

Statistik i kliniske forsøg: Vigtige begreber

Introduktion

Statistiske metoder leverer en formel redegørelse for årsager til forskellighed i patienters respons på behandling. Den kliniske forsker kan bruge statistik til at udlede rimelige og nøjagtige følgeslutninger ud fra indsamlet information og træffe fornuftige beslutninger om et usikkert forløb. Statistik er nøglen til at undgå fejl og bias i medicinsk forskning. Denne artikel dækker vigtige statistiske begreber og deres anvendelse i kliniske forsøg.

Test af hypotese

En hypotese er en antagelse eller en serie af antagelser, der enten a) fremsætter en foreløbig påstand, som har til formål at lede en videnskabelig undersøgelse, eller b) bekræfter noget som højst sandsynligt set i lyset af veletablerede fakta.

I denne forbindelse er vi interesserede i den hypotese, der fremsætter en påstand om noget – f.eks. at en ny behandling af en sygdom er bedre end den eksisterende standardbehandling. Hvis den nye behandling kaldes "B", og standardbehandlingen kaldes "A", erklærer hypotesen, at "B" er bedre end "A".

Man kunne tro, at forskere vil forsøge at bevise denne hypotese, men det er ikke tilfældet. I stedet beskæftiger de sig indirekte med denne målsætning. I stedet for at prøve at bevise hypotese B antager den videnskabelige metode, at det faktisk er A, der er sand – at der ikke er nogen forskel på standardbehandlingen og den nye behandling. Det kaldes "nulhypotesen". Forskerne forsøger derefter at modbevise A.

Det kaldes også at tilbagevise nulhypotesen. Hvis de kan det – altså bevise, at hypotese A er falsk, og at standardbehandlingen ikke er bedre end den nye behandling – kan det udledes, at B er sand, og at den nye behandling er bedre end standardbehandlingen.

Hvorfor gør man det?

Der findes ikke et simpelt svar på det. Det er den bredt accepterede metode, der har udviklet sig inden for moderne videnskab, men det kan måske hjælpe at bruge en juraanalogi. Nulhypotesen dækker vores nuværende situation eller viden (hvilket i en retssag svarer til, at “den anklagede er uskyldig”), som vi er nødt til at stole på, medmindre vi får en tilstrækkelig mængde beviser for det modsatte. Hvis vi søgte at bevise, at den “alternative hypotese” (dvs. det modsatte af “nulhypotesen”), vil det i bund og grund betyde, at “den anklagede er skyldig”.

En anden og måske lettere måde at formidle denne pointe på er at citere Albert Einstein:

“End ikke alverdens forsøg kan bevise, at jeg har ret; Et enkelt eksperiment kan bevise, at jeg ikke har det.”

Det antyder, at når man prøver at bevise, at nulhypotesen er falsk eller forkert, er det en strengere og mere opnåelig proces end at prøve at bevise, at den alternative hypotese er sand. Bemærk, at dette IKKE tilfredsstillende forklarer, hvorfor videnskaben anvender denne proces, men det kan måske gøre det lettere for os at forstå og acceptere et indviklet begreb.

Type I- og Type II-fejl

I tabellen nedenfor kan du se forskellen på Type I-fejl (falsk positive) og Type II-fejl (falsk negative).

	Nulhypotesen er sand	Nulhypotesen er falsk
Afvis nulhypotesen	Type I-fejl "Falsk positiv"	Korrekt udfald "Sandt positiv"
Nulhypotesen kan ikke afvises	Korrekt udfald "Sandt negativ"	Type II-fejl "Falsk negativ"

Dette er stadig meget forvirrende, så for at udtrykke det på en mere enkel måde følger her et meget drastisk eksempel:

- Type I-fejl kan slå en patient ihjel – forestil dig en undersøgelse, der fejlagtigt har konkluderet, at standardbehandlingen ikke var bedre end den nye behandling og som følge heraf gav nye behandlinger til folk, hvilket fik katastrofale følger. Når der begås Type I-fejl, registreres der fejlagtigt en effekt, der ikke er reel.
- Type II-fejl betyder, at potentielt værdifuld forskning går til spilde. Måske kunne denne forskning have været meget nyttig, men da der ikke foretages flere undersøgelser, lider patienterne ingen overlast. Når der begås Type II-fejl, registreres der ikke en effekt, der er reel.

Det fremgår således, at Type I-fejl er mere alvorlige for patienterne end Type II-fejl.

Signifikansniveau

Signifikansniveauet er sandsynligheden for at begå en Type I-fejl. Dette påvirkes af stikprøvestørrelsen og af testens "statistiske validitet".

Statistisk validitet

"Validiteten" af en statistisk test er sandsynligheden for, at

den vil føre til en korrekt afvisning af en nulhypotese – eller med andre ord testens evne til at registrere en effekt, hvis en sådan effekt rent faktisk findes. En anden måde at beskrive dette på er at sige, at “validiteten” af en test er sandsynligheden for, at der IKKE begås en Type II-fejl.

P-værdier

P-værdier, eller “probability”-værdier, afvejer styrken af bevismaterialet på en skala fra 0 til 1. En lav p-værdi (normalt under 0,05 eller 5 %) angiver, at der er stærke beviser *mod* nulhypotesen, som kan føre til, at du afviser nulhypotesen, mens en høj p-værdi (over 0,05) angiver det modsatte.

Korrelation versus kausalitet

Når resultaterne fra et forsøg analyseres, er det vigtigt at huske, at korrelation ikke er det samme som kausalitet. Korrelation er, når to variabler er knyttet til hinanden på en eller anden måde. Det betyder dog ikke, at den ene er årsag til den anden (der er association mellem begge variabler). Et eksempel på dette er hormonerstatningsbehandling og koronar hjertesygdom, hvor kvinder, der fik hormonerstatningsbehandlinger, havde lavere risiko for at få koronar hjertesygdom. Dette skyldtes dog ikke selve processen omkring hormonerstatningsbehandlingen, men det faktum, at den gruppe mennesker, der får hormonerstatningsbehandlinger, havde tendens til at komme fra en højere socioøkonomisk gruppe, hvor deres kost og træningsvaner var bedre end gennemsnittet.

Kausalitet kan observeres, når en faktor er årsag til et udfald. En kausal faktor er ofte en delvis årsag til et udfald. For at skelne mellem korrelation og kausalitet er det vigtigt at registrere så mange oplysninger som muligt om deltagerne i forsøg. Det er også nødvendigt at være omhyggelig med at bruge den videnskabelige metode i kliniske

forsøgsdesigns og vurdere den mulige bias i forsøget.

Datamanipulering

Datamanipulering opstår, når der selektivt rapporteres forkerte data eller fabrikeres falske resultater. Et eksempel på dette er, når data, der ikke stemmer overens med det forventede resultat, bevidst bliver kasseret for at øge andelen af resultater, som bekræfter den erklærede hypotese. Hvis en forsker fjerner afvigelserne (resultater, der er meget større eller mindre end det nærmeste resultat) fra resultaterne, er det vigtigt at bekræfte, at disse rent faktisk er afvigelser og ikke bare resultater, som skiller sig ud fra de forventede eller ønskede resultater. Et andet eksempel på datamanipulering er, når en dataindsamler genererer et helt randomiseret datasæt ud fra en enkelt indsamlet patientmåling.

Datatransformation

Datatransformation er anvendelsen af en matematisk formel på data, der er indsamlet via et forsøg. Det bruges ofte til at gøre præsentation af data mere overskuelig eller lettere at forstå. Hvis man f.eks. måler bilers brændstoffektivitet, er det naturligt at måle effektiviteten i formatet "kilometer pr. liter". Hvis man skulle vurdere, hvor meget ekstra brændstof der skal bruges til at øge den tilbagelagte afstand, udtrykkes det dog som "liter pr. kilometer". Hvis der anvendes en forkert formel til at sætte den nye præsentation af disse data op i dette tilfælde, påvirker det de samlede resultater af forsøget.

Datafletning

Datafletning betyder, at der kombineres data fra flere undersøgelser for at få en bedre forståelse af situationen. En

af de mest almindelige former for dette er metaanalyse, hvor resultaterne fra flere offentliggjorte forsøg sammenlægges med henblik på aggregering og sammenligning. Under udførelse af en metaanalyse er det vigtigt at kontrollere, at forsøgenes metoder er ens eller sammenlignelige. Der skal tages højde for enhver forskel i forsøgsdesign, så der ikke er nogen underliggende forskellige variabler (konfunderende variabler). Et eksempel på forkert datafletning kan være at aggregere data fra flere forsøg med forskellige musearter som dyretest.

A2-4.33.3-V1.1