

Klinische relevantie van onderzoeksuitkomsten

Wat schiet de patiënt ermee op?

drs K. van Deventer, dr H.J.E.M. Janssens

Nr 1 | 2019 (53) | Pagina 1-11 | Thema-artikel

Tags

klinische relevantie irrelevante uitkomstmaten non-inferioriteitsonderzoek

Bij onderzoek naar geneesmiddelen en medische hulpmiddelen is het voor arts, apotheker of andere zorgverleners niet altijd zomaar duidelijk of in de werkelijkheid van alledag zijn of haar patiënt wat opschiet met vaak 'fraai en glossy' gepresenteerde onderzoeksuitkomsten. Ondanks dat uitkomsten gebaseerd kunnen zijn op ingewikkelde statistieken en analyses, mag de patiënt hierover een kritisch rationele opstelling verwachten van zorgverleners die de geneesmiddelen of medische hulpmiddelen voorschrijven of afleveren. Bij veel onderzoek zijn statistische grootspraak en gebrek aan echt klinisch relevante bevindingen namelijk geen uitzondering. Dit komt bijvoorbeeld omdat ongeschikte of irrelevante uitkomstmaten worden gebruikt, grote aantallen patiënten nodig zijn om geringe effecten aan te tonen en onderzochte patiënten vaak niet overeenkomen met de gewone patiënten in de dagelijkse praktijk. Ook bij onderzoeken gepubliceerd in vooraanstaande vakbladen of aangeboden om te bepalen of geneesmiddelen op de markt toegelaten mogen worden, komt dit voor. Dit Ge-Bu-artikel probeert aan de hand van voorbeelden duidelijkheid te verschaffen over hoe van wetenschappelijke onderzoeksuitkomsten de klinische relevantie (het nut voor de patiënt) beoordeeld kan worden zonder een 'ervaren geschoolde onderzoeker' te hoeven zijn.

Ge-Bu Plaatsbepaling

- **Bij onderzoek naar geneesmiddelen of medische hulpmiddelen is het veel bij marketing gebruikte begrip 'statistische significantie' onvoldoende om te kunnen beoordelen of de patiënt werkelijk wat opschiet met gevonden behandel-effecten.**
- **Of statistisch significante onderzoeksresultaten uiteindelijk zinvol zijn voor een patiënt in apotheek of spreekkamer hangt af van de detecteerbaarheid van een behandel-effect bij onderzoeksdeelnemers (power), het bij hen kunnen vaststellen van de effectgrootte van de klinisch meest relevante uitkomstmaat en hun representativiteit voor de patiënt.**
- **In goed wetenschappelijk onderzoek naar geneesmiddelen of medische hulpmiddelen is vooraf een powerberekening gemaakt van de belangrijkste (primaire) uitkomstmaat en is vooraf bepaald welk behandel-effect voor de patiënt van alledag klinisch relevant is.**
- **Bij uitkomsten van een non-inferioriteitsonderzoek dient men zich extra af te vragen wat de patiënt in apotheek of spreekkamer eraan heeft, zeker als van een interventie alleen wordt aangetoond dat deze niet slechter is dan een bestaande.**
- **Gezondheidsautoriteiten en richtlijnontwikkelaars hebben een verantwoordelijkheid om te checken of officieel te registreren of te adviseren geneesmiddelen of medische hulpmiddelen gebaseerd zijn op klinisch relevante uitkomstmaten.**
- **Zorgverleners betrokken bij het voorschrijven of afleveren van geneesmiddelen en medische hulpmiddelen zouden zoveel mogelijk bekend moeten zijn met begrippen als statistische significantie versus klinisch relevante behandel-effecten, de number-needed-to-treat, als ook met de kwaliteit van onderzoeksuitkomstmaten (hard, surrogaat of samengesteld). Alleen dan zijn patiënten juist te informeren over (vaak rooskleurige) onderzoeksresultaten.**

- **Van zorgverleners mag worden verwacht dat zij ‘glossy’ marketing voor geneesmiddelen en medische hulpmiddelen voorzien, dat zij zich blijven concentreren op de werkelijke klinische relevantie van gepresenteerde onderzoeksuitkomsten en dat zij hiernaar handelen in een direct patiëntencontact.**

Inleiding

Bij wetenschappelijk onderzoek naar geneesmiddelen en medische hulpmiddelen is het voor arts, apotheker of andere zorgverleners niet altijd zomaar duidelijk of in de complexe werkelijkheid van de medische praktijk zijn of haar patiënt wat opschiet met gestelde onderzoeksvragen, gevonden onderzoeksuitkomsten of conclusies. Met andere woorden, de klinische relevantie van onderzoek blijft vaak ongewis voor de individuele patiënt. Dit artikel probeert meer duidelijkheid te verschaffen over hoe van wetenschappelijke onderzoeksuitkomsten de klinische relevantie (het nut voor de patiënt) beoordeeld kan worden zonder een ‘ervaren geschoolde onderzoeker’ te hoeven zijn. We proberen dit op een compacte manier te doen, met behulp van een aantal onderzoeksmethodologische begrippen die gebruikt worden om onderzoeksuitkomsten te kunnen presenteren of interpreteren. De begrippen zullen kort worden uitgelegd aan de hand van voorbeelden van gepubliceerde wetenschappelijke onderzoeken, zonder verder uitgebreid aandacht te besteden aan statistische methoden en analyses. Ervaren onderzoekers zullen mogelijk details of onderdelen missen en bepaalde bewoordingen als incompleet of te beperkt beschouwen. Het artikel beoogt kennisvergroting voor iedereen die dagelijks als zorgverlener direct betrokken is bij het voorschrijven en afleveren van geneesmiddelen en medische hulpmiddelen en daarover (vaak ‘veel belovende’) onderzoeksresultaten hoort of leest.

De verschillende onderzoeksmethodologische begrippen (tussen haken in onderstaande opsomming) die in dit artikel aan bod komen, zijn toegewezen aan één van de volgende vijf domeinen:

- I. Het verschil in of de gelijkwaardigheid van een te onderzoeken interventie ten opzichte van geen of een alternatieve interventie (*statistisch significant verschil, klinisch relevant verschil, superioriteit, non-inferioriteit*)
- II. De detecteerbaarheid van een effect (*power of statistische zeggingskracht, primaire uitkomstmaat, secundaire uitkomstmaat*).
- III. De kwantificeerbaarheid van een effect (*Number Needed to Treat, Number Needed to Harm*)
- IV. De representativiteit van de gebruikte uitkomstmaat voor werkzaamheid (*harde uitkomstmaat, surrogaatuitkomstmaat, samengestelde uitkomstmaat*)
- V. De representativiteit van de onderzochte patiënten of van de interventie (*inclusiecriteria, exclusiecriteria, verrijkingfase*)

Toewijzing aan één van de domeinen hoeft niet te betekenen dat het begrip niet ook (deels) van toepassing kan zijn op één van de andere domeinen.

In dit artikel wordt uitgegaan van voorbeelden van gerandomiseerd dubbelblind onderzoek, de gouden standaard om in geneesmiddelenonderzoek werkzaamheid van interventies aan te tonen, waar de patiënt wat aan heeft. Bevindingen uit onderzoeken waarbij patiënten gedurende een bepaalde tijd (cohortonderzoek) of op een bepaald moment (dwarsdoorsnedeonderzoek) met controlepersonen worden vergeleken komen niet aan de orde. In dit soort onderzoek gaat het meestal om associaties, waarmee een oorzakelijk verband hooguit aannemelijk gemaakt kan worden. Vanwege de onduidelijkheid over causaliteit is het dan moeilijk te bepalen of gevonden verbanden of verschillen, na bijvoorbeeld blootstelling aan wel of geen belasting, voor een patiënt direct van belang zouden kunnen zijn. Dit geldt zeker als het gaat over het toepassen of achterwege laten van geneesmiddelen of medische hulpmiddelen.

I. Verschil of gelijkwaardigheid van een interventie

Statistisch significant verschil

In wetenschappelijk onderzoek worden hypothesen getoetst die geformuleerd zijn op basis van een onderzoeksvraag. In dit

verband betreft het hypothesen over het wel of niet effectief zijn van een interventie met een geneesmiddel of medisch hulpmiddel. Een dergelijke hypothese moet voorafgaand aan het onderzoek duidelijk zijn met als doel deze na afloop van het onderzoek te kunnen verwerpen of aannemen. Er zijn in principe twee hypothesevormen: een nulhypothese en een alternatieve hypothese. Een nulhypothese gaat ervan uit dat een interventie geen effect heeft, bijvoorbeeld: geneesmiddel A zal de duur van griepverschijnselen niet verkorten. Een alternatieve hypothese verwacht wel effect. Dit is bijvoorbeeld het geval als bij de introductie van een nieuw diabetesgeneesmiddel X getoetst moet worden of dit nieuwe middel werkzaam is dan de standaardbehandeling met geneesmiddel Y. Ook kunnen beide hypothesevormen gebruikt worden om te toetsen of er wel of niet een negatief effect van een interventie uitgaat, bijvoorbeeld het ontstaan van bijwerkingen.

In het geval van een superioriteitsonderzoek waarbij onderzoekers een superieur effect van een interventie willen aantonen ten opzichte van een andere interventie of geen interventie (placebo), moeten zij uitgaan van de nulhypothese dat er geen superioriteit is. Deze hypothese blijft waar tot het tegendeel bewezen is. Met behulp van statistische toetsen wordt dan ook nagegaan of de nulhypothese kan worden verworpen. Dit zou niet kunnen als een gevonden verschil louter of zeer waarschijnlijk op toeval berust. De nulhypothese wordt in de praktijk verworpen als de waarschijnlijkheid dat een gevonden verschil in effectgrootte op toeval berust klein is. Deze waarschijnlijkheid op toeval wordt uitgedrukt met de p-waarde, naar het Engelse 'probability' waarvoor vaak de grenswaarde 0,05 wordt gehanteerd. Als de p-waarde kleiner is dan 0,05 wordt de nulhypothese verworpen en wordt de alternatieve hypothese, dat het middel wel een effect heeft, aangenomen. Er wordt dan gesproken van een statistisch significant effect met een p-waarde kleiner dan 0,05. De kans dat dit effect op toeval berust en dat de nulhypothese onterecht verworpen wordt is 5% of lager. Dit zou kunnen blijken als onderzoekers de interventiestudie zouden herhalen.

Als onderzoekers in een non-inferioriteitsonderzoek gelijkwaardigheid of non-inferioriteit van twee interventies willen aantonen gaan zij op een vergelijkbare manier te werk als hierboven beschreven. De nulhypothese is in dit geval echter dat de te onderzoeken interventie inferieur is aan een controlebehandeling (placebo of een bestaande behandeling). Pas als zij kunnen aantonen dat dit niet het geval is, wordt de alternatieve hypothese aangenomen. Dit houdt dan in dat op statistische gronden gelijkwaardigheid of non-inferioriteit waarschijnlijk is. Exacte gelijkwaardigheid zal nooit gevonden worden. Daarom moeten onderzoekers vooraf aangeven welk effectverschil zij in principe wel of niet nog als gelijkwaardig of als niet-minder-werkzaam beschouwen (de zgn. equivalentie- of non-inferioriteitsmarge; zie voorbeeld **A**). Het voordeel van dergelijk onderzoek is dat er over het algemeen in vergelijking met een superioriteitsonderzoek minder patiënten ingesloten hoeven worden. Bij gelijkwaardigheid van een nieuwe interventie ten opzichte van een bestaande interventie dient wel de klinisch relevante winst ervan bekend te zijn, zoals een aanzienlijke kostenbesparing, minder bijwerkingen, verbeterd gebruikersgemak of minder contra-indicaties.¹ Op een nieuwe interventie die slechts aantoonbaar niet-slechter is dan een bestaande zit geen patiënt te wachten. Waarom bijvoorbeeld een kind een NSAID geven dat niet slechter of beter werkzaam is dan paracetamol (voorbeeld **A**)?

Het vinden van een statistisch verschil tussen het effect van de ene en de andere interventie zegt nog niets over de grootte van het verschil (behandeleffect; zie domein III) en ook niet of het verschil door de patiënt als zinvol of klinisch relevant wordt ervaren (zie hieronder).

Klinisch relevant verschil

In wetenschappelijke publicaties wordt nogal eens de nadruk gelegd op statistisch significante verschillen met een bepaalde p-waarde. Maar als onderzoeksuitkomsten statistisch significant in het voordeel van een bepaalde interventie werken, betekent dit niet zonder meer dat ze voor de patiënt van belang zijn, of dat ze op termijn van invloed zullen zijn op een gunstig ziektebeloop.² Belangrijker is om na te gaan of een statistisch significant behandelingseffect of het ontbreken ervan voor de patiënt betekenisvol of klinisch relevant is. Als het bijvoorbeeld gaat om de duur van de overleving in geneesmiddelenonderzoek bij maligne aandoeningen is het de vraag of een statistisch significante verlenging van de overleving van enkele weken of maanden relevant voor de patiënt is.³ Of bij het aantal laesies bij de behandeling van acne vulgaris, is dat relevant bij een afname of niet verdwijnen van 3, 10 of pas bij meer dan 20 laesies?⁴ Of hoe klinisch relevant is 5, 10, 20 mm op een pijnschaal van 100 mm (voorbeeld **A**) of het aantal milliliters van het FEV₁ (geforceerd expiratoire volume in 1 seconde (voorbeeld **B** en **C**)?⁵ Hoe relevant is 1 kg afvallen bij een BMI van 40 of bij een BMI van 15 bij anorexiapatiënten (voorbeeld **I**)? De grens tussen een wel of niet klinisch relevant verschil is vaak, zoals ook in deze voorbeelden, in zekere mate arbitrair, maar kan op kwalitatief onderzoek naar patiëntbeleving gebaseerd worden.⁶

Statistisch significante effecten, die wel of niet gevonden worden na goed opgezet onderzoek van geneesmiddelen en medische hulpmiddelen, zijn bijna altijd van waarde voor de medische wetenschap. Voor de patiënt is het echter van belang of statistische significantie ook klinisch relevant het verschil kan maken.

Voorbeeld A: visueel analoge schaal bij analgetica

Als onderzoek naar de pijnstillende effecten van NSAID's bij kinderen al statistische verschillen laat zien met paracetamol- of placebobehandeling zijn deze zelden of nooit klinisch relevant volgens vooraf gedefinieerde criteria. ⁷ Geen pijn meer is de meest relevante uitkomst van een medicamenteuze pijnbehandeling. Bij volwassenen wordt een absolute pijnreductie van 30 mm aangegeven op een Visueel Analoge Schaal (VAS) van 100 mm of een relatieve afname van minstens 55% als klinisch relevant beschouwd op basis van validatieonderzoeken met patiënttevredenheid als referentie. ⁶⁻⁸⁻⁹ Bij non-inferioriteitsonderzoek wordt als marge voor geen klinisch relevant verschil in werkzaamheid (non-inferioriteitsmarge) 13 mm aangehouden op een VAS van 100 mm. ⁶⁻⁸⁻⁹ Bij een onderzoek naar pijnreductie van ibuprofen bij kinderen, met een trauma van nek, armen of benen met schade van weke delen of een fractuur, ten opzichte van paracetamol werd vooraf een verschil van 15 mm als klinisch relevant beschouwd. Na behandeling met ibuprofen was de pijnreductie -24 mm (-29 - -20) en met paracetamol -12 mm (-16 - -8): een statistisch verschil ten voordele van ibuprofen, dat niet klinisch relevant was. ¹⁰

Voorbeeld B: combinatiepreparaat bij COPD geen relevante verbetering FEV₁

Recent is een inhalatiegeneesmiddel beschikbaar gekomen waarin het inhalatiecorticosteroïd beclometason gecombineerd wordt met twee luchtwegverwijders, het langwerkende beta-2-sympathicomimeticum formoterol en het parasymphaticolyticum glycopyrronium (Trimbrow®). In één van de twee onderzoeken (TRILOGY) dat als uitgangspunt gebruikt werd om het middel op de markt toe te laten, werden drie primaire uitkomstmaten gebruikt, waarvan twee surrogaat, het geforceerd expiratoire volume in 1 seconde (FEV₁) voor en na inhalatie, en een wat 'hardere' uitkomstmaat, de door de patiënt gerapporteerde benauwdheid. ¹¹ Alleen patiënten met een FEV₁ minder dan 50% werden ingesloten. Na 26 en 52 weken toonden beide surrogaatuitkomstmaten statistisch significante, maar klinisch irrelevante FEV₁-verbeteringen van 63 tot 117 ml ten opzichte van de controlebehandeling (beclometason-formoterol). ⁵ De door de patiënt ervaren benauwdheid verschilde niet statistisch significant, waarmee de klinische relevantie van de drie primaire uitkomsten voor de patiënt ver te zoeken is. Voor de secundaire uitkomstmaat het gemiddeld aantal exacerbaties per jaar, die bepaald werden door de behandelaars als zij een indicatie stelden voor corticosteroïden, antibiotica of een ziekenhuisopname, was er een gering statistisch voordeel (0,1 exacerbatie/jaar). De auteurs van het onderzoek gepubliceerd in de Lancet beweren dat zij bewijs leveren voor de "klinische voordelen" voor deze triple-therapie. ¹¹

Voorbeeld C: combinatiepreparaat bij COPD minimaal minder exacerbaties

In het tweede onderzoek (TRINITY) gebruikt voor de toelating (zie ook voorbeeld B) van het nieuwe combinatie-inhalatiegeneesmiddel beclometason met formoterol en glycopyrronium (Trimbrow®) was het aantal exacerbaties de primaire uitkomstmaat. De werkzaamheid werd vergeleken met tiotropium als monotherapie. ¹² Ook nu bepaalde de behandelaar zelf of er sprake was van een exacerbatie, namelijk wanneer deze het nodig vond dat de patiënt of corticosteroïden of antibiotica moest krijgen of opgenomen moest worden in een ziekenhuis. De resultaten voor deze matig harde uitkomstmaat toonden een statistisch significant minimaal verschil in het gemiddelde aantal exacerbaties per jaar (0,5 vs. 0,6 exacerbaties) ten voordele van de combinatiebehandeling, met onduidelijkheid over de klinische relevantie ervan. FEV₁-verbetering gemeten voor inhalatie was nu een secundaire uitkomst (de FEV₁ na inhalatie werd niet vermeld). Deze bleek na 1 jaar slechts 60 ml voordeel op te leveren. Als secundaire uitkomstmaat is deze bevinding indicatief, maar niet bewijzend voor werkzaamheid (zie domein III). Samen met het minimale voordeel van de primaire uitkomstmaat blijft er dan nauwelijks winst voor de patiënt over die al tiotropium gebruikt. Het minimale bewijs dat er misschien is, geldt niet voor patiënten met een milde COPD (zie domein V). Ook in deze studie werden namelijk alleen patiënten met een FEV₁ minder dan 50% ingesloten. ¹² Ook de auteurs van dit onderzoek beweren de klinische winst van deze triple-therapie aan te tonen, zoals ook te lezen valt in vele advertenties in medische vakbladen, o.a. gericht op huisartsen. Huisartsen zullen echter zelden patiënten tegenkomen die worden gerepresenteerd (zie domein V) door de deelnemers van deze onderzoeken: COPD-patiënten met een FEV₁ minder dan 50% die niet reageren op tiotropium als monotherapie.

II. Detecteerbaarheid van een effect

Bij het ontwerp van een geneesmiddelenonderzoek moet van tevoren met een zogenoemde powerberekening bepaald zijn hoeveel patiënten met het onderzoek mee moeten doen om bij hen een statistisch significant effect van een interventie te kunnen detecteren. Dit aantal moet altijd terug te vinden zijn bij publicatie van het onderzoek. Dat geldt ook voor de uitkomstmaat waarop de powerberekening van toepassing is. Meestal gaat het hierbij om één uitkomstmaat: de primaire uitkomstmaat. Deze meestal belangrijkste uitkomstmaat wordt verwoord in de vooraf geformuleerde te bewijzen of te verwerpen nulhypothese, of alternatieve hypothese. De powerberekening wordt niet gemaakt voor secundaire uitkomstmaten, uitkomstmaten die volgens de onderzoekers ook van belang kunnen zijn. De patiënt heeft het meest aan een klinisch relevante primaire uitkomstmaat die maximaal betrouwbaar gedetecteerd kan worden.

Bewijs voor wel of geen effect volgens een bepaalde uitkomstmaat is nooit 100%, ook als een onderzoek volgens de

powerberekening voldoende patiënten lijkt te hebben ingesloten. De berekening kan onjuist of ongeschikt zijn geweest, bijvoorbeeld als toch te weinig patiënten werden ingesloten om een effect te detecteren. Maar als er (veel) meer patiënten nodig blijken te zijn dan oorspronkelijk berekend werd, wordt de kans dat een individuele patiënt er wat mee opschiet kleiner, omdat het zogenoemde Number Needed to Treat (NNT) toeneemt (zie domein III; voorbeeld J). Het kan ook zo zijn dat het aantal patiënten voldoende is om non-inferioriteit van een interventie aan te tonen, maar (vooralsnog) onvoldoende om in dezelfde studie superioriteit ervan te detecteren, terwijl dat het impliciete doel van de onderzoekers is (voorbeeld J).

In publicaties van geneesmiddelenonderzoek wordt nogal eens afgeweken van het oorspronkelijke onderzoeksprotocol bij het presenteren van de resultaten. In het bijzonder geldt dit voor secundaire uitkomstmaten.¹³ Selectief of disproportioneel veel aandacht besteden aan statistisch significante secundaire uitkomstmaten, omdat het effect op de primaire uitkomstmaat bijvoorbeeld tegenvalt of niet bewezen wordt, zoals bij insluiten van te weinig patiënten, kan marketingtechnisch gezien interessant zijn zonder dat de patiënt er wat aan heeft (voorbeeld B).

Gebaseerd op de powerberekening voor de primaire uitkomstmaat en niet op een eigen berekening is de statistische zeggingskracht van bevindingen voor secundaire uitkomstmaten gering. Klinisch relevante voor- of nadelen van secundaire uitkomstmaten zijn zo niet detecteerbaar en blijven uiteindelijk onbewezen. Ze zijn hypothese genererend en niet effect bewijzend. Secundaire onderzoeksresultaten kunnen aanleiding geven tot of richtinggevend zijn voor een vervolgonderzoek. In het geval van bijwerkingen, die in geneesmiddelenonderzoek bijna altijd bestudeerd worden als secundaire uitkomstmaten, zijn bevindingen erover alvast waarschuwend en mogelijk van belang voor de directe patiëntzorg (voorbeeld F).

III. Kwantificeerbaarheid van een effect

Een bij onderzoek gevonden statistisch significant (klein of groot) behandel-effect zegt alleen iets over de onwaarschijnlijke kans dat het effect op toeval berust (zie domein I). Het zegt niets over de grootte van dat effect. Voor de kwantificeerbaarheid van het effect is, bijvoorbeeld in het geval dat geneesmiddel X vergeleken wordt met geneesmiddel Y, informatie nodig over 'het risico' op genezing bij gebruik van middel X in vergelijking met middel Y. Maten voor dit 'genesrisico' zijn odds-ratio, hazard-ratio, het relatieve risico of het absolute risico, gespecificeerd met zogenoemde 95% betrouwbaarheidsintervallen. Deze meetgrootheden lenen zich niet makkelijk om klinische relevantie in bijvoorbeeld spreekkamer of apotheek aan patiënten uit te leggen. Het 'Number Needed to Treat' (NNT) en het 'Number Needed to Harm' (NNH), afgeleid van het absolute risico, zijn maten waarmee artsen, apothekers of andere zorgverleners en patiënten waarschijnlijk gemakkelijker uit de voeten kunnen om de grootte van een behandel-effect te kwantificeren. Het NNT is het aantal patiënten dat moet worden behandeld gedurende een bepaalde periode om bij één patiënt een gunstige onderzoeksuitkomst te verkrijgen of een ongunstige gebeurtenis te voorkomen. Het NNH is daarentegen het aantal patiënten dat behandeld wordt om bij één patiënt een ongunstige uitkomst of schade te veroorzaken. Het zijn met andere woorden maten voor respectievelijk het aantal patiënten dat 'de pech' heeft vergeefs te worden behandeld omdat de interventie bij hen geen verschil uitmaakt en voor het aantal patiënten dat 'het geluk' heeft geen nadeel of bijwerkingen opgelopen te hebben van de behandeling.

De patiënt schiet het meest op met een zo laag mogelijk NNT en een zo hoog mogelijk NNH (voorbeeld D, E en J). Hoe laag of hoe hoog is arbitrair, maar dient besproken te worden met de patiënt en zou altijd punt van aandacht moeten zijn voor gezondheidsautoriteiten en richtlijnontwerpers.¹⁴ De NNT en NNH kunnen vrij eenvoudig berekend worden. (zie begrippenlijst <https://www.ge-bu.nl/artikel/placebos-en-placebo-effecten-ij>). Voor het berekenen van NNT en NNH zijn makkelijk te gebruiken calculators op internet beschikbaar.

Voorbeeld D: hoge NNT bij rosuvastatine

In het JUPITER-onderzoek naar de werkzaamheid van rosuvastatine voor primaire preventie bij personen zonder hyperlipidemie werd de samengestelde uitkomstmaat (zie domein IV) bestaande uit myocardinfarct, CVA of cardiovasculaire sterfte gebruikt om de werkzaamheid te kwantificeren.¹⁵ In de rosuvastatinegroep (8.901 patiënten) waren 83 incidenten en in de placebogroep (8.901 patiënten) 157. Het absolute risico daalde van 1,76% (157/8901) in de placebogroep naar 0,93% (83/8901) met rosuvastatine, een statistisch significant verschil. Omgerekend kwam dit neer op een NNT van 120. Dat betekent dat 120 deelnemers gemiddeld 1,9 jaar moesten worden behandeld om één ernstig incident te voorkomen. De klinische relevantie van deze uitkomst is twijfelachtig te noemen, zoals eerder in het Ge-Bu geconcludeerd. ¹⁶ Binnen een termijn van 2 jaar is voor een individuele patiënt de kans namelijk erg groot dat deze voor niets behandeld wordt. Of de patiënt er nadien nog wat aan heeft, blijft met dit soort

onderzoek ongewis.

Voorbeeld E: hoge NNT versus lage NNH bij de DOAC rivaroxaban

In de COMPASS-studie werd de werkzaamheid van de DOAC rivaroxaban in combinatie met acetylsalicylzuur vergeleken met alleen acetylsalicylzuur bij patiënten met atherosclerotisch vaatlijden.¹⁷ De primaire samengestelde uitkomstmaat (zie domein IV) was overlijden ten gevolge van een cardiovasculaire aandoening of het krijgen van een beroerte (ischemisch, hemorragisch of onbekend) of een myocardinfarct. Het onderzoek werd recent in het Ge-Bu kritisch tegen het licht gehouden.¹⁸ Er kon een NNT (voorkomen van één van de uitkomsten uit de samengestelde uitkomstmaat) en een NNH (bloedingen) uitgerekend worden. Deze waren beiden ongeveer 80. Dat zou betekenen dat zo'n 80 patiënten gedurende de onderzoeksperiode niets aan de behandeling hebben gehad, maar dat onder hen zeker één patiënt een bloeding opliep. Het ontstaan van een hemorragische beroerte (vaker voorkomend bij patiënten behandeld met rivaroxaban) telde merkwaardigerwijs wel mee bij het berekenen van het NNT en niet bij het berekenen van het NNH, dat daarmee nog lager had kunnen uitkomen. Het relatief hoge NNT samen met een relatief laag NNH lijken in dit geval in klinisch relevant opzicht weinig goeds op te leveren voor de atherosclerotische patiënt die al acetylsalicylzuur gebruikt, de enorme marketing voor rivaroxaban voor allerlei (nieuwe) indicaties ten spijt.

IV. Representativiteit uitkomstmaat

Harde of surrogaatuitkomstmaat

Harde uitkomstmaten zijn uitkomstmaten die van direct en merkbaar belang zijn voor de patiënt, zoals mortaliteit, cardiovasculaire aandoeningen of relevante overlevingstijd bij maligne aandoeningen. Het kunnen ook door de patiënt gerapporteerde relevante veranderingen zijn, zoals meer, minder of geen pijn (voorbeeld **F**). Een surrogaatuitkomstmaat is geen echte of vanzelfsprekende maat voor een klinisch effect. Surrogaatuitkomstmaten, waarbij de patiënt weinig hoeft te merken van veranderingen ervan (voorbeeld **F** en **G**), hebben veelal op pathofysiologische gronden slechts een indirect verband met harde uitkomstmaten. Ze kunnen de onderzoeker, arts of patiënt laten zien of een interventie wel of niet in de gewenste richting gaat en geven hooguit een indirect bewijs of de patiënt er uiteindelijk wat mee opschiet. Een surrogaatuitkomstmaat is vaak eenvoudig te meten, bijvoorbeeld de bloeddruk, het geglycolyseerde hemoglobine (HbA_{1c})-gehalte (voorbeeld **G**), het geforceerde expiratoire volume in 1 seconde (FEV₁, voorbeeld **B**) of het seruminezuur (voorbeeld **F**).²

Onderzoek naar geneesmiddelen of medische hulpmiddelen waarbij men gebruik maakt van surrogaatuitkomstmaten is sneller en goedkoper in vergelijking met onderzoek op harde uitkomstmaten.² Ze worden gebruikt als maat voor het uiteindelijke effect van de interventie op de gewenste harde uitkomstmaten. Wanneer er echter geen directe relatie is aangetoond tussen een surrogaatuitkomstmaat en een klinisch relevante harde uitkomstmaat, is de bruikbaarheid van zo'n surrogaatuitkomstmaat beperkt (voorbeeld **H**). Daarnaast dienen onderzoekers rekening te houden met het feit dat behalve de interventie ook fysiologische of andere factoren surrogaatuitkomstmaten kunnen beïnvloeden. Ook kunnen twee verschillende interventies met effect op een zelfde surrogaatuitkomstmaat verschillende gevolgen hebben op een harde uitkomstmaat²

De verleiding blijkt groot om effecten van weinig harde uitkomstmaten en van surrogaatuitkomstmaten in wetenschappelijke publicaties als 'hard' te presenteren (voorbeeld **B**, **C** en **G**) of te benadrukken (voorbeeld **F** en **I**). Dit is voor de patiënt klinisch irrelevant, maar marketingtechnisch voor een geneesmiddelenfabrikant interessant. Voor de patiënt is het doorgaans het relevantst als de arts een behandelkeuze maakt op basis van informatie op de meest harde uitkomstmaten.

Voorbeeld F: seruminezuurverlaging en jichtaanvallen

Allopurinol of andere urinezuurverlagende geneesmiddelen worden veelvuldig toegepast om bij jicht (recidief)aanvallen te voorkomen. Het minderen of liever nog stoppen van aanvallen is de belangrijkste uitkomstmaat vanuit het perspectief van de patiënt (klinisch relevantst).¹⁹ In de drie non-inferioriteitsonderzoeken ten behoeve van markttoelating van het seruminezuurverlagende middel febuxostat werd in slechts één onderzoek de aanvalsfrequentie onderzocht.²⁰ Als secundaire uitkomstmaat werd in deze studie de proportie van patiënten bepaald die om een behandeling verzochten voor minimaal één aanval.²¹ Er bleek geen statistisch significant verschil tussen febuxostat en allopurinol. De primaire uitkomstmaat in de drie onderzoeken was een surrogaatuitkomstmaat: de proporties patiënten met een seruminezuurverlaging tot minimaal 0.36 mmol/l. Bij het gebruik van de seruminezuurconcentratie als uitkomstmaat gaat men uit van een pathofysiologisch concept, waarbij verondersteld wordt dat verlaging van de urinezuurconcentratie zal leiden tot minder vorming van urinezuurkristallen. Deze kristallen zijn inflammatoir actief tijdens een jichtaanval. Er zijn echter volgens systematische literatuuronderzoeken geen gerandomiseerde dubbelblinde onderzoeken te vinden die aantonen dat statistisch

significante verlaging van de serumurinezuurconcentratie met geneesmiddelen leidt tot minder of geheel uitblijven van jichtaanvallen. **22-23** Desondanks wordt door onderzoekers en richtlijnontwikkelaars zeer veel waarde gehecht aan de noodzaak van verlaging van de serumurinezuurconcentratie, zelfs tot onder het fysiologische niveau. **24** Logisch dat hierover dan controverse ontstaat, waarbij gepleit wordt voor onderzoek met hard bewijs. **25** Opmerkelijk is het overigens dat in de drie febuxostat-registratieonderzoeken allopurinol, met een maximale dosis van 300 mg per dag, niet optimaal gedoseerd kon worden. Richtlijnen adviseren maximale dagdoseringen van 600 tot 900 mg bij uitblijven van voldoende resultaat. Patiënten in de dagelijkse praktijk worden dan niet vertegenwoordigd door de onderzoeksdeelnemers (domein V). De gevonden non-inferioriteit van febuxostat ten opzichte van allopurinol voor de surrogaatuitkomstmaat serumurinezuurverlaging kon dan wel eens overschat zijn. Het bewijs dat een jichtpatiënt die febuxostat gebruikt uiteindelijk tevreden wordt over minder recidiveren van jichtaanvallen blijft daarmee zeer ongewis.

Voorbeeld G: HbA_{1c} onvoldoende als uitkomstmaat van sitagliptine

Ruim 10 jaar geleden kwam de dipeptidylpeptidase(DPP)-4-remmer sitagliptine als nieuw oraal bloedglucoseverlagend middel beschikbaar. In een Ge-Bu-overzichtsartikel uit 2010 werden drie gerandomiseerde onderzoeken samengevat die ten tijde van de registratie waren gepubliceerd. **26** Deze onderzoeken waren primair opgezet om een verschil in daling van het HbA_{1c}-gehalte aan te tonen. **27-28-29** De onderzoeksuitkomsten waren alle positief, namelijk een statisch significante daling van HbA_{1c}-gehalte in het voordeel van sitagliptine. Voor de registratieautoriteiten waren de bescheiden dalingen van het HbA_{1c}-gehalte voldoende voor markttoelating. Het Ge-Bu gaf destijds een pilwaardering '+/-' af, "vanwege gebrek aan onderzoek op harde eindpunten". **26** In een recent Ge-Bu-overzichtsartikel (10 jaar later) met aandacht voor de onderzoeken die in loop van het laatste jaar werden gepubliceerd, werd geconcludeerd dat er eigenlijk geen plaats is voor dit middel bij de behandeling van diabetes mellitus. **30** De pilwaardering werd '-' want "op cardiovasculaire uitkomstmaten werkt sitagliptine niet beter dan placebo en de werkzaamheid op microvasculaire uitkomstmaten is nog onbewezen", naast ook nog eens een ongunstig bijwerkingsprofiel. Een HbA_{1c}-daling bij gebruik van sitagliptine blijkt in dit geval, als surrogaat, voor de patiënt niets op te leveren.

Voorbeeld H: pulmonale arteriële hypertensie en de zes-minutenlooptest

Epoprostenol was het eerste geneesmiddel dat werd geregistreerd voor de behandeling van pulmonale arteriële hypertensie (PAH). De Amerikaanse registratieautoriteit Food and Drug Administration (FDA) stelde als eis dat de primaire uitkomstmaat een maat voor de inspanningstolerantie zou zijn of dat primair de overleving zou worden onderzocht. **31** Omdat voor het onderzoeken van de overleving meer patiënten nodig zouden zijn, kozen onderzoekers voor een maat om de inspanningstolerantie te kwantificeren met behulp van een zes-minutenlooptest. Dit is een test waarbij gedurende 6 minuten de afgelegde loopafstand in meters wordt gemeten. Het was niet bekend of deze zes-minuten-loopafstand wel een harde uitkomstmaat was en, zo ja, wat de klinische relevantie ervan was. Er ontstond achteraf discussie over de uiteindelijke werkzaamheid van epoprostenol, omdat de veranderingen in de zes-minutenlooptest slechts voor een gering deel het klinisch relevante behandelingseffect bleken te verklaren. **32-33** Het lijkt daarmee voor de patiënt slechts een matig bruikbare surrogaatuitkomstmaat. **34**

Voorbeeld I: relevantie gewichtsafname door afslankmiddelen

Gewichtsafname lijkt een bruikbare uitkomstmaat voor onderzoek naar afslankmiddelen bij mensen met overgewicht. In het verleden zijn diverse afslankmiddelen op de markt gekomen en ook weer uit de handel genomen. **35** De werkzaamheid van één van deze middelen, sibutramine, werd in een meta-analyse onderzocht ten opzichte van placebo. Het betrof de resultaten van tien onderzoeken met 2.347 patiënten, die gedurende minimaal 1 jaar een BMI van > 32 kg/m² hadden. De gewichtsafname was met 4,2 kg statistisch significant groter dan bij gebruik van een placebo. **36** Absoluut gezien lijkt deze geringe afname echter nauwelijks (klinisch) relevant, noch in cosmetisch of fysiek opzicht, noch wat betreft het verlagen van het hogere risico op harde uitkomsten als cardiovasculaire aandoeningen, bepaalde vormen van kanker en overlijden. Op deze harde uitkomstmaten, die men bij een verhoogd BMI kan verwachten, zijn afslankmiddelen nooit onderzocht, waardoor er geen bewezen klinisch relevant nut van afslankmiddelen voor de patiënt overblijft. **37**

Samengestelde uitkomstmaten

Onderzoekers maken nogal eens gebruik van samengestelde uitkomstmaten, bijvoorbeeld omdat meerdere eindpunten van belang kunnen zijn, of als er meerdere uitkomstmaten nodig zijn om tot een adequaat oordeel te komen over de voor- en nadelen van een behandeling. Toepassing van samengestelde uitkomstmaten kan ook, onderzoekstechnisch gezien, een praktische reden hebben wanneer een uitkomstmaat, zoals sterfte, onvoldoende voorkomt. Door sterfte te combineren met uitkomstmaten zoals myocardinfarct en CVA, is het makkelijker om statistische significantie aan te tonen voor het overall-effect van deze combinatie dan voor elk van de componenten. Voordeel van het gebruik van een samengestelde uitkomstmaat is dat minder onderzoekstijd nodig is en minder patiënten ingesloten hoeven te worden, wat aanzienlijk scheelt in de kosten. Het risico bestaat wel dat de effectgrootte van een samengestelde uitkomstmaat verschilt van de effectgrootte van de afzonderlijke componenten. Het is niet uitzonderlijk dat een interventie wel een statistisch significant effect heeft op de samengestelde uitkomstmaat, maar niet op elke afzonderlijke uitkomstmaat of zelfs op geen enkele van de afzonderlijke uitkomstmaten (voorbeeld J). Bovendien is niet elke component van een samengestelde uitkomstmaat voor elke patiënt even relevant. Een kleine kans om opgenomen te

worden op verdenking van een acuut coronair syndroom of een kans op een goed te behandelen myocardinfarct zijn iets anders dan een levenslange halfzijdige verlamming na een CVA.

Harde uitkomstmaten, zoals sterfte en myocardinfarct, worden vaak gecombineerd met objectief minder harde uitkomstmaten, zoals een ongeplande revascularisatieprocedure of angina pectoris. Dit maakt gevonden resultaten moeilijker interpreteerbaar, omdat minder relevante uitkomstmaten meestal meer bijdragen aan het overall-effect dan de hardere (voorbeeld **K**). Sterfte blijkt in cardiologisch onderzoek bijvoorbeeld de meest (98%) gebruikte afzonderlijke uitkomstmaat in samengestelde uitkomsten, maar draagt relatief weinig bij aan de schatting van het overall-effect, namelijk slechts 27%.² Al met al is een overall-effect van een samengestelde uitkomstmaat als getal moeilijk te begrijpen voor arts en patiënt. Het effect wordt gemakkelijk overschat, iets waar de sponsorende industrie wel iets mee kan, maar de patiënt vaak niet.

Voorbeeld J: samengestelde of aparte harde uitkomstmaten van canagliflozine

De toelating op de Amerikaanse markt van de SGLT-2-remmer canagliflozine voor de behandeling van diabetes type 2, werd gebaseerd op een non-inferioriteitsonderzoek naar de cardiovasculaire veiligheid van dit geneesmiddel, wanneer het toegevoegd werd aan de bestaande behandeling bij al cardiovasculair belaste diabetespatiënten.³⁸ Op het moment van een tussentijds mogelijk veiligheidsrisico werd de onderzoeksopzet aangepast, wat de mogelijkheid bood om bij bewezen non-inferioriteit (niet onveilig dan de bestaande behandeling) ook superioriteit (veilig dan bestaande behandeling) aan te kunnen tonen.³⁹ De samengestelde uitkomstmaat mortaliteit door cardiovasculaire oorzaken, niet-fataal hartinfarct, niet-fatale beroerte (ischemisch, hemorragisch en onbekend, mits neurologische uitval > 24 uur) kwam statistisch significant minder vaak voor (hazard ratio 0.86) bij toevoeging van canagliflozine. Maar voor geen van de afzonderlijke uitkomstmaten was superioriteit statistisch aan te tonen. Daarmee wordt het niet eenvoudig de individuele diabetespatiënt bekend met een cardiovasculaire aandoening te overtuigen van de klinische relevantie van het geneesmiddel, zeker als deze ook nog het hoge NNT van circa 200 uitgelegd krijgt.⁴⁰ De sterke aanwijzing voor een hoger risico op bijwerkingen (bijvoorbeeld amputaties), die als secundaire uitkomstmaten onderzocht werden,⁴⁰ maakt een en ander in de spreekkamer nog moeilijker, zo niet onmogelijk.

Voorbeeld K: zachte en harde uitkomstmaten van de DOAC apixaban

In de gerandomiseerde onderzoeken naar de werkzaamheid van de directwerkende orale anticoagulantia zijn samengestelde uitkomstmaten gebruikt. Deze bevatten in voorkomende gevallen niet alleen harde uitkomstmaten, maar ook minder harde, zoals asymptomatische trombose, waar de patiënt geen last van heeft. Dit gold bijvoorbeeld voor drie onderzoeken naar de werkzaamheid van apixaban bij orthopedische ingrepen op basis waarvan het middel op de markt werd toegelaten. In deze onderzoeken werd apixaban vergeleken met enoxaparine.⁴¹⁻⁴²⁻⁴³ De primaire uitkomstmaat in deze gerandomiseerde dubbelblinde non-inferioriteitsonderzoeken was samengesteld uit asymptomatische trombose, symptomatische diepveneuze trombose, longembolie en overlijden ongeacht de oorzaak. Het combineren van uitkomstmaten kan leiden tot een te positieve schatting van het overall-effect, zeker als een vaker voorkomende minder harde maat één van de uitkomstmaten is.⁴⁴ Het blijft dan onduidelijk of en in welk opzicht de vooruitzichten van de patiënt daadwerkelijk verbeteren.

V. Representativiteit onderzochte patiënten

Op het moment dat geneesmiddelen op de markt worden toegelaten, zijn ze onderzocht in gespecificeerde en relatief homogene patiëntpopulaties. Door de vaak strikte in- en uitsluitcriteria komen ingesloten patiënten vaak niet overeen met de patiënten in de dagelijkse praktijk. De zogenaamde externe validiteit van de onderzoeksuitkomsten wordt daarmee beperkt, want het is maar de vraag in hoeverre het geneesmiddel werkzaam en veilig is als het wordt voorgeschreven aan patiënten buiten het onderzoek. Wanneer kinderen, ouderen (voorbeeld **L**), patiënten met veel co-morbiditeit, of patiënten met een ernstige vorm van de te onderzoeken aandoening worden uitgesloten, is het lastig om de onderzoeksuitkomsten te generaliseren naar alle patiënten in de dagelijkse praktijk (voorbeeld **G**). Dit geldt ook als bijvoorbeeld alleen patiënten ingesloten worden die in een aan het onderzoek voorafgaande zogenoemde verrijkingfase bewezen hebben onderzoeksmedicatie goed te verdragen (voorbeeld **H**, **M** en **N**), in hoge mate therapietrouw blijken te zijn (voorbeeld **I**), een ruimere behandelindicatie blijken te hebben dan dat officieel volgens de markttoelatingseisen de bedoeling is (voorbeeld **H** en **M**) of als controlepersoon voor de bestaande behandeling niet behandeld werden volgens de gangbare standaard of met een onvoldoende dosering (voorbeeld **F**).

Uitgebreide in- en uitsluitcriteria niet overeenkomend met de dagelijkse medische praktijk komen bij onderzoek naar geneesmiddelen en medische hulpmiddelen vaak voor.⁴⁵ Daarom is het van groot belang dat artsen, apothekers en andere zorgverleners, die praktische conclusies verbinden aan een onderzoeksbevinding, zich steeds moeten afvragen of de onderzochte patiënten overeenkomen met hun eigen patiënt (voorbeeld **C**). Anders valt nauwelijks te beoordelen of hun patiënt er

werkelijk wat mee opschiet.

Voorbeeld L: registratieonderzoek dabigatran bij niet-representatieve patiënten

Toen voor het antistollingsmiddel dabigatran een handelsvergunning werd aangevraagd, legde de fabrikant vier gerandomiseerde onderzoeken voor aan de autoriteiten, die beslisten over marktoelating. **46-47-48-49** De auteurs van deze vier onderzoeken concluderen dat dabigatran niet minder werkzaam is dan enoxaparine en geen hoger bloedingsrisico heeft bij patiënten die een gewrichtsvervangende operatie ondergaan. Maar de generaliseerbaarheid van deze uitkomsten wordt beperkt door de gebrekkige representativiteit van de onderzochte patiënten. **50** De onderzoekers hanteerden namelijk zeer strikte uitsluitcriteria, zoals een verhoogde bloedingsneiging, verminderde nierfunctie en geneesmiddelen die het bloedingsrisico vergroten. De kans is echter groot dat een patiënt, bij wie een gewrichtsvervangende operatie in de praktijk nodig is, 'behept is met deze exclusiecriteria', alleen al op grond van een oudere leeftijd. Het is dan twijfelachtig of de patiënt van alledag wat opschiet met onderzoek waarin niet-representatieve ideale proefpersonen meedoen.

Voorbeeld M: Guanfacine bij ADHD: geen bewijs voor werkzaamheid bij de doelgroep

In de onderzoeken op basis waarvan guanfacine, een selectieve α_2A -adrenerge receptoragonist, geregistreerd werd voor de behandeling van ADHD werd een statistisch significante en een volgens vooraf gedefinieerde criteria klinisch relevante werkzaamheid vastgesteld ten opzichte van een placebobehandeling. **51-52-53** Voor deze onderzoeken werden echter patiënten ingesloten die niet voldeden aan de in Nederland vereiste indicatiecriteria, te weten het falen van de behandeling met centraal stimulerende middelen of het bestaan van een contra-indicatie voor deze middelen. Ook is onduidelijk of een (mislukte) niet-medicamenteuze behandeling, zoals ook in Nederland aangewezen is, vooraf is gegaan aan de onderzoeksdeelname of onderdeel uitmaakte van de behandeling. Daarmee zijn de onderzochte patiënten niet representatief voor de ADHD-patiënt die in Nederland voor guanfacine in aanmerking komt. Er is nog een andere reden om aan de representativiteit te twijfelen. In het onderzoek naar de bijwerkingen werden namelijk alleen patiënten ingesloten die een voorafgaand onderzoek naar de werkzaamheid van guanfacine volledig hadden afgerond. De patiënten die al eerder bijwerkingen ondervonden en daarom met de onderzoeken waren gestopt werden uitgesloten (een verrijkingfase). Hierdoor bestaat er mogelijk een onderschatting van het werkelijke aantal bijwerkingen. Het blijft al bij al erg onzeker of een patiënt met ADHD werkelijk wat zal hebben aan dit nieuwe middel, zoals het Ge-Bu eerder concludeerde. **54**

Voorbeeld N: verrijkingfasen in gerandomiseerd onderzoek naar saxagliptine

In gerandomiseerd onderzoek naar de werkzaamheid van saxagliptine bij de behandeling van diabetes mellitus maakten onderzoekers gebruik van zogenoemde verrijkingfasen. **55-56-57** Dit waren één of enkele weken durende enkelblinde placebo 'lead in'-fasen, waarin de therapietrouw van patiënten werd onderzocht. Alleen patiënten met een hoge mate van therapietrouw (>80%) mochten verder aan het onderzoek deelnemen. Logischerwijs geeft dit een overschatting van het behandelingseffect, waardoor de generaliseerbaarheid van de resultaten beperkt is. In de dagelijkse praktijk is lang niet iedere patiënt perfect of vergelijkbaar met een goed geïnstrueerde onderzoeksdeelnemer als het om therapietrouw gaat. **58**

Beschouwing

De centrale vraag in dit artikel luidt: wat schiet de individuele patiënt op met in wetenschappelijk onderzoek verkregen uitkomsten over de behandelingseffecten van geneesmiddelen en medische hulpmiddelen? Onafhankelijk van het gegeven of patiënten zelf deze vraag stellen, zal deze in de praktijk van alledag beantwoord worden, of moeten worden beantwoord door artsen, apothekers en andere zorgverleners die de geneesmiddelen of medische hulpmiddelen voorschrijven of afleveren. Dit artikel beoogt hen hierbij te helpen aan de hand van het bespreken van een aantal begrippen uit de wetenschappelijke onderzoeksmethodologie.

Een statistisch significant verschil vinden in behandelingseffecten tussen twee interventies op een onderzoeksuitkomstmaat zegt nog weinig voor de patiënt, als deze niet weet of het over een klinisch relevant effect gaat. Daarvoor is meer informatie nodig, met name over een valide drempelwaarde waarbij een effect werkelijk door patiënten als verbetering, verslechtering of als geen-effect wordt ervaren (domein I). Of een behandelingseffect werkelijk gedetecteerd had kunnen worden in een medisch-wetenschappelijk onderzoek is de moeite van het overwegen waard (domein II). Om klinische relevantie nog beter in beeld te krijgen moet de patiënt geïnformeerd kunnen worden over de uiteindelijke grootte van een bij onderzoek gevonden uitkomstmaat (domein III). Uitkomstmaten moeten zo 'hard' mogelijk zijn en, in klinisch relevant opzicht, duidelijk en merkbaar voor de patiënt. Het moet niet zomaar een maat zijn die een harde uitkomstmaat vertegenwoordigt op basis van een mogelijke of ogenschijnlijke associatie ermee (domein IV). Tot slot moet de patiënt zich kunnen herkennen in de deelnemers aan een onderzoek naar het behandelingseffect van bijvoorbeeld een geneesmiddel dat deze moet gaan gebruiken (domein V).

Zorgverleners in het veld hebben een verantwoordelijkheid bij het doorgronden en verduidelijken van de klinische relevantie van onderzoeksuitkomsten. Maar zij zouden hierover ook maximale duidelijkheid mogen verwachten van gezondheidszorgautoriteiten

en auteurs van behandelrichtlijnen. Dat dit geen vanzelfsprekendheid is bij marktintroductie van bijvoorbeeld nieuwe geneesmiddelen blijkt uit de in dit artikel gepresenteerde voorbeelden.

De enorme lobby en invloed van de producenten van geneesmiddelen en medische hulpmiddelen is evident. Winstvergroting is tenslotte voor hen een primair uitgangspunt. Statistische grootspraak en gebrek aan klinisch relevante uitkomstmaten zijn bij hen daarom geen uitzondering, zoals blijkt uit dit artikel. Zorgverleners in het veld hebben hierbij een extra verantwoordelijkheid. Van hen mag verwacht worden dat zij de 'glossy' marketing doorzien, dat ze zich blijven concentreren op de werkelijke klinische relevantie van 'fraai' gepresenteerde onderzoeksuitkomsten en dat zij hiernaar handelen in een direct patiëntencontact. Een recept is snel uitgeschreven of afgeleverd op basis van een ronkende reclame of gelikte presentatie. Nadenken over of de patiënt hier werkelijk wat mee opschiet vergt meer, namelijk een kritisch rationele opstelling. De patiënt mag dit verwachten van een academisch opgeleide zorgverlener, zonder dat deze daarvoor een ervaren wetenschappelijke onderzoeker hoeft te zijn. Het Ge-Bu stelt zich al meer dan 50 jaar ten doel artsen, apothekers, arts ondersteunende zorgverleners, specialistisch verpleegkundigen en allen, die daarvoor in opleiding zijn, daarbij behulpzaam te zijn door middel van toegankelijke artikelen over rationele farmacotherapie en het rationeel gebruik van medische hulpmiddelen.

Literatuurreferenties

1. Dekkers OM. Het non-inferioriteitsonderzoek. [Gebu. 2015;49\(3\):27-34.](#)
2. Anoniem. Keuze en interpretatie van uitkomstmaten in gerandomiseerd onderzoek. [Ge-bu. 2014;48\(7\):71-8.](#)
3. Beex LVAM, van Laarhoven HWM. Palliatieve behandeling van gemetastaseerd mammacarcinoom. [Gebu. 2010;44\(7\):74-81.](#)
4. Kraag DE. Behandeling van acne vulgaris. [Gebu. 2014;48\(9\):95-103.](#)
5. Anoniem. Combinatiepreparaten van luchtwegverwijders voor de behandeling van COPD. [Gebu. 2016;50\(2\):15-21.](#)
6. Todd KH, Funk KG, Funk JP, Bonacci R. Clinical significance of reported changes in pain severity. *Ann Emerg Med.* 1996;27(4):485-9.
7. de Vries TW, Janssens HJ. NSAID's bij kinderen. Behandeling van pijn. [Gebu. 2018;52\(6\):49-56.](#)
8. Janssens HJ, Janssen M, van de Lisdonk EH, van Riel PL, van Weel C. Use of oral prednisolone or naproxen for the treatment of gout arthritis: a double-blind, randomised equivalence trial. *Lancet.* 2008;371(9627):1854-60.
9. Rainer TH, Cheng CH, Janssens HJ, Man CY, Tam LS, Choi YF, et al. Oral Prednisolone in the Treatment of Acute Gout: A Pragmatic, Multicenter, Double-Blind, Randomized Trial. *Ann Intern Med.* 2016;164(7):464-71.
10. Clark E, Plint AC, Correll R, Gaboury I, Passi B. A randomized, controlled trial of acetaminophen, ibuprofen, and codeine for acute pain relief in children with musculoskeletal trauma. *Pediatrics.* 2007;119(3):460-7.
11. Singh D, Papi A, Corradi M, Pavlisova I, Montagna I, Francisco C, et al. Single inhaler triple therapy versus inhaled corticosteroid plus long-acting beta2-agonist therapy for chronic obstructive pulmonary disease (TRILOGY): a double-blind, parallel group, randomised controlled trial. *Lancet.* 2016;388(10048):963-73.
12. Vestbo J, Papi A, Corradi M, Blazhko V, Montagna I, Francisco C, et al. Single inhaler extrafine triple therapy versus long-acting muscarinic antagonist therapy for chronic obstructive pulmonary disease (TRINITY): a double-blind, parallel group, randomised controlled trial. *Lancet.* 2017;389(10082):1919-29.
13. van den Bogert CA, Souverein PC, Brekelmans CTM, Janssen SWJ, Koeter GH, Leufkens HGM, et al. Primary endpoint discrepancies were found in one in ten clinical drug trials. Results of an inception cohort study. *J Clin Epidemiol.* 2017;89:199-208.
14. van der Graaf Y. Kiezen tussen wel of niet behandelen. *Ned Tijdschr Geneeskd.* 2018;162:B1524.
15. Ridker PM, Danielson E, Fonseca FA, Genest J, Gotto AM, Jr., Kastelein JJ, et al. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. *N Engl J Med.* 2008;359(21):2195-207.
16. Kerst AJFA. JUPITER, een planeet te ver: rosuvastatine op geleide van C-reactief proteïne voor primaire preventie van hart- en vaatziekten. [Gebu. 2009;43\(1\):9-10.](#)
17. Eikelboom JW, Connolly SJ, Bosch J, Dagenais GR, Hart RG, Shestakovska O, et al. Rivaroxaban with or without Aspirin in Stable Cardiovascular Disease. *N Engl J Med.* 2017;377(14):1319-30.
18. van der Heijden S, Janssens HJEM. Recent onderzoek naar rivaroxaban. Geen reden voor uitbreiding indicatiestelling. [Gebu. 2018;52\(7-8\):64-70.](#)
19. Morillon MB, Stamp L, Taylor W, Fransen J, Dalbeth N, Singh JA, et al. Using serum urate as a validated surrogate end

- point for flares in patients with gout: protocol for a systematic review and meta-regression analysis. *BMJ Open*. 2016;6(9):e012026.
20. Ramdhan A. Febuxostat. [Gebu 2015;49\(4\):46-8](#).
 21. Becker MA, Schumacher HR, Jr., Wortmann RL, MacDonald PA, Eustace D, Palo WA, et al. Febuxostat compared with allopurinol in patients with hyperuricemia and gout. *N Engl J Med*. 2005;353(23):2450-61.
 22. Kydd AS, Seth R, Buchbinder R, Edwards CJ, Bombardier C. Uricosuric medications for chronic gout. *Cochrane Database Syst Rev*. 2014(11):CD010457.
 23. Shekelle PG, Newberry SJ, FitzGerald JD, Motala A, O'Hanlon CE, Tariq A, et al. Management of Gout: A Systematic Review in Support of an American College of Physicians Clinical Practice Guideline. *Ann Intern Med*. 2017;166(1):37-51.
 24. Richette P, Doherty M, Pascual E, Barskova V, Becce F, Castaneda-Sanabria J, et al. 2016 updated EULAR evidence-based recommendations for the management of gout. *Ann Rheum Dis*. 2017;76(1):29-42.
 25. Janssens HJEM. Controverse over profylactische urinezuurverlaging bij jicht. [Gebu. 2018;52\(1-2\):9-10](#).
 26. Bijl D. Sitagliptine (Januvia®), behandeling diabetes mellitus. [Gebu. 2007;41\(11\):117-9](#).
 27. Charbonnel B, Karasik A, Liu J, Wu M, Meininger G, Sitagliptin Study G. Efficacy and safety of the dipeptidyl peptidase-4 inhibitor sitagliptin added to ongoing metformin therapy in patients with type 2 diabetes inadequately controlled with metformin alone. *Diabetes Care*. 2006;29(12):2638-43.
 28. Aschner P, Kipnes MS, Lunceford JK, Sanchez M, Mickel C, Williams-Herman DE, et al. Effect of the dipeptidyl peptidase-4 inhibitor sitagliptin as monotherapy on glycemic control in patients with type 2 diabetes. *Diabetes Care*. 2006;29(12):2632-7.
 29. Raz I, Hanefeld M, Xu L, Caria C, Williams-Herman D, Khatami H, et al. Efficacy and safety of the dipeptidyl peptidase-4 inhibitor sitagliptin as monotherapy in patients with type 2 diabetes mellitus. *Diabetologia*. 2006;49(11):2564-71.
 30. Nieuwhof MAE. Tien jaar sitagliptine. [Gebu. 2018;52\(1-2\):11-5](#).
 31. Rich S. The 6-minute walk test as a primary endpoint in clinical trials for pulmonary hypertension. *J Am Coll Cardiol*. 2012;60(13):1202-3.
 32. Savarese G, Paolillo S, Costanzo P, D'Amore C, Cecere M, Losco T, et al. Do changes of 6-minute walk distance predict clinical events in patients with pulmonary arterial hypertension? A meta-analysis of 22 randomized trials. *J Am Coll Cardiol*. 2012;60(13):1192-201.
 33. Gabler NB, French B, Strom BL, Palevsky HI, Taichman DB, Kawut SM, et al. Validation of 6-minute walk distance as a surrogate end point in pulmonary arterial hypertension trials. *Circulation*. 2012;126(3):349-56.
 34. Anoniem. Pulmonale hypertensie. [Gebu. 2013;47\(11\):123-30](#).
 35. Poobalan AS, Aucott LS, Smith WC, Avenell A, Jung R, Broom J. Long-term weight loss effects on all cause mortality in overweight/obese populations. *Obes Rev*. 2007;8(6):503-13.
 36. Rucker D, Padwal R, Li SK, Curioni C, Lau DC. Long term pharmacotherapy for obesity and overweight: updated meta-analysis. *BMJ*. 2007;335(7631):1194-9.
 37. Anoniem. De medicamenteuze behandeling van overgewicht. [Gebu. 2015;49\(10\):111-8](#).
 38. Neal B, Perkovic V, Mahaffey KW, de Zeeuw D, Fulcher G, Erond N, et al. Canagliflozin and Cardiovascular and Renal Events in Type 2 Diabetes. *N Engl J Med*. 2017;377(7):644-57.
 39. Neal B, Perkovic V, Mahaffey KW, Fulcher G, Erond N, Desai M, et al. Optimizing the analysis strategy for the CANVAS Program: A prespecified plan for the integrated analyses of the CANVAS and CANVAS-R trials. *Diabetes Obes Metab*. 2017;19(7):926-35.
 40. Nieuwhof MAE, Janssens HJEM. Cardiovasculaire veiligheid van canagliflozine. De risico's vergeleken met placebo. [Gebu. 2018;52\(7-8\):57-9](#).
 41. Lassen MR, Raskob GE, Gallus A, Pineo G, Chen D, Portman RJ. Apixaban or enoxaparin for thromboprophylaxis after knee replacement. *N Engl J Med*. 2009;361(6):594-604.
 42. Lassen MR, Gallus A, Raskob GE, Pineo G, Chen D, Ramirez LM, et al. Apixaban versus enoxaparin for thromboprophylaxis after hip replacement. *N Engl J Med*. 2010;363(26):2487-98.
 43. Lassen MR, Raskob GE, Gallus A, Pineo G, Chen D, Hornick P, et al. Apixaban versus enoxaparin for thromboprophylaxis after knee replacement (ADVANCE-2): a randomised double-blind trial. *Lancet*. 2010;375(9717):807-15.
 44. Levi MM. Nieuwe antistollingsmiddelen. [Gebu. 2005;39\(3\):25-32](#).
 45. Dekkers OM, Vandenbroucke JP. Generaliseerbaarheid van therapeutisch onderzoek. *Ned Tijdschr Geneesk* 2007.

- 2007;151:2203-7.
46. Eriksson BI, Dahl OE, Rosencher N, Kurth AA, van Dijk CN, Frostick SP, et al. Dabigatran etexilate versus enoxaparin for prevention of venous thromboembolism after total hip replacement: a randomised, double-blind, non-inferiority trial. *Lancet*. 2007;370(9591):949-56.
 47. Eriksson BI, Dahl OE, Rosencher N, Kurth AA, van Dijk CN, Frostick SP, et al. Oral dabigatran etexilate vs. subcutaneous enoxaparin for the prevention of venous thromboembolism after total knee replacement: the RE-MODEL randomized trial. *J Thromb Haemost*. 2007;5(11):2178-85.
 48. Ginsberg JS, Davidson BL, Comp PC, Francis CW, Friedman RJ, Huo MH, et al. Oral thrombin inhibitor dabigatran etexilate vs North American enoxaparin regimen for prevention of venous thromboembolism after knee arthroplasty surgery. *J Arthroplasty*. 2009;24(1):1-9.
 49. Eriksson BI, Dahl OE, Huo MH, Kurth AA, Hantel S, Hermansson K, et al. Oral dabigatran versus enoxaparin for thromboprophylaxis after primary total hip arthroplasty (RE-NOVATE II*). A randomised, double-blind, non-inferiority trial. *Thromb Haemost*. 2011;105(4):721-9.
 50. Bijl D. Nieuwe orale anticoagulantia: niet vergoed, wel voorgeschreven. [Gebu. 2012;46\(5\):58-60](#).
 51. Hervas A, Huss M, Johnson M, McNicholas F, van Stralen J, Sreckovic S, et al. Efficacy and safety of extended-release guanfacine hydrochloride in children and adolescents with attention-deficit/hyperactivity disorder: a randomized, controlled, phase III trial. *Eur Neuropsychopharmacol*. 2014;24(12):1861-72.
 52. Wilens TE, Robertson B, Sikirica V, Harper L, Young JL, Bloomfield R, et al. A Randomized, Placebo-Controlled Trial of Guanfacine Extended Release in Adolescents With Attention-Deficit/Hyperactivity Disorder. *J Am Acad Child Adolesc Psychiatry*. 2015;54(11):916-25 e2.
 53. Newcorn JH, Harpin V, Huss M, Lyne A, Sikirica V, Johnson M, et al. Extended-release guanfacine hydrochloride in 6-17-year olds with ADHD: a randomised-withdrawal maintenance of efficacy study. *J Child Psychol Psychiatry*. 2016;57(6):717-28.
 54. Stolk LML. Nieuw geneesmiddel bij ADHD: guanfacine. [Gebu. 2018;52\(11-12\):90-4](#).
 55. DeFronzo RA, Hissa MN, Garber AJ, Luiz Gross J, Yuyan Duan R, Ravichandran S, et al. The efficacy and safety of saxagliptin when added to metformin therapy in patients with inadequately controlled type 2 diabetes with metformin alone. *Diabetes Care*. 2009;32(9):1649-55.
 56. Jadzinsky M, Pflutzner A, Paz-Pacheco E, Xu Z, Allen E, Chen R, et al. Saxagliptin given in combination with metformin as initial therapy improves glycaemic control in patients with type 2 diabetes compared with either monotherapy: a randomized controlled trial. *Diabetes Obes Metab*. 2009;11(6):611-22.
 57. Chacra AR, Tan GH, Apanovitch A, Ravichandran S, List J, Chen R, et al. Saxagliptin added to a submaximal dose of sulphonylurea improves glycaemic control compared with uptitration of sulphonylurea in patients with type 2 diabetes: a randomised controlled trial. *Int J Clin Pract*. 2009;63(9):1395-406.
 58. Onzenoort HAW. Therapietrouw. [Gebu. 2012;46\(5\):49-55](#).
-

Gerelateerde artikelen

- [Het non-inferioriteitsonderzoek](#)
02-04-2015 | dr O.M. Dekkers, onder medeverantwoordelijkheid van de redactiecommissie
- [Geneesmiddeleninteracties en de rol van P-glycoproteïne](#)
27-11-2014
- [Keuze en interpretatie van uitkomstmaten in gerandomiseerd onderzoek](#)
14-08-2014