

## ONDERZOEK

# Het ene oordeel is het andere niet

## KWANTITATIEVE ANALYSE VAN DE VARIATIE BIJ IGZ-INSPECTEURS

Saskia M. Tuijn, Frans J.G. Janssens, Huub van den Bergh en Paul B.M. Robben

- DOEL** Onderzoek naar de variatie in oordelen tussen inspecteurs van de Inspectie voor de Gezondheidszorg (IGZ) in het systeem van gefaseerd toezicht op de verpleeghuiszorg en naar de relatie tussen het type oordeel en de aan- of afwezigheid van onderbouwing daarbij.
- OPZET** Descriptief, kwantitatief, retrospectief.
- METHODE** In totaal werden 4914 oordelen met bijbehorende onderbouwingen van 26 inspecteurs uit 182 toezichtrapporten geanalyseerd. De oordelen waren in 2005 en 2006 gegeven op 25 criteria voor verantwoorde zorg in verpleeghuizen. Om de inspecteurs en hun oordelen over verschillende instellingen te kunnen vergelijken, werd met covariantieanalyse statistisch gecorrigeerd voor instellingskenmerken.
- RESULTATEN** Er waren statistisch significante verschillen in beoordeling tussen de inspecteurs. Na correctie voor instellingskenmerken bleven voor 14 van de 25 criteria significante verschillen bestaan. Tevens bleek de aanwezigheid van een onderbouwing bij een oordeel zowel afhankelijk te zijn van de inspecteur als van het gegeven oordeel.
- CONCLUSIE** Beoordelaarsverschillen tussen inspecteurs spelen een rol in de tweede fase van het gefaseerd toezicht op de verpleeghuizen. De IGZ zal resultaten van dit onderzoek gebruiken om haar toezicht verder te ontwikkelen.

Als persoonlijke opvattingen van inspecteurs een te grote rol spelen in het toezicht op de gezondheidszorg, doet dat afbreuk aan de geloofwaardigheid en het gezag van de Inspectie voor de Gezondheidszorg (IGZ) en daarmee aan het maatschappelijk vertrouwen in de gezondheidszorg. Daarom zijn betrouwbaarheid van en zorgvuldigheid in de oordelen van inspecteurs van groot belang.<sup>1</sup> De IGZ onderkent dit belang en besteedt aandacht aan de uniformering van de werkwijze van haar inspecteurs ([www.igz.nl/117334/07-11\\_meerjarenplan.pdf](http://www.igz.nl/117334/07-11_meerjarenplan.pdf)). Het standaardiseren van werkwijzen is in diverse beroepsgroepen dé manier om overeenstemming tussen professionals te verhogen. Dat gebeurt bijvoorbeeld in medische beroepen door het gebruik van onderzoeksprotocollen, de standaardisatie van criteria, consensusbijeenvakomen en beslissingsondersteunende software.

De inspectie maakt in haar toezicht gebruik van 3 verschillende methodieken: het proactieve risicogestuurde toezicht ('gefaseerd toezicht'), het thematisch toezicht en het interventietoezicht. Het gefaseerd toezicht verloopt in 3 fasen ([www.igz.nl/loketzorgaanbieders/indicatoren/toelichting-gt](http://www.igz.nl/loketzorgaanbieders/indicatoren/toelichting-gt)). Voordat de 1e fase begint, worden in overleg met veldpartijen normenkaders en kwaliteitsindicatoren opgesteld. In de 1e fase maakt de inspectie op basis van deze indicatoren of criteria een risicoselectie. In de 2e fase bezoeken inspecteurs instellingen waarvan wordt vermoed dat er risico's zijn in de zorg. Op basis van het oordeel van de inspecteurs moeten instellingen, indien nodig, maatregelen treffen om de zorg te verbeteren. In

*Inspectie voor de Gezondheidszorg,*

*Kenniscentrum, Utrecht.*

*Drs. S.M. Tuijn, communicatiekundige;*

*dr. P.B.M. Robben, arts.*

*Universiteit Twente,*

*faculteit Gedragwetenschappen,*

*afd. Onderwijsmanagement*

*en -organisatie, Enschede.*

*Prof.dr. F.J.G. Janssens, onderwijskundige.*

*Universiteit Utrecht,*

*faculteit Geesteswetenschappen,*

*afd. Taalbeheersing, Utrecht.*

*Prof.dr. H. van den Bergh, psycholoog*

*(tevens: Universiteit van Amsterdam, Instituut*

*voor de Lerarenopleiding, Amsterdam).*

*Contactpersoon: drs. S.M. Tuijn (s.tuijn@igz.nl).*

**TABEL 1** Overzicht van de 27 criteria die de Inspectie voor de Gezondheidszorg in 2005 en 2006 hanteerde bij toezicht op de verpleeghuiszorg

zorgzwaartemeting	klachtenregeling	bekwaamheden vereist volgens de Wet BIG
decubitus melden incidenten aan commissie veiligheid van materiaal	systematisch werken met zorgplannen registratie van afwijkingen van het individuele zorgplan regeling van de verantwoordelijkheid voor de inhoud en de coördinatie van de uitvoering van het zorgplan	professioneel handelen bejegening van cliënten meerjarenbeleidplan
tevredenheid van cliënten cliëntenraad algemene informatievoorziening voor cliënten passende hulp bij eten en drinken	regeling voor multidisciplinaire cliëntbespreking cliëntenrechten met betrekking tot het zorgplan scholing in het opstellen en gebruiken van een zorgplan beschikbaarheid en deskundigheid van zorgverleners	jaarwerkplan organisatiestructuur managementinformatiesysteem systeem voor het bewaken van een verantwoord zorgniveau privacy
adequaat toezicht en permanente alertheid	scholingsplan	

Wet BIG: Wet op de Beroepen in de Individuele Gezondheidszorg.

de 3e fase van gefaseerd toezicht kan de inspectie bestuursrechtelijke sancties opleggen en strafrechtelijke maatregelen nemen.

Standaardisering van werkwijzen leidt niet zonder meer tot uniform handelen van inspecteurs van de IGZ. Dit bleek uit onderzoek naar de beroepsuitoefening van inspecteurs en hun visie op toezicht. Inspecteurs verschillen in werkwijze en in zienswijze over de rol van de inspectie.<sup>2</sup> Deze resultaten onderstrepen de noodzaak om nader onderzoek te doen naar de vraag of inspecteurs in gelijke situaties tot gelijke oordelen komen. Wij analyseerden hiervoor gegevens van de procedure voor gefaseerd toezicht. Hiermee wilden wij antwoorden krijgen op 3 vragen: (a) zijn er verschillen in beoordeling tussen inspecteurs? (b) zijn deze verschillen toe te schrijven aan de inspecteurs of aan instellingskenmerken? (c) onderbouwen inspecteurs hun oordeel en in hoeverre is de aanwezigheid van een onderbouwing afhankelijk van de inspecteur en van het gegeven oordeel?

## METHODEN

Het onderzoek richtte zich op de overeenkomsten en verschillen in inspecteuroordelen binnen de 2e fase van het gefaseerd toezicht op de verpleeghuiszorg in 2005 en 2006 (n = 645). Inspecteurs (n = 31) maakten toen gebruik van een toetsingskader bestaande uit 27 criteria voor de beoordeling van verantwoorde zorg, vastgelegd in een inspectie-instrument. Deze criteria zijn weergegeven in tabel 1. De criteria zijn een combinatie van structuur-, proces- en uitkomstmaten. Bij het criterium 'decubitus' bijvoorbeeld registreert men de prevalentie van decubitus (procesmaat) en de aanwezigheid van een decubitus-protocol (structuurmaat). Tijdens de toezichtbezoeken beoordeelden inspecteurs de criteria op een vierpuntschaal oplopend van een negatief naar een positief oor-

deel: 'afwezig', 'aanwezig', 'operationeel', 'geborgd'. In het inspectie-instrument stond per criterium aangegeven wanneer welk oordeel van toepassing was. Ter illustratie zijn deze beoordelingsnormen voor het criterium 'melden incidenten commissie' weergegeven in tabel 2. Bij score 'afwezig' of 'aanwezig', werd er van de instelling een plan van aanpak gevraagd om deze punten te verbeteren.

Voor dit onderzoek gebruikten wij de oordelen van inspecteurs die minimaal 7 rapporten hadden geschreven over hun toezichtbezoeken in 2005 en 2006 die voldeden aan het IGZ-format. In deze rapporten vinken zij oordelen aan in tabellen. Met dit inclusiecriterium werd gegarandeerd dat er voldoende waarnemingen per inspecteur waren en tegelijkertijd dat een zo groot mogelijk aantal inspecteurs in het onderzoek betrokken konden

**TABEL 2** Overzicht van de normen en oordelen bij het criterium 'melden incidenten commissie', toegepast door de Inspectie voor de Gezondheidszorg bij het toezicht op de verpleeghuiszorg

norm	oordeel
<ul style="list-style-type: none"> <li>de organisatie hanteert procedures voor het registreren en evalueren van incidenten en ongevallen</li> <li>op basis van deze registratie worden zo nodig corrigerende en preventieve maatregelen genomen</li> <li>er is een systeem aanwezig binnen de organisatie voor het registreren en behandelen van meldingen van incidenten in de zorg- en dienstverlening</li> <li>van het kwaliteitssysteem maakt in ieder geval de incidentenregistratie onderdeel uit</li> </ul>	<p>'afwezig' er worden geen incidenten geregistreerd en geëvalueerd</p> <p>'aanwezig' er worden incidenten geregistreerd</p> <p>'operationeel' de resultaten van incidentenanalyse worden gebruikt ten behoeve van verbeteringen</p> <p>'geborgd' het systeem voor melding van incidenten wordt systematisch geëvalueerd en zo nodig bijgesteld</p>



FOTO: ENVERHIRSCH/LAIF / HOLLANDE HOOGTE

worden (26 van de 31). Uit de rapporten van de geselecteerde inspecteurs over hun bezoeken (minimaal 9, maximaal 41 per inspecteur) namen wij aselekt 7 rapporten per inspecteur. In totaal werden 182 rapporten geanalyseerd (123 rapporten van verpleeghuizen, 42 van zorgcentra, 11 van verzorgingshuizen, 1 van een verpleegunit in een zorgcentrum en 5 van verpleegunits in een verzorgingshuis). De verschillende typen instellingen waren daarbij willekeurig verdeeld over de inspecteurs. Omdat al deze zorginstellingen langdurige zorg boden en op dezelfde criteria werden beoordeeld, verwijzen wij in de verdere rapportage naar deze groep met de term 'verpleeghuizen'.

#### STATISTISCHE ANALYSE

De in totaal 4914 oordelen uit de 182 rapporten werden samengebracht in een databestand. Voor elke inspecteur werd per criterium het gemiddelde van zijn of haar oordelen over de 7 verpleeghuizen berekend. Met een variantieanalyse onderzochten wij of er verschillen in beoordeling waren tussen inspecteurs en in hoeverre deze toe te schrijven waren aan instellingskenmerken. Bij de resultaten van deze analyse gaven wij 80%-betrouw-

baarheidsintervallen weer.<sup>3</sup> Ook onderzochten wij met een variantieanalyse of er significante verschillen bestonden tussen inspecteurs in het onderbouwen van oordelen en of de interactie tussen de onafhankelijke variabelen 'inspecteur' en 'het oordeel' invloed had op de afhankelijke variabele 'de onderbouwing van het oordeel'. Bij alle analyses hanteerden wij een tweezijdige overschrijdingskans van 5% (enkelvoudige variantieanalyse, SPSS versie 15). Van de 27 criteria waren er 2 door het merendeel van de inspecteurs niet beoordeeld ('privacy' en 'scholing zorgplan'); deze hebben we daarom niet in de analyse betrokken.

Om te onderzoeken in welke mate beoordelaarsverschillen waren toe te schrijven aan de inspecteur of aan de beoordeelde instelling, was het beter geweest als meerdere inspecteurs dezelfde instellingen hadden beoordeeld.<sup>3</sup> Dit is echter in de praktijk van het toezicht niet gebruikelijk. Daarom corrigeerden wij in onze analyse voor instellingskenmerken, als benadering van de ideale situatie. Hiertoe namen wij de oordelen op 24 van de 25 criteria mee als covariaten in de analyse. Per inspecteur werd het gemiddelde oordeel op het onderzochte criterium gecorrigeerd op basis van de oordelen van alle inspecteurs op de andere 24 criteria die als covariaat waren meegenomen. Hierdoor werd de heterogene groep instellingen op 24 criteria gehomogeniseerd en werden de oordelen van inspecteurs zo veel mogelijk vergelijkbaar.

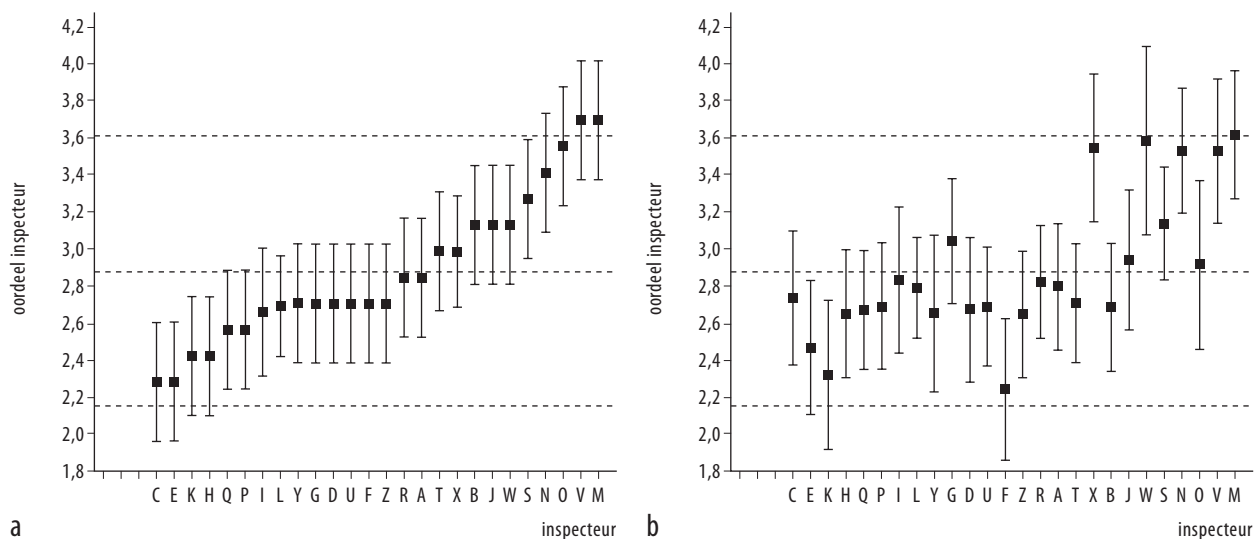
Hetzelfde gold voor eventuele verschillen tussen inspecteurs in het onderbouwen van het oordeel. Ook daarvoor hadden idealiter verschillende inspecteurs een oordeel over dezelfde instelling moeten geven. Deze situatie benaderden wij eveneens door het oordeel op het betreffende criterium als covariaat in de analyse mee te nemen.

#### RESULTATEN

Uit de analyses blijkt dat de oordelen van inspecteurs op 22 van de 25 criteria significant verschilden.

De proportie verklaarde variantie of effectgrootte ( $\eta^2$ ) varieerde van 0,2 tot 0,4; dit wil zeggen dat tussen de 20 en 40% van de totale variantie was toe te schrijven aan verschillen tussen inspecteurs. Na correctie voor instellingskenmerken waren de verschillen nog steeds statistisch significant voor 14 van de 25 criteria, met een middelmatig effect ( $\eta^2 = 0,3-0,4$ ). Een van de criteria waarvoor dit gold, was 'melden incidenten commissie' ( $F = 1,652$ ;  $p = 0,041$ ;  $\eta^2 = 0,3$ ;  $df_1 = 25$ ;  $df_2 = 133$ ).

Om te visualiseren hoe de gemiddelde oordelen van de inspecteurs zich tot elkaar verhielden, berekenden wij per inspecteur een betrouwbaarheidsinterval. Figuur 1 geeft van elke inspecteur het gemiddelde oordeel met het



**FIGUUR** Variatie in de oordelen van inspecteurs over verpleeghuizen vóór en na correctie voor kenmerken van de instellingen. Weergegeven is het gemiddelde per inspecteur van zijn of haar oordelen over 7 verpleeghuizen voor het criterium 'melding incidenten commissie', met het individuele 80%-betrouwbaarheidsinterval (a). De horizontale lijnen geven het gemiddelde (2,89) en de standaarddeviatie weer van de oordelen van alle inspecteurs. Na correctie per inspecteur voor kenmerken van de beoordeelde instellingen zijn er nog steeds verschillen tussen de inspecteurs in de gemiddelde oordelen over het criterium 'melding incidenten commissie' (b).

80%-betrouwbaarheidsinterval weer voor het criterium 'melden incidenten commissie'. Uit figuur 1a blijkt dat er grote verschillen waren in oordelen; het gemiddelde oordeel verschilt van inspecteur tot inspecteur. De correctie voor instellingskenmerken had een groot effect op de oordelen op dit criterium, zoals te zien in figuur 1b. Door de correctie deden zich verschuivingen voor in de rangorde van de inspecteurs, in hun gemiddelden en in de grootte van de betrouwbaarheidsintervallen.

Voor alle criteria waren er significante verschillen tussen inspecteurs in het wel of niet geven van onderbouwingen. Na correctie voor instellingskenmerken bleven deze verschillen bestaan voor 22 van de 25 criteria. Het bleek dus van de inspecteur af te hangen of er een onderbouwing van het oordeel was. Eén van de criteria waarvoor dit gold, was 'professioneel handelen' ( $F = 6,699$ ;  $p < 0,001$ ;  $df_1 = 25$ ;  $df_2 = 78$ ). De proportie verklaarde variantie was groot ( $\eta^2 = 0,5$ ). Bij 9 van deze 22 criteria was er sprake van een interactie-effect. Dit betekent dat de aanwezigheid van een onderbouwing mede afhankelijk was van het type oordeel dat de inspecteur gaf. Dit gold bijvoorbeeld voor 'zorgzwaartemeting' ( $F = 2,047$ ;  $p < 0,001$ ;  $df_1 = 75$ ;  $df_2 = 78$ ). De proportie verklaarde variantie was groot ( $\eta^2 = 0,5$ ).

## BESCHOUWING

Uit de analyses bleek dat beoordelaarsverschillen in het gefaseerd toezicht een rol spelen. Het effect van deze verschillen was middelmatig tot groot ( $\eta^2 \geq 0,3$ ) en was aanwezig bij het merendeel van de gehanteerde criteria. De verschillen kunnen niet verklaard worden doordat bepaalde inspecteurs alleen of vooral instellingen met kwalitatief minder goede zorg hebben beoordeeld, aangezien gecorrigeerd werd voor instellingskenmerken. Het is aannemelijk dat de gevonden verschillen gedeeltelijk berusten op verschillen tussen inspecteurs.

Voor 22 van de 25 criteria was de aanwezigheid van een onderbouwing bij een oordeel eveneens afhankelijk van de inspecteur. Daarbij bleek bij 9 van de 25 criteria een interactie-effect. Positievare oordelen werden minder vaak onderbouwd, maar ook dit verschilde van inspecteur tot inspecteur.

Het bestaan van beoordelaarsverschillen is niet uniek voor inspecteurs van de IGZ. Ook bij de controleurs van de Voedsel en Waren Autoriteit,<sup>4</sup> de inspecteurs van de Inspectie van het Onderwijs<sup>3</sup> en binnen andere beroepsgroepen zoals rechters,<sup>5-7</sup> docenten,<sup>8</sup> verzekeringsartsen<sup>9,10</sup> en medische specialisten<sup>11-14</sup> bestaan beoordelaarsverschillen en onderzoekt men hoe deze verschillen verkleind kunnen worden.<sup>15-18</sup>



## LEERPUNTEN

- Verschillen tussen beoordelaars komen voor bij verschillende beroepsgroepen in de medische sector.
- Ook tussen de inspecteurs van de Inspectie voor de Gezondheidszorg zijn er verschillen in de oordelen over verpleeghuiszorg.
- Positievare oordelen worden minder vaak onderbouwd, maar dit verschilt van inspecteur tot inspecteur.
- De resultaten van het onderzoek hiernaar worden gebruikt om de verschillen in beoordelingen te verkleinen.

## KANTTEKENINGEN BIJ DIT ONDERZOEK

Voor een juiste interpretatie van de resultaten zijn enkele kritische opmerkingen van belang. In dit onderzoek behandelden wij ordinale gegevens als discrete gegevens, dat wil zeggen: de semantische categorieën 'afwezig', 'aanwezig', 'operationeel' en 'geborgd' werden omgezet in een score oplopend van 1 naar 4, waardoor een interval-schaal ontstond. Het berekenen van een gemiddelde op deze schaal was problematisch, omdat het gemiddelde niet in de 4 semantische categorieën past. Het behandelen van ordinale gegevens als discrete of continue gegevens is echter een veelgebruikte manier om ordinale data te analyseren; over het algemeen leidt dit tot slechts zeer geringe vertekeningen.<sup>2,3</sup>

Een ander kritiekpunt is de statistische methode. In dit onderzoek is gebruikgemaakt van inspecteursoordelen die in de werkelijkheid zijn toegekend aan zorginstellingen. Het voordeel hiervan is dat de ecologische validiteit van het onderzoek hoog is ('ecologische validiteit' is de mate waarin de onderzoeksresultaten overeenkomen met de alledaagse praktijk). Het nadeel van deze methode is dat er achteraf statistisch gecorrigeerd moet worden voor instellingskenmerken.

Daarnaast beïnvloedt de definitie van de criteria en de methode van beoordelen de betrouwbaarheid van een beoordeling.<sup>9</sup> Inspecteurs kunnen verschillende criteria gebruikt hebben om tot een oordeel te komen, doordat ze verschillende definities van de indicatoren hanteerden. Een validiteitsprobleem kan mede oorzaak zijn van de beoordelaarsverschillen.

Het instrument waarmee de inspecteurs werkten, werd verder ontwikkeld in de periode dat de bezoeken plaatsvonden. De criteria bleven onveranderd, maar kleine wijzigingen in de tekst van het instrument kunnen beoordelaarsverschillen veroorzaakt hebben. De gevonden verschillen zouden dus mede verklaard kunnen worden door wijzigingen in de methoden van beoordelen.

Voor dit onderzoek maakten wij gebruik van oordelen van inspecteurs over verschillende soorten instellingen. In de toekomst zullen wij met een zogenaamde vignetenstudie de beoordeling van instellingen van hetzelfde

type door verschillende inspecteurs onderzoeken. Daarbij krijgen inspecteurs gefingeerde beschrijvingen van instellingen voorgelegd.

## IMPLICATIES ONDERZOEKSRESULTATEN

Het bestaan van beoordelaarsverschillen in het gefaseerde toezicht impliceert dat een positief of negatief oordeel over de instelling niet alleen afhangt van haar eigen kenmerken op vastgestelde criteria, maar ook van de inspecteur die de instelling bezoekt. Door misclassificatie kan het gebeuren dat ten onrechte geen maatregelen ter verbetering gevraagd worden op plaatsen waar wel risico's in de zorg zijn. Instellingen waar intensiever toezicht nodig is, krijgen daardoor in de praktijk te weinig toezicht. Ook kunnen instellingen teveel toezicht krijgen en kan ten onrechte om verbetermaatregelen gevraagd zijn. Met nadruk wijzen wij erop dat de inspectie nooit zware sanctiemaatregelen oplegt aan instellingen alleen op grond van de oordelen op criteriumniveau.

De afwezigheid van onderbouwingen is vooral een factor van betekenis bij negatieve oordelen. De onderbouwing is van belang voor de acceptatie van het oordeel door instellingen. Als een onderbouwing ontbreekt, kan dat de acceptatie van een oordeel beïnvloeden. Zonder onderbouwing zijn er ook geen concrete aanknopingspunten voor de verbetering van de zorg.

## NADER ONDERZOEK

Dit onderzoek maakt deel uit van een omvangrijk onderzoek naar beoordelaarsverschillen. Daarin onderzoeken wij of oordelen van inspecteurs valide zijn en of bijvoorbeeld organisatieomstandigheden en persoonsgebonden factoren de beoordelaarsverschillen verklaren. Inzicht in de verklarende factoren gebruiken wij om een interventie te ontwerpen die de beoordelaarsverschillen kan verkleinen. Deze interventie is een aanvulling op bestaande initiatieven die gericht zijn op een hogere overeenstemming tussen beoordelaars, zoals scholingsprogramma's, studiedagen en intercollegiale toetsing. Met een naming onderzoek wij of de interventie effectief is.

Drs. J.A.H. van Veen, hoofdinspecteur Verpleging en Chronische Zorg (IGZ), en dr. C.A.J. Ketelaars, coördinerend inspecteur Ketenzorg (IGZ) gaven commentaar op het artikel.

Belangenconflict: geen gemeld. Financiële ondersteuning: geen gemeld.

Aanvaard op 19 november 2008

Citeer als: Ned Tijdschr Geneesk. 2009;153:B63

➤ [Literatuur op www.ntvg.nl/onderzoek](http://www.ntvg.nl/onderzoek)