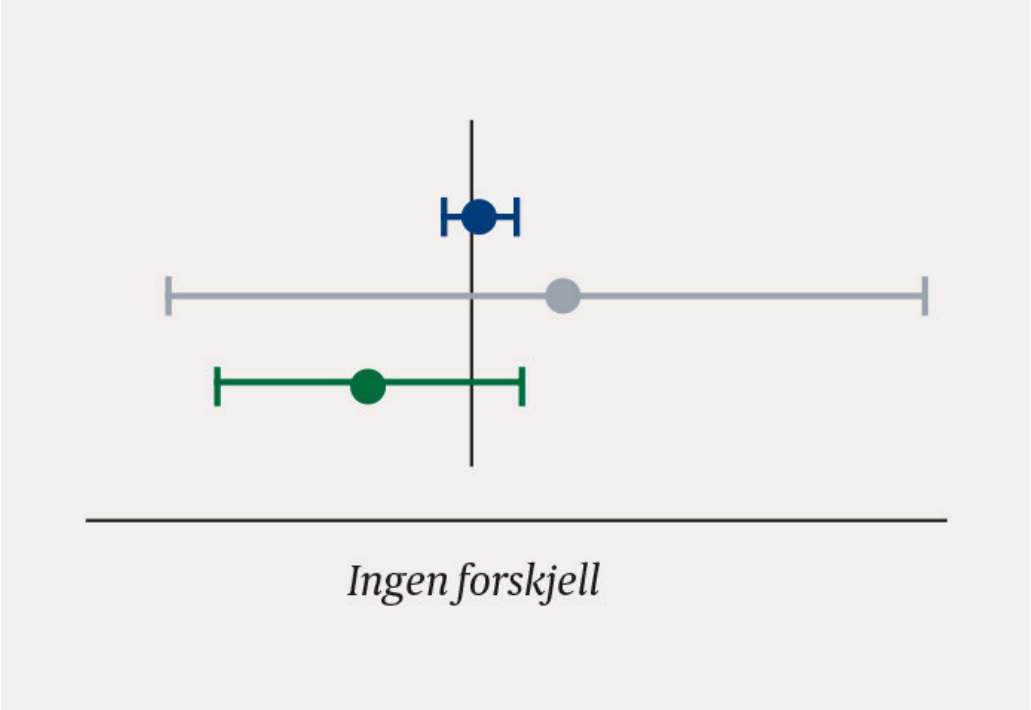
En essensiell del av planleggingen av en klinisk studie er å anslå hvor mange pasienter som må inkluderes. I beregningen baserer man seg gjerne på teststyrke; sannsynlighet for å avdekke en forskjell som statistisk signifikant, gitt at denne forskjellen faktisk eksisterer. Det er vanlig å kreve at en studie har minst 80 % teststyrke. Det innebærer at man aksepterer en sannsynlighet på 20 % for ikke å avdekke en bestemt forskjell selv om den i virkeligheten finnes.

Et vanlig problem i klinisk forskning er at det er vanskelig å inkludere et tilstrekkelig antall pasienter, og en studie kan dermed være for liten til at resultatet blir statistisk signifikant selv om det i virkeligheten er en klinisk relevant forskjell mellom to behandlinger. Det er vist at mer enn 50 % av studier der ingen signifikant forskjell i effekt ble påvist, trolig skyldtes lav teststyrke (4–6). I slike tilfeller er det ikke korrekt å hevde at den eksperimentelle behandlingen har like god (eller dårlig) effekt som standardbehandlingen.

Effektestimater og konfidensintervall

En p-verdi alene gir ikke tilstrekkelig dokumentasjon om effekt av behandling. Det er mer hensiktsmessig å se på et 95 % konfidensintervall for effekt enn å fokusere på p-verdien. Er konfidensintervallet vidt, tyder det på at studien kanskje er for liten. Hvis konfidensintervallet derimot er smalt, kan man trekke en holdbar konklusjon om effekt dersom studien ellers er av høy kvalitet.

Figur 1 viser eksempler på studier der det ikke er påvist statistisk signifikant forskjell mellom to behandlinger. Det øverste konfidensintervallet kommer fra en studie med et stort antall pasienter og høy presisjon. Her er det godt dokumentert at behandlingene har temmelig lik effekt og at en eventuell forskjell vil være så liten at den neppe har klinisk betydning. Det midterste intervallet viser resultatet av en studie med få pasienter og stor usikkerhet i effektestimatet. Her synes det urimelig å hevde at de to behandlingene er like. Det nederste intervallet viser en tendens til dårligere effekt av den eksperimentelle behandlingen, men forskjellen er ikke statistisk signifikant.



Figur 1 Eksempler på 95 % konfidensintervaller for effektforskjell. Øverst (i blått): en studie med et stort antall pasienter og høy presisjon. I midten (i grått): stor usikkerhet i effektestimatet. Nederst (i grønt): dårligere effekt av den eksperimentelle behandlingen, men ikke statistisk signifikant

Alle de tre eksemplene er «ikke-signifikante» ($p > 0,05$), men det er stor forskjell i hvordan de bør fortolkes.

LITTERATUR

1. Pripp AH. Antalls- og styrkeberegninger i medisinske studier. Tidsskr Nor Legeforen 2017. [CrossRef]

2. Scott C, Wasserman T. When is a negative study not negative? Int J Radiat Oncol Biol Phys 1997; 39: 859 - 61. [PubMed][CrossRef]

3. Brody BA, Ashton CM, Liu D et al. Are surgical trials with negative results being interpreted correctly? J Am Coll Surg 2013; 216: 158 - 66. [PubMed] [CrossRef]

4. Costa LJM, Xavier ACG, del Giglio A. Negative results in cancer clinical trials–equivalence or poor accrual? Control Clin Trials 2004; 25: 525 - 33. [PubMed][CrossRef]

5. Keen HI, Pile K, Hill CL. The prevalence of underpowered randomized clinical trials in rheumatology. J Rheumatol 2005; 32: 2083 - 8. [PubMed]

6. Bedard PL, Krzyzanowska MK, Pintilie M et al. Statistical power of negative randomized controlled trials presented at American Society for Clinical Oncology annual meetings. J Clin Oncol 2007; 25: 3482 - 7. [PubMed] [CrossRef]

Publisert: 17. april 2018. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.18.0170

Opphavsrett: © Tidsskriftet 2026 Lastet ned fra tidsskriftet.no 18. juni 2026.