

Statistik i kliniska prövningar: nyckelbegrepp

Inledning

Statistiska metoder utgör ett formellt sätt att förklara källorna till variabilitet när det gäller patienternas svar på behandling. Med hjälp av statistik kan den kliniska forskaren dra rimliga och noggranna slutsatser utifrån insamlad information och fatta välgrundade beslut när osäkerhet råder. Statistik är nyckeln till att förhindra felaktigheter och bias vid medicinsk forskning. I den här artikeln beskrivs några av statistikens nyckelbegrepp och hur de tillämpas vid kliniska prövningar.

Hypotesprövning

En hypotes är ett antagande, eller en uppsättning antaganden, som antingen a) hävdar något på provisorisk grund för att vägleda en vetenskaplig undersökning eller b) bekräftar att något är mycket sannolikt mot bakgrund av vedertagna fakta.

I detta sammanhang är vi intresserade av hypoteser som hävdar något – till exempel att en ny behandling för en sjukdom är bättre än den befintliga standardbehandlingen. Om vi kallar den nya behandlingen för B och standardbehandlingen för A säger hypotesen att B är bättre än A.

Du kanske förutsätter att forskarna sätter igång med att bevisa denna hypotes, men så är inte fallet. I stället närmar man sig målet indirekt. I stället för att försöka bevisa hypotes B förutsätter den vetenskapliga metoden i själva verket att A är sann – att det inte finns någon skillnad mellan standardbehandlingen och den nya behandlingen. Detta kallas nollhypotesen. Forskarna försöker sedan att motbevisa A.

Detta kallas även att bevisa att nollhypotesen är falsk. Om de lyckas med detta – att bevisa att hypotes A är falsk och att standardbehandlingen inte är bättre än den nya behandlingen – innebär det att B är sann och att den nya behandlingen är bättre än standardbehandlingen.

Varför gör man så här?

Det finns inget enkelt svar på detta, det är den allmänt accepterade metod som har utvecklats inom modern vetenskap, men det kan vara till hjälp att dra en juridisk parallell. Nollhypotesen innefattar den aktuella situationen eller kunskapen (vid jämförelse med en rättssal innebär det alltså att "den anklagade är oskyldig"), vilken vi måste godta om vi inte har tillräckliga bevis för något annat. Om vi däremot försöker bevisa den "alternativa hypotesen" (det vill säga motsatsen till "nollhypotesen") innebär det i själva verket att "den anklagade är skyldig".

Ett annat kanske enklare sätt att komma fram till detta är att citera Albert Einstein:

"Inga försök i världen kan bevisa att jag har rätt, men ett enda kan bevisa att jag har fel."

Detta verkar betyda att det är ett striktare och mer uppnåbart mål att försöka bevisa att nollhypotesen är falsk eller felaktig än att försöka bevisa att den alternativa hypotesen är sann. Observera att detta egentligen INTE förklarar varför vetenskapen använder den här strategin, men det kanske kan hjälpa oss att lättare förstå och acceptera en komplicerad princip.

Fel av typ I och typ II

Om du tittar på tabellen nedan kan du se vad skillnaden mellan fel av typ I (falskt positiva) och fel av typ II (falskt negativa) är.

	Nollhypotesen är sann	Nollhypotesen är falsk
Förkasta nollhypotesen	Fel av typ I "falskt positivt"	Korrekt utfall "sant positivt"
Nollhypotesen kunde inte förkastas	Korrekt utfall "sant negativt"	Fel av typ II "falskt negativt"

Detta är fortfarande mycket förvirrande, så här följer ett tydligt exempel för att uttrycka det på ett enklare sätt:

- Fel av typ I kan döda en patient – föreställ dig en studie som felaktigt kommer fram till att standardbehandlingen inte är bättre än den nya behandlingen, och människor därför får den nya behandlingen med katastrofala resultat. Fel av typ I innebär att man felaktigt upptäcker en effekt som inte finns.
- Fel av typ II innebär att potentiellt värdefull forskning sker till ingen nytta. Forskningen kanske kunde ha varit verkligt användbar, men eftersom ingen ytterligare studie äger rum sker ingen skada på patienterna. Fel av typ II innebär att man inte upptäcker en effekt som finns.

Det är alltså uppenbart att fel av typ I är allvarligare för patienterna än fel av typ II.

Signifikansnivå

Signifikansnivån är sannolikheten för att begå ett fel av typ I. Detta påverkas av urvalets storlek och av testets "statistiska styrka".

Statistisk styrka

Styrkan hos ett statistiskt test är sannolikheten för att det

korrekt leder till att nollhypotesen förkastas – eller med andra ord testets möjlighet att upptäcka en effekt, om denna effekt faktiskt finns. Ett annat sätt att beskriva detta är att säga att styrkan hos ett test är sannolikheten för att ett fel av typ II INTE inträffar.

P-values

P-värden, eller sannolikhetsvärden, väger evidensens styrka på en skala mellan 0 och 1. Ett lågt p-värde (vanligtvis lägre än 0,05, eller 5 procent) tyder på att det finns starka bevis *mot* nollhypotesen, som skulle kunna leda till att nollhypotesen förkastas, medan ett högt p-värde (högre än 0,05) tyder på motsatsen.

Korrelation kontra kausalitet

Vid analys av resultaten från en prövning är det viktigt att komma ihåg att korrelation inte är samma sak som kausalitet. Korrelation är när två variabler hänger ihop på något sätt, men det betyder inte att den ena är orsaken till den andra (det finns ett samband mellan de två variablerna). Ett exempel på detta är hormonersättningsbehandling och kranskärlssjukdom, där kvinnor som får hormonersättningsbehandling har lägre risk för kranskärlssjukdom. Detta beror dock inte på själva hormonersättningsbehandlingen, utan snarare på att den grupp av människor som får hormonersättningsbehandling tenderar att tillhöra en högre socioekonomisk grupp, med bättre kost- och motionsvanor än genomsnittet.

Kausalitet kan observeras när en faktor orsakar ett resultat. En kausalitetsfaktor är ofta en delorsak till ett resultat. För att skilja mellan korrelation och kausalitet är det viktigt att registrera så mycket information som möjligt om deltagarna i prövningar. Det är också nödvändigt att noggrant tillämpa vetenskaplig metodik vid utformning av

kliniska provningar och att bedöma möjlig bias i provningen.

Datamanipulering

Datamanipulering är när man selektivt rapporterar data felaktigt eller skapar falska resultat. Ett exempel på detta är när data som inte överensstämmer med det förväntade resultatet avsiktligt förkastas för att öka andelen av resultaten som bekräftar den hävdade hypotesen. När en forskare avlägsnar de avvikande värdena (resultat som är väldigt mycket större eller mindre än det närmaste resultatet) från resultaten är det viktigt att verifiera att de är verkliga avvikande värden och inte bara resultat som skiljer sig från de förväntade eller önskade resultaten. Ett annat exempel på datamanipulering är när en datainsamlare slumpmässigt genererar en hel uppsättning data ur en enda insamlad patientmätning.

Datatransformation

Datatransformation är tillämpning av en matematisk formel på vissa data som erhållits genom en provning. Detta används ofta för att presentationen av data ska bli tydligare eller lättare att förstå. Exempelvis vid mätning av bränsleeffektiviteten hos bilar är det naturligt att mäta effektiviteten i form av kilometer per liter. Om man däremot skulle bedöma hur mycket ytterligare bränsle som skulle krävas för att öka det tillryggalagda avståndet skulle det uttryckas som liter per kilometer. Om en felaktig formel användes för att erhålla den nya presentationen av data i det här fallet skulle det påverka provningens totala resultat.

Datasammanslagning

Datasammanslagning innebär att man kombinerar data från flera studier för att få en bättre förståelse av situationen. En av de vanligaste formerna av detta är metaanalys, där resultaten

från flera publicerade prövningar sammanställs för att slås ihop och jämföras. När man utför en metaanalys är det viktigt att noga kontrollera att prövningarnas metodik är densamma eller jämförbar. Man måste ta hänsyn till eventuella skillnader i utformningen, så att det inte finns några underliggande variabler som skiljer sig åt (förväxlingsvariabler). Ett exempel på felaktig datasammanslagning skulle kunna vara hopslagning av data från flera prövningar på olika musarter till ett djurtest.

Bilagor

A2-4.33.3-V1.1