

# Bewijskracht: gerandomiseerd of observationeel onderzoek?

Olaf M. Dekkers

2020;54(1):1-8 doi:10.35351/gebu.nl.2020.1.1 24-01-2020

Auteurs hebben het '[Formulier belangenverstrengeling](#)' geheel ingevuld en ondertekend met de verklaring "geen belangenverstrengeling".

Gerandomiseerd gecontroleerd onderzoek (RCT) geldt binnen de geneeskunde als de gouden standaard om te bewijzen of veronderstelde effecten van medicamenteuze of niet-medicamenteuze behandelingen causaal zijn. Om financiële, praktische of ethische redenen is het echter niet altijd mogelijk om voor elke relevante behandeling een RCT uit te voeren. Observationeel onderzoek is daarom tegenwoordig een veel gebruikt alternatief om effecten en bijwerkingen van, met name, geneesmiddelen te schatten. Dit is mede ingegeven door de steeds betere beschikbaarheid van grote hoeveelheden data. Vanuit wetenschapsfilosofisch oogpunt is observationeel onderzoek uitsluitend hypothese-genererend te gebruiken. Bij gepresenteerde bevindingen dient men daarom altijd rekening te houden met methodologische beperkingen, zoals 'confounding by indication', het 'healthy cohort effect' en 'immortal time bias'. Steeds zal de afweging moeten worden gemaakt of voor deze zaken voldoende kan worden gecorrigeerd om bevindingen van observationeel onderzoek als valide te beschouwen voor een causaal effect of verband.

## Ge-Bu Plaatsbepaling

- Om een causaal effect van een geneesmiddel te schatten, wordt het gerandomiseerd dubbelblind design gezien als de gouden standaard.
- Gerandomiseerd onderzoek is niet voor alle vraagstellingen aangaande therapie-effecten mogelijk; om die reden is het relevant te bedenken of, en onder welke voorwaarden, observationeel onderzoek een valide alternatief kan zijn.
- Het belangrijkste wetenschappelijke bezwaar tegen het gebruik van observationeel onderzoek voor het schatten van een causaal therapie-effect is dat de te vergelijken groepen vaak verschillend zijn. Voor deze verschillen kan vaak niet volledig worden gecorrigeerd waardoor een vertekend resultaat wordt verkregen.
- Voor sommige vraagstellingen is confounding beperkt en kan observationeel onderzoek een goed alternatief zijn voor gerandomiseerd onderzoek; dit geldt met name voor onderzoek naar zeldzame bijwerkingen van geneesmiddelen. Dit is mede relevant omdat gerandomiseerd onderzoek vaak een te korte follow-up heeft en te weinig mensen includeert om zeldzame bijwerkingen te kunnen detecteren.

## RCT de gouden standaard

Het verkrijgen van wetenschappelijk bewijs voor een causaal verband in geneeskundig onderzoek, bijvoorbeeld of een geneesmiddel werkt, is moeilijk. Desondanks zijn er veel onderzoekers die uitsluitend gerandomiseerd onderzoek erkennen als hard bewijs voor effectiviteit van geneesmiddelen. Ook het Ge-Bu beschouwt gerandomiseerd gecontroleerd onderzoek (RCT) als de gouden standaard voor het bewijs van werkzaamheid van geneesmiddelen of medische hulpmiddelen.<sup>1</sup>

In de klinische praktijk zijn er echter vraagstellingen die niet met gerandomiseerd onderzoek beantwoord kunnen worden.<sup>2</sup> Hier biedt observationeel onderzoek wellicht een uitkomst.

In dit artikel wordt ingegaan op de (on)mogelijkheden om observationele data te gebruiken voor het schatten van therapie-effecten. Het fundamentele verschil tussen gerandomiseerd en observationeel onderzoek zal worden besproken, inclusief de consequenties die dit heeft voor de validiteit van de bevindingen uit observationeel onderzoek.

## Kritiek op RCT

Het is onmogelijk om iedere vraag die voor de klinische praktijk relevant is te beantwoorden met een RCT. Met de onderstaande kritiekpunten wordt de roep om gebruik van observationele data om therapie-effecten te schatten steeds luider.

### Vertraagde vertaling van wetenschap naar praktijk?

Een veel gehoord argument tegen (tijdrovende) RCT's is dat resultaten vertraagd hun weg vinden naar de dagelijkse praktijk. Een voorbeeld is een recente RCT bij patiënten met membraaneuze nefropathie, waarin werd aangetoond dat een behandeling met het geneesmiddel rituximab een hogere kans geeft op het bereiken en onderhouden van remissie ten opzichte van behandeling met het al langer toegepaste cyclosporine.<sup>3</sup> In een bijbehorend commentaar werd gewezen op het feit dat reeds 20 jaar voor publicatie van de RCT, observationeel onderzoek op basis van casuïstiek al aantoonde dat rituximab effectief zou kunnen zijn voor patiënten met membraaneuze nefropathie.<sup>4 5</sup> Hoeveel patiënten zijn jarenlang met de destijds gangbare alkylerende middelen (met een hoog bijwerkingenrisico) behandeld, vroegen de auteurs van het commentaar zich af. Snelle implementatie van kennis over geneesmiddeleffecten zou verhinderd kunnen worden omdat bij het wetenschappelijk beoordelen van werkzaamheid van geneesmiddelen het gerandomiseerd onderzoek tot gouden standaard is verheven. Door de huidige mogelijkheden van observationele ('big') data kan het langdurige en kostbare proces van vergaring van klinische kennis worden versneld. Maar ondanks deze mogelijke versnelling blijft de fundamentele vraag onder welke aannames observationeel onderzoek een valide schatting kan geven van de effecten van interventies.

### Beperkingen door studieopzet

Vaak wordt er gewezen op de beperkingen die de gerandomiseerde studieopzet met zich meebrengt. Niet-representatieve patiënten, te korte follow-up van de deelnemers, geen klinisch relevante uitkomstmaten (surrogaatuitkomsten) en ongelijkwaardige interventies zijn factoren die belangrijke beperkingen kunnen opleveren voor de juiste interpretatie van een RCT. Voor meer informatie hierover zie paragraaf Achtergrondinformatie. Gerandomiseerd onderzoek kan in de opzet deels aan deze beperkingen tegemoetkomen, bijvoorbeeld met representatieve in- en exclusiecriteria, voldoende lange follow-up en klinisch relevante uitkomsten.

### Ethische bezwaren

Ook is gerandomiseerd onderzoek niet voor alle behandelingen mogelijk. Soms zijn er onoverkomelijke ethische bezwaren, bijvoorbeeld bij onderzoek in minderjarige of wilsonbekwame populaties of als de belasting voor de proefpersonen niet in verhouding staat tot de relevantie van de onderzoeksvraag. Bovendien moet bij een RCT voldoende onzekerheid bestaan over de superioriteit van de interventie. Het is namelijk niet ethisch om een deel van de proefpersonen bloot te stellen aan een behandeling waarvan gereede aanwijzingen bestaan dat deze minder goed is dan de behandeling waarmee wordt vergeleken.

### Praktische beperkingen

Een voorbeeld van een praktische belemmering om een optimaal valide RCT uit te voeren is dat bij chirurgische behandelingen blinding vaak niet goed mogelijk is. Een ander voorbeeld van een belemmering is onderzoek bij zeer zeldzame aandoeningen waarbij simpelweg niet voldoende patiënten beschikbaar zijn om voldoende statistische power te verkrijgen.

### Behandelingen al toegepast zonder RCT-bewijs

Vanuit het oogpunt van 'evidence based medicine' is het prima om een gebrek aan gerandomiseerd onderzoek te constateren. Artsen moeten echter behandelkeuzes doen in de afwezigheid van gerandomiseerd onderzoek. Hoewel er, bijvoorbeeld, veel mensen in Nederland schildklierhormoon gebruiken, is er geen bewijs uit gerandomiseerd onderzoek dat antwoord geeft op de vraag wat de optimale TSH-waarde is voor de behandeling van hypothyreoïdie.

## Essentie gerandomiseerd onderzoek

Het centrale kenmerk van een gerandomiseerd onderzoek is, zoals de naam al aangeeft, randomisatie. Randomisatie maakt dat deelnemers aan het onderzoek op basis van toeval tussen twee of meer behandelstrategieën verdeeld worden. Deze randomisatie brengt met zich mee dat deelnemers van elk van de groepen naar verwachting vergelijkbare kenmerken hebben. Anders geformuleerd: in een gerandomiseerd onderzoek zullen de onderzochte groepen dezelfde prognose hebben bij de start van de studie. Deze fundamentele eigenschap van gerandomiseerd onderzoek is de basis voor het standpunt dat een verschil in uitkomst toe te schrijven is aan een verschil in behandeling.

### Geen garantie gelijkheid onderzoeksdeelnemers

Er is overigens geen garantie dat de deelnemers in de verschillende onderzoeksgroepen daadwerkelijk vergelijkbaar zijn in alle kenmerken. Immers, worden de groepen op basis van toeval ingedeeld, dan kunnen toevallige verschillen optreden. De claim 'de groepen zijn gelijk want er is gerandomiseerd' is dus niet correct.<sup>6</sup> Naarmate de vergeleken groepen groter zijn, is de kans op een toevallig verschil in prognose kleiner. In grote studies die vaak een klein effect zoeken, kunnen echter kleine verschillen tussen groepen relevant zijn. Er kan door middel van stratificatie worden gestuurd op een gelijke verdeling van de op papier belangrijkste kenmerken over de onderzoeksgroepen.

### Meer nodig dan alleen randomisatie

Voor de validiteit van gerandomiseerd onderzoek is meer nodig dan enkel randomisatie.<sup>7</sup> Andere factoren die ook bepalend zijn voor de validiteit zijn blinderen van de toewijzing ('concealment of allocation'), blinderen van patiënten, artsen en beoordelaars, complete follow-up en eerlijke en complete rapportage. Voor uitleg van deze begrippen zie paragraaf Achtergrondinformatie.

## Essentie observationeel onderzoek

Observationeel onderzoek beïnvloedt niet de keuze van de behandeling. Deze keuze is reeds gemaakt tijdens de

reguliere interactie tussen arts en patiënt. Dit is in tegenstelling tot gerandomiseerd onderzoek, waarbij als experiment patiënten behandelingen of zelfs placebo kunnen krijgen, die ze in de dagelijkse praktijk niet zonder meer zouden hebben gehad.

### Niet het toeval bepalend, zoals bij RCT

In een observationele studie wordt de keuze voor een behandeling, onafhankelijk van de onderzoekers, bepaald door zorgverleners in de (gangbare) medische praktijk. Hoe en waarom voor een bepaalde behandeling is gekozen, zijn factoren waarop onderzoekers geen invloed hebben. Wel moet met deze factoren rekening worden gehouden bij de analyse en interpretatie van observationele studies. Twee van deze factoren, 'confounding by indication' en 'healthy cohort effect' worden hierna besproken.

#### Confounding by indication

Laten we uitgaan van observationeel onderzoek waarbij twee antibiotica (A en B) worden vergeleken. Het onderzoek is gebaseerd op gegevens van het gebruik van deze antibiotica uit de routinepraktijk. Het is dan de verwachting dat patiënten die antibioticum A voorgeschreven kregen, zullen verschillen van patiënten met antibioticum B voor wat betreft hun klinische kenmerken. Dat er verschillen zijn is niet verbazend. Artsen zullen immers, bewust of onbewust, een reden hebben om de ene groep patiënten antibioticum A te geven en de andere groep antibioticum B. Onder de aanname dat antibioticum A nieuwer en als iets sterker werkzaam op de markt werd gebracht dan B, zullen met A behandelde patiënten gemiddeld mogelijk iets zieker zijn dan de patiënten in groep B. Doordat de keuze voor het antibioticum in dat geval gerelateerd is aan de prognose van de patiënt zal een directe vergelijking van deze twee middelen, A versus B, niet zomaar een valide schatting van het therapie-effect geven. Dit probleem wordt 'confounding by indication' genoemd.

#### Healthy cohort effect

Het is niet altijd zo dat de groep die met het als sterker geldende middel behandeld wordt zieker is. Zou dit wel zo zijn, dan zou observationeel onderzoek standaard een behandel-effect schatten dat de ondergrens van het daadwerkelijk effect weergeeft. Denk aan het vergelijken van chemotherapie voor vergevorderde maligniteiten: nog wel met chemotherapie behandelde patiënten zijn gemiddeld iets fitter dan de niet meer behandelde groep met een vergelijkbaar tumorstadium. Dit wordt het 'healthy cohort effect' genoemd.

### Verschillen patiëntkenmerken corrigeren

In observationeel onderzoek kan het veronderstelde causale effect van een behandeling niet geschat worden door een directe vergelijking omdat de vergeleken groepen verschillend zijn. Er bestaan methoden om voor een aantal verschillen in kenmerken te corrigeren. Mogelijke methoden (matching, regressieanalyse, propensity scores) maken gebruik van gemeten variabelen en proberen statistisch het effect van de verschillen tussen groepen op de uitkomst te corrigeren. Voor meer informatie over deze methoden zie paragraaf Achtergrondinformatie.

#### Confounding adequaat gecorrigeerd?

De vraag of in een specifiek observationeel onderzoek adequaat voor confounding gecorrigeerd is, is primair een inhoudelijk oordeel. Het gaat hierbij in eerste instantie om de vraag: zijn alle belangrijke kenmerken of factoren (zoals bijvoorbeeld roken) die tot confounding kunnen leiden gemeten? Onderzoekers zouden bij voorkeur moeten beschrijven welke variabelen belangrijke confounders zijn, en daarnaast of die factoren adequaat en precies gemeten zijn. Dit vraagt primair om (medisch) inhoudelijke expertise. Hierna volgt de vraag: is een goede statistische benadering gekozen om voor confounding te corrigeren? Deze laatste vraag vraagt om statistische expertise.

De simpele conclusie: dit vergelijkend onderzoek is niet valide want het is observationeel en het kent dus confounding, is te kort door de bocht. Omgekeerd, het observationele onderzoek is valide, want er is gecorrigeerd voor confounding, is eveneens niet zonder meer waar.

### Moelijk te corrigeren confounders

Het ligt voor de hand te denken dat voor confounders die makkelijk te meten zijn (leeftijd, aanwezigheid van diabetes) relatief goed te corrigeren is. Lastiger ligt dit voor variabelen die moeilijker te meten zijn (mate van vaatschade bij diabetes, gedetailleerde informatie over leefstijl). Denk bijvoorbeeld aan 'kwetsbaarheid' ('frailty'), een begrip dat zowel het fysiek als het psychosociaal functioneren van ouderen omvat. Hoewel 'kwetsbaarheid' wordt meegewogen door behandelaars bij behandelvoorstellen ('is deze patiënt fit genoeg voor chemotherapie?'), is dit patiëntkenmerk moeilijk te meten. Daarnaast wordt het niet standaard gerapporteerd. In een dergelijke situatie is dus belangrijke informatie niet beschikbaar en kan hier niet voor worden gecorrigeerd in observationeel onderzoek.

## Gerandomiseerd versus observationeel onderzoek

### Statines gerandomiseerd onderzocht

Stel, een gerandomiseerd onderzoek vergelijkt atorvastatine 40 mg met simvastatine 40 mg. Het eindpunt van de studie is het aantal hartinfarcten. Alle gerandomiseerde patiënten zullen dan starten met atorvastatine 40 mg of simvastatine 40 mg. Afgesproken kan zijn dat indien er spierklachten zijn de dosis aangepast kan worden. De vergelijkende analyse zal echter een startdosering van 40 mg in beide groepen betreffen. Een dergelijk gerandomiseerd onderzoek zal dus antwoord geven op de vraag: geeft starten met 40 mg atorvastatine meer, minder of evenveel hartinfarcten in vergelijking met 40 mg simvastatine. In gerandomiseerd onderzoek is meestal direct duidelijk welke behandelingen worden vergeleken.

### Statines observationeel onderzocht

In observationeel onderzoek is het aan de onderzoeker om de vergelijking te bepalen en vervolgens de data die bij de vraagstelling horen te selecteren. Er kan een observationele studie worden opgezet met als eindpunt het optreden van hartinfarcten, waarin simvastatinegebruikers met atorvastatinegebruikers worden vergeleken. Geïnccludeerde patiënten zullen echter binnen één groep verschillende doseringen gebruiken. Zelfs als de studie laat zien dat één van de twee statines superieur is, dan geeft het nog geen antwoord op de vraag welke dosering optimaal is. Een directe vertaling van dergelijke onderzoeksresultaten naar de klinische praktijk is dus niet vanzelfsprekend. Een voor de hand liggende optie is om in dit observationeel onderzoek twee vaste doseringen te vergelijken. Als er een vertaalslag naar de praktijk moet worden gemaakt, moet in observationeel onderzoek een interventie precies gedefinieerd zijn ('well-defined') in het onderzoeksprotocol.<sup>7</sup>

### Observationeel onderzoek alsof er gerandomiseerd is

Het kan behulpzaam zijn om over observationeel onderzoek na te denken alsof het een gerandomiseerd onderzoek betreft: hoe zou de trial eruitzien voor de behandeling die ik observationeel wil gaan onderzoeken? Een dergelijke denkbeeldige trial (dit wordt in epidemiologische termen een geëmuleerde trial genoemd<sup>7</sup>) maakt direct duidelijk dat er geen plaats is voor niet goed gedefinieerde interventies. Het is immers ondenkbaar om in een gerandomiseerd onderzoek atorvastatine met simvastatine te vergelijken zonder de dosis te specificeren.

### Niet-medicamenteuze interventies extra lastig

Met name voor niet-geneesmiddeleninterventies wordt soms over het hoofd gezien dat de interventie helder gedefinieerd moet zijn. Denk aan een onderzoek dat de effecten op overgewicht van sporten wil onderzoeken. Hier geldt dat een dergelijk onderzoek zowel de interventie-arm (sporten) als de controle-arm (niet sporten) goed moet worden gedefinieerd. Bij enig nadenken wordt snel duidelijk dat er oneindig veel mogelijkheden zijn om de interventie 'sporten' te definiëren. Het kan gaan om minimaal 3x30 minuten per week, in vergelijking met mensen die minder dan 3x30 minuten per week sporten; maar ook om minimaal 1 uur per week versus minder dan 1 uur per week. Enig nadenken maakt twee problemen duidelijk: ten eerste laat deze definitie toe dat mensen die 10 uur sporten per week op eenzelfde wijze als 'sporters' worden geclassificeerd als mensen die

precies 1 uur per week sporten. Ten tweede is de vraag of het gerechtvaardigd is alle sporten op een hoop te gooien. Heeft 1 uur tennissen eenzelfde effect op overgewicht als 1 uur intensief hardlopen? En als sporten een positief effect op BMI blijkt te hebben, welke sport gaan we dan aanbevelen en voor hoe lang? En zelfs als de onderzoekers de interventie en de vergelijking precies definiëren, is de vraag of de dataset genoeg gedetailleerde informatie over sporten zal bevatten om de vraag goed te beantwoorden. In een gerandomiseerd onderzoek naar het effect van sporten zal dit probleem minder spelen, omdat in het protocol precies omschreven zal worden hoe de sportinterventie gedefinieerd is. Ook voor observationeel onderzoek naar effecten van sporten kan het dus zeer behulpzaam zijn om de vraag te stellen: hoe zou het gerandomiseerde onderzoek eruitzien?

### **‘Real world’ = geen garantie validiteit**

Een veelgehoord argument voor het gebruik van observationeel onderzoek is dat dit een middel is om ‘real-world-effecten’ te schatten. Dit idee berust op het feit dat in gerandomiseerd onderzoek vaak niet-representatieve patiënten zijn geïncludeerd en de behandelsetting ook afwijkt van de dagelijkse praktijk.<sup>8</sup> In de dagelijkse praktijk, bijvoorbeeld, belt er geen onderzoeksverpleegkundige dagelijks met de vraag of de studiemedicatie is ingenomen. Hoewel deze punten relevant zijn, is het feit dat observationeel onderzoek de dagelijkse praktijk dichter benadert, geen enkel argument voor de validiteit van observationeel onderzoek. Kort gezegd: wanneer observationeel onderzoek weliswaar de dagelijkse praktijk reflecteert, maar methodologisch niet valide is, dan zijn de studieresultaten van generlei waarde.

### **RCT referentie voor validiteit observationele studie?**

Het meeste observationele onderzoek wordt niet uitgevoerd met als doel om te zien of de resultaten overeenkomen met een al bekende trial. Wanneer er echter een goed uitgevoerd gerandomiseerd onderzoek bestaat om uitkomsten van observationeel onderzoek mee te vergelijken, maakt dat de beoordeling van de validiteit van het observationeel onderzoek eenvoudiger, dan wanneer dit niet het geval is. Het zal juist dan van waarde kunnen zijn wanneer er geen data uit gerandomiseerd onderzoek beschikbaar zijn.

Er is veel onderzoek gedaan naar de overeenstemming van effecten uit gerandomiseerd onderzoek en observationeel onderzoek. Daarbij zijn resultaten van RCT's vergeleken met grote observationele onderzoeken met dezelfde vraagstelling.

### **Resultaten vergelijkbaar**

Case-studies laten zien dat in een aantal gevallen resultaten vergelijkbaar zijn. Een voorbeeld is een studie waarbij de resultaten van gerandomiseerd onderzoek en observationeel onderzoek voor de behandeling met anti-retrovirale therapie werd vergeleken.<sup>9</sup>

### **Resultaten niet vergelijkbaar**

Aan de andere kant zijn er voorbeelden waar observationeel onderzoek heeft geleid tot verkeerde conclusies, zoals bij chronische hormoonsuppletie. In 2002 werd een RCT gepubliceerd waaruit duidelijk werd dat chronische suppletie met een combinatie van oestrogenen en progestagenen het risico verhoogde op coronairlijden, borstkanker, trombo-embolieën en beroertes bij gezonde postmenopauzale vrouwen.<sup>10</sup> Deze behandeling werd eerder verondersteld een preventief effect te hebben op cardiovasculaire aandoeningen op basis van observationeel onderzoek.<sup>11, 12, 13</sup> Een nieuwe analyse van de observationele data volgens de hoogste epidemiologische standaard liet zien dat de discrepantie terug te voeren was op het incorrect analyseren van de observationele data en residuele confounding vanwege een disbalans in socio-economische patiëntkenmerken.<sup>14, 15, 16</sup>

### **Verschillen resultaten te verklaren**

Het hoeft overigens niet zo te zijn dat de resultaten uit observationeel onderzoek en gerandomiseerd onderzoek voor dezelfde interventie-uitkomst identiek zouden moeten zijn. Het zou zo kunnen zijn dat een verschil op reële effectmodificatie (zie <http://www.minerva-ebm.be/Results/Glossary/1379>) berust, bijvoorbeeld doordat er een iets andere patiëntenpopulatie is geïncludeerd. Bijvoorbeeld: het zou kunnen zijn dat de effecten van



antistollingsbehandeling in observationele studies een ander effect laten zien dan effecten in een RCT, omdat de studiepopulatie meer comorbiditeit heeft dan de populatie in een RCT.

## Observationeel onderzoek geschikt voor bijwerkingen?

Er is reden te denken dat bijwerkingen vaak valide geschat kunnen worden met observationeel onderzoek. Zoals hierboven uiteengezet is het grote probleem van het schatten van therapie-effecten met observationeel onderzoek 'confounding by indication'. Denk opnieuw aan het observationeel onderzoek waarbij twee antibiotica (A en B) worden vergeleken. Onder de aanname dat antibioticum A nieuwer is en iets sterker werkt dan B, zijn verschillen tussen de twee groepen te verwachten, omdat A aan iets ziekere patiënten zal worden voorgeschreven. Maar nu wil een onderzoeker dezelfde database gebruiken om een vergelijking tussen A en B te maken, maar nu voor de kans op een allergische reactie. De kans op een allergische reactie is niet afhankelijk van de werkingssterkte en er zijn geen patiëntkenmerken die een (nog niet bekende) allergische reactie kunnen voorspellen. Er is dus geen verwachte confounding voor de vergelijking tussen deze twee middelen wanneer het om een allergische reactie gaat. Meer algemeen kan worden gesteld dat voor onverwachte en onbedoelde effecten de confounding zeer beperkt zal zijn.<sup>17</sup> Dit is door empirische vergelijking van schattingen uit observationeel onderzoek en gerandomiseerd onderzoek voor dezelfde bijwerking bevestigd.<sup>18</sup> Ook bij onderzoek naar bijwerkingen kan echter een geldende communis opinio van meer veiligheid een bron zijn voor 'confounding by indication'.

## Achtergrondinformatie gerandomiseerd onderzoek

### Validiteit van gerandomiseerd onderzoek

#### Blinderen van de toewijzing

Inadequaat blinderen van de toewijzing ('concealment of allocation') van een te onderzoeken behandeling kan leiden tot onvergelykbaarheid tussen onderzoeksgroepen en daarmee de validiteit van de bevindingen van een RCT beperken. Een gerandomiseerde studie kan bijvoorbeeld zo zijn opgezet dat patiënten, na te zijn gerandomiseerd, weten welke behandeling zij gaan krijgen. Wanneer na de randomisatie, maar voor start behandeling, sommige deelnemers toch besluiten niet meer mee te doen, kan vertekening van de resultaten (bias) ontstaan. Stel een ongeblindeerde gerandomiseerde studie waarin medicamenteuze pijnstilling met fysiotherapie wordt vergeleken. Wanneer nu in de fysiotherapiegroep de patiënten met relatief weinig pijn besluiten toch maar niet mee te doen ontstaat onvergelykbaarheid van groepen. Blindering van de toewijzing is in een placebogecontroleerde, of dubbel-dummy gecontroleerde trial, gegarandeerd door de opzet van het onderzoek.

#### Blinderen van patiënten, artsen en beoordelaars

Wanneer tijdens de follow-up van een studie bekend is welke behandeling een patiënt heeft, dan kan de zorg gaan verschillen tussen de groepen. Dit wordt 'performance bias' genoemd.<sup>7</sup> Wanneer de beoordelaars van de eindpunten niet geblindeerd zijn kan dit ook bias geven ('detection bias').<sup>7</sup> In een niet-geblindeerde RCT naar het effect van rosiglitazon op hart- en vaatziekten bleek dat de beoordeling van eindpunten vaak in het voordeel van rosiglitazon uitviel.<sup>19</sup>

Blinderen is niet altijd mogelijk, bijvoorbeeld in gerandomiseerd onderzoek naar een vergelijking tussen medicamenteuze pijnstilling en fysiotherapie.

#### Complete en vergelijkbare follow-up

Wanneer de follow-up tussen de groepen verschillend is, dan kan dit bias geven, vooral wanneer uitval uit de

studie gerelateerd is aan de prognose van de ziekte. Een voorbeeld is een gerandomiseerd onderzoek naar het effect op overleving bij behandeling met chemotherapie. Wanneer de mensen bij wie het middel niet lijkt te werken, en die ook nog bijwerkingen hebben uit de studie verdwijnen, dan zal dit tot een overschatting van het effect van chemotherapie leiden. Deze vorm van bias wordt 'attrition-bias' of selectiebias genoemd.<sup>7</sup>

### **Eerlijke en complete rapportage**

Om bias tegen te gaan, dient een gerandomiseerd onderzoek uitgevoerd en gerapporteerd te worden volgens protocol. Wanneer immers achteraf het primaire eindpunt wordt gekozen of aangepast, hebben de onderzoekers meer mogelijkheden om een positief resultaat te verkrijgen. Dit is een reden waarom veel tijdschriften eisen dat alleen gerandomiseerde onderzoeken worden geaccepteerd waarvan het onderzoeksprotocol is gepubliceerd in één van de daarvoor beschikbare registers voordat met het onderzoek gestart werd. Ondanks de eis tot vooraf publiceren van het protocol is veranderen van- of selectief rapporteren van uitkomsten nog steeds niet uitgebannen.<sup>20</sup>

Waar op internationaal niveau veel wet- en regelgeving wordt ingevoerd om de transparantie van RCT's te verbeteren, is bij observationeel onderzoek de rapportage nog minder goed controleerbaar. Er is geen wettelijke verplichting voor het prospectief publiekelijk registreren van een studie.

### **Beperkingen van gerandomiseerd onderzoek**

#### **Ingesloten patiënten niet representatief**

Een veel gehoord kritiekpunt van gerandomiseerd onderzoek is de beperkte generaliseerbaarheid.<sup>8</sup> Gerandomiseerd onderzoek wordt vaak uitgevoerd met relatief gezonde patiënten. Voor jonge kinderen en ouderen geldt dat ze vaak niet deel mogen nemen, of dat ze weliswaar deel mogen nemen, maar dat ze maar een heel klein deel van de onderzoekspopulatie vormen. Studies die een vergelijking maken tussen patiënten die daadwerkelijk deelnemen aan een gerandomiseerd onderzoek, en patiënten die geïncludeerd hadden kunnen worden, maar toch niet deelnamen, laten zien dat de daadwerkelijk geïncludeerde patiënten een veel betere prognose hebben.<sup>21</sup> De in RCT's ingesloten patiënten zijn dus lang niet altijd representatief voor de patiënten uit de dagelijkse praktijk.<sup>1</sup> Dit versterkt de wens om onderzoeksdata te verzamelen waarin patiënten uit de dagelijkse praktijk worden geïncludeerd.

#### **Beperkte follow-up**

Een tweede punt van zorg is dat gerandomiseerd onderzoek vaak een beperkte follow-up heeft. Dit heeft twee nadelen. Ten eerste kan de langetermijnveiligheid van geneesmiddelen daarmee niet worden onderzocht. Wanneer bijvoorbeeld de vraag is of een geneesmiddel carcinogeen is, dan zijn onderzoeken met een looptijd van ongeveer een jaar weinigzeggend. Ook langetermijneffecten van geneesmiddelen (denk aan immuniteit op de lange termijn na vaccinaties) kunnen bij een korte looptijd van gerandomiseerd onderzoek niet worden onderzocht.

#### **Surrogaateindpunten**

In dezelfde lijn ligt het kritiekpunt dat in gerandomiseerd onderzoek met enige regelmaat surrogaateindpunten worden onderzocht. Het zijn eindpunten die voor de patiënt niet zonder meer klachten geven.<sup>1</sup> Het probleem is dat een effect op surrogaateindpunten (bloeddruk, glucoseregulatie) niet altijd goed correleert met een effect op klinisch relevante eindpunten (hartinfarcten, nierfalen).<sup>21</sup> Een bekend voorbeeld is het middel rosiglitazon. Er werd weliswaar een verbetering van de glucoseregulatie aangetoond, maar uit een meta-analyse van RCT's bleek dat rosiglitazon zelfs slechtere resultaten gaf op cardiovasculaire eindpunten vergeleken met glucoseverlagende behandelingen zonder rosiglitazon.<sup>22</sup>

#### **Lastig te randomiseren interventies**

Sommige interventies lenen zich niet goed voor een gerandomiseerd design. Denk aan de vergelijking van twee chirurgische technieken; het kan zo zijn dat niet alle chirurgen die deelnemen aan het onderzoek beide



technieken even goed beheersen. In een standaard gerandomiseerd onderzoek waarbij patiënten worden gerandomiseerd tussen twee chirurgische technieken zou dan een aantal patiënten suboptimaal geopereerd worden. Dat is ten eerste niet ethisch en geeft ten tweede geen goede schatting van de effectiviteit van beide chirurgische technieken.

Voor onderzoek naar voedingsinterventies wordt al snel duidelijk dat het aantal mogelijk uit te denken gerandomiseerd onderzoek oneindig is. Stel onderzoekers willen gerandomiseerd onderzoeken of broccoli gezond is. Hoeveel broccoli per dag? Hoe lang duurt de interventie? Moet de broccoli worden gekookt of gestoomd? Wat krijgt de vergelijkingsgroep? Het zal snel duidelijk zijn dat gerandomiseerd onderzoek naar voeding snel tegen zijn eigen beperkingen gaat aanlopen.

## Achtergrondinformatie observationeel onderzoek

### Validiteit effectschatting

Er is de afgelopen decennia veel methodologisch onderzoek gedaan naar voorwaarden voor de validiteit van effectschattingen in observationeel onderzoek. Ten dele zijn deze methodologische voorwaarden zeer technisch. Sommige methodologische problemen zijn niet fundamenteel anders dan voor gerandomiseerd onderzoek (misclassificatie, eindpuntbeoordeling, missende data), andere problemen zijn specifiek voor observationeel onderzoek.

### Immortal time bias

Een probleem dat specifiek is voor observationeel onderzoek, is 'immortal time bias'. Deze vorm van bias treedt op wanneer de definitie van behandeling gebaseerd is op gegevens van ná de start van de follow-up.<sup>23</sup>

#### Immortal time van Oscarwinnaars

Ter illustratie, onderzoek liet zien dat Oscarwinnaars gemiddeld ouder werden dan mensen die nooit een Oscar hadden gewonnen.<sup>24</sup> Dit komt omdat voor het winnen van een Oscar doorgaans een zekere leeftijd bereikt dient te worden (immortal time). Door elke Oscarwinnaar te matchen met een controlepersoon die even oud was als de Oscarwinnaar op het moment van het winnen van de Oscar, kon voor deze vertekening worden gecorrigeerd en bleek de conclusie niet te kloppen.

#### Door immortal time bias niet-plausibel resultaat

Een voorbeeld van een studie met door 'immortal time bias' niet-plausibele resultaten, is een vergelijking van het mortaliteitsrisico tussen patiënten met en patiënten zonder tweede tumor na bestraling van het KNO-gebied. De resultaten van de studie lieten zien dat patiënten met een tweede tumor langer leefden dan patiënten zonder tweede tumor.<sup>25</sup> De reden ligt voor de hand: patiënten moesten een bepaalde tijd hebben overleefd alvorens een tweede tumor te hebben ontwikkeld. De tijd tussen start van de follow-up en het optreden van de tweede tumor wordt 'immortal time' genoemd.

#### Door immortal time bias overschatting resultaten

In observationeel onderzoek waarbij verschillende interventies worden onderzocht, kan 'immortal time bias' optreden. Denk aan een observationeel onderzoek waarbij na operatie voor een maligniteit de overleving van patiënten met en zonder chemotherapie wordt vergeleken. De chemotherapie zal niet de eerste dag na operatie worden gegeven, maar vaak pas worden gestart na een paar weken, als de patiënt redelijk van de operatie is hersteld. De patiënten in de chemotherapiegroep hebben dus per definitie de eerste postoperatieve weken overleefd, en werden daarnaast op het moment van de chemotherapie door de arts fit genoeg bevonden voor de chemotherapie. Omdat de patiënten in de groep met chemotherapie dus een deel van de follow-up 'immortal' is, zal het effect van chemotherapie worden overschat.

Een voorbeeld van een therapie-effect dat door deze vorm van bias overschat werd, is een farmaco-

epidemiologische studie naar het effect van beta-blokkers in het jaar na een hartinfarct. Ze (b)leken de mortaliteit te verlagen.<sup>26</sup>

### **Geen rol immortal time bias in RCT**

In gerandomiseerd onderzoek speelt 'immortal time bias' geen rol. De reden is dat in dergelijk onderzoek de intentie tot behandeling wordt vergeleken, ook wanneer die behandeling pas een paar weken na randomisatie start. Zo zal in het bovengenoemde voorbeeld van chemotherapie versus geen chemotherapie in een gerandomiseerd onderzoek, in beide groepen de follow-up na operatie beginnen. Stel nu, een patiënt in de chemotherapiegroep overlijdt voor start van de chemotherapie. In een gerandomiseerd onderzoek zal deze patiënt geanalyseerd worden in de chemotherapie groep, maar in observationeel onderzoek in de niet-chemotherapie groep. Dit heeft te maken met het gegeven dat in observationeel onderzoek de intentie tot behandeling veelal niet geregistreerd is, en dus een intention-to-treat-analyse heel veel lastiger is.

### **Methoden om verschillen in kenmerken te verdisconteren**

#### **Matching**

Bij matching wordt voor elke patiënt één (of meerdere) controlepatiënt(en) gezocht met vergelijkbare kenmerken. Er kan bijvoorbeeld gematcht worden voor geslacht, leeftijd en de aanwezigheid van hypertensie. Voor variabelen die niet zijn gemeten (bijvoorbeeld gegevens over kwetsbaarheid) kan niet worden gematcht.

#### **Regressieanalyse**

Bij regressieanalyse wordt statistisch de relatie tussen een behandeling en een uitkomst gemodelleerd, waarbij het statistische model verschillen tussen groepen verdisconteert. Er zijn verschillende vormen van regressie, zoals logistische regressie, Cox-regressie of Poisson-regressie. Er zit geen ranking in de verschillende regressietechnieken, wel kan een bepaalde analyse passender zijn dan een ander, gegeven de data. Voor variabelen die niet zijn gemeten of niet objectief te meten zijn, kan in een regressieanalyse niet worden gecorrigeerd.

#### **Propensity scores**

Een observationeel onderzoek gebaseerd op een 'propensity score' maakt gebruik van twee regressiestappen. Eerst wordt op basis van alle relevante factoren de kans op behandeling geschat. Dit levert een 'propensity score' op voor iedere patiënt. Deze score wordt vervolgens in een regressieanalyse gebruikt om verschillen tussen groepen te verdisconteren (hetzij door matching op 'propensity score', of door corrigeren). Hoewel de 'propensity score' door sommige auteurs wordt gezien als een alternatief voor gerandomiseerd onderzoek, geldt ook hier de belangrijke voorwaarde: voor variabelen die niet zijn gemeten, kan niet worden gecorrigeerd door een 'propensity score'.

#### **Self-controlled designs**

In self-controlled designs worden binnen één patiënt periodes met en periodes zonder het te onderzoeken geneesmiddel onderzocht. Dit heeft het voordeel dat voor een aantal confoundingfactoren door de studieopzet al is gecontroleerd. De patiënt wordt immers met zichzelf vergeleken.<sup>27</sup> Een voorbeeld is een studie waarin bij clopidogrelgebruikers het risico op een hartinfarct werd onderzocht in periodes dat de patiënten wel en in periodes dat patiënten geen protonpompremmers gebruikten. Protonpompremmers verhoogden de kans op een hartinfarct.<sup>28</sup>

### **Mendeliaanse randomisatie**

Om in een observationeel onderzoek effecten van geneesmiddelen te schatten kan genetische informatie gebruikt worden.<sup>29</sup> Voorwaarde is dat het biologisch mechanisme van het geneesmiddel bekend is, en dat genetische varianten bekend zijn die dezelfde biologische effecten hebben. Deze genetische varianten kunnen vervolgens in de analyse gerelateerd worden aan klinische uitkomsten.

#### **Randomiseren o.b.v. genetische informatie, een voorbeeld**

Er zijn genetische varianten bekend die het LDL-cholesterol verlagen. Hoe kunnen deze varianten worden gebruikt om positieve geneesmiddeleffecten te schatten op hart- en vaatziekten door (zonder geneesmiddel) het LDL-cholesterol te verlagen? Statines hebben deze effecten ook. Zoals eerder uiteengezet, is het voornaamste doel van randomiseren in RCT's dat groepen een vergelijkbare verdeling van risicofactoren hebben op baseline, en dat eventuele verschillen op toeval berusten. Zogenoemde mendeliaanse-randomisatiestudies werken op vergelijkbare wijze. Een genetische variant die het LDL-cholesterol verlaagt, is onafhankelijk van genen die andere risicofactoren voor hartziekten beïnvloeden. Vergelijken we in een degelijke (observationale) studie mensen met deze genetische variant met mensen zonder deze variant, dan is te verwachten dat de andere mogelijk versturende risicofactoren gelijk verdeeld zijn en confounding tot een minimum beperkt kan worden. Onder genoemde aannames kan zo in een observationele studie een genetische variant indirect het effect van een geneesmiddel schatten.

### **Effect genetische variant verschilt van geneesmiddeleffect**

Mendeliaanse-randomisatiestudies schatten om verschillende redenen een iets ander effect dan geneesmiddelen. Ten eerste gaat het bij dergelijke studies om een levenslang effect van een genetische variant op een biologische variabele, terwijl geneesmiddelen later in het leven worden gegeven. Ten tweede zijn de biologische effecten van de genetische variant vaak geringer dan het effect van het geneesmiddel. En ten derde kent de genetische variant geen therapieontrouw.

### **Mendeliaanse-randomisatiestudies voor bijwerkingen**

Deze genetische benadering is ook interessant voor mogelijke bijwerkingen van geneesmiddelen. Veel trials hebben te weinig patiënten om met zekerheid iets over bijwerkingen te kunnen zeggen. De follow-up is relatief kort en sommige langetermijnbijwerkingen (kanker bijvoorbeeld) kunnen vaak niet worden onderzocht. Als genetische varianten onderzocht zouden kunnen worden, kunnen de bijwerkingen worden geschat, voor zover deze verlopen via de biologische variabele waarop de genetische variant van invloed is.

## **Literatuurreferenties**

1. Van Deventer KR, Janssens JEM. Klinische relevantie van onderzoeksuitkomsten. [Gebu. 2019;53\(1\):1-11.](#)
2. RVS Zonder context geen bewijs. Over de illusie van evidence-based practice in de zorg. Publicatie 17-05. RVS, Den Haag 2017. ISBN: 987-90-5732-2679.
3. Fervenza FC, Appel GB, Barbour SJ, Rovin BH, Lafayette RA, Aslam N, et al. Rituximab or Cyclosporine in the Treatment of Membranous Nephropathy. *The New England journal of medicine.* 2019;381(1):36-46.
4. Ruggenti P, Remuzzi G. A First Step toward a New Approach to Treating Membranous Nephropathy. *N Engl J Med.* 2019; 381(1):86-88.
5. Remuzzi G, Chiurciu C, Abbate M, Brusegan V, Bontempelli M, Reggenenti P. Rituximab for idiopathic membranous nephropathy. *Lancet.* 2002; 360(9337):923-924.
6. Senn S. Seven myths of randomisation in clinical trials. *Statistics in medicine.* 2013;32(9):1439-50.
7. Mansournia MA, Higgins JP, Sterne JA, Hernan MA. Biases in Randomized Trials: A Conversation Between Trialists and Epidemiologists. *Epidemiology (Cambridge, Mass).* 2017;28(1):54-9.
8. Dekkers OM, von Elm E, Algra A, Romijn JA, Vandenbroucke JP. How to assess the external validity of therapeutic trials: a conceptual approach. *International journal of epidemiology.* 2010;39(1):89-94.
9. Lodi S, Phillips A, Lundgren J, Logan R, Sharma S, Cole SR, et al. Effect Estimates in Randomized Trials and Observational Studies: Comparing Apples With Apples. *American journal of epidemiology.* 2019;188(8):1569-77.
10. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. *JAMA.* 2002; Jul 17;288(3):321-33.

11. Mendelsohn ME, Karas RH. The Protective Effects of Estrogen on the Cardiovascular System. *N Engl J Med*. 1999; 340(23):1801-1811.
12. Stampfer MJ, Colditz GA, Willett WC, Manson JE, Rosner B, Speizer FE, et al. Postmenopausal estrogen therapy and cardiovascular disease. Ten-year follow-up from the Nurses' Health Study. *N Engl J Med* 1991;325(11):756-62.
13. Grady D, Rubin SM, Petitti DB, Fox CS, Black D, Ettinger B, et al. Hormone therapy to prevent disease and prolong life in postmenopausal women. *Ann Intern Med* 1992;117:1016-37.
14. Hernán MA, Alonso A, Logan R, Grodstein F, Michels KB, Willett WC, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology* 2008;19(6):766-79.
15. Vandenbroucke JP. The HRT controversy: observational studies and RCTs fall in line. *Lancet*. 2009;373(9671):1233-5.
16. Rosén M, Axelsson S, Lindblom J. Observational studies versus RCTs: what about socioeconomic factors? *Lancet*. 2009;373(9680):2026.
17. Vandenbroucke JP. When are observational studies as credible as randomised trials? *Lancet* (London, England). 2004;363(9422):1728-31.
18. Papanikolaou PN, Christidi GD, Ioannidis JP. Comparison of evidence on harms of medical interventions in randomized and nonrandomized studies. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*. 2006;174(5):635-41.
19. Psaty BM, Prentice RL. Minimizing bias in randomized trials: the importance of blinding. *Jama*. 2010;304(7):793-4.
20. Dwan K, Gamble C, Williamson PR, Kirkham JJ. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. *PloS one*. 2013;8(7):e66844.
21. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet* (London, England). 2005;365(9453):82-93.
22. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *The New England journal of medicine*. 2007;356(24):2457-71.
23. Suissa S. Immortal time bias in pharmaco-epidemiology. *American journal of epidemiology*. 2008;167(4):492-9.
24. Via: <https://www.ncbi.nlm.nih.gov/pubmed/16954361>
25. Rennemo E, Zatterstrom U, Boysen M. Impact of second primary tumors on survival in head and neck cancer: an analysis of 2,063 cases. *The Laryngoscope*. 2008;118(8):1350-6.
26. Rochon PA, Tu JV, Anderson GM, Gurwitz JH, Clark JP, Lau P, et al. Rate of heart failure and 1-year survival for older people receiving low-dose beta-blocker therapy after myocardial infarction. *Lancet* (London, England). 2000;356(9230):639-44.
27. Petersen I, Douglas I, Whitaker H. Self controlled case series methods: an alternative to standard epidemiological study designs. *BMJ (Clinical research ed)*. 2016;354:i4515.
28. Douglas IJ, Evans SJ, Hingorani AD, Grosso AM, Timmis A, Hemingway H, et al. Clopidogrel and interaction with proton pump inhibitors: comparison between cohort and within person study designs. *BMJ (Clinical research ed)*. 2012;345:e4388.
29. Burgess S, Butterworth A, Malarstig A, Thompson SG. Use of Mendelian randomisation to assess potential benefit of clinical intervention. *BMJ (Clinical research ed)*. 2012;345:e7325.

Auteurs hebben het 'Formulier belangenverstrengeling' geheel ingevuld en ondertekend met de verklaring "geen belangenverstrengeling".

