

Statistiek in klinische onderzoeken: belangrijkste concepten

Inleiding

Statistische methoden bieden een formele verklaring voor wat de variabiliteit van responsen van patiënten op een behandeling veroorzaakt. Door statistiek toe te passen kan de klinische onderzoeker redelijke en correcte conclusies trekken uit verzamelde informatie en goede beslissingen nemen indien er sprake is van onzekerheid. Statistiek is de sleutel tot het voorkomen van fouten en bias bij medisch onderzoek. In dit artikel worden enkele van de belangrijkste concepten van statistiek besproken en de toepassing ervan in klinische onderzoeken.

Testen van een hypothese

Een hypothese is een aanname of set van aannames die a) op voorlopige basis iets beweert op grond waarvan wetenschappelijk onderzoek wordt gestuurd of b) iets bevestigt als zeer waarschijnlijk in het licht van vastgestelde feiten.

Voor ons doel zijn we geïnteresseerd in de hypothese die iets beweert – bijvoorbeeld dat een nieuwe behandeling voor een ziekte beter is dan de bestaande standaardbehandeling. Als een nieuwe behandeling 'B' wordt genoemd, en de standaardbehandeling heet 'A', dan luidt de hypothese dat 'B' beter is dan 'A'.

U denkt misschien dat wetenschappers zich zouden inzetten om deze hypothese te bewijzen, maar dat is niet zo. In plaats daarvan wordt dit doel indirect benaderd. In plaats van de B-

hypothese bewijzen gaat de wetenschappelijk methode ervan uit dat A in feite waar is – dat er geen verschil is tussen de standaardzorg en de nieuwe behandeling. Dit wordt de ‘nulhypothese’ genoemd. De wetenschappers proberen vervolgens te bewijzen dat A niet juist is. Dit wordt ook aangeduid als ‘bewijzen dat de nulhypothese onjuist is’. Als hun dit lukt – aantonen dat hypothese A onjuist is, en dat de standaardzorg niet beter is dan de nieuwe behandeling – volgt daaruit dat B juist is, en dat de nieuwe behandeling beter is dan de standaardbehandeling.

Waarom wordt dit gedaan?

Hierop is geen eenvoudig antwoord te geven. Dit is de algemeen geaccepteerde methode die zich in de moderne wetenschap heeft ontwikkeld, maar het kan helpen een juridische analogie te gebruiken. De nulhypothese heeft betrekking op onze huidige situatie of kennis, die we moeten vertrouwen tenzij we voldoende ander bewijs hebben (dus naar analogie in een rechtszaal houdt dit in dat de ‘verdachte onschuldig’ is tot het tegendeel bewezen is). Maar als we de ‘alternatieve hypothese’ trachtten te bewijzen (zoals deze bekend is, tegengesteld aan de ‘nulhypothese’) dan geldt ‘de verdachte is schuldig’.

Een andere, misschien gemakkelijker manier om dit te bereiken is door Albert Einstein te citeren:

“Hoeveel experimenten ook, niets kan mijn gelijk aantonen; één enkel experiment kan bewijzen dat ik ongelijk heb.”

Dit lijkt erop te duiden dat proberen te bewijzen dat de nulhypothese onjuist of fout is, een zorgvuldiger en haalbaarder doel is dan om te proberen te bewijzen dat de alternatieve hypothese juist is. Merk op dat dit NIET goed verklaart waarom de wetenschap deze benadering toepast, maar misschien kan het ons hier helpen een moeilijk concept gemakkelijker te begrijpen en accepteren.

Type I- en type II-fouten

In onderstaande tabel kunt u zien wat het verschil is tussen type I-fouten (foutpositieven) en type II-fouten (foutnegatieven).

	Nulhypothese is juist	Nulhypothese is onjuist
De nulhypothese verwerpen	Type I-fout 'Foutpositief'	Juiste uitkomst 'Echt positief'
De nulhypothese kan niet worden verworpen	Juiste uitkomst 'Echt negatief'	Type II-fout 'Foutnegatief'

Dit is nog steeds erg verwarrend; om dit in eenvoudiger bewoordingen uit te leggen volgt nu een zeer sterk voorbeeld:

- Door type I-fouten zouden patiënten kunnen overlijden – stel u voor dat in een onderzoek onterecht werd gevonden dat de standaardzorg niet beter was dan de nieuwe behandeling en dat de nieuwe behandeling daarom aan mensen zou worden gegeven met rampzalige gevolgen. Type I-fouten detecteren ten onrechte een effect dat er niet is.
- Type II-fouten betekenen dat mogelijk waardevol onderzoek verloren gaat. Dit onderzoek zou misschien echt zinvol zijn geweest, maar aangezien er geen verder onderzoek wordt uitgevoerd, heeft het de patiënten geen schade berokkend. Door type II-fouten te maken wordt er geen effect gedetecteerd dat wel bestaat.

Het is dan duidelijk dat voor patiënten type I-fouten ernstiger zijn dan type II-fouten.

Significantieniveau

Significantieniveau is de waarschijnlijkheid dat er een type I-fout wordt gemaakt. Dit wordt beïnvloed door de omvang van de steekproef en door het 'statistisch onderscheidend vermogen' ('statistical power') van de test.

Statistisch onderscheidend vermogen ('power')

De 'power' van een statistische test is de waarschijnlijkheid dat deze correct leidt naar verwerping van een nulhypothese – of met andere woorden, het vermogen van de test een effect op te sporen als dat effect werkelijk bestaat. Een andere manier om dit te beschrijven is zeggen dat de 'power' van een test de waarschijnlijkheid is om GEEN type II-fout te maken.

P-waarden

P-waarden, of 'waarschijnlijkheidswaarden', wegen de sterkte van het bewijs af op een schaal tussen 0 en 1. Een kleine p-waarde (meestal kleiner dan 0,05 of 5%) verwijst naar sterk bewijs *tegen* de nulhypothese, wat ertoe kan leiden dat u de nulhypothese verwerpt, terwijl een grotere p-waarde (groter dan 0,05) het tegenovergestelde aangeeft.

Correlatie versus oorzakelijkheid

Wanneer de resultaten van een onderzoek worden geanalyseerd, is het belangrijk te bedenken dat correlatie niet hetzelfde is als oorzakelijkheid. Correlatie is wanneer twee variabelen op en of andere manier zijn gekoppeld, maar dit betekent niet dat de een de ander veroorzaakt (er is een verband tussen beide variabelen). Een voorbeeld hiervan is hormoonsuppletie therapie (HST) en coronaire hartziekte (CHZ), waarbij vrouwen die HST innemen minder risico hadden op CHZ. Dit was echter niet het

gevolg van de HST-behandeling zelf, maar meer van het feit dat de groep mensen die HST kregen, behoorden tot een hogere sociaaleconomische groep, met beter dan gemiddelde voeding en lichaamsbeweging.

Oorzakelijkheid kan worden waargenomen wanneer een factor een uitkomst veroorzaakt. Een oorzakelijke factor is vaak een gedeeltelijke oorzaak van een uitkomst. Om correlatie en oorzakelijkheid van elkaar te onderscheiden is het belangrijk zo veel mogelijk informatie te verzamelen over de deelnemers in onderzoeken. Ook moet de wetenschappelijke methodologie in de opzet van klinische onderzoeken zorgvuldig worden toegepast en de mogelijke bias in het onderzoek worden beoordeeld.

Datamanipulatie

Datamanipulatie is de praktijk van selectief onjuiste gegevens rapporteren of onjuiste resultaten verzinnen. Een voorbeeld hiervan zou zijn wanneer gegevens die niet in overeenstemming zijn met het verwachte resultaat, opzettelijk terzijde worden geschoven om het aandeel bevestigende resultaten te vergroten. Wanneer een onderzoeker de uitbijters (of uitschieters, een resultaat dat veel groter of kleiner is dan het volgende dichtstbij gelegen resultaat) weglaat, is het belangrijk dat dit echte uitbijters zijn en niet alleen resultaten die afwijken van de verwachte of gewenste resultaten. Een ander voorbeeld van datamanipulatie is wanneer degene die de gegevens verzamelt, willekeurig een hele reeks gegevens genereert afkomstig van de meting bij één patiënt.

Datatransformatie

Datatransformatie is de toepassing van een wiskundige formule op enkele, tijdens een onderzoek verkregen gegevens. Dit wordt vaak gedaan om de presentatie van de gegevens te verduidelijken of gemakkelijker te begrijpen. Zo is het gebruikelijk om bij beoordeling van het brandstofverbruik van

auto's de efficiëntie te berekenen als 'kilometer per liter'. Maar als u beoordeelt hoeveel extra brandstof er nodig zou zijn om de afgelegde afstand te vergroten, zou het worden uitgedrukt als 'liters per kilometer'. Toepassing in dit geval van een onjuiste formule om de nieuwe presentatie van deze gegevens te verkrijgen, zou de algehele resultaten van het onderzoek beïnvloeden.

Gegevens samenvoegen

Gegevens samenvoegen is het combineren van gegevens afkomstig van meerdere onderzoeken om meer inzicht te krijgen in de situatie. Een van de gebruikelijkste vormen hiervan is de meta-analyse waarbij de resultaten van verscheidene gepubliceerde onderzoeken worden samengevoegd en vergeleken. Bij het uitvoeren van een meta-analyse is het belangrijk zorgvuldig te controleren of de onderzoeksmethodologieën wel identiek of vergelijkbaar zijn. Er moet rekening worden gehouden met eventuele verschillen in de opzet, zodat er geen onderliggende versturende variabelen ('confounders') zijn. Een voorbeeld van onjuist samenvoegen van gegevens kan zijn het bijeenvoegen van gegevens van verscheidene onderzoeken met verschillende soorten muizen tot één dieronderzoek.

A2-4.33.3-V1.1