

Analyse van resultaten van klinische onderzoeken

Inleiding

Wanneer farmaceutische bedrijven klinische onderzoeken uitvoeren, worden de medische gegevens (maar niet hun identiteit) van de deelnemende patiënten verzameld in een computerdatabank samen met de resultaten van eventuele uitgevoerde metingen. Vervolgens worden er statistische analyses uitgevoerd om de uitkomsten van het onderzoek formeel te beoordelen.

Analyses van de resultaten van klinische onderzoeken omvatten drie aandachtsgebieden:

- demografische gegevens en baseline-informatie
- werkzaamheid
- veiligheid

Deze gebieden worden hieronder verder beschreven. Het type en het ontwerp van het klinische onderzoek zijn een belangrijke factor bij de interpretatie van de statistisch analyses.

Demografische gegevens en baseline-informatie

Wie namen er aan het onderzoek deel? De effecten van een geneesmiddel kunnen tussen verschillende groepen patiënten aanzienlijk verschillen. Daarom is het belangrijk op de hoogte te zijn van de kenmerken van alle patiënten in het onderzoek, zoals:

- leeftijd
- geslacht

- etniciteit
- ernst van hun ziekte

In het algemeen geldt dat hoe beter de overeenkomst tussen een onderzoeksgroep en een populatie waarnaar de belangstelling uitgaat, hoe relevanter de bevindingen.

Werkzaamheid

Hoe goed werkte het onderzoeksmiddel? Dit onderdeel van de analyse wordt gebaseerd op vooraf gedefinieerde 'eindpunten'. Dit zijn specifieke metingen gerelateerd aan de ziekte in kwestie. Eindpunten worden vooraf in het onderzoeksprotocol gedefinieerd (het document dat uitvoerig beschrijft hoe het onderzoek wordt uitgevoerd).

Eindpunten kunnen in het algemeen worden ingedeeld als:

- 'harde' eindpunten – eindpunten in de vorm van getallen met een intrinsiek klinisch belang. Bijvoorbeeld hoelang de patiënt overleefde of welk percentage patiënten herstelde van een infectie;
- 'zachte' eindpunten – eindpunten die mogelijk worden beïnvloed door de meting zelf of met twijfelachtige reproduceerbaarheid, bijvoorbeeld een kwaliteit-van-leven-vragenlijst of de beschrijving van de stemming van de patiënt op een bepaald moment. Voor statistische analyse moeten zachte eindpunten worden omgezet in een numeriek formaat. Deze procedure kan controversieel zijn aangezien deze vaak gebaseerd is op subjectieve gegevens en in potentie onderhevig is aan inconsistenties;
- 'surrogaateindpunten' – eindpunten die op zich geen deel uitmaken van de ziekte-ervaring van de patiënt, maar er sterk aan verwant kunnen zijn, bijvoorbeeld de uitslagen van laboratoriumtests.

In het algemeen hebben harde eindpunten de voorkeur boven zachte en surrogaateindpunten. Zachte en surrogaateindpunten

moeten zorgvuldig worden beoordeeld in het licht van hoe goed ze een afspiegeling vormen van de ziekte die wordt onderzocht.

De keuze van de eindpunten hangt sterk af van de aard van de ziekte die wordt bestudeerd. Zo biedt kanker duidelijke harde eindpunten in de vorm van overleving, terwijl een beoordeling van depressie onvermijdelijk zachtere eindpunten moet omvatten. Voor andere ziekten, zoals diabetes, gelden goed vastgesteld surrogaateindpunten als bloedsuikerspiegels.

Veiligheid

Welke bijwerkingen had het geneesmiddel? Wanneer de arts die een klinisch onderzoek uitvoert een patiënt ziet, wordt aan hem/haar gevraagd of de patiënt iets negatiefs heeft ervaren. Informatie over deze 'ongewenste voorvallen' wordt verzameld en later geanalyseerd om meer inzicht te verkrijgen in een mogelijk oorzakelijk verband met het bestudeerde geneesmiddel. Als er een dergelijk oorzakelijk verband wordt vastgesteld, wordt een ongewenst voorval een 'bijwerking'. Bijzondere aandacht wordt besteed aan 'ernstige' bijwerkingen – bijwerkingen die levensbedreigend zijn of gepaard gaan met overlijden, ziekenhuisopname of geboortefwijkingen.

Type klinisch onderzoek

Klinische onderzoeken variëren aanzienlijk in omvang, duur en opzet. Deze factoren spelen een belangrijke rol bij de interpretatie van de onderzoeksresultaten.

De meest informatieve opzet van een klinisch onderzoek is de 'dubbelblinde gerandomiseerde vergelijking', waarbij sommige patiënten het nieuwe geneesmiddel krijgen en andere patiënten een alternatieve behandeling. De alternatieve behandeling, ook wel 'controle' genoemd, kan zijn:

- A **placebo** – een onwerkzame 'nepbehandeling'

- een **werkzaam vergelijkingsmiddel** – meestal een algemeen erkende behandeling voor de ziekte die wordt bestudeerd.

De deelnemers worden willekeurig (door het 'lot') bij elke onderzoeksgroep ingedeeld. Het onderzoek is zodanig opgezet dat terwijl het onderzoek loopt, noch de arts noch de patiënt weet wie welke behandeling krijgt. Een dergelijke onderzoeksopzet wordt 'dubbelblind' genoemd. Dubbelblindering verkleint de kans op vertekening (bias) in de resultaten.

In dergelijke onderzoeken worden de resultaten gepresenteerd als het verschil tussen de groep die het nieuwe geneesmiddel kreeg en de groep met de controlebehandeling.

- Wanneer met een placebo wordt vergeleken, is dit verschil een maat voor het werkelijke effect van het nieuwe geneesmiddel.
- Wanneer wordt vergeleken met een werkzaam vergelijkingsmiddel, geeft het verschil inzicht in hoe het nieuwe geneesmiddel het doet vergeleken met de actuele medische praktijk.

In beide gevallen worden er waarschijnlijk twee aspecten van het verschil gerapporteerd:

- **Omvang:** dit wordt vaak gemeld als het feitelijke verschil gerapporteerd in een specifiek onderzoek samen met een '95%-betrouwbaarheidsinterval'. Dit is de spreiding waarbinnen we voor 95% zeker zijn dat het echte verschil ligt voor de populatie. Hoewel er een statistisch significant verschil kan worden gevonden, hoeft dit niet klinisch relevant te zijn. In het algemeen geldt dat hoe groter dit verschil, hoe groter de kans dat het klinisch relevant is (de overleving met een jaar verlengen heeft een grotere klinische relevantie dan verlenging met een dag).
- **Statistische significantie:** omdat sommige mensen beter op behandeling reageren dan andere, is er altijd een

risico dat het waargenomen verschil tussen groepen in een klinisch onderzoek toeval is. Als bijvoorbeeld alle personen die inherent een goede respons vertonen toevallig bij één groep waren ingedeeld en alle personen met een slechte respons in de andere. Statistici kunnen berekenen hoe groot de kans is dat dit scenario zich heeft voorgedaan in een bepaald klinisch onderzoek en ze drukken hun resultaat uit als een 'p-waarde'.

Een p-waarde van 0,05 betekent dat er een kans is van 5% of van 1 op 20 dat het verschil een toevalsbevinding was. Deze waarde wordt gewoonlijk aangehouden als de drempelwaarde voor accepteren van de resultaten als 'statistisch significant'. Het is belangrijk te bedenken dat het woord 'significant' op deze manier gebruikt niets zegt over het medische belang van de resultaten – het geeft alleen de zekerheid dat het resultaat zeer waarschijnlijk niet toevallig is. Een voorbeeld: van een toename van één meter bij een zes-minutenlooptest kan, in een onderzoek van voldoende omvang, worden aangetoond dat dit statistisch significant is (d.w.z. geen toevalligheid), maar het zou door een patiënt met hartfalen of zijn/haar arts nooit als van enige klinische waarde worden beschouwd.

Een tweede belangrijke groep klinische onderzoeken, vaak uitgevoerd om de veiligheid op lange termijn te onderzoeken, zijn de open-labelonderzoeken. Deze onderzoeken hebben geen controlegroep – iedereen wordt met het nieuwe geneesmiddel behandeld en hun ervaring wordt vastgelegd. Tussen de groepen kunnen geen verschillen optreden (ofwel toevallig of door echte therapeutische effecten) en dus kan er niet worden getoetst op significantie. Afgezet tegen deze tekortkomingen omvatten open-labelonderzoeken vaak grote aantallen patiënten (tot wel ettelijke duizenden) die gedurende lange tijd worden onderzocht (in sommige gevallen enkele jaren). Aan de hand van deze onderzoeken kunnen zeldzame bijwerkingen en bijwerkingen die zich pas na lange tijd ontwikkelen, gemakkelijker worden

opgespoord.

De resultaten van dergelijke onderzoeken worden gepresenteerd als duidelijke tabellen waarin de verschillende bijwerkingen worden vermeld en hoe vaak deze werden waargenomen.