
Analyser underveis i randomiserte studier

MEDISIN OG TALL

EVA SKOVLUND

eva.skovlund@ntnu.no

Eva Skovlund er professor i medisinsk statistikk ved Institutt for samfunnsmedisin og sykepleie, NTNU.

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

Inklusjon av pasienter i randomiserte studier kan ta lang tid, og det kan være hensiktsmessig å stoppe en studie hvis forskjell i behandlingseffekt kan påvises tidlig. For å unngå økt risiko for falskt positive funn må man definere stopperegler på forhånd.

Mens man inkluderer og følger opp pasienter i en kontrollert klinisk studie, akkumuleres informasjon om effekt av behandlingen. Av både etiske og økonomiske årsaker kan det være av interesse å stoppe studien straks man har tilstrekkelig informasjon til å konkludere at en behandling er bedre enn en annen. Analyser som gjøres mens studien pågår, såkalte interimanalyser, vil kunne redusere antallet pasienter det er nødvendig å inkludere.

Signifikansnivå

Når man gjør en signifikanstest, sammenligner man p -verdien med et valgt signifikansnivå, som regel satt til 5 %. Det vil si at man aksepterer inntil 5 % sannsynlighet for å forkaste en nullhypotese som egentlig er sann (1). Hver gang en test blir utført på 5 %-nivå, er sannsynligheten 5 % for å konkludere med at det er forskjell mellom de to behandlingene, selv om det i virkeligheten ikke er det. Hvis man utfører flere tester, vil risikoen for at minst én av dem leder til et falskt positivt funn, bli større enn 5 %. Det er vist at med én ekstra analyse i et voksende datasett, vil den sanne risikoen øke til 8 %, og med maksimalt fem analyser vil den bli 14 % (2).

Stopperegler

For å unngå problemer knyttet til multiple signifikanstester må man planlegge på forhånd hvorvidt det skal utføres interimanalyser, og eventuelt hvor mange. Forskjellige metoder har til felles at hver enkelt interimanalyse utføres med et lavt signifikansnivå for å sikre at den totale sannsynligheten for et falskt positivt funn blir maksimalt 5 %. Samme nivå kan benyttes for hver test, for eksempel 2,9 %-nivå med maksimalt to tester eller 1,8 %-nivå med maksimalt fire (3). O'Brien og Fleming (4) benytter i stedet et svært lavt nivå for den første analysen og deretter økende nivåer (tabell 1). Begge tilnærmingene sikrer at det totale signifikansnivået blir korrekt (5 %).

Tabell 1

Signifikansnivå for hver enkelt test som sikrer et totalt signifikansnivå på 5 % etter O'Brien og Flemings metode (4).

Maksimalt antall analyser	Analysenummer	Signifikansnivå for hver test	
	1	1	0,05
	2	1	0,0051
		2	0,0475
4	1	1	0,00004
		2	0,0039
		3	0,0184
		4	0,0411

Et fast signifikansnivå gir større sannsynlighet for å kunne stoppe tidlig, men gir samtidig tap av teststyrke og kan gjøre det nødvendig å inkludere flere pasienter enn i en studie uten interimanalyser. O'Brien og Flemings metode påvirker nesten ikke teststyrken siden det meste av nivået spares til den siste analysen, der alle pasientene er inkludert. Dette er antagelig hovedgrunnen til at denne metoden er mye brukt i praksis. Det kan også ses som en fordel at man ikke stopper en studie basert på et lite antall pasienter, med mindre den estimerte forskjellen i effekt mellom to behandlinger viser seg å være stor.

Å avslutte en nytteløs studie

De fleste formelle interimanalyser utføres for å oppdage forskjell i effekt tidlig. Men hva hvis tidlige resultater synes å indikere at forskjellen er for liten til å ha klinisk betydning? For å spare ressurser kan det også være hensiktsmessig å avslutte en studie på grunn av «nytteløshet» (eng. *futility*). Beslutningen om å

stoppe baseres gjerne på at den predikerte sannsynligheten for å vise forskjell i effekt er lav ut fra de foreløpige resultatene. Denne tilnærmingen vil ikke øke sannsynligheten for et falskt positivt funn slik som gjentatte signifikanstester gjør. Slike «nytteløshetsanalyser» kan være planlagt på forhånd, men er nok ofte ikkedet. Uansett om de er forhåndsplanlagt eller ikke, er det en risiko for at man avslutter en studie som egentlig kunne ha vist forskjell i effekt, og dermed taper teststyrke.

Åpenhet?

Åpenhet om resultater i pågående studier kan få konsekvenser for rekruttering av pasienter. En tidlig interimanalyse som viser liten forskjell mellom behandlinger, vil kunne gi problemer med å rekruttere pasienter. En positiv, men ikke signifikant trend vil også kunne påvirke inklusjon, og hvis studien ikke er dobbeltblind, vil frafallet typisk kunne øke i den ene behandlingsgruppen. For å opprettholde kvaliteten på en studie er det viktig å tenke igjennom hvem som får informasjon om når det skal gjøres interimanalyser, og hvem som skal ha tilgang til resultatene (5).

Den endelige avgjørelsen om å stoppe en randomisert studie tidlig bør neppe baseres kun på en statistisk stopperegel. Forekomst av bivirkninger, en vurdering av forholdet mellom nytte og risiko og ny kunnskap som er frembrakt utenfor den aktuelle studien, må også tas hensyn til.

LITTERATUR

1. Lydersen S. Type I-feil og type II-feil. Tidsskr Nor Legeforen 2021; 141. doi: 10.4045/tidsskr.21.0013. [PubMed][CrossRef]
2. Armitage P, McPherson CK, Rowe BC. Repeated significance tests on accumulating data. J R Stat Soc [Ser A] 1969; 132: 235–44. [CrossRef]
3. Pocock SJ. Group sequential methods in the design and analysis of clinical trials. Biometrika 1977; 64: 191–9. [CrossRef]
4. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics 1979; 35: 549–56. [PubMed][CrossRef]
5. European Medicines Agency. Guideline on Data Monitoring Committees. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-data-monitoring-committees_en.pdf Lest 14.9.2021.

Publisert: 8. november 2021. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.21.0660
Opphavsrett: © Tidsskriftet 2026 Lastet ned fra tidsskriftet.no 4. juni 2026.