

Statistikk er et viktig hjelpemiddel i vitenskapelige studier, men brukes for ofte som erstatning for klinisk vurdering av resultatene

Statistisk signifikans og klinisk relevans

I desember 2005 skjedde det noe uventet i det europeiske legemiddelverket EMEA (European Medical Agency). EMEAs komité for humane legemidler (CHMP) avviste en søknad fra Novartis om markedsføringstillatelse for legemidlet tegaserod (1). Saken fikk ingen oppmerksomhet i Norge, men den er av stor betydning for fortolkningen av vitenskapelige studier.

Tegaserod er en 5-hydroksytryptamin (5-HT) type 4-reseptoragonist for pasienter med irritable tarm-syndrom med forstoppelse. Ved tidspunktet for CHMPs avslag på søknaden om markedsføringstillatelse forelå det flere store kliniske studier der man hadde sammenliknet tegaserod med placebo. Disse resultatene hadde ført til at midlet allerede var godkjent og i utstrakt bruk i de fleste ikke-europeiske land, deriblant Canada og USA. CHMPs avgjørelse var derfor overraskende både for fagpersoner og myndigheter i landene der det var blitt godkjent. I flere kliniske miljøer i Mellom-Europa var skuffelsen stor, da tegaserod allerede var blitt hyppig brukt på registreringsfritak i påvente av snarlig godkjenning. Begrunnelsen for avslaget var at effekten av legemidlet på plager med irritable tarm ikke var klinisk signifikant forskjellig fra placebo (1). Dette til tross for at de primære effektvariablene (magesmerter og oppblåsthet) viste en statistisk høysignifikant effekt av tegaserod mot placebo ($p < 0,001$) i de aller fleste fremlagte studier (1).

Lave p-verdier betegnes i medisinsk vitenskap ofte som signifikante. Terskelverdien for signifikans er ikke gitt og kan legges vilkårlig av den som utfører en klinisk studie, men sedvanen er å bruke signifikansbegrepet for p-verdier under 0,05. Det er imidlertid viktig å være klar over at man her kun snakker om statistisk signifikans. Dessverre tolkes en statistisk signifikant p-verdi ofte som ensbetydende med en stor, viktig forskjell mellom behandlingsgrupper, for eksempel i favør av et nytt legemiddel mot et gammelt eller mot placebo. Man snakker ofte om «en signifikant effekt» eller om at «tegaserod var signifikant bedre enn placebo ($p < 0,05$)», for å sitere forfatterne av en artikkel om midlet (2). Statistisk signifikans kan imidlertid være veldig forskjellig fra klinisk signifikans, som CHMPs noe uvanlige, men kloke og retningsgivende avgjørelse er et godt eksempel på.

P-verdier angir ikke mer og ikke mindre enn sannsynligheten for at en observert forskjell mellom behandlinger (f.eks. at legemiddel A er bedre enn legemiddel B) skyldes tilfeldighet og ikke reell forskjell (3). Med andre ord tallfester p-verdien sannsynligheten for å «opplage» en effekt som ikke er reell. Hvor stor forskjellen er, sier p-verdien ikke noe om. For å vurdere størrelsen av effekten angir man et punkttestimat (f.eks. gjennomsnittlig forskjell) og et tilhørende konfidensintervall. For eksempel ligger den sanne forskjellen mellom A og B med 95 % sannsynlighet innenfor et 95 % konfidensintervall.

I kliniske studier finnes det imidlertid, noe forenklet sagt, to årsaker til at man får lave, «signifikante», p-verdier. Den mest opplagte er at det er stor forskjell i effekt hos behandlings- og kontrollgruppe. Men p-verdien vil også bli lavere når man øker antallet pasienter i en studie. Selv om forskjellen i effekt mellom behandling og placebo er lik i to studier, kan forskjellen være statistisk signifikant i en studie med mange deltakere, mens den ikke er statistisk signifikant i en mindre studie. Bare ved å øke antall deltakere i kliniske

studier øker altså sjansen for signifikante p-verdier. Statistisk signifikans må derfor ikke tolkes som ensbetydende med klinisk viktig forskjell mellom behandlingsgrupper. Det er et paradoks at store studier i medisinsk vitenskap ofte i seg selv anses som kvalitetsstempel – når man vet at store, klinisk signifikante effekter kun krever mindre studier og at bakgrunnen for å lage en stor studie noen ganger kan være å få lave p-verdier for klinisk mindre betydelige forskjeller mellom to behandlinger.

Statistikk er et viktig hjelpemiddel for å vurdere validiteten av en observert forskjell. Men statistikk må ikke erstatte en grundig medisinsk vurdering av relevansen av de observerte forskjellene i vitenskapelige studier. Medisinskvitenskapelige artikler bør inneholde en klinisk vurdering av de observerte resultatene som setter statistikken i perspektiv. Denne vurderingen krever vitenskapelig skolerte klinikere, noe Ivar Følling påpeker i en interessant kronikk i dette nummer av Tidsskriftet (4). I Tidsskriftet mener vi at den kliniske tolkingen av statistiske forskjeller er en sentral del av ethvert manuskript. Vi ønsker derfor en medisinsk vurdering og tolking i alle artikler som omhandler originale data og i oversiktsartikler. Det er viktig å være klar over at denne medisinske vurderingen alltid vil være subjektiv. Men er den velbegrunnet, gjennomtenkt og tuftet på solid klinisk vitenskapelig erfaring, er den en naturlig og viktig del av kunnskapsbasert medisin.

Vurderingen av klinisk signifikans er altfor ofte direkte koblet til størrelsen på p-verdiene. CHMPs vedtak om ikke å gi markedsføringstillatelse til tegaserod var derfor en viktig og modig avgjørelse fordi den brøt med denne uheldige tradisjonen. Avgjørelsen ble tatt etter at komiteen hadde innhentet råd hos erfarne klinikere som uttalte seg om den kliniske signifikansen (som altså var svak i dette tilfellet) av de statistisk høysignifikante effektene. Dette skiller seg fra den vanlig forekommende overfortolkning av p-verdier som ofte gjøres i medisinsk forskning. Denne utviklingen ønskes velkommen av Tidsskriftet.

Det hører med til historien at tegaserod nylig ble trukket fra markedet i landene der det var godkjent, pga. mistanke om økt risiko for alvorlige kardiovaskulære bivirkninger. Denne risikoen var ukjent da CHMP tok sin avgjørelse i desember 2005.

Michael Bretthauer
michael.bretthauer@rikshospitalet.no

Michael Bretthauer (f. 1969) er dr.med., lege ved Gastromedisinsk avdeling, Rikshospitalet, forsker ved Kreftregisteret og medisinsk redaktør i Tidsskriftet.

Oppgitte interessekonflikter: Bretthauer var ekstern rådgiver for CHMP i deler av behandlingen av markedsføringstillatelsen for tegaserod i 2005.

Litteratur

1. www.emea.europa.eu/humandocs/PDFs/EPAR/zelnorm/H-621-RAR.pdf (20.01.2008).
2. Novick J, Miner P, Krause R et al. A randomized, double-blind, placebo-controlled trial of tegaserod in female patients suffering from irritable bowel syndrome with constipation. *Aliment Pharmacol Ther* 2002; 16: 1877–88.
3. Lindbæk M, Skovlund E. Kontrollerte kliniske forsøk – jakten på sann effekt av behandling. *Tidsskr Nor Lægeforen* 2002; 122: 2631–5.
4. Følling I. Énprosentmedisinen – en kilde til overbehandling. *Tidsskr Nor Lægeforen* 2008; 128: 334–6.