

Analyse af kliniske forsøgsresultater

Introduktion

Når medicinalvirksomheder foretager kliniske forsøg, indsamles der medicinske oplysninger om de deltagende patienter (men ikke deres identitet) i en computerdatabase sammen med resultaterne af eventuelle målinger, der er udført. Der udføres derefter statistiske analyser for at foretage en formel vurdering af forsøgets udfald.

Analyser af kliniske forsøgsresultater dækker tre interesseområder:

- Demografiske oplysninger og baselineoplysninger
- Virkning
- Sikkerhed

Disse områder beskrives mere detaljeret nedenfor. Det kliniske forsøgs type og design spiller en vigtig rolle i fortolkningen af de statistiske analyser.

Demografiske oplysninger og baselineoplysninger

Hvem deltog i forsøget? Effekten af et lægemiddel kan variere markant mellem forskellige patientgrupper. Derfor er det vigtigt at kende detaljerne for alle patienterne i forsøget, f.eks.:

- Alder
- Køn
- Etnisk oprindelse
- Sværhedsgraden af deres sygdom

Generelt gælder det, at jo bedre et match der er mellem en forsøgsgruppe og en interessepopulation, desto mere relevant vil resultaterne være.

Virkning

Hvor godt virkede forsøgslægemidlet? Denne del af analysen er baseret på foruddefinerede "slutpunkter". Dette er specifikke målinger, der vedrører den relevante sygdom. Slutpunkterne angives i forvejen i forsøgsprotokollen (det dokument, der indeholder en beskrivelse af, hvordan forsøget udføres).

Slutpunkter kan generelt kategoriseres som:

- "Hårde" slutpunkter – de slutpunkter, der har form af numeriske fakta, som er meget vigtige rent klinisk. Det kan f.eks. være, hvor lang tid patienten har haft sygdommen, eller hvor stor en andel af patienterne der er blevet raske efter en infektion.
- "Bløde" slutpunkter – dem, der potentielt påvirkes af målingsprocessen eller har tvivlsom reproducerbarhed. Det kan f.eks. være et spørgeskema om livskvalitet eller en beskrivelse af patientens humør på et bestemt tidspunkt. For at kunne analysere bløde slutpunkter statistisk skal de konverteres til et numerisk format. Denne proces kan være kontroversiel, da den ofte anvender subjektive data og potentielt er sårbar over for uoverensstemmelser.
- "Surrogat"-slutpunkter – dem, der ikke i sig selv er en del af patientens oplevelse af sygdommen, men kan være tæt forbundet med den. Det kan f.eks. være resultatet af laborietest.

Hårde slutpunkter foretrækkes generelt fremfor bløde slutpunkter og surrogat slutpunkter. Bløde slutpunkter og surrogat slutpunkter skal evalueres omhyggeligt i forhold til, hvor godt de repræsenterer den sygdom, der undersøges.

Valget af, hvilke slutpunkter der skal bruges, afhænger i høj grad af, hvilken sygdom der undersøges. Kræft har f.eks. åbenlyse hårde slutpunkter i form af overlevelse, hvorimod en evaluering af depression uundgåeligt involverer blødere slutpunkter. Andre sygdomme, f.eks. diabetes, er knyttet til veletablerede surrogatslutpunkter som f.eks. blodsukkerniveau.

Sikkerhed

Hvilke bivirkninger havde lægemidlet? Når den læge, der udfører et klinisk forsøg, har en konsultation med en patient, bliver patienten spurgt om, hvorvidt denne har oplevet noget usædvanligt. Oplysningerne om disse "uønskede hændelser" indsamles og analyseres senere i forløbet for at give et fingerpeg om en mulig kausal relation med det lægemiddel, der undersøges. Hvis en sådan kausal relation fastslås, bliver den uønskede hændelse til en "bivirkning". Der holdes særligt godt øje med "alvorlige" bivirkninger – som er livstruende eller knyttet til dødsfald, hospitalsindlæggelse eller medfødte misdannelser.

Klinisk forsøgstype

Kliniske forsøg varierer betragteligt i størrelse, varighed og design. Disse faktorer spiller en vigtig rolle i fortolkningen af forsøgsresultaterne.

Det mest informative kliniske forsøgsdesign er "dobbelblindet randomiseret sammenligning", hvor nogle patienter får det nye lægemiddel, mens andre får en anden behandling. Den anden behandling, som af og til kaldes "kontrolbehandlingen", kan enten være:

- Et **placebo** – en inaktiv, "virkningsløs" behandling
- En **aktiv komparator** – generelt en veletableret behandling for den sygdom, der undersøges.

Deltagerne bliver fordelt tilfældigt i undersøgelsesgrupperne.

Forsøget er struktureret, så hverken lægen eller patienten ved, hvem der får hvilken behandling, mens undersøgelsen er i gang. Sådant en forsøgsstruktur kaldes for "dobbeltblindet". Dobbeltblinding reducerer risikoen for bias i resultaterne.

I denne form for forsøg præsenteres resultaterne i forhold til forskellen på den gruppe, der får det nye lægemiddel, og den gruppe, der får kontrolbehandlingen:

- Når sammenligningen er med et placebo, er denne forskel en måling af det nye lægemiddels reelle effekt.
- Når sammenligningen er med en aktiv komparator, giver forskellen indsigt i, hvordan det nye lægemiddel klarer sig i forhold til den nuværende almindelige behandling.

I begge tilfælde er det sandsynligt, at to aspekter af forskellen bliver rapporteret:

- **Størrelse:** Dette rapporteres ofte som den faktiske forskel, der registreres i et bestemt forsøg, sammen med et "95 % konfidensinterval". Det er det interval, hvor vi kan være 95 % sikre på, at dette ville være den reelle forskel for populationen. Selvom du kan registrere en statistisk signifikans, er den muligvis ikke klinisk relevant. Generelt gælder det, at jo større forskellen er, desto mere sandsynligt er det, at den er klinisk relevant (det har større klinisk relevans at øge overlevelsen med et år end at øge den med en dag).
- **Statistisk signifikans:** Da visse personer reagerer bedre på behandlingen end andre, er der altid en risiko for, at den observerede forskel mellem grupper i et klinisk forsøg kan være opstået ved en tilfældighed. Det kan f.eks. være, at alle dem, der endte med at reagere positivt på behandlingen, blev randomiseret i en af grupperne, mens dem med en negativ reaktion blev placeret i den anden. Statistikere kan beregne, hvor sandsynligt det er, at dette scenarie kan være opstået i et bestemt klinisk forsøg, og de udtrykker deres

resultat som "p-værdi".

En p-værdi på 0,05 betyder, at der er 5 % eller 1/20 sandsynlighed for, at forskellen er opstået ved en tilfældighed. Det opfattes traditionelt set som grænsen for, hvornår man accepterer resultater som "statistisk signifikante". Det er vigtigt at være klar over, at når man bruger ordet "signifikant" på denne måde, siger det ikke noget om resultaternes medicinske vigtighed – det giver bare vished om, at resultatet sandsynligvis ikke er fremkommet ved en tilfældighed. I et forsøg, der er stort nok, kan en øget afstand på en meter på en seks minutters gåtur påvises at være statistisk signifikant (dvs. det er usandsynligt, at det er opstået ved en tilfældighed), men dette vil aldrig blive tillagt nogen reel klinisk værdi af en patient med hjertefejl eller dennes læge.

En anden vigtig gruppe af kliniske forsøg, som ofte udføres for at undersøge sikkerheden på lang sigt, har form af et "ikke-blindet" forsøg. I disse er der ingen kontrolgruppe – alle bliver behandlet med det nye lægemiddel, og deres oplevelser registreres. Der kan ikke opstå nogen forskelle mellem grupperne (hverken utilsigtet eller via reelle terapeutiske virkninger), og derfor er der ikke grund til at foretage signifikanstest. Som modvægt til disse mangler har ikke-blindede forsøg ofte et stort antal patienter (op til flere tusinde), der undersøges i lange perioder (i nogle tilfælde i flere år). Disse forsøg gør det derfor lettere at registrere sjældne bivirkninger og de bivirkninger, som først udvikles efter en lang periode.

Resultaterne af disse forsøg præsenteres som enkle tabeller, der viser de forskellige uønskede hændelser og angiver, hvor ofte de blev observeret.