

Geloofwaardigheid en inflatie van de effectgrootte in meta-analyses

P. Chevalier

NNT (number needed to treat) en ARR (absolute risicoreductie)

Minerva wees reeds eerder op de noodzaak om NNT's voorzichtig te interpreteren¹. Een NNT geeft aan hoeveel personen men gedurende een bepaalde periode (de studieduur) moet behandelen om één extra geval van een bepaalde ziekte te voorkomen of te genezen.

NNT's op basis van verschillende studies zijn maar valabel als men in de verschillende studies dezelfde behandeling vergelijkt, eenzelfde comparator gebruikt, dezelfde uitkomstmaat hanteert, het effect meet bij patiënten in hetzelfde stadium van de aandoening (met hetzelfde initiële risico) en wanneer men deze patiënten gedurende eenzelfde periode opvolgt. We moeten de NNT's van meta-analyses even omzichtig interpreteren. Een vergelijking van NNT's van verschillende meta-analyses is vanwege de bovenvermelde redenen niet zeer betrouwbaar².

Een NNT wordt berekend op basis van de absolute wijziging van het risico (dikwijls een absolute risicoreductie – ARR) voor de gekozen uitkomstmaat. De absolute risicoreductie geeft een goed beeld van de effectgrootte van de behandeling.

Is de effectgrootte van een behandeling die men vaststelt in een meta-analyse betrouwbaar?

Geloofwaardigheid of inflatie?

Pereira en Ioannidis onderzochten in 2011 de geloofwaardigheid en de omvang van het effect in meta-analyses van de Cochrane Collaboration³. De auteurs selecteerden 461 meta-analyses of systematische reviews in de Cochrane Database of Systematic Reviews van 2005. Inclusiecriteria waren minstens vier geïncludeerde studies, gebruik van dichotome uitkomstmaten en statistisch significante odds ratio's (OR) als resultaat ($p < 0,05$). In 2010 zochten ze in dezelfde databank naar herziene versies en vonden op die manier 80 updates van de oorspronkelijke 461 meta-analyses.

Op basis van deze gegevens onderzochten de auteurs de geloofwaardigheid van de oorspronkelijke 461 meta-analyses, de evolutie van de effectgrootte en de geloofwaardigheid in de updates van 2010.

Geloofwaardigheid is een concept uit de Bayesiaanse statistiek. De geloofwaardigheid is de a posteriori kans op een echt positief effect gebaseerd op een complexe berekening van een Bayesiaanse factor die gebaseerd is op de resultaten van een meta-analyse. Uitgaande van die analyse zou 63 tot 84% van de 461 meta-analyses volgens verschillende vooraf gekozen hypothesen, een terecht positief resultaat hebben terwijl 16 tot 37% ondanks een statistisch significant effect vals positief is. Wanneer men in 2010 kijkt naar de effectgrootte bij de updates, dan stelt men een significante daling vast. De gemiddelde daling is de oorspronkelijke OR $\times 0,85$ (IQR van 0,66 tot 1,06). De wijziging in effectgrootte was omgekeerd gecorreleerd aan

het gewicht van de meta-analyses, het aantal geïncludeerde studies, het aantal patiënten en het aantal gerapporteerde gebeurtenissen. Voor de meta-analyses met een groter gewicht ($N=40$) bedroeg de mediane verandering van de OR in de herziene meta-analyses (2010) $\times 0,88$ ten opzichte van de oorspronkelijke meta-analyses en voor de meta-analyses met lager gewicht was dit $\times 0,65$. Dat is opnieuw een illustratie van inflatie van de effectgrootte bij kleine steekproeven. De auteurs verklaren deze vaststelling o.a. door het fenomeen van 'de vloek van de winnaar' (het winners' curse). Dat fenomeen kan zich voordoen als men de resultaten selecteert op basis van hun superioriteit ten opzichte van een significantiedrempel en men tegelijkertijd poogt een schatting te maken van de effectgrootte. Mathematisch gezien zal dit doorgaans leiden tot inflatie van de resultaten⁴.

Ook andere vormen van bias kunnen inflatie veroorzaken: vroegtijdige resultaten zijn soms meer belovend dan achteraf blijkt, eerdere studies hebben soms meer beperkingen op het methodologische vlak, in eerdere studies zijn soms meer hoogrisicopopulaties opgenomen met een meer uitgesproken effect van de behandeling, kleinere studies met positieve resultaten zetten aan tot het uitvoeren van grotere studies om de resultaten te bevestigen... wat niet altijd het geval is.

Om al deze redenen pleiten de auteurs voor een continue monitoring van de resultaten van meta-analyses door sequentieel nieuwe studies toe te voegen (cumulatieve meta-analyses). Het belang van deze aanpak is in 2004 aangetoond door Juni et al.⁵: de cardiovasculaire risico's van rofecoxib kwamen reeds duidelijk naar voor in een cumulatieve meta-analyse, lang voor men dit risico erkende en men rofecoxib van de markt haalde (grafieken: zie website).

In het onderzoek van Pereira en Ioannidis wees de Bayes factor op een betere geloofwaardigheid in 56 van de 80 herziene meta-analyses.

Bayesiaanse onderzoekstechnieken tonen aan dat meta-analyses vals-positieve resultaten en/of resultaten met inflatie van de effectgrootte kunnen weergeven, vooral als het over een klein aantal patiënten gaat. Regelmatige herziening is dus noodzakelijk.

Referenties

1. Chevalier P. Number Needed to Treat. *Minerva* 2009;8(1):12.
2. Chevalier P. Meta-analyses: hetzelfde onderwerp, uiteenlopende resultaten... *Minerva* 2008;7(10):156.
3. Pereira TV, Ioannidis JP. Statistically significant meta-analyses of clinical trials have modest credibility and inflated effects. *J Clin Epidemiol* 2011;64:1060-9.
4. Ioannidis JP. Why most discovered true associations are inflated. *Epidemiology* 2008;19:640-8.
5. Juni P, Nartey L, Reichenbach S, et al. Risk of cardiovascular events and rofecoxib: cumulative meta-analysis. *Lancet* 2004;364:2021-9.