

Continue variabelen analyseren met ANCOVA

Tom Poelman, Vakgroep Huisartsgeneeskunde en Eerstelijnsgezondheidszorg, Universiteit Gent

In een recente RCT onderzochten MacPherson et al. het effect van acupunctuur of counseling op de ernst van depressieve symptomen bij patiënten met depressie^{1,2} (zie bespreking in dit nummer op blz. 101). Om de ernst te meten gebruikten de onderzoekers de **PHQ-9** score (variërend van 0 tot 27, met 5-9 als milde, 10-14 als matige, 15-19 als matig ernstige en 20 of hoger als ernstige depressie). Deze score bepaalden ze na randomisatie en na 3 maanden follow-up bij alle deelnemers:

PHQ-9	Acupunctuur + gebruikelijke zorg	Counseling + gebruikelijke zorg	Alleen gebruikelijke zorg
Begin studie	15,3 (SD 5,33)	16,6 (SD 5,27)	16,2 (SD 5,09)
Na 3 maanden	9,4 (SD 6,33)	10,9 (SD 6,45)	12,7 (SD 6,47)

Op basis van bovenstaande cijfers konden de onderzoekers het effect van acupunctuur op 2 verschillende manieren analyseren:

- ofwel konden ze de gemiddelde eindscore (score na follow-up) van acupunctuur + gebruikelijke zorg met deze van gebruikelijke zorg alleen vergelijken. Dat zou geresulteerd hebben in een verschil in PHQ van $9,4 - 12,7 = -3,3$ punten. In dat geval zouden we dus kunnen zeggen dat na 3 maanden de gemiddelde PHQ 3,3 punten lager lag in de groep met acupunctuur dan in de groep met alleen gebruikelijke zorg;
- ofwel konden ze de gemiddelde verschillingscore (score na follow-up verminderd met de beginscore) van acupunctuur + gebruikelijke zorg met deze van gebruikelijke zorg alleen vergelijken. Dat zou dan geresulteerd hebben in een verschil in PHQ van $(9,4 - 15,3) - (12,7 - 16,2) = -2,4$ punten. In dat geval zouden we dus kunnen zeggen dat na 3 maanden de gemiddelde PHQ met 2,4 punten meer was gedaald in de groep met acupunctuur dan in de groep met alleen gebruikelijke zorg.

Uit het bovenstaande kunnen we afleiden dat het geschatte behandelingseffect op basis van beide methodes alleen gelijk zal zijn wanneer ook de gemiddelde beginscores in beide groepen gelijk zijn. Voor beide methodes zal de statistische significantie van deze schatting dan echter nog afhangen van de correlatie tussen de begin- en de eindwaarden. In het geval van een lage correlatie zal een significant verschil tussen de gemiddelde eindscores gemakkelijker en tussen de gemiddelde verschillingscores moeilijker aan te tonen zijn. Het omgekeerde zal gelden wanneer er een hoge correlatie bestaat tussen de begin- en de eindwaarden. Het is methodologisch onjuist wanneer de onderzoekers uit deze 2 analysetechnieken post-hoc kiezen voor de techniek met de meest significante resultaten.

Aangezien de beginwaarde van PHQ tussen beide groepen in ons voorbeeld verschillend was, is het niet mogelijk om het behandelingseffect op basis van de gemiddelde eindscores correct in te schatten. Maar, zelfs gebruik makend van de gemid-

delde verschillingscores, moeten we rekening blijven houden met een belangrijke vorm van bias als gevolg van **'regression to the mean'**. Door dit fenomeen zullen de scores van patiënten met hogere beginscores globaal meer dalen dan deze van personen met lagere beginscores^{3,4}. Omdat in de groep met gebruikelijke zorg meer personen met hogere beginscores zijn opgenomen, zal het effect van acupunctuur dus onderschat worden.

Met een 'analysis of covariance' (ANCOVA) kunnen we deze vertekening voorkomen⁴. In feite gaat het hier om een regressieanalyse die de eindscore van iedere patiënt corrigeert met zijn of haar beginscore. Op die manier wordt de inschatting van het effect dus onafhankelijk van een verschil in gemiddelde beginscore tussen de onderzoeksgroepen en wordt tegelijk ook het fenomeen van 'regression to the mean' uitgeschakeld. Met deze techniek vonden de onderzoekers in ons voorbeeld een verschil van -2,46 (95% BI van -3,72 tot -1,21) punten tussen de groep met acupunctuur en de groep met alleen gebruikelijke zorg. Deze waarde komt goed overeen met de inschatting van het effect op basis van de gemiddelde verschillingscores (zie hoger). Dat komt omdat de correlatie tussen de beginscores en de follow-up scores in ons voorbeeld wel voldoende groot was (SD van PHQ-9 verschilt weinig tussen het begin van de studie en na 3 maanden follow-up). Het fenomeen van 'regression to the mean' speelde dus een kleine rol in ons voorbeeld. Misschien kozen de onderzoekers toch voor ANCOVA omwille van de grotere statistische power om een verschil tussen onderzoeksgroepen aan te tonen⁴. Omdat ANCOVA een vorm van multiële regressieanalyse is, heeft het gebruik van deze techniek ook als voordeel dat men tegelijk ook kan corrigeren voor bepaalde prognostische variabelen zoals leeftijd,...

Besluit

ANCOVA is een multiële regressieanalyse die toelaat om een effect van de interventie correct in te schatten tussen twee onderzoeksgroepen, waarbij de uitkomstmeting voor en na de interventie op correcte wijze in rekening wordt gebracht.

Referenties

1. MacPherson H, Richmond S, Bland M, et al. Acupuncture and counseling for depression in primary care: a randomised controlled trial. *PLoS Med* 2013;10:e1001518.
2. Declercq T, Poelman T. Acupunctuur en counseling als aanpak van depressie in de eerste lijn: een gerandomiseerde gecontroleerde studie. *Minerva* 2014;13(8):101-2.
3. Bland JM, Altman DG. Some examples of regression towards the mean. *BMJ* 1994;309:780.
4. Vickers AJ, Altman DG. Statistic notes: Analysing controlled trials with baseline and follow up measurements. *BMJ* 2001;323:1123-24.