

## RCT met 'enriched enrollment' Het risico van bias en beperkte veralgemeenbaarheid

Pierre Chevalier

Een RCT bij patiënten met chronische lagerugpijn toonde aan dat transdermaal buprenorfine vergeleken met placebo statistisch significant meer pijnstilling gaf na twaalf weken ( $p=0,010$ )<sup>1</sup>. Voor deze RCT selecteerde men 1466 patiënten, waarvan er 1024 initieel zonder blinding behandeld werden met buprenorfine gedurende maximum 27 dagen. De studie zelf gebeurde uiteindelijk met de 541 patiënten die in de initiële fase positief reageerden op transdermaal buprenorfine.

De auteurs pasten hier een 'enriched enrollment' of een 'verrijking' toe. Hoe moeten we de resultaten van een dergelijke studie interpreteren?

Wanneer men een steekproef neemt uit een populatie om de werkzaamheid en/of de veiligheid van een therapeutische optie aan te tonen, wil men conclusies trekken voor de volledige populatie. Dat is niet mogelijk als men selectiecriteria toepast die een belangrijk deel van de populatie systematisch uitsluiten. Het gaat hier niet om correcte exclusiecriteria. De exclusie van bijvoorbeeld zwangere vrouwen bij de evaluatie van een nieuw geneesmiddel met een onduidelijk teratogeen effect is verantwoord en ethisch verplicht. Een ander verantwoord exclusie criterium is bv. nier- of leverinsufficiëntie bij de evaluatie van een (nieuw) geneesmiddel dat nog onvoldoende gekend is of dat mogelijke risico's inhoudt.

De CLASS-studie vergeleek de gastro-intestinale veiligheid van celecoxib met klassieke NSAID's bij patiënten met artrose of reumatoïde artritis<sup>2</sup>. De auteurs excludeerden de patiënten met actieve gastro-intestinale pathologie, nierinsufficiëntie (zonder duidelijke drempelwaarde in de publicatie!), leverinsufficiëntie en coagulatieproblemen. Voor de evaluatie van niet-selectieve NSAID's zijn deze exclusiecriteria wel aanvaardbaar.

Als we studieresultaten proberen toe te passen op een bepaalde patiënt, moeten we de karakteristieken van deze patiënt vergelijken met de karakteristieken van de geïncludeerde (en geëxcludeerde!) studiepopulatie. Met andere woorden: zou 'mijn' patiënt ook in aanmerking komen om deel te nemen aan deze studie?

De CLASS-studie excludeerde patiënten met nierinsufficiëntie. Roger is 75 jaar en komt op raadpleging voor een zeer pijnlijke gonartrose sedert twee dagen; bij de laatste bloedcontrole bedroeg de creatinineklaring 55 ml/min. Hij wordt behandeld met amlodipine en een ACE-inhibitor voor hypertensie die goed onder controle is. We kunnen de gunstige resultaten van de CLASS-studie niet veralgemenen naar de situatie van Roger. Door het gebruik van een NSAID, al of niet selectief, riskeert hij dat de nierinsufficiëntie verergert.

Gebruik maken van een 'enriched enrollment' of een 'verrijking' bij de inclusie van patiënten zorgt voor een groter probleem. Bij deze methode selecteert men alleen die patiënten die positief reageren op de behandeling en deze ook goed verdragen. Alleen deze patiënten randomiseren bij een therapeutische vergelijking, leidt tot een cohorteffect. Dat vormt een risico van 'bias van verwijdering ten aanzien van de nulhypothese' (het verschil is 'kunstmatig' vergroot) en maakt de extrapolatie naar een bredere populatie onmogelijk.

Een cohorteffect (of selectie van een cohort) is duidelijk aangetoond in de studies over het effect van lithium als preventie van recidieven van bipolaire stoornis.

In 2005 verscheen een literatuuroverzicht over het effect van lithium als preventie van recidieven van bipolaire stoornis<sup>3</sup>. In de gepoolde resultaten van twee studies 'zonder verrijking' bedroeg de OR voor lithium versus placebo 3,2 (95% BI van 0,65 tot 15,46). In drie studies maakte men gebruik van een verrijkingfase voor lithium door alleen patiënten te includeren die vóór de randomisatie goed reageerden op lithium en dit goed konden verdragen. De OR in deze drie studies bedroeg 22,0 (7,0-68,7). In twee studies includeerde men alleen patiënten die goed reageerden op lamotrigine: de OR bedroeg 1,9 (1,2-2,8), en was dus 11 maal kleiner dan in de studies met een verrijkingfase voor lithium. Dat verschil kan moeilijk aan het toeval te wijten zijn ( $p<0,05$  op basis van de betrouwbaarheidsintervallen). Het is dus duidelijk dat het hier om een cohorteffect gaat.

Tsai et al. deden een systematische review naar het effect van aripiprazol voor de preventie van recidieven van een bipolaire stoornis<sup>4</sup>. Ze vonden slechts één onderzoek (twee publicaties van Keck et al.<sup>5,6</sup>) met 567 patiënten die in een open fase van 6 tot 18 weken behandeld werden met aripiprazol en waarna de 161 patiënten (28%) die positief reageerden op de behandeling, gerandomiseerd werden over aripiprazol of placebo.

Dit soort selectie vormt een risico van bias. Een wetenschappelijk klinisch experiment toetst het verschil tussen twee hypothesen. Conventioneel kiest men als nulhypothese de uitkomst alsof de interventie niets uithaalt. De alternatieve hypothese formuleert dan de eventuele winst die de interventie realiseert. Een OR van 1 is een goed voorbeeld voor de bevestiging van de nulhypothese (er is geen verschil bewezen). Wanneer men echter bij het samenstellen van de steekproef voor de randomisatie alleen patiënten selecteert die gunstig zullen reageren op de interventie, dan creëert men als het ware bij de start van de studie een groter verschil tussen de nul- en de alternatieve hypothese. Dat is de bias die binnensluit in studies met een 'verrijkte selectie', namelijk de 'bias van verwijdering ten aanzien van de nulhypothese'.

Het ligt dus voor de hand dat we de resultaten van een studie met 'enriched enrollment' niet zomaar kunnen veralgemenen naar een bredere populatie.

In de studie van Steiner et al.<sup>1</sup> had transdermaal buprenorfine statistisch gezien meer effect dan placebo op het vlak van pijnverlichting over twaalf weken behandeling... bij de patiënten die gedurende maximum 27 dagen gunstig reageerden en de progressieve dosis verdroegen, en dat was alleen het geval bij 53% van de patiënten... Een faalpercentage van 47% voor een pijnbehandeling, voorgesteld als nieuwe behandeling van matige tot ernstige chronische pijn, is niet erg geruststellend. De patiënten die in de open fase goed reageerden hadden in de gerandomiseerde fase meer pijnverlichting dan de placebogroep. Het echte placebo-effect van deze behandeling kunnen we echter niet meer bepalen. Uiteindelijk kunnen we dus de werkzaamheid van deze behandeling niet veralgemenen naar alle patiënten met chronische lagerugpijn.

Referenties: zie [www.minerva-ebm.be](http://www.minerva-ebm.be)