





Faktorer för en **trovärdig** studie

Det finns inget enkelt svar på hur man bedömer om en studie har stor trovärdighet, i stället får man titta på en rad faktorer, till exempel studiens storlek, hur medicinskt möjliga resultaten är och om det finns en kontrollgrupp att jämföra med. **Anna Törner**, statistiker och verkställande direktör i Scandinavian Development Services, skriver här om vad som gör en studie trovärdig.

De regulatoriska myndigheterna kräver i regel minst två pivotala (registrerings-) studier som visar samstämmiga resultat för att godkänna ett nytt läkemedel. Två studier, även om de tillsammans har färre patienter, har oftast högre bevisvärde än en enskild större studie. Att enkelt abdikera från en intelligent utvärdering av studier och snabbt använda etiketter som "statistiskt signifikant", "fas II har låg evidensgrad" och liknande är ett onyanserat sätt att förhålla sig till medicinsk forskning.

Kontrollgrupp ger en referenspunkt

För att utvärdera effekten av en ny behandling bör det finnas en kontrollgrupp i en klinisk studie. Många av de statistiska testen bygger på att jämföra effekten av den aktuella behandlingen med en kontrollgrupp. Kontrollgruppen kan vara placebo (som får verkningslösa sockerpiller eller inte behandlas alls) eller en grupp som får en annan aktiv behandling. För studier där kontrollgruppen också får aktiv behandling är

det viktigt att den jämförande behandlingen är relevant och ges i rätt doser och med rätt behandlingstid.

Att använda sig av en kontrollgrupp som behandlats tidigare (historiska data) kan vara svårt eftersom vården av patienter förändras sig över tid. Då finns det risk för att en bättre överlevnad i undersökningsgruppen jämfört med historiska kontroller speglar förbättringar inom vården i stället för en bättre effekt.

I princip krävs det alltid randomiserade och kontrollerade studier för att få marknadsföringsgodkännande för ett läkemedel av regulatoriska myndigheter. I situationer där det inte finns etablerade behandlingar jämför man ofta med placebo, annars är det naturliga valet att jämföra med den bästa etablerade behandlingen för den aktuella indikationen.

Randomisering är den enskilt viktigaste faktorn

Den enskilt viktigaste faktorn som gör att kliniska studier är mer trovärdiga jämfört med epidemiologiska studier

är randomisering (slumpmässigt urval). Det får alltså inte ligga några systematiska val bakom vilka patienter som får de olika behandlingarna.

Randomisering garanterar däremot inte att två grupper i en klinisk studie blir identiska med avseende på prognostiska faktorer. Genom att randomisera undviker man däremot systematiska fel (bias) som är svårt att hantera. Vi säkrar också att okända, potentiellt viktiga faktorer fördelas slumpvis genom att randomisera.

När man bedömer evidensgraden (bevisvärdet) av olika behandlingar är det i stort sett bara randomiserade studier som klassas med den högsta evidensgraden. Enkelt kan man säga att en grundförutsättning för att statistiska test och p-värden ska ha den tolkning vi vill att de ska ha är att studierna är randomiserade.

Större studier är mer trovärdiga

Stora studier är nästan alltid mer trovärdiga än mindre studier. Tekniskt sett betyder ett p-värde $<0,05$ exakt samma sak i en mindre studie som i



en större studie (att det är mindre än 5 procents chans att observera en så stor effekt av ren slump när det egentligen inte är någon skillnad i effekt). Trots detta är en större studie nästan alltid mer trovärdig.

Man kan vara mer säker på att det faktiskt är en skillnad i behandlingseffekt om resultaten är statistiskt signifikanta. Det beror på att en mindre studie är mer utsatt för andra typer av systematiska fel, till exempel att patientpopulationen är starkt selekterad eller att det föreligger så kallad publikationsbias. Publikationsbias innebär att studier med signifikanta resultat har en större sannolikhet att accepteras för publicering. Urvalet av publicerade studier är därför skevt och risken för att studien visar ett falskt positivt resultat (typ I-fel) är större än det nominella (angivna) p-värdet.

” De regulatoriska myndigheterna kräver i regel minst två pivotala (centrala) studier som visar samstämmiga resultat för att godkänna ett nytt läkemedel.

Är fas III-studier mer trovärdiga än fas II-studier?

Ja och nej. Så enkelt är det oftast inte. En väl designad fas II med ett lite större patientantal där effekterna är tydliga och resultaten fogar sig väl till tidigare kända egenskaper för läkemedlet kan vara nog så trovärdig som en fas III-studie. Men i en fas II-studie har man oftast begränsade kunskaper om effekten vid den aktuella indikationen. Det betyder att resultaten av en fas II-studie bör tolkas utifrån vad som är känt för det aktuella läkemedlet och liknande substanser och hur biologiskt plausibla resultaten är.

Ofta fäster man lite för mycket vikt vid om studien är en fas II eller fas III-studie. Andra egenskaper som är viktigare för att bedöma studiens trovärdighet är om studien är randomiserad och väl kontrollerad, tillräckligt dimensionerad och att resultaten är tydliga. En väl designad fas II-studie kan därför vara mer intressant och övertygande än en fas III-studie av lägre kvalitet med otydliga resultat.

Storlek på effekt och p-värde

Det är kanske uppenbart att en större observerad effekt (kopplat till ett mindre p-värde, till exempel $<0,01$) gör att man oftast kan fästa större tillit till en enskild studie. Om effekten är stor och p-värdet litet är det mindre sannolikt att skillnaden i effekt kan förklaras av andra faktorer eller potentiella systematiska fel.

Detta är också något myndigheter beaktar när de undantagsvis godkänner läkemedel baserat på en pivotal studie i stället för minst två studier som bekräftar varandra. Oftast krävs då ett betydligt lägre p-värde än 0,05.

En stor signifikant effekt i en liten studie är potentiellt intressant. Det är faktiskt så att effekterna behöver vara mycket större i en liten studie för att de ska nå statistisk signifikans. Dock blir effekten mer osäker eftersom studien är mindre och det kan finnas systematiska fel.

Hårda effektmått ökar trovärdigheten

En studie som använt kliniskt relevanta och hårda effektmått är generellt sett mer trovärdig. Hårda effektmått är effekter som både är relevanta för patienten och som kan utvärderas noggrant objektivet (till exempel död).

Trovärdigheten ökar också om man mäter PFS (progression-free-survival, överlevnad utan att sjukdomen förvärras) och dödlighet jämfört med en studie där man sett på respons rate, det vill säga hur bra patienten svarar på behandlingen. Generellt kan man säga att ju mer orienterade effektmåtten är mot resultat hos patienterna, desto mer värdefull är studien för att kunna bedöma nyttan av en ny behandling. Det finns exempel på behandlingar som minskat antal tumörer eller ökat mängden antikroppar mot en specifik infek-

tion utan att patientens överlevnad eller symptom påverkats i mätbar utsträckning.

Blindning förebygger systematiska fel vid utvärdering av effekt

I de fall där det är möjligt att göra blinda kliniska studier, det vill säga maskera vilken behandling patient får, så kan man generellt säga att detta ökar värdet av studien. Då kan man nämligen undvika systematiska fel vid utvärdering av behandlingseffekter. Ibland är det inte möjligt att maskera behandlingen genom att använda placebo, till exempel när man jämför behandlingar med olika administrationsvägar (intravenös och peroralt) eller om den ena behandlingen har uttalade biverkningar. Då är det ännu viktigare att man kan utvärdera effekterna så objektivet som möjligt.

Godkännande av läkemedel baserat på en enda studie

Det finns riktlinjer inom den regulatoriska världen för hur man värderar om det är tillräckligt med en enda studie som grundlag för att godkänna ett nytt läkemedel. Några faktorer som nämns är just de som diskuteras ovan men också intern och extern validitet. Intern validitet handlar om hur säkra man kan vara att de skillnader i effekt som observerats faktiskt kan relateras till läkemedelsbehandlingen, med andra ord ett kvalitetsmått. Extern validitet handlar om generaliserbarhet av studieresultatet. De effekter man observerar ska vara representativa för en mer heterogen patientgrupp utanför ramarna för en klinisk studie, där man ofta gjort ganska snäva urval. Myndigheter är generellt väldigt strikta när det gäller att godkänna läkemedel som är baserat på en pivotal studie och kraven är höga.



ANNA TÖRNER
statistiker, verkställande direktör
Scandinavian Development Services