

STAND VAN ZAKEN

'Non-inferiority'-studies: mogelijkheden en kanttekeningen

METHODOLOGIE VAN ONDERZOEK 3

Darius Soonawala en Olaf M. Dekkers

- 'Non-inferiority'-studies hebben tot doel om aan te tonen dat een nieuwe behandeling niet minder effectief is dan een standaardbehandeling, ofwel niet-inferieur.
- De non-inferioritymarge dient voor aanvang van een studie te worden vastgelegd. Deze marge bepaalt hoe groot het verschil in werkzaamheid maximaal mag zijn om te mogen concluderen dat een nieuwe behandeling niet minder werkzaam is dan de standaardbehandeling.
- Vaak is er een goede reden voor een non-inferioritydesign, bijvoorbeeld omdat de nieuwe behandeling makkelijker in gebruik is of een kostenvoordeel of vermindering van bijwerkingen oplevert.
- De lezer van een non-inferioritytrial dient een aantal aspecten kritisch te bekijken: (a) de reden van het gebruik van deze studieopzet, (b) het veronderstelde voordeel van de nieuwe behandeling, (c) de onderbouwing van de keuze voor de non-inferioritymarge, (d) de keuze van de standaardbehandeling waarmee de nieuwe behandeling wordt vergeleken.

In de rubriek Stand van zaken verschijnen de komende periode regelmatig bijdragen over methoden die gebruikt worden bij het opzetten van wetenschappelijk onderzoek. De artikelen in deze serie illustreren op begrijpelijke wijze wat een bepaalde methode behelst, zonder dat hier uitvoerige methodologische kennis voor nodig is. Zowel oude als nieuwe methodologische principes worden zo inzichtelijk gemaakt voor artsen die klinische onderzoeken goed willen interpreteren.

Meestal willen we van een nieuw medicijn weten of het beter is dan de standaardtherapie. In sommige gevallen stellen we ons in onderzoek echter tevreden met de vraag of een nieuw middel even goed is als de standaardtherapie. Een voorbeeld:

Het levende verzwakte vaccin tegen Japanse encefalitis wordt sinds 2005 niet meer geproduceerd. Bij 95% van de mensen leidde vaccinatie hiermee tot seroconversie, en bij 91% geeft dit bescherming tegen Japanse encefalitis. Het vaccin veroorzaakte sporadisch overgevoeligheidsreacties en neurologische bijwerkingen. Het vaccin was niet geregistreerd in Europa, maar werd onder bepaalde voorwaarden aangeboden aan reizigers die risico liepen. Er is sinds enkele jaren een nieuw geïnactiveerd vaccin beschikbaar, met minder bijwerkingen.¹

Het is vrijwel onmogelijk om door middel van gerandomiseerd onderzoek te bewijzen dat het nieuwe vaccin bij een nog hoger percentage tot seroconversie leidt. Om een verschil in seroconversie van 1% aan te tonen in een stu-

*Leids Universitair Medisch Centrum,
afd. Klinische epidemiologie, Leiden.*

*Drs. D. Soonawala, internist-nefroloog
in opleiding; dr. O.M. Dekkers,*

*internist-endocrinoloog en klinisch epidemioloog
(tevens LUMC, afd. Endocrinologie
en Metabole ziekten).*

*Contactpersoon: dr. O.M. Dekkers
(o.m.dekkers@lumc.nl).*

die met adequate power zouden heel veel proefpersonen nodig zijn: circa 7.000 per behandelarm. Dit maakt het onderzoek nagenoeg onbetaalbaar en daarmee onuitvoerbaar. Daarnaast is het niet aannemelijk dat het nieuwe vaccin seroconversie veroorzaakt bij meer dan 95% van de proefpersonen. Superioriteit van het nieuwe vaccin zal dus zelfs bij een hele grote trial waarschijnlijk niet worden bewezen.

Het alternatief is middels onderzoek aan te tonen dat het nieuwe vaccin niet slechter is. Dit vraagt om een specifieke onderzoeksopzet: namelijk, een equivalentietrial of een 'non-inferiority'-studie. Deze onderzoeksopzet wordt sinds de jaren '90 van de vorige eeuw steeds vaker toegepast.² Dit artikel gaat in op de karakteristieken van non-inferioritystudies en bespreekt een aantal aspecten waar lezers alert op dienen te zijn bij het interpreteren van zulke studies.

SUPERIORITEIT, EQUIVALENTIE EN NON-INFERIORITEIT

In gerandomiseerd therapeutisch onderzoek wordt een nieuwe behandeling vergeleken met placebo of met een reeds gevalideerde behandeling. Meestal is het doel om aan te tonen dat de nieuwe therapie beter is dan placebo of dan de standaardtherapie. Voor een dergelijke studie wordt tevoren bepaald wat men het kleinste relevante verschil acht (bijvoorbeeld 1% meer seroconversie voor een nieuw vaccin) en hoeveel statistische zekerheid men wenst te hebben over het resultaat.

Een equivalentiestudie of non-inferioritystudie wordt opgezet om aan te tonen dat een nieuwe behandeling gelijkwaardig is aan de standaardbehandeling. Een equivalentiestudie wil aantonen dat het nieuwe middel niet beter én niet slechter is dan de bestaande therapie, terwijl een non-inferioritystudie alleen wil aantonen dat het nieuwe middel niet slechter is.^{3,4} In de praktijk worden de termen vaak als synoniemen van elkaar gebruikt.

In feite is het onmogelijk om aan te tonen dat 2 behandelingen exact even effectief zijn. Dit zou statistisch gezien een oneindige steekproefgrootte vergen. In een equivalentiestudie definieert men daarom van tevoren binnen welke marges men mag spreken van een gelijkwaardig effect. Men spreekt bijvoorbeeld af, dat van een gelijkwaardige effectiviteit sprake is, zolang het verschil in seroconversie tussen het nieuwe vaccin en het bestaande vaccin niet groter is dan 4%. Formeel gaat men daarbij uit van de onderste en bovenste limiet van het 95%-betrouwbaarheidsinterval voor het verschil in seroconversie tussen de 2 vaccins: er is sprake van equivalentie wanneer die limieten niet groter zijn dan 4%.

In een non-inferioritystudie is men primair geïnteresseerd in de vraag of de nieuwe behandeling niet minder effectief is dan de standaardbehandeling, ofwel niet-inferieur, 'non-inferior'. Men kijkt daarom slechts naar één kant van

het betrouwbaarheidsinterval. Bijvoorbeeld: men kijkt of de limiet van het betrouwbaarheidsinterval voor het verschil in seroconversie niet groter is dan 4% ten nadele van de nieuwe behandeling. Deze vooraf gedefinieerde marge van 4% wordt de non-inferioritymarge genoemd, of Δ (delta). De omvang van de non-inferioritymarge dient te berusten op een combinatie van statistische en klinische overwegingen en is van invloed op de steekproefgrootte.^{5,6} Wanneer de grens van het betrouwbaarheidsinterval de non-inferioritymarge niet overschrijdt, mag worden gesteld dat de nieuwe behandeling niet-inferieur is. Dit is weergegeven in de figuur (situatie A, B en C).⁶ Als het betrouwbaarheidsinterval de non-inferioritymarge wel overschrijdt, is het resultaat inconclusief (situatie D en E). De figuur laat nog een opvallend aspect zien van non-inferioritytrials: de mogelijkheid bestaat dat bij een resultaat waarbij de nieuwe behandeling statistisch gezien inferieur is, het eindoordeel toch zou mogen zijn dat de nieuwe behandeling niet-inferieur is aan de oude behandeling (situatie C).

Stel nu dat we een non-inferioritystudie doen waarin we een nieuw middel X vergelijken met de standaardtherapie S, en dat de uitkomst is zoals in situatie C. Hoewel middel X statistisch slechter is dan middel S, is de boodschap toch dat het niet-inferieur is. Vervolgens wordt in een nieuwe non-inferioritystudie ditzelfde middel X als referentie voor een nieuw middel Y gebruikt. De uitkomst is weer situatie C. Doorgeredeneerd kunnen we dan te maken krijgen met middelen die steeds slechter presteren ondanks schijnbare non-inferiority. Dit fenomeen wordt 'biocreep' genoemd. Wanneer dit fenomeen zich inderdaad zou voordoen, is te verwachten dat in een grote meta-analyse van allerlei non-inferioritystudies, de experimentele therapie gemiddeld iets slechter zou zijn dan de standaardbehandeling. Dit blijkt echter niet het geval.²

DE NON-INFERIORITYMARGE

De 'Consolidated standards of reporting trials' (CONSORT)-richtlijn voor non-inferioritytrials en de European Medicines Agency (EMA) geven richting aan de keuze van de omvang van de non-inferioritymarge.^{5,6} Toch blijft er een arbitrair element in de keuze zitten. Hoe groter de marge, hoe makkelijker het is om voor een nieuwe behandeling non-inferiority te claimen. Een ruime non-inferioritymarge kan er bovendien toe leiden dat een minder effectieve nieuwe behandeling ten onrechte als even effectief wordt beschouwd.^{7,8} Maar een ruim gekozen marge zal de lezer niet makkelijk kunnen overtuigen.

HET NON-INFERIORITYDESIGN: EEN PLAATSBEPALING

Er kunnen verschillende redenen zijn om te kiezen voor een non-inferioritydesign.

De eerste is de situatie waarbij de standaardbehandeling

nagenoeg 100% effectief is, zoals in het eerder genoemde voorbeeld van het Japanse encefalitisvaccin.¹ Zelfs als het mogelijk zou zijn om een nog effectiever vaccin te ontwikkelen, is het nagenoeg ondoenlijk om middels gerandomiseerd onderzoek aan te tonen dat het nieuwe vaccin effectiever is.

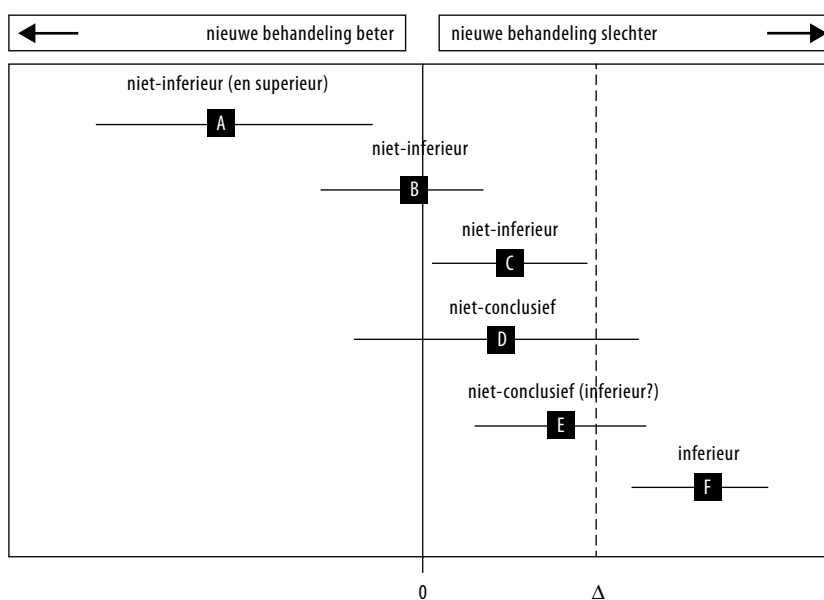
Een tweede reden is wanneer verondersteld wordt dat een nieuwe behandeling een ander voordeel biedt dan een grotere effectiviteit. Bijvoorbeeld minder bijwerkingen of lagere kosten of een groter gebruiksgemak. Hoewel de behandeling van de ziekte van Hodgkin zeer effectief is, zijn er bijwerkingen op lange termijn, zoals een verhoogde kans op maligniteiten en hart- en vaatziekten. Gebruik van een lagere dosering chemo- of radiotherapie kan de kans op deze bijwerkingen verminderen. Middels een non-inferioritystudie kan men bepalen of de effectiviteit in belangrijke mate wordt verminderd door gebruik van een lagere dosering.⁹ Hetzelfde geldt voor het gebruik van lagere doseringen teneinde een kostenbesparing te

bereiken of omdat de productiecapaciteit beperkt is.

Voor zeldzame ziekten waarbij een nieuwe behandeling mogelijk iets beter is dan een bestaande behandeling, kan het moeilijk zijn om genoeg proefpersonen in een trial te includeren om superioriteit aan te tonen. Dit is een derde reden om voor een non-inferioritydesign te kiezen. Om een kleine mate van superioriteit aan te tonen zijn namelijk meer proefpersonen nodig dan om gelijkwaardigheid aan te tonen binnen de bandbreedte van een non-inferioritymarge. Een voorbeeld is onderzoek naar anidulafungine voor de behandeling van candidemie.¹⁰

Een vierde reden is de situatie waarbij een nieuw middel waarschijnlijk even effectief is als een bestaand middel, maar het voordeel biedt dat het de ziekte op een andere wijze bestrijdt. Bijvoorbeeld een nieuw klasse antidiabeticum of antiretroviraal middel om een infectie met hiv te behandelen.^{11,12}

Als laatste geldt dat in zeldzame gevallen in een non-inferioritydesign zelfs placebo getest kan worden ten



FIGUUR Mogelijke uitkomsten voor verschil in behandel-effecten van 'non-inferiority'-studies (uitgedrukt als risicoverschil). De figuur is gebaseerd op de CONSORT-richtlijn voor non-inferioritytrials.⁶ De lijnen geven het 95%-betrouwbaarheidsinterval weer. De Δ geeft de non-inferioritymarge weer. De interpretatie van een non-inferioritytrial hangt af van de vraag hoe het betrouwbaarheidsinterval zich tot deze non-inferioritymarge verhoudt. Bevindt zich het gehele betrouwbaarheidsinterval binnen de marge, dan is het nieuwe middel niet-inferieur (A, B en C). Wanneer de marge in het betrouwbaarheidsinterval valt (D en E), dan is de uitkomst van de studie inconclusief. Wanneer het hele betrouwbaarheidsinterval aan de andere kant van de marge valt is het nieuwe middel inferieur (F).

Het is goed te realiseren dat bovengenoemde interpretaties enkel berusten op de relatie tussen betrouwbaarheidsinterval en marge, hetgeen in een aantal situaties (C en E) niet intuïtief is. Immers, we kijken bij een studie ook altijd met een schuin oog naar de vraag wat de uitkomst van een standaard statistische test zou zijn geweest. Deze test berust in de figuur op de relatie tussen het betrouwbaarheidsinterval en de nul (geen effect). Volgens een standaardtest zou in geval A het middel superieur zijn; terwijl C, E en F inferieur zijn. Het kan dus voorkomen dat een middel (C) zowel niet-inferieur is (ten opzichte van de marge) alsook inferieur is (ten opzichte van de nul).

LEERPUNTEN

- ‘Non-inferiority’-trials hebben tot doel aan te tonen dat een nieuwe behandeling niet minder werkzaam is dan de standaardtherapie.
- Een non-inferioritydesign is geschikt als een nieuwe behandeling een voordeel biedt op een ander vlak dan werkzaamheid, bijvoorbeeld een grotere mate van gebruiksgemak, lagere kosten of minder bijwerkingen.
- Voor aanvang van de studie moet de marge gedefinieerd zijn waarbinnen het verschil in werkzaamheid moet vallen om te mogen concluderen dat een nieuwe behandeling niet minder werkzaam is dan de standaardbehandeling.
- Bij het beoordelen van de resultaten van non-inferioritytrials is het belangrijk kritisch te kijken naar de standaardbehandeling waarmee de nieuwe behandeling werd vergeleken.

opzichte van een standaardbehandeling. Dit betreft de situatie waarbij men zich afvraagt of een veelgebruikte maar niet evidencebased behandeling veilig kan worden weggelaten. Zo werd darmlavage voorafgaand aan colonchirurgie jarenlang gepropageerd zonder hard bewijs voor de werkzaamheid. In een non-inferioritystudie werd bepleit dat het veilig is om darmlavage achterwege te laten.¹³

KRITIEK OP NON-INFERIORITYSTUDIES

Gebruik van het non-inferioritydesign is aan kritiek onderhevig. Tegenstanders stellen dat deze studieopzet wordt misbruikt om geneesmiddelen op de markt te brengen die niet beter zijn dan bestaande middelen, zogenaamde ‘me-too drugs’, en die geen aangetoond voordeel hebben, zoals een vermindering van het aantal bijwerkingen of een kostenvoordeel of een hogere mate van gebruiksgemak.⁸ Men kan zich afvragen of veronderstelde voordelen van een nieuwe behandeling niet eveneens onderzocht moeten zijn. Bijvoorbeeld middels een aparte analyse van data over bijwerkingen, of door het gebruik van een gecombineerde uitkomstmaat, of middels onderzoek naar de effectiviteit en veiligheid van nieuwe, reeds geregistreerde behandelingen nadat ze op de markt zijn gekomen (‘post-marketing surveillance’). Door nadruk te leggen op de vraag naar bewijs voor vermeende voordelen kan worden voorkomen dat patiënten worden blootgesteld aan nieuwe en vaak duurdere behandelingen die geen daadwerkelijk voordeel opleveren.

INTERPRETATIE VAN NON-INFERIORITYSTUDIES

Er zijn een aantal vragen die een lezer zich zou moeten stellen bij de interpretatie van non-inferioritystudies. De belangrijkste vraag is waarom de onderzoekers hebben gekozen voor een non-inferioritydesign. En als het nieuwe

middel inderdaad even effectief blijkt als de bestaande therapie, biedt de nieuwe behandeling dan wel een voordeel dat niet op het vlak van effectiviteit ligt? Daarbij is het van belang om na te gaan in hoeverre het vermeende voordeel vanzelfsprekend is (bijvoorbeeld pillen versus injecties), dan wel wordt ondersteund door data.

Net als bij superioriteitsstudies dient te lezer zich de vraag te stellen in hoeverre de standaardtherapie optimaal is, zowel in de keuze van het middel als in de toedieningsvorm en dosering. Wanneer niet is aangetoond dat een standaardbehandeling beter is dan placebo, kan het voorkomen dat men in een non-inferioritystudie in feite onderzoekt of een nieuwe behandeling niet onderdoet voor placebo.¹⁴

Zowel de primaire uitkomstmaat als de non-inferioritymarge dienen vooraf in een protocol te zijn vastgelegd. De marge is immers de scheidsrechter voor de vraag of een nieuwe behandeling niet-inferieur is. Gezien de centrale rol die de non-inferioritymarge speelt is het opvallend dat de registratie van trials nog niet eist dat de non-inferioritymarge wordt genoemd.¹⁵ Auteurs dienen in hun artikel de keuze voor de marge te onderbouwen en lezers moeten deze keuze kritisch bekijken. Wanneer bijvoorbeeld in een onderzoek gesteld wordt dat een middel niet-inferieur genoemd kan worden wanneer het mortaliteitsrisico niet meer dan 2 keer groter is dan van een standaardtherapie, zal niemand een eventuele non-inferioriteitsclaim serieus nemen.

Tot slot nog een meer technisch aspect. In een superioriteitsstudie wordt het resultaat van de ‘intention-to-treat’-analyse beschouwd als de standaard voor het aantonen van een behandel-effect. Een intention-to-treatanalyse geeft immers de meest conservatieve schatting, omdat alle personen die zijn geïnccludeerd in de studie worden meegenomen in de analyse, los van de vraag of ze daadwerkelijk de studiemedicatie hebben ingenomen en of ze tot het eind in de studie zijn gebleven. In een non-inferioritystudie daarentegen geeft het resultaat in de ‘per-protocol’-analyse meestal de meer conservatieve schatting van het behandel-effect, omdat hierbij alleen gerekend wordt met personen die het onderzoek hebben voltooid.⁶ De per-protocolpopulatie is kleiner dan de intention-to-treatpopulatie, en geeft dus bredere betrouwbaarheidsintervallen. Een breder betrouwbaarheidsinterval maakt het waarschijnlijker dat de non-inferioritymarge wordt overschreden. Daarom mag de per-protocolanalyse als een strengere analyse worden beschouwd voor het non-inferioritydesign.

CONCLUSIE

Gerandomiseerd onderzoek met een non-inferioritydesign heeft tot doel om aan te tonen dat een nieuwe

behandeling niet slechter is dan een bestaande behandeling. De onderliggende aanname is dat de nieuwe behandeling op een ander vlak een voordeel biedt. De lezer van een non-inferioritystudie dient een aantal aspecten kritisch te bekijken: (a) de reden van het gebruik van deze studie-opzet, (b) het veronderstelde voordeel van de nieuwe behandeling, (c) de onderbouwing van de keuze voor de non-inferioritymarge en (d) de keuze van de standaardbehandeling waarmee de nieuwe behandeling wordt vergeleken.

Belangenconflict: geen gemeld. Financiële ondersteuning: geen gemeld.

Aanvaard op 28 maart 2012

Citeer als: Ned Tijdschr Geneeskd. 2012;156:A4665

 [Meer op www.ntvg.nl/klinischepraktijk](http://www.ntvg.nl/klinischepraktijk)

LITERATUUR

- 1 Tauber E, Kollaritsch H, Korinek M, et al. Safety and immunogenicity of a Vero-cell-derived, inactivated Japanese encephalitis vaccine: a non-inferiority, phase III, randomised controlled trial. *Lancet*. 2007;370:1847-53.
- 2 Soonawala D, Middelburg RA, Egger M, Vandenbroucke JP, Dekkers OM. Efficacy of experimental treatments compared with standard treatments in non-inferiority trials: a meta-analysis of randomized controlled trials. *Int J Epidemiol*. 2010;39:1567-81.
- 3 Dunnett CW, Gent M. An alternative to the use of two-sided tests in clinical trials. *Stat Med*. 1996;15:1729-38.
- 4 The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). E9: Statistical principles for clinical trials. *Fed Regist*. 1998;63:49583.
- 5 Committee for Medicinal Products for Human Use (CHMP) guideline on the choice of the non-inferiority margin. *Stat Med*. 2006;25:1628-38.
- 6 Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA*. 2006;295:1152-60.
- 7 Gøtzsche PC. Lessons from and cautions about noninferiority and equivalence randomized trials. *JAMA*. 2006;295:1172-74.
- 8 Garattini S, Bertele V. Non-inferiority trials are unethical because they disregard patients' interests. *Lancet*. 2007;370:1875-77.
- 9 Dühmke E, Franklin J, Pfreundschuh M, et al. Low-dose radiation is sufficient for the noninvolved extended-field treatment in favorable early-stage Hodgkin's disease: long-term results of a randomized trial of radiotherapy alone. *J Clin Oncol*. 2001;19:2905-14.
- 10 Reboli AC, Rotstein C, Pappas PG, et al. Anidulafungin versus fluconazole for invasive candidiasis. *N Engl J Med*. 2007;356:2472-82.
- 11 Gallwitz B, Bohmer M, Segiet T, et al. Exenatide twice daily versus premixed insulin aspart 70/30 in metformin-treated patients with type 2 diabetes: a randomized 26-week study on glycemic control and hypoglycemia. *Diabetes Care*. 2011;34:604-6.
- 12 Cooper DA, Heera J, Goodrich J, et al. Maraviroc versus efavirenz, both in combination with zidovudine-lamivudine, for the treatment of antiretroviral-naïve subjects with CCR5-tropic HIV-1 infection. *J Infect Dis*. 2010;201:803-13.
- 13 Contant CM, Hop WC, van't Sant HP, et al. Mechanical bowel preparation for elective colorectal surgery: a multicentre randomised trial. *Lancet*. 2007;370:2112-7.
- 14 The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). E10: Choice of control group and related issues in clinical trials. *Fed Regist*. 2001;66:24390-1.
- 15 Dekkers OM, Soonawala D, Vandenbroucke JP, Egger M. Reporting of noninferiority trials was incomplete in trial registries. *J Clin Epidemiol*. 2011;64:1034-8.