



Cohort- onderzoek (deel 7)

Dit is het laatste artikel uit een serie van zeven artikelen over het beoordelen van diverse vormen van medisch-wetenschappelijke publicaties.¹⁻⁶⁾ De reden voor deze serie is om verpleegkundigen te helpen bij het efficiënt en kritisch beoordelen van wetenschappelijke literatuur op hun vakgebied. Het bijhouden van vakliteratuur is voor velen lastig omdat hiervoor eigenlijk (te) weinig tijd is. Daarnaast vertrouwt men vaak op de eigen ervaring en routine. Toch is het goed om te realiseren dat we, ondanks onze opgebouwde expertise, vele malen per dag keuzes maken of onzekerheden hebben over hoe iets komt of waarom sommige ziektes bij groepen kinderen of volwassenen vaker voorkomen. Ook kunt u zich soms afvragen of er een verband bestaat tussen een ziekte en een mogelijke oorzaak.

dr. Hester Vermeulen

projectleider afdeling Kwaliteit en Proces Innovatie van het AMC Amsterdam, docent EBP en lid lectoraat EBP bij de Amsterdam School of Health Professionals, Hogeschool van Amsterdam

dr. Dirk Ubbink

programmaleider afdeling Kwaliteit en Proces Innovatie van het AMC Amsterdam

» Enkele voorbeelden: Is er een relatie tussen lichaamsbeweging en de ontwikkeling van borstkanker?⁷⁾ Is er een relatie tussen een te laag geboortegewicht en het op latere leeftijd ontwikkelen van suikerziekte of hart- en vaatziekten (www.hongerwinter.nl)? Is het schadelijk voor de baby om tijdens de zwangerschap antidepressiva te gebruiken?⁸⁾ Of is er een relatie tussen een hoog geboortegewicht en het later optreden van reuma?⁹⁾ Zulke vragen zijn niet te onderzoeken door middel van een gerandomiseerd onderzoek, want het is niet ethisch om mensen te randomiseren voor blootstelling aan potentieel schadelijke effecten en het duurt vaak erg lang voordat de aandoening optreedt.¹⁾ Voor dit soort vragen

is cohortonderzoek beter geschikt. Over de kritische beoordeling van deze vorm van onderzoek gaat dit artikel.

Cohortonderzoek

Het doel van deze 'toolbox' is om de waarde van een cohortonderzoek kritisch te beoordelen. In cohortonderzoek wordt een afgebakende groep patiënten gedurende een bepaalde tijd gevolgd en wordt gekeken of een bepaalde uitkomst (bijvoorbeeld borstkanker of reuma) optreedt. Cohortonderzoek kan niet alleen vraagstellingen betreffen op etiologisch (= oorzakelijk) gebied of op het gebied van bijwerkingen (ook wel *Harm* genoemd), maar ook prognostische vraagstellingen. In cohortonderzoek wordt de relatie tussen een determinant (een etiologische factor, een bepaalde interventie of een prognostische factor, in dit voorbeeld geboortegewicht) en de uitkomst (in dit voorbeeld het optreden van reuma) onderzocht. Hierbij wordt een vergelijking gemaakt tussen een groep deelnemers met een bepaalde determinant (= risicofactor) en een groep deelnemers zonder deze determinant. In dit artikel zal gesproken worden over deelnemers die wel of niet een bepaalde determinant hadden.

Cohortonderzoeken kunnen prospectief zijn (de blootstelling aan een risicofactor is gedefinieerd en de deelnemers zijn geselecteerd voordat de uitkomst optrad) of retrospectief (historisch cohortonderzoek; de blootstelling in het verleden is bekeken, meestal in medische en verpleegkundige statussen, nadat de uitkomst al bekend was en wordt daarna de uitkomst ervan onderzocht). Retrospectieve cohortonderzoeken zijn gevoeliger voor bias (= vertekening). In cohortonderzoek kan sprake zijn van twee te onderscheiden situaties:

De relatie tussen één centrale determinant en de uitkomst wordt onderzocht. Deze situatie komt voor in etiologisch onderzoek of onderzoek naar bijwerkingen. Soms is men bij prognostisch onderzoek geïnteresseerd in slechts één determinant. Kenmerk van deze situaties

is dat een schatting van de sterkte van de relatie tussen de determinant en de uitkomst wordt gemaakt die gecorrigeerd wordt voor allerlei mogelijke andere factoren (confounders).

De relatie tussen meerdere determinanten en de uitkomst wordt onderzocht. Deze methode wordt vooral toegepast in prognostisch onderzoek. Hierbij wordt een prognostisch model geschat en wordt gekeken wat de invloed is van die verschillende determinanten op de uitkomst.

In cohortonderzoek worden dus groepen mensen over een langere periode op gezette tijden onderzocht op het optreden van de uitkomst. Deze groepen delen vaak iets met elkaar. Zijn ze bijvoorbeeld allen geboren in de Hongerwinter, hebben ze allen hetzelfde beroep of komen ze allen uit dezelfde woonplaats? De groepen krijgen dan bijvoorbeeld om de twee jaar een vragenlijst toegestuurd over onder andere hun eet-, sport- en rookgewoontes. Na verloop van tijd kan men op basis van de gegevens over deze groepen een theorie toetsen en bijvoorbeeld tot de conclusie komen dat mensen die roken eerder sterven. Dit type onderzoek heeft tot nu toe zeer interessante fundamentele inzichten opgeleverd en geeft aanwijzingen om verder te zoeken.

Cohortonderzoek heeft een aantal voordelen: het is relatief goedkoop, het betreft vaak een grote representatieve groep (wat de kans op het vinden van toevallige verbanden vermindert) en de grote set gegevens kan onverwacht interessante nevenconclusies opleveren. Nadelen zijn er echter ook: het heeft vaak een lange duur en het geeft wel inzicht of er een relatie met de uitkomst is, maar niet waarom die relatie er is. De kritische beoordeling van een willekeurige studie, en dus ook van een cohortstudie, bestaat altijd uit drie belangrijke punten. Ten eerste of de informatie in een studie valide is (dat wil zeggen: antwoord kan geven op de vraagstelling), ten tweede wat, hoe groot en hoe precies de uitkomsten van de studie zijn en ten derde of deze toepasbaar zijn in uw eigen situatie.

Checklist

Checklists om deze drie punten te beoordelen zijn onder andere te vinden op de site van het Dutch Cochrane Centre¹⁰⁾, onderdeel van de Cochrane Collaboration. De checklist is bestemd voor het beoordelen van patiëntcontroleonderzoek (zie tabel 1).

Ter illustratie van deze checklist gebruiken wij een cohortonderzoek van Mandl et al. over de relatie die er zou bestaan tussen geboortegewicht en het later optreden van reuma.⁹⁾ Dit is een illustratie van een mogelijk verband tussen een determinant (geboortegewicht) en een uitkomst (ontwikkelen van reuma (RA)), wat niet in een gerandomiseerd onderzoek bepaald kan worden.¹⁾ Maar het is wel belangrijk om te weten of deze relatie bestaat. U zult gelijkenissen zien met de vragen die er ook gesteld worden voor patiëntcontroleonderzoek.⁶⁾

Tabel 1.

Checklist voor de beoordeling van een patiëntcontroleonderzoek

	Ja	Nee	?
1 Zijn de te vergelijken onderzoeksgroepen duidelijk gedefinieerd?	x		
2 Kan selectiebias voldoende worden uitgesloten?	x		
3 Is de blootstelling duidelijk gedefinieerd en is de methode voor beoordeling van blootstelling adequaat?	x		
4 Is de uitkomst duidelijk gedefinieerd en is de methode voor beoordeling van de uitkomst adequaat?	x		
5 Is de uitkomst blind voor de blootstellingstatus bepaald?			x
6 Is er een voldoende lange follow-up?	x		
7 Kan selectieve loss-to-follow-up voldoende worden uitgesloten?	x		
8 Zijn de belangrijkste confounders of prognostische factoren geïdentificeerd en is er adequaat rekening mee gehouden in het ontwerp van het onderzoek of in de analyse?	x		



Validiteit

1 Definitie van de studiegroepen

De belangrijkste kenmerken van de blootgestelde en niet blootgestelde groep dienen duidelijk beschreven te zijn. Bij prognostisch onderzoek dient er sprake te zijn van een groep (cohort) die is samengesteld uit patiënten op eenzelfde tijdstip in het beloop van hun ziekte (inceptie-cohort).

In het voorbeeld is duidelijk beschreven wat wordt verstaan onder blootgesteld en niet blootgesteld. Beide groepen zijn duidelijk beschreven. De groepen met en zonder een hoog geboortegewicht zijn allebei afkomstig uit dezelfde populatie, namelijk een cohort van 121.700 getrouwde vrouwen in de leeftijd van 30-55 jaar die de eerste vragenlijst in 1976 hadden ingevuld.

2 Selectiebias

Om te beoordelen of sprake is van selectiebias moeten de inclusie- en exclusiecriteria van de onderzoeksgroepen duidelijk beschreven zijn. Selectiebias kan bijvoorbeeld ontstaan als retrospectief een cohort wordt samengesteld na kennis over de uitkomst. De beslissing om deelnemers te includeren kan dan afhankelijk zijn van de uitkomst. Hierdoor wordt de kans om als 'zieke' blootgestelde geselecteerd te worden in het cohort verschillend van de kans om als 'zieke' niet-blootgestelde geselecteerd te worden. Selectiebias kan ook ontstaan wanneer er maar een deel van het oorspronkelijke cohort geselecteerd wordt, bijvoorbeeld op basis van een onbewuste voorkennis over een relatie tussen bepaalde factoren en de uitkomst. Hierdoor kunnen belangrijke verschillen ontstaan tussen de te vergelijken onderzoeksgroepen (naast de blootstelling) die samenhangen met de uiteindelijke uitkomst waarvoor niet gecorrigeerd kan worden. *In het voorbeeld gaat het om een prospectief cohort dat werd gekozen op het moment dat nog niet duidelijk was wie later RA zou krijgen. De in- en exclusiecriteria staan duidelijk beschreven: vrouwen met kanker of een bind-*

weefselziekte bij het begin van de studie of kanker gedurende de follow-up werden uitgesloten, evenals vrouwen die zeiden dat ze RA hadden gekregen, zonder dat dit bevestigd kon worden uit de medische gegevens.

3 Blootstelling

De blootstelling dient duidelijk gedefinieerd te zijn en er moet aangeven zijn met welk instrument, wanneer en onder welke omstandigheden de determinant is vastgesteld. Dit dient in beide onderzoeksgroepen op dezelfde wijze gedaan te zijn.

Hieraan wordt in de studie duidelijk voldaan. De 'blootstelling' is hier het geboortegewicht, dat is ingedeeld in vijf groepen (zie hiervoor tabel 1 in het artikel van Mandl et al.⁹). Dit geboortegewicht was weliswaar een door de vrouwen zelf gerapporteerde schatting (door middel van een vragenlijst in 1992), maar uit een andere studie was al gebleken dat deze eigen schatting goed overeenkomt met de werkelijkheid.

4 Uitkomst

De uitkomst en gebruikte criteria voor het vaststellen van de uitkomst dienen duidelijk gedefinieerd te zijn, zodanig dat dit door andere onderzoekers herhaald kan worden. De uitkomst dient op een valide manier gemeten te zijn.

De diagnose RA werd gesteld uit de medische gegevens op basis van criteria van de American College of Rheumatology, door twee ervaren reumatologen. Er moest aan minimaal vier van de zeven criteria voor deze diagnose zijn voldaan om definitief te kunnen spreken van RA. Eventuele meningsverschillen werden door overleg bijgelegd. Dit beoordelen we daarom als valide.

5 Geblindeerde uitkomstmeting

Blinding wordt bereikt door zonder kennis over de blootstellingstatus de uitkomst te meten. Door de blinding wordt voorkomen dat de uitkomstmeting wordt

beïnvloed door kennis over de blootstelling. Indien er geen blinding heeft plaatsgevonden, zijn er dan aanwijzingen over de invloed van deze kennis over de blootstellingstatus op de beoordeling van de uitkomst?

In het voorbeeldartikel wordt niet expliciet aangegeven of de reumatologen op de hoogte waren van het geboortegewicht van de vrouwen bij de beoordeling of ze RA hadden.

6 Follow-up

De follow-upduur dient zodanig lang te zijn dat de bestudeerde uitkomst erin op kan treden.

Het onderzoek naar het optreden van RA vindt 30 jaar later (in 2002) plaats, wanneer de vrouwen op een leeftijd zijn gekomen dat deze aandoening in deze lange follow-upperiode had kunnen optreden. Deze follow-upperiode is ruim voldoende.

7 Selectieve loss-to-follow-up

Er mogen geen belangrijke verschillen in uitval van deelnemers tussen de onderzoeksgroepen te zijn. Aantallen en redenen voor uitval dienen gerapporteerd te zijn.

Ook als er geen uitvallers waren, moet dit zijn beschreven. Het is niet mogelijk om op voorhand per indicatiegebied aan te geven welk percentage uitval nog acceptabel is. Relatief grote uitval maakt een cohortonderzoek gevoelig voor selectieve uitval. Dit kan het geval zijn als de absolute aantallen uitvallers en de redenen van uitval tussen de onderzoeksgroepen verschillen.

De auteurs vermelden dat na 16 jaar follow-up (in 1992) nog gegevens van 90 procent van het originele cohort vrouwen verkregen kon worden. Dit is dus ruim voldoende en er zijn geen aanwijzingen om te denken dat de resterende 10 procent allemaal wel, of allemaal géén afwijkend geboortegewicht of RA hadden.

8 Confounders/prognostische factoren

Voor het beantwoorden van de vraag 'Zijn de belangrijkste confounders of prognostische factoren geïdentificeerd en is er adequaat rekening mee gehouden in het ontwerp van het onderzoek of in de analyse?' is wat uitleg over confounding nodig. Confounding is een verstoring van een relatie tussen blootstelling en uitkomst door een andere factor die zowel met de blootstelling als de uitkomst geassocieerd is. Bijvoorbeeld: een verhoogde consumptie door ouderen van koolzuurhoudende dranken zou leiden tot een verhoogd aantal doden. Echter, de consumptie van dit soort dranken is hoger bij hoge zomerse temperaturen. Deze verhoogde zomerse temperatuur heeft invloed op zowel de consumptie van koolzuurhoudende dranken als ook op het overlijden van oudere mensen.

In een cohortonderzoek zal vrijwel altijd sprake zijn van een ongelijke verdeling van confounders over de groepen. Klinisch inzicht zal moeten helpen om te beoordelen of alle belangrijke confounders in overweging zijn genomen. Er dient te zijn aangegeven welke confounders

er overwogen zijn en hoe daar in het ontwerp van het onderzoek en in de analyse rekening mee is gehouden. Ook kan achteraf in de analyse hiervoor gecorrigeerd worden.

Diverse mogelijke confounders rond de geboorte van de vrouwen zijn in het voorbeeldartikel onderzocht: vroeggeboorte, hun moeders leeftijd, sociaaleconomische status, roken, diabetes, lichaamsvorm op 50-jarige leeftijd, en de regio waar ze woonden. Ook confounders (risicofactoren) die te maken kunnen hebben met het krijgen van RA werden meegenomen: het aantal rookjaren, leeftijd van het begin van de menstruatie, pilgebruik, hormoongebruik na de menopauze, totaal aantal jaren dat ze borstvoeding hebben gegeven, de BMI op 18-jarige leeftijd en diabetes. Hiervoor wordt in de analyse ook gecorrigeerd.

Algemeen oordeel

Hier wordt een inschatting gevraagd van de validiteit (Is dit een goed onderzoek?) en toepasbaarheid (Gaat dit onderzoek over mijn PICO?). Let hierbij ook op eventuele fouten in het onderzoek die funest zijn voor de validiteit ervan (*red flags, fatal flaws*). Er zijn geen regels te geven voor welke punten positief gescoord moet worden of hoeveel punten ten minste positief gescoord moeten worden. Dit is deels afhankelijk van de 'state-of-the-art' met betrekking tot het betreffende onderwerp. Het gaat er hier om het samenvattend oordeel van wat de beoordelaar van het artikel zou willen meedelen over de bruikbaarheid van het artikel voor de besluitvorming. *Op bijna alle punten scoort deze cohortstudie goed. Het oordeel in deze is daardoor gemakkelijk: ja, het is een valide cohortstudie.*

Door het beantwoorden van bovengenoemde punten heeft u dus een indruk gekregen hoe degelijk ('valide') de studie is uitgevoerd. Als er minpunten zijn gevonden, moet u zich afvragen of de uitvoering van het onderzoek beter had gekund.

Zijn er ernstige tekortkomingen gevonden in de validiteit die wel beter hadden gekund en gemoeten, dan zijn de resultaten vals of vertekend en dus niet betrouwbaar. In dat geval hoeft men niet verder te lezen. Het is dan beter om een andere publicatie te zoeken die misschien niet evenveel bewijskracht heeft, maar wel goed en dus betrouwbaar is uitgevoerd.

Zijn er geen (ernstige) tekortkomingen gevonden in de validiteit, dan kunnen de resultaten veilig worden geïnterpreteerd.

Resultaten

Noteer dan de bestudeerde uitkomst en de eenheid van blootstelling waarvoor het effect berekend is. Noteer ook de confounders die in de multivariate analyse (zie hierna) zijn opgenomen, en de follow-upduur.

De uitkomsten van een cohortonderzoek kunnen dichotoom zijn (wel/geen schade, wel/geen bijwerking, wel/geen genezing, wel/geen overleving) of continu

(bloeddruk, cholesterolwaarde). Bij dichotome uitkomsten kan het effect van de blootstelling uitgedrukt worden in een relatief risico (RR), odds ratio (OR) of absolute risicoreductie (ARR). Bij continue uitkomsten wordt het effect uitgedrukt in een verschilscore (Engels: *mean difference* (MD)). Omwille van de lengte van dit artikel laten we hier diverse andere analysemogelijkheden buiten beschouwing. U kunt die nalezen op www.cochrane.nl.

De resultaten uit het voorbeeldartikel van Mandl et al. zijn in verschillende tabellen weergegeven. In tabel 1 van het artikel is te zien dat een vrouw die bij haar geboorte een gewicht had van meer dan 4,54 kg, een tweemaal zo hoge kans heeft op het krijgen van RA ten opzichte van een geboortegewicht tussen de 3,2 en 3,85 kg. Het bijbehorende 95% betrouwbaarheidsinterval loopt van 1,3 tot 3,0, dus is hier sprake van een significant effect.

Met deze gegevens kunt u nu zelf dit relatieve risico terugrekenen naar een absolute risicotoename (ARI) om een number needed to harm (NNH) uit te kunnen rekenen: Het aantal vrouwen dat de onderzoekers vonden met RA was 619, op een totaal van 87.077 vrouwen van wie de gegevens beschikbaar waren.

De kans op het krijgen van RA in de hele groep was dus $619/87.077 = 0.007$. Dit is bij benadering het 'achtergrond-risico' op het krijgen van RA in de algemene populatie.

De kans op het krijgen van RA in de groep vrouwen met een hoog geboortegewicht was tweemaal zo hoog, dus 0.014 . Dus het risicoverschil (ARI) tussen beide groepen is $0.014 - 0.007 = 0.007$.

Het number needed to harm is dan $1/ARI = 140$. Dus zal bij elke 140ste vrouw met een hoog geboortegewicht 1 extra patiënt met RA voorkomen.

Toepasbaarheid

Eerst hebt u een algemeen oordeel over de validiteit gegeven. Daarna hebt u de resultaten nader beschouwd. De laatste stap is het beoordelen of de gevonden resultaten ook toepasbaar zijn in de eigen Nederlandse situatie of in een andere populatie (zoals algemene bevolking, eerste lijn, tweede lijn, academisch ziekenhuis). Beide vragen zijn belangrijk om met uw collega's te bespreken. Dit dient plaats te vinden voordat u met uw collega's de resultaten uit dit cohortonderzoek kunt gaan toepassen.

Bij deze discussie is tevens een globale samenvatting van het eindoordeel over het artikel gewenst. Probeer, indien aanwijzingen bestaan voor vertekening van de resultaten, ten minste een inschatting te maken van de richting van de vertekening (overschatting of onderschatting van de bestudeerde associatie) en ook over de grootte van de vertekening. Eventuele aanwijzingen voor mogelijke belangenverstrengeling van de auteurs met belanghebbende opdrachtgevers, kunt u hier ook rapporteren. Ook is het verstandig ingezonden brieven en/of redactionele

commentaren op het hier door u beoordeelde onderzoek te raadplegen bij het formuleren van uw conclusie.

In deze is de discussie redelijk eenvoudig. Dit is een mooi uitgevoerde cohortstudie, die ook voor ons land geldig is.

Samenvatting

Een cohortstudie is een geschikte methode om relaties tussen mogelijke oorzaken en een ziekte te kunnen opsporen. Het onderzoek is gericht op het vinden van zo'n relatie, maar niet waarom een bepaalde uitkomst dan vaker optreedt bij een bepaalde blootstelling. Vaak kan een cohortonderzoek daarom leiden tot vervolgonderzoek. Er zijn voor- en nadelen aan cohortonderzoek, denk hierbij aan de lange follow-upduur, maar ook aan de geringe kosten.

Met behulp van de checklist kunt u beoordelen of het onderzoek goed uitgevoerd is en met uw collega's een onderbouwde discussie aangaan of de resultaten relevant en toepasbaar zijn op de patiëntengroep waarmee u te maken hebt. <

Literatuur

- 1) Ubbink D, Vermeulen H, Scholten R. Toolbox deel 1: Gerandomiseerd klinisch onderzoek. Nederlands Tijdschrift voor Evidence Based Practice 2007; 6:8-12.
- 2) Vermeulen H, Ubbink DT, Hooft L. Toolbox deel 2: Systematisch literatuuronderzoek. Nederlands Tijdschrift voor Evidence Based Practice 2008; 1:4-8.
- 3) Goossens A, Ubbink D, Vermeulen H. Toolbox deel 3: Evidence-based richtlijnen. Nederlands Tijdschrift voor Evidence Based Practice 2008; 2:18-23.
- 4) Leeftang M, Vermeulen H, Hooft L. Toolbox deel 4: Diagnostisch onderzoek. Nederlands Tijdschrift voor Evidence Based Practice 2008; 3:18-23.
- 5) Vermeulen H, Ubbink DT, Hooft L. Toolbox deel 5: Systematisch literatuuronderzoek van diagnostisch onderzoek. Nederlands Tijdschrift voor Evidence Based Practice 2008; 4:18-22.
- 6) Vermeulen H, Korevaar J, Ubbink DT. Toolbox deel 6: Patiënt-controleonderzoek. Nederlands Tijdschrift voor Evidence Based Practice 2008; 5:18-23.
- 7) Suzuki S, Kojima M, Tokudome S, Mori M, Sakauchi F, Fujino Y, Wakai K, Lin Y, Kikuchi S, Tamakoshi K, Yatsuya H, Tamakoshi A; for the Japan Collaborative Cohort Study Group. Effect of physical activity on breast cancer risk: findings of the Japan Collaborative Cohort Study. *Cancer Epidemiol Biomarkers Prev* 2008; 17:3396-3401.
- 8) Maschi S, Clavenna A, Campi R, Schiavetti B, Bernat M, Bonati M. Neonatal outcome following pregnancy exposure to antidepressants: a prospective controlled cohort study. *BJOG* 2008; 115:283-289.
- 9) Mandl LA, Costenbader KH, Simard J, Karlson EW. Is birthweight associated with risk of rheumatoid arthritis? Data from a large prospective cohort study. *Ann Rheum Dis* 2008; Jul 1.
- 10) www.cochrane.nl.

