



# Hvorfor brukes sjelden ensidige hypotesetester?

---

## MEDISIN OG TALL

STIAN LYDERSEN

E-post: [stian.lydersen@ntnu.no](mailto:stian.lydersen@ntnu.no)

Stian Lydersen er dr.ing. og professor i medisinsk statistikk ved Regionalt kunnskapssenter for barn og unge – psykisk helse og barnevern (RKBU Midt-Norge) ved Institutt for psykisk helse, NTNU. Forfatteren har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

---

Mange hypoteser i medisinsk forskning er i utgangspunktet ensidige, for eksempel i et randomisert kontrollert forsøk der man vil undersøke om en ny type medisinsk behandling har bedre effekt enn vanlig behandling (*treatment as usual*). Hvorfor bruker man da tosidige hypotesetester?

La oss for eksempel anta at man registrerer antall suksesser, altså antall som blir friske av sykdommen, i hver av to behandlingsgrupper. Nullhypotesen ( $H_0$ ) er at sannsynligheten for suksess er den samme i hver gruppe. Men hva er alternativhypotesen? Dette er en studie der man vil undersøke om den nye behandlingen har bedre effekt enn standard behandling, altså en overlegenhetsstudie. Da kunne man tenke seg at alternativhypotesen er nettopp dette. Dette kalles en ensidig alternativhypotese, og den tilhørende hypotesetesten og p-verdien kalles ensidig. Hvis man derimot velger en tosidig alternativhypotese, dvs. at den nye behandlingen har en effekt som er forskjellig fra vanlig behandling, får man en tosidig hypotesetest og en tilhørende tosidig p-verdi.

## Høyere styrke i ensidige tester?

Argumentet som kunne brukes for å velge en ensidig test, er at den har høyere statistisk styrke enn den tilsvarende tosidige testen. La oss tenke oss at man planlegger et randomisert kontrollert forsøk, og ønsker høy sannsynlighet for å påvise en forskjell i effekt dersom sannsynligheten for suksess ved vanlig behandling og ny behandling er henholdsvis 0,6 og 0,8. Hvis man planlegger å bruke en tosidig test, vil det trenge 82 pasienter i hver gruppe for å få en statistisk styrke på 80 % ved signifikansnivå 0,05. Hvis man derimot planlegger en ensidig test, er det tilstrekkelig med 64 pasienter i hver gruppe.

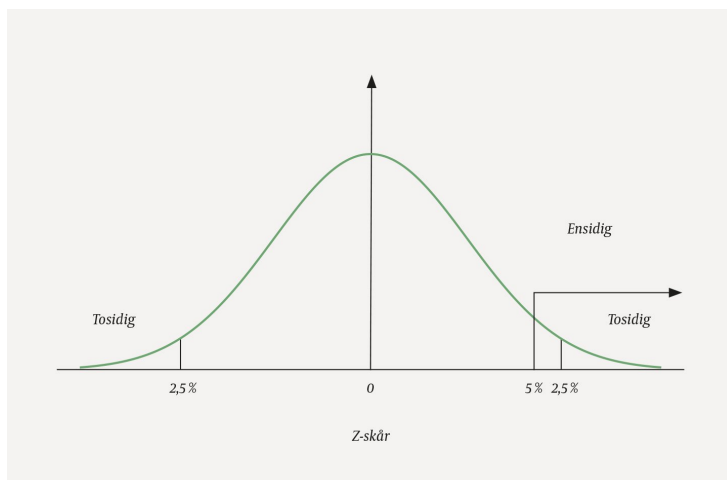
La oss tenke oss at forsøket deretter ble gjennomført med 100 pasienter i hver av gruppene. I gruppen med vanlig behandling ble 64 friske, mens i gruppen med ny type behandling ble 76 friske. Estimert forskjell i sannsynlighet for suksess blir  $76/100 - 64/100 = 0,12$ . Pearsons khikvadrattest gir en tosidig p-verdi på 0,064, altså er ikke forskjellen statistisk signifikant ved signifikansnivå på 0,05. Men hvis alternativhypotesen var ensidig, ville p-verdien bli halvparten av dette, altså 0,032. En tosidig p-verdi er generelt lik to ganger den tilsvarende ensidige p-verdien.

Det var en viss diskusjon om valg av ensidige versus tosidige tester i medisinsk statistikk rundt 1990-tallet (1, 2). Men en sak har alltid vært udiskutabel: Man må bestemme på forhånd om man skal bruke ensidig eller tosidig hypotesetest. Her har det antakelig vært syndet mye. Altman skrev i sin lærebok i 1991: «De få ensidige testene som jeg har sett rapportert i publiserte artikler har vanligvis gitt p-verdier mellom 0,025 og 0,05, slik at resultatet ville vært ikke-signifikant ved en tosidig test. Jeg tviler på at de fleste av disse var forhåndsplanlagt som ensidige tester» (3, s. 171, egen oversettelse).

## Effekt i begge retninger er mulig

Man kunne argumentere for å bruke en ensidig test bare hvis en effekt i motsatt retning er umulig eller uten interesse. Men man kan sjelden være sikker på at en effekt i motsatt retning er umulig. Da ville det ikke være behov for å gjennomføre studien (3, s. 171). Imidlertid finnes det en type studie der en effekt i motsatt retning er uten interesse, nemlig en ikke-underlegenhetsstudie. Den nye behandlingen vil bli tatt i bruk dersom den er minst like god som standard behandling. Den behøver ikke være bedre, og den *relevante* hypotesen er ensidig (4).

Kunne man bruke en ensidig test i en overlegenhetsstudie hvis dette er bestemt på forhånd? Det ville bli problematisk hvis effekten faktisk skulle vise seg å være i motsatt retning, altså at den nye behandlingen viste dårligere effekt enn standard behandling. Da måtte dette i prinsippet tilskrives tilfeldigheter, uansett hvor stor forskjellen var. Ensidige tester har riktignok høyere statistisk styrke i den ene retningen, men utelukker muligheten for å kunne påvise en eventuell effekt i motsatt retning. Dette er illustrert i Figur 1. Dette, samt andre argumenter for å bruke tosidige tester, er nærmere utdypet i (1).



**Figur 1** Ensidig eller tosidig test i en overlegenhetsstudie.

## Konsensus om tosidige tester

Det synes i dag å være konsensus om å bruke tosidige tester innen medisinsk forskning. Det gjelder både i intervensjonsstudier og observasjonelle studier. Det eneste viktige unntaket er ikke-underlegenhetsstudier, der det er riktig å bruke ensidige tester.

---

### LITTERATUR:

1. Moyé LA, Tita ATN. Defending the rationale for the two-tailed test in clinical research. *Circulation* 2002; 105: 3062–5. [PubMed][CrossRef]
2. Bland JM, Altman DG. One and two sided tests of significance. *BMJ* 1994; 309: 248. [PubMed][CrossRef]
3. Altman DG. *Practical statistics for medical research*. London: Chapman and Hall, 1991.
4. Skovlund E. Hvordan vise likhet? *Tidsskr Nor Legeforen* 2017; 137. doi: 10.4045/tidsskr.17.0668.

Publisert: 7. juni 2021. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.21.0111  
© Tidsskrift for Den norske legeforening 2020. Lastet ned fra tidsskriftet.no