



Steekproefgrootte

PRAKTIJKPROBLEEM

Het is opvallend dat bij gerandomiseerde trials het aantal deelnemende patiënten sterk kan verschillen. Soms worden er 2000 patiënten ingesloten, soms zijn groepen van 25 patiënten kennelijk al groot genoeg om een eventueel verschil aan te tonen. Wat is de verklaring hiervoor?

ACHTERGROND

Bij een te verwachten groot verschil tussen twee therapeutische opties gaan we ervan uit dat dit verschil al met kleine groepen aangetoond kan worden. Indien echter een klein verschil al consequenties voor de behandeling zou kunnen hebben, moeten er veel grotere steekproeven worden gebruikt.

De steekproefgrootte is afhankelijk van een aantal factoren.^{1,2}

- In de eerste plaats moeten we rekening houden met het vinden van een verschil terwijl dat in feite niet bestaat. Dit wordt een 'type-1-fout' genoemd en wordt aangeduid als alfa (α). Onderzoekers hanteren vrijwel altijd een alfa van 5%.
- Daarnaast moeten we ook rekening houden met het missen van een wél bestaand relevant verschil. Dit wordt een 'type-2-fout' genoemd en aangeduid als bèta (β). De power of het onderscheidingsvermogen van een onderzoek is het complement van bèta (dus $1 - \beta$). Meestal houden onderzoekers een bèta op 10% of 20%.
- Ook is het van belang voor welke uitkomst de clinicus het huidige beleid wil veranderen. Op basis van kennis uit de praktijk of uit de literatuur geeft hij aan welk verschil er werkelijk toe doet. Dit noemt men een minimaal klinisch relevant verschil. Bijvoorbeeld: met een bepaald geneesmiddel wordt bij 50% genezing bereikt. De onderzoekers kunnen dan beredeneren dat een nieuw geneesmiddel pas beter is als het 10% meer mensen van hun klacht afhelpt.

UITWERKING

Met een fictief voorbeeld lichten we toe hoe groot de groepen zijn bij verschillende keuzes van de bèta en/of het minimaal klinisch relevante verschil. Bij ernstige consequenties (mortaliteit) is een kleine afname al klinisch relevant, terwijl bij behandelingen waar de consequenties minder groot zijn grotere verschillen pas relevant zijn [voorbeeld]. De gebruikte (eenvoudige) formules staan toegelicht in de geciteerde literatuur.^{1,2}

De serie Praktische epidemiologie laat zien dat er een wetenschappelijke onderbouwing bestaat voor veel handelingen die de huisarts in de dagelijkse praktijk intuïtief uitvoert. Aan de hand van een herkenbaar praktisch gegeven in de praktijk geven we kort aan hoe de wetenschap achter dit praktijkprobleem in elkaar zit. Correspondentie: j.eekhof@nhg.org

Voorbeeld

Onderzoekers willen nagaan of bij fasciitis plantaris de toevoeging van een corticosteroïd aan de lokale injectie met een anestheticum op de langere termijn (bijvoorbeeld 3 maanden) een beter effect heeft (minder klachten of klachtenvrij). Men weet uit de literatuur dat een gunstig effect met een lokaal anestheticum op de lange termijn slechts 30% is en wil een verbetering met 20% kunnen vaststellen. Met een power van 80% (bèta 0,20) en met een significantie van 5% kan worden uitgerekend dat er ten minste 90 personen in beide groepen moeten worden ingesloten. Wil je een power van 90%, dan moeten er 120 personen in beide groepen zitten. Om 30% verbetering te kunnen aantonen, zijn er met een power van 90% 53 mensen nodig; met een power van 80% zijn er 40 personen per groep nodig. Vind je het pas echt relevant als er 40% verbetering wordt aangetoond, dan zijn er met een power van 90% 28 mensen per groep nodig; met een power van 80% 21 personen.

BETEKENIS VOOR DE PRAKTIJK

Hoeveel patiënten in een onderzoek nodig zijn is eenvoudig te berekenen met een formule. Vanuit de praktijk wordt berekend welk gevonden verschil klinisch relevant is. Wanneer ook de kans op een type-1- en type-2-fout is gekozen, is te berekenen hoeveel patiënten minimaal in een onderzoek moeten worden ingesloten om een effect aan te tonen. Er zijn overigens altijd nog meer patiënten nodig omdat je rekening moet houden met uitvallers tijdens het onderzoek. Het is ook begrijpelijk dat men concessies moet doen indien in een pilot van een onderzoek de inclusie van proefpersonen zeer moeizaam blijkt te zijn. In dergelijke gevallen wordt dan met name de power vaak aangepast van 90% naar 80%, zodat de steekproef kleiner kan zijn. De onderzoekers nemen dan wel bewust het risico dat een wel bestaand verschil niet wordt gevonden. ■

LITERATUUR

- 1 Knuistingh Neven A, Jamin RH, Van Geldrop WJ (redactie). Vraagbaak statistiek en methodologie. Een leidraad met praktische oefeningen opgesteld door de Commissie Wetenschappelijk Onderzoek NHG. Utrecht: NHG, 2003.
- 2 Pocock SJ. Clinical trials, a practical approach. Chichester: John Wiley & Sons, 1995.