



Methodes om de betrouwbaarheid van observationele studies te vergroten: wat is de plaats van een instrumentele-variabele-analyse?

Tom Poelman, Vakgroep Huisartsgeneeskunde en Eerstelijnsgezondheidszorg, UGent

Voor de meeste lezers trappen we een open deur in wanneer we zeggen dat gerandomiseerde gecontroleerde studies de hoeksteen vormen om de voor- en nadelen van interventies te onderzoeken. Daartegenover staat echter ook dat we in onze dagelijkse praktijkvoering herhaaldelijk ervaren dat klinisch relevante behandelvragen niet beantwoord kunnen worden met RCT's. Ofwel omdat de bestaande studies onbruikbaar zijn (te weinig patiënten, ondermaatse methodologie), of nog niet zijn uitgevoerd of... omdat ze niet uitvoerbaar zijn. Om bijvoorbeeld de effecten van een behandeling op (zeer) lange termijn te evalueren, of om (zeer) zeldzame ongewenste effecten van een behandeling op te sporen, zouden we duizenden deelnemers gedurende meerdere jaren moeten randomiseren, wat vaak praktisch, financieel en ethisch onmogelijk is. In dergelijke gevallen vormt observationeel onderzoek een belangrijk alternatief. Sommigen gaan nog verder en wijzen op de nadelen van RCT's, zoals de lage generaliseerbaarheid en de hoge kostprijs om observationeel onderzoek op de voorgrond te schuiven (1). Waar we observationeel onderzoek ook positioneren, de vraag blijft: zijn observationele gegevens over het effect van behandelingen betrouwbaar?

Het centrale probleem van observationeel onderzoek is de aanwezigheid van versturende factoren of **confounders**. In tegenstelling tot interventieel onderzoek zal bij observationeel onderzoek over het effect van een behandeling het wel of niet krijgen van de behandeling afhangen van de keuze van de patiënt en/of de arts zonder tussenkomst van de onderzoekers. Zo ontstaat er tussen de blootgestelde en de niet-blootgestelde groep, of eenvoudiger gezegd tussen de 'behandelings- en de controlegroep', een zeker onevenwicht in prognostische factoren dat het te onderzoeken effect kan beïnvloeden. Als we erin slagen om confounders, zoals leeftijd, geslacht, ziekte-ernst, comorbiditeit, op een accurate manier te meten, kunnen we vervolgens hun invloed op de uitkomst met verschillende statistische methodes neutraliseren. Sommige kwamen reeds aan bod in Minerva (2,3). Een recent artikel biedt ons een logisch overzicht van de meest gebruikte technieken (4). We hebben ze voor u hieronder samengevat.

Bij **stratificatie** verdeelt men alle onderzochte personen in homogene groepen (strata) voor telkens één prognostische factor (bijvoorbeeld <50 jaar en ≥ 50 jaar). Vervolgens berekent men het verschil in effect tussen de behandelings- en de controlegroep (geen behandeling of een andere behandeling) in alle strata om nadien al deze verschillen te combineren.

Stel: in een cohortonderzoek worden 100 personen wel (behandelingsgroep) en 100 personen niet (controlegroep) blootgesteld aan een behandeling tijdens een bepaald tijdsinterval. 80% van de behandelingsgroep is <50 jaar en 80% van de controlegroep is ≥ 50 jaar.

In de behandelingsgroep sterven 12 personen (8 zijn <50 jaar en 4 zijn ≥ 50 jaar) terwijl er in de controlegroep 18 personen (2 zijn <50 jaar en 16 zijn ≥ 50 jaar) sterven. Dat komt neer op een 'ongecorrigeerde' RR van $12\%/18\%=0,67$ of een 'ongecorrigeerde' RRR van 33% op overlijden in de behandelgroep.

Maar, wanneer we de 200 personen van het cohortonderzoek verdelen in een stratum <50 jaar en een stratum ≥ 50 jaar zien we:

- in het stratum <50 jaar: 8/80 overlijdens in de behandelingsgroep en 2/20 overlijdens in de

controlegroep of een RR_1 van $10\%/10\%=1$

- in het stratum ≥ 50 jaar: $4/20$ overlijdens in de behandelingsgroep en $16/80$ overlijdens in de controlegroep of een RR_2 van $20\%/20\%=1$

De combinatie van RR_1 en RR_2 levert een ‘gecorrigeerde’ RR van 1 op. Rekening houdend met het verschil in leeftijd tussen de behandelingsgroep en de controlegroep leidt de behandeling dus niet tot een daling in mortaliteit.

Hoe meer prognostische factoren waarvoor men wil corrigeren, hoe meer strata men zal moeten voorzien (aantal strata=aantal factoren²). Vandaar dat het op een bepaald moment eenvoudiger wordt om gebruik te maken van een **multivariate regressieanalyse** (2). Door in deze vergelijking de waarde van alle confounders constant te houden kunnen we het eigenlijke effect van een interventie op de uitkomst inschatten. Wanneer de uitkomstmaat (of afhankelijke variabele) dichotoom is, spreken we van een **logistische regressieanalyse**. Voor continue uitkomstmaten gebruiken we een lineaire regressieanalyse en wanneer men wil weten in hoeverre de interventie een bepaalde gebeurtenis in de tijd uitstelt, past men een **Cox proportional hazards model** toe. Respectievelijk wordt het verband dan uitgedrukt in odds ratio (OR), absoluut risicoverschil (AR) en hazard ratio (HR).

Als we in verhouding tot het aantal observaties voor te veel confounders gaan corrigeren, zal de multivariate regressieanalyse tot weinig valide resultaten leiden. Dat probleem kan men opvangen door vooreerst aan elke patiënt een propensity score (0 tot 1) toe te kennen (3). Deze score, die de kans uitdrukt om een behandeling te krijgen, zal afhangen van andere prognostische factoren (bijvoorbeeld: ernstig zieke ouderen zullen vlugger behandeld worden dan minder ernstig zieke jongeren). Door te corrigeren voor deze score zullen we dus meteen voor veel prognostische factoren corrigeren zonder aan validiteit te moeten inboeten. Maar, hoeveel we ook corrigeren, toch zullen nog altijd ‘onbekende’ prognostische factoren voor een onevenwicht blijven zorgen. Houdt het verhaal hier op? Nee... In een observationele studie kunnen factoren voorkomen die samenhangen met het wel of niet krijgen van een behandeling maar niet geassocieerd zijn met de prognose. Deze instrumentele variabele zorgt er in feite voor dat de behandeling in zekere mate door toeval toegewezen werd (pseudorandomisatie) waardoor ook voor onbekende prognostische factoren een evenwicht in de onderzochte groepen kon ontstaan (5).

Stel: in een cohortstudie wordt 80% van de patiënten van centrum X behandeld met geneesmiddel A en 20% met geneesmiddel B terwijl in centrum Y 40% wordt behandeld met geneesmiddel A en 60% met geneesmiddel B. Als we aannemen dat patiënten toevallig in centrum X en Y terechtkomen en als andere aspecten van de behandeling niet verschillend zijn tussen centrum X en Y, dan kunnen we het verschil van 40% in gebruik van geneesmiddel A of B tussen de patiënten van beide centra beschouwen als een pseudorandomisatie.

Stel nu dat 48% van de patiënten uit centrum X klachtenvrij is versus 42% uit centrum Y, dan kunnen we het verschil van 6% in effect voor 40% toeschrijven aan het gebruik van geneesmiddel A in plaats van geneesmiddel B. Stel dat alle patiënten geneesmiddel A in plaats van geneesmiddel B hadden gekregen (100%), dan waren er naar schatting 15% ($6\%/0,40$) meer patiënten klachtenvrij geweest met geneesmiddel A in plaats van met geneesmiddel B.

Uit het bovenstaande fictieve voorbeeld kunnen we enkele belangrijke voorwaarden voor een betrouwbare instrumentele variabele afleiden:

- de onderzoekers moeten duidelijk rapporteren hoeveel patiënten de interventie wel en niet gekregen hebben
- er moet een substantieel verschil zijn in de status van de instrumentele variabele (bijvoorbeeld: duidelijk minder of meer kans op het krijgen van een interventie in ziekenhuis A versus ziekenhuis B)
- de instrumentele variabele zelf mag de uitkomst niet beïnvloeden (bijvoorbeeld: het mag niet zijn dat men in ziekenhuis A naast de onderzochte behandeling ook meer bijkomende behandelingen geeft die de genezing van patiënten kunnen beïnvloeden)

- de prognostische factoren moeten gelijk verdeeld zijn naargelang de status van de instrumentele variabele (bijvoorbeeld: het mag niet zijn dat ziekenhuis A vooral oudere patiënten opvangt).

Als we met deze voorwaarden rekening houden, kunnen er vijf soorten van instrumentele variabelen voorkomen: regionale verschillen in zorg, verschillende hospitaalgewoontes, verschillen in (huis-) artspraktijken (alle drie analoog met clusterrandomisatie), alsook de voorgeschiedenis van de patiënt die niet samenhangt met de onderzochte uitkomstmaat en de kalendertijd (zoals tijd tot goedkeuring van een geneesmiddel).

Omdat een instrumentele-variabele-analyse ook kan corrigeren voor niet-gemeten en onmeetbare confounders, zal de schatting dus dicht bij de waarheid liggen dan met andere technieken. Maar, hoe dicht? Dat kan dan weer alleen met gerandomiseerd onderzoek achterhaald worden...

Besluit

Er bestaat een breed arsenaal van statistische methodes om de resultaten van observationeel onderzoek betrouwbaarder te maken. Omdat een instrumentele-variabele-analyse ook corrigeert voor niet-gemeten en onmeetbare confounders, zullen de resultaten ervan meer aanleunen bij deze van een RCT.

Referenties

1. Goodman SN, Schneeweiss S, Baiocchi M. Using design thinking to differentiate useful from misleading evidence in observational research. *JAMA* 2017;317:705-7. DOI: 10.1001/jama.2016.19970
2. Poelman T. Het verschil tussen regressie en correlatie. *Minerva* 2016;15(2):51-3.
3. Poelman T. Propensity Score Matching. *Minerva* 2013;12(8):103.
4. Agoritsas T, Merglen A, Shah ND, et al. Adjusted analyses in studies addressing therapy and harm. *Users' guides to the medical literature. JAMA* 2017;317:748-59. DOI: 10.1001/jama.2016.20029
5. Boef AG, le Cessie S, Dekkers OM. Instrumentele-variabele-analyse. *Ned Tijdschr Geneeskd* 2013;157: A5481.