



# Antalls- og styrkeberegninger i medisinske studier

---

## MEDISIN OG TALL

ARE HUGO PRIPP

E-post: [apripp@ous-hf.no](mailto:apripp@ous-hf.no)

Are Hugo Pripp (f. 1971) er forsker og biostatistiker ved Oslo senter for biostatistikk og epidemiologi, Forskningsstøtteavdelingen, Oslo universitetssykehus, og professor II ved Fakultet for helsefag, Høgskolen i Oslo og Akershus.

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

---

Antallet deltagere i en studie påvirker kostnader, tidsbruk og arbeidsmengde, men også hvilken kunnskap og innsikt studien kan gi.

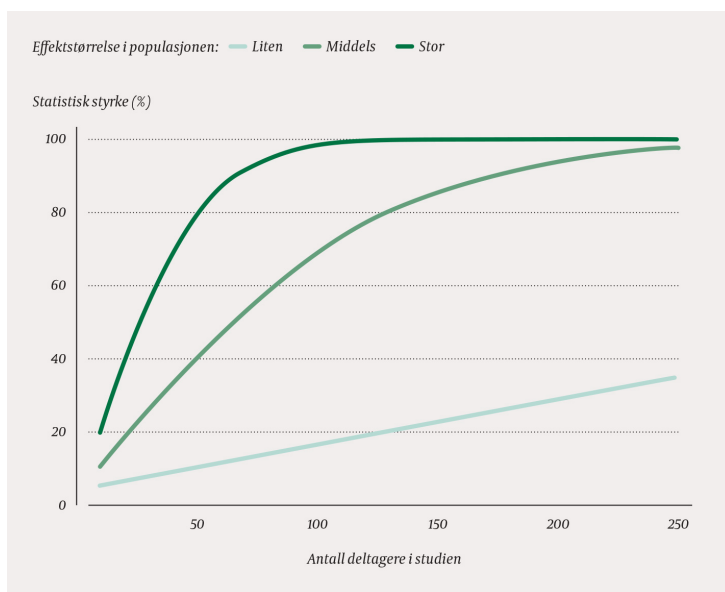
Mange forbinder statistiske analyser i medisinen med beskrivende statistikk, signifikanstesting og regresjonsmodeller. Analyser av observerte målinger gir oss medisinsk og vitenskapelig innsikt.

Antalls- og styrkeberegninger er på sett og vis den motsatte prosessen. Vi anslår på forhånd hvor stor en effekt enten kan forventes å være eller hvor stor den bør være for at den skal ha klinisk betydning og beregner sannsynligheten for at resultatet av en studie med et gitt antall pasienter blir statistisk signifikant.

Den statistiske styrken er sannsynligheten for et statistisk signifikant resultat, gitt at våre antagelser er korrekte. I statistisk terminologi er styrken sannsynligheten for at den statistiske testen forkaster nullhypotesen hvis den alternative hypotesen er sann. Den matematiske og statistiske teorien er klart definert for de fleste statistiske metoder, men de etiske og forskningsmetodologiske implikasjonene av antalls- og styrkeberegninger er mer omdiskutert.

## Hva gir statistisk styrke?

Antall deltagere i en studie, effektstørrelsen, utvalgets homogenitet og risikoen for feilaktige statistiske konklusjoner påvirker den statistiske styrken (fig 1) (1). Jo større effektstørrelse, desto større statistisk styrke – og desto færre deltagere trengs for å finne statistisk signifikante resultater.



**Figur 1** Sammenhengen mellom antall deltagere i studien og statistisk styrke ved liten, middels eller stor effektstørrelse, Cohens  $d$  lik henholdsvis 0,2, 0,5 eller 0,8 (1). Økt antall deltagere i studien gir økt statistisk styrke og større sannsynlighet for å kunne påvise en statistisk signifikant forskjell mellom to utvalg ( $p$ -verdi  $< 0,05$ )

Utvalgets homogenitet uttrykker hvor like deltagerne er. Mer homogene deltagere gir lavere standardavvik og økt statistisk styrke.

Som ellers i statistikkfaget må vi vurdere sannsynligheten for å ta feil. Statistikere anbefaler ofte 5% sannsynlighet for å hevde effekt når denne ikke er sann (ofte kalt  $\alpha$  eller signifikansnivået) og 20% sannsynlighet for å hevde ingen effekt når det faktisk er en sann effekt (ofte kalt  $\beta$ ). Den statistiske styrken er  $1 - \beta$ , noe som da gir en statistisk styrke på 80% (2).

## Statistisk og klinisk signifikans

Statistisk og klinisk signifikans er ikke det samme. Statistisk signifikans er matematisk definert, mens definisjonen av klinisk signifikans på ingen måte er like entydig. Ideelt sett burde det være enighet i det medisinske fagmiljøet om hvor stor en effekt skal være for at den skal ha klinisk betydning, men det er ikke alltid slik.

Resultater fra tidligere studier kan være til hjelp. Eller man kan bruke generelle statistiske uttrykk for effektstørrelse (3). Cohens  $d$  er et eksempel på en statistisk effektstørrelse. Den uttrykker forskjellen i gjennomsnitt mellom to grupper relativt til standardavviket og angir liten, middels eller stor effekt (1).

## Statistisk styrke og etikk

Vi kan bruke feil antagelser i beregningene, benytte en uriktig metode, unngå å utføre antalls- og styrkeberegninger eller ikke ta hensyn til slike beregninger ved gjennomføringen av en studie. Alt dette kan medføre at enten for mange eller for få deltagere blir inkludert.

For mange inkluderte medfører unødvendig bruk av tid og ressurser og at deltagerne kan bli utsatt for unødig stor risiko og byrde. Hvis det er for få deltagere i studien, blir det ofte resultater som tyder på en reell medisinsk effekt, men som ikke er statistisk signifikant. En  $p$ -verdi lavere enn 0,05 kan være forskjellen mellom å kunne hevde noe med «sikkerhet» eller kun å kunne si at flere studier er nødvendig (4).

Hva er egentlig verst – for få eller for mange deltagere i en studie? Spesielt er studier med lav statistisk styrke på grunn av for få inkluderte deltagere kritisert. Det er sagt at studier med lav statistisk styrke er uetiske, fordi de utsetter deltagerne for unødig stor risiko og byrde

uten at de kan gi tilstrekkelig vitenskapelig og medisinsk innsikt (3). Andre hevder derimot at det etiske forholdet mellom deltagerens byrde og studiens vitenskapelige verdi kan være mindre fordelaktig for store enn for små studier. De mener at studier med lav statistisk styrke ikke nødvendigvis er uetiske og at en statistisk styrke på minst 80% ikke er et generelt krav til en etisk forsvarlig studie (5).

Etter min erfaring er antalls- og styrkeberegninger i anvendt medisinsk forskning basert på en ikke ubetydelig del kvalifiserte gjetninger, og tilgjengelige ressurser avgjør i stor grad antallet inkluderte deltagere i studien. Antalls- og styrkeberegningene kan dels være blitt påvirket av hva som er praktisk og økonomisk mulig. Selv i slike situasjoner er det nyttig med slike beregninger. Da vet man i hvert fall hva studien ikke kan vise.

---

#### LITTERATUR:

1. Cohen J. The Effect Size Index: d. Statistical Power Analysis for the Behavioral Sciences. 2. utg. Hillsdale, NJ: Lawrence Earlbaum Associates, 1988: 20-7.
2. Burmeister E, Aitken LM. Sample size: how many is enough? Aust Crit Care 2012; 25: 271 - 4. [PubMed][CrossRef]
3. Halpern SD, Karlawish JH, Berlin JA. The continuing unethical conduct of underpowered clinical trials. JAMA 2002; 288: 358 - 62. [PubMed][CrossRef]
4. Pripp AH. Hvorfor p-verdien er signifikant. Tidsskr Nor Legeforen 2015; 135: 1462 - 4. [PubMed][CrossRef]
5. Bacchetti P, Wolf LE, Segal MR et al. Ethics and sample size. Am J Epidemiol 2005; 161: 105 - 10. [PubMed][CrossRef]

---

Publisert: 18. september 2017. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.17.0414

© Tidsskrift for Den norske legeforening 2020. Lastet ned fra tidsskriftet.no