

De relatie tussen normen en oordelen in het toezicht op de gezondheidszorg

S.M. Tuijn,¹ H. van den Bergh,² P.B.M. Robben,¹ F.J.G. Janssens³

In dit onderzoek hebben we de nauwkeurigheid van oordelen van inspecteurs van de Inspectie voor de Gezondheidszorg (IGZ) onderzocht. Het gaat om oordelen over vier criteria van verantwoorde zorg die gegeven zijn in toezichtsbezoeken in verpleeghuizen in 2005/2006. Deze bezoeken vonden plaats in het kader van het Gefaseerd Toezicht (GT). We hebben de onderbouwingen van 615 oordelen over de criteria “decubitus”, “voldoende hulp bij eten en drinken”, “permanent toezicht huiskamers” en “zorgzwaartemeting” geanalyseerd. Onderzocht is in hoeverre de argumenten uit de onderbouwingen van de oordelen overeenkomen met de normen uit het toezichtskader van de IGZ. Hierdoor kon worden nagegaan of de door de inspecteurs feitelijk gegeven oordelen overeenkomen met de oordelen die op basis van de aanwezige argumenten bij een strikte toepassing van de IGZ-normen gegeven hadden moeten worden. De nauwkeurigheid van de feitelijke oordelen zijn door twee onafhankelijke beoordelaars geanalyseerd. Het onderzoek laat zien dat er sprake is van een validiteitsprobleem: de betekenis van de feitelijke oordelen verschilt regelmatig. Bij de vier onderzochte criteria leidt dit tot 52% vals-positieve oordelen en 1% vals-negatieve oordelen. Er is geen verband tussen het percentage vals-positieve oordelen en het gemiddelde oordeel over de vier criteria. Vals-positieve oordelen komen voor bij zowel inspecteurs die lager dan gemiddeld oordelen, als bij inspecteurs die hoger dan gemiddeld oordelen. Het geven van een te positief oordeel vergeleken met de IGZ-norm, komt bij alle in het onderzoek betrokken inspecteurs voor en is daarmee een situatiegebonden kenmerk van deze inspecteurs. Het percentage vals-positieve oordelen is afhankelijk van het criterium dat wordt beoordeeld. Daarmee zijn vals-positieve oordelen ook afhankelijk van de kwaliteit van het gehanteerde instrument,

Trefwoorden: toezicht, beoordelaarsverschillen, validiteit, evaluatieonderzoek

INLEIDING

Niet alle inspecteurs van de Inspectie voor de Gezondheidszorg (IGZ) komen in vergelijkbare situaties tot gelijke oordelen. Dit blijkt uit eerder onderzoek naar beoordelaarsverschillen in het Gefaseerd Toezicht (GT) op de verpleeghuiszorg.¹ Het oordeel over (aspecten van) een instelling is niet alleen afhankelijk van kenmerken van die instelling, maar ook afhankelijk van de inspecteur; sommige inspecteurs oordelen strenger dan anderen. Onbekend is nog hoe nauwkeurig deze oordelen zijn. Als inspecteurs een criterium niet beoordelen volgens de geoperationaliseerde IGZ-normen, dan zijn deze oordelen als onnauwkeurig te beschouwen omdat de betekenis van de oordelen dan afwijkt van de normen die bij het toezicht gebruikt worden. In dit onderzoek hebben we de nauwkeurigheid van inspecteursoordelen onderzocht.

Beoordelingscriteria in het Gefaseerd Toezicht

Het GT bestaat uit drie fasen: informatie verzamelen, oordelen en interveniëren (zie: www.igz.nl/loketzorgaanbieders/indicatoren/toelichting-gt). Inspecteurs beoordelen instellingen die zij in de tweede fase bezoeken op een aantal van te voren vastgestelde criteria. Op basis van deze oordelen moeten instellingen, indien nodig, maatregelen treffen om de zorg te verbeteren. Wanneer deze maatregelen niet binnen de gestelde termijn getroffen zijn, start de derde fase van GT. In deze fase kan de inspectie bestuurlijke sancties opleggen en strafrechtelijke en tuchtrechtelijke procedures starten. In het GT op de verpleeghuiszorg maakten inspecteurs in 2005 en 2006 gebruik van 27 criteria voor kwaliteit van zorg die zij beoordeelden op een vierpuntschaal: “afwezig”, “aanwezig”, “operationeel”, “geborgd”. De beoordelingschaal op deze criteria loopt op van een negatief naar een positief oordeel, waarbij “afwezig” het meest negatieve en “geborgd” het meest positieve oordeel is. Als een instelling op criteria het oordeel “afwezig” of “aanwezig” krijgt, wordt de instelling gevraagd maatregelen te treffen om deze punten te verbeteren.

¹ Kenniscentrum Inspectie voor de Gezondheidszorg, Utrecht

² Faculteit Geesteswetenschappen, afdeling Taalbeheersing, Universiteit Utrecht en Universiteit van Amsterdam, Instituut voor de lerarenopleiding

³ Faculteit Gedragswetenschappen, afdeling Onderwijsmanagement en -organisatie, Universiteit Twente, Enschede

Bij de beoordeling van de criteria maken inspecteurs gebruik van IGZ-normen, die beschrijven onder welke omstandigheden welk oordeel gegeven moet worden. Volgens de IGZ-norm moet bijvoorbeeld het criterium “decubitus” in het toezicht op de verpleeghuiszorg als “afwezig” gescoord worden als er geen decubitusregistratie is in de instelling (tabel 1). Als inspecteurs bij het ontbreken van een decubitusregistratie het oordeel “aanwezig” geven, wijkt de betekenis van dit oordeel af van de IGZ-norm. Het oordeel is dan niet nauwkeurig en is er sprake van een validiteitsprobleem.²⁻⁴

Dit onderzoek is een vervolg op een eerder onderzoek naar interbeoordelaarsverschillen bij inspecteurs in het gefaseerd toezicht op zorg in verpleeghuizen. In dat onderzoek zijn de inspecteursoordelen over 25 van de 27 criteria voor verantwoorde zorg in verpleeghuizen onderzocht.¹ Uit dat onderzoek bleek dat er significante interbeoordelaarsverschillen waren bij 14 van de 25 criteria. We zijn toen niet ingegaan op verklaringen voor deze verschillen. Eén van die verklaringen kan zijn dat inspecteurs een verschillende betekenis toekennen aan de IGZ-normen die zij hanteren bij het beoordelen van criteria. Wanneer de feitelijke oordelen afwijken van de IGZ-normen ontstaat een validiteitsprobleem. Het oordeel is dan niet meer eenduidig omdat onduidelijk is wat het oordeel “aanwezig” of “operationeel” betekent, omdat het een andere, wellicht inspecteursafhankelijke, betekenis gekregen heeft.

METHODEN

Om te achterhalen in hoeverre een validiteitsprobleem in de oordelen over verpleeghuizen in 2005/2006 een rol speelden, hebben we dezelfde dataset gebruikt als bij het onderzoek naar beoordelaarsverschillen.¹ In deze dataset zijn alleen de oordelen van inspecteurs (N = 26) opgenomen die minimaal negen rapporten hadden geschreven over hun toezichtsbezoeken in verpleeghuizen in 2005/2006. Voorwaarde was verder dat deze rapporten voldeden aan het IGZ-format. Dit format houdt onder andere in dat inspecteurs de oordelen hebben gerapporteerd in de daarvoor bestemde tabellen en dat ze de oordelen onderbouwd hebben.

Per inspecteur zijn at random zeven rapporten geselecteerd. In totaal zijn 182 rapporten geanalyseerd. Van de 27 criteria die in 2005/2006 gebruikt werden, hebben we de oordelen over de criteria “decubitus”, “voldoende hulp bij eten en drinken”, “permanent toezicht huiskamers” en “zorgzwaarte meting” en de bijbehorende on-

Kernpunten

- De inspecteursoordelen over vier van 27 te beoordelen criteria uit het gefaseerd toezicht op de verpleeghuiszorg zijn onderzocht.
- De onderzochte oordelen zijn in 53% van de gevallen onnauwkeurig; 52% van de oordelen is te positief, 1% is te negatief. Er is sprake van een validiteitsprobleem.
- Vals positieve oordelen komen bij alle onderzochte inspecteurs voor en hangen niet samen met de hoogte van het gegeven oordeel.
- Het percentage vals positieve oordeel verschilt in hoogte bij de vier onderzochte criteria en is daarmee criterium afhankelijk.
- Nader onderzoek is nodig om achterliggende oorzaken te achterhalen.

derbouwing van deze oordelen onderzocht. Tabel 1 beschrijft deze criteria en de operationalisering van de IGZ-normen op basis waarvan de inspecteurs deze criteria moeten beoordelen. We hebben gekozen voor deze criteria omdat zij een belangrijke rol spelen in het toezicht op de patiëntveiligheid. Ook bepaalden deze criteria in belangrijke mate het maatschappelijke beeld van de verpleeghuiszorg in Nederland in 2005/2006. In totaal hebben we 615 onderbouwingen en oordelen onderzocht.

Data-analyse

De onderbouwingen van de oordelen over deze vier criteria zijn per criterium geanonimiseerd en a-select in een tekstdocument verwerkt. Voor elke onderbouwing uit dit tekstdocument hebben twee (onafhankelijk werkende) beoordelaars geanalyseerd welke argumenten uit de IGZ-normen door de betrokken inspecteurs zijn genoemd om tot hun feitelijke oordeel te komen. De eerste beoordelaar, de eerste auteur van dit artikel, heeft wel onderzoekservaring maar geen inspecteurservaring. De tweede beoordelaar is als inspecteur werkzaam geweest bij het toezicht op de verpleeghuiszorg en was bekend met de gebruikte criteria en bijbehorende IGZ-normen. De beoordelaars hebben vervolgens bepaald welk oordeel de inspecteur bij deze argumenten had moeten geven als de IGZ-norm strikt gehanteerd zou zijn (voorgeschreven oordeel). Dit maakt een vergelijking mogelijk tussen het feitelijke oordeel zoals door de inspecteurs gegeven en het voorgeschreven oordeel zoals dat op basis van de argumentatie en de IGZ-normen toegekend zou moeten worden.

Zo moet volgens de IGZ-normen het oordeel over “decubitus” alleen als “afwezig” gescoord worden als er geen registraties van decubitus plaats vinden (tabel 1). Het oordeel “aanwezig” wordt gegeven als er registratie van decubitus plaatsvindt en de instelling een decubitusprotocol heeft. Het oordeel “operationeel” wordt gegeven als er een decubitusprotocol is, als registratie van decubitus plaatsvindt, als de instelling personeel schoolt in de omgang met decubitus en als het protocol bekend is bij personeel. Het oordeel “geborgd” wordt gegeven als de instelling een decubitusprotocol heeft, decubitus registreert en het protocol en de systematiek van registreren systematisch evalueert en zo nodig bijstelt (tabel 1). Als in de onderbouwing in een inspectierapport staat dat “er met betrekking tot decubitus een geplastificeerde kaart is opgesteld, maar dat deze kaart nog niet bekend is bij diverse gesprekspartners”, dan is dit door de beoordelaars gescoord als “afwezig” omdat er in de onderbouwing geen argumenten staan over de aanwezigheid van een

Tabel 1 Overzicht van de vier onderzochte criteria die in 2005/2006 gebruikt werden voor het toezicht op de verpleeghuiszorg en de bijbehorende IGZ-normen.

Criteria	Beschrijving	Afwezig	Aanwezig	Operationeel	Geborgd
Zorgzwaartemeting	De organisatie hanteert een registratiesysteem voor de zorgzwaarte	Er is geen zicht op de zorgzwaarte en/of de zorgzwaarte wordt niet volgens een vaste methodiek geregistreerd	De zorgzwaarte wordt volgens een vaste methodiek geregistreerd	De zorgzwaartegegevens worden gebruikt als sturingsinstrument	De zorgzwaartegegevens worden gebruikt als sturingsinstrument én het zorgzwaarteregistratiesysteem wordt systematisch geëvalueerd en zo nodig bijgesteld
Decubitus	De instelling heeft een protocol voor de preventie, behandeling en registratie van decubitus. Het protocol voldoet aan de CBO richtlijn. Het protocol en de methodiek van registratie wordt periodiek geëvalueerd en bijgesteld.	Decubitus wordt niet geregistreerd	Decubitus wordt geregistreerd en er bestaat een decubitusprotocol	Decubitus wordt geregistreerd metingen vinden plaats. Het protocol is bekend en scholing vindt plaats	De systematiek van de metingen én het protocol worden systematisch geëvalueerd en zo nodig bijgesteld
Permanent toezicht	Er is permanente aanwezigheid van een geïnstrueerd persoon voor toezicht t.b.v. psychogeriatrisch geïndiceerde cliënten in huiskamers.	Er is geen richtlijn aanwezig over adequaat toezicht op PG geïndiceerde cliënten in de huiskamers door medewerkers en/of geïnstrueerde personen	Er is/zijn richtlijnen aanwezig. En er is een richtlijn over de instructie aan diegenen die toezicht houden. En deze richtlijn is niet bekend bij medewerkers of van deze richtlijn wordt meerdere keren (meer dan 10% van de tijd) afgeweken. Of vrijwilligers/familieleden die worden belast met het toezicht op cliënten in de huiskamer krijgen geen instructie	Er is/zijn richtlijnen aanwezig. En deze richtlijn is bekend bij medewerkers of van deze richtlijn (wordt zelden (minder dan 10% van de tijd) afgeweken. En vrijwilligers/familieleden die worden belast met het toezicht op cliënten in de huiskamer krijgen instructie	De richtlijnen zelf én de uitvoering van de richtlijnen worden periodiek geëvalueerd en zo nodig bijgesteld
Voldoende geïnstrueerde hulp bij eten en drinken	Er zijn voldoende geïnstrueerde medewerkers/familieleden/vrijwilligers aanwezig voor de hulp bij eten en drinken van cliënten.	Er is geen richtlijn over de aanwezigheid van voldoende geïnstrueerde medewerkers/familieleden/vrijwilligers voor hulp bij eten en drinken van cliënten	Er is/zijn richtlijnen over de aanwezigheid van voldoende geïnstrueerde medewerkers/vrijwilligers/familieleden voor hulp bij eten en drinken van cliënten. De richtlijnen zijn niet bekend bij medewerkers en/of de bezetting voldoet structureel (meer dan 10%) niet aan de richtlijn. Of vrijwilligers/familieleden die betrokken zijn bij het helpen van cliënten bij het eten en drinken krijgen geen instructie.	Er is/zijn bij medewerkers richtlijnen bekend over de aanwezigheid van voldoende geïnstrueerde medewerkers/familieleden/vrijwilligers voor hulp bij eten en drinken van cliënten. En er wordt niet of minder dan 10% afgeweken van de richtlijn	De richtlijnen en de uitvoering van de richtlijn met betrekking tot de aanwezigheid van voldoende geïnstrueerde medewerkers/vrijwilligers/familieleden voor hulp bij eten en drinken van cliënten wordt periodiek geëvalueerd en zo nodig bijgesteld

decubitusprotocol en over registratie van decubitus. Op deze manier hebben de beoordelaars per onderbouwing van de oordelen voor de vier criteria gescoord welke argumenten uit de IGZ-normen aanwezig zijn en welk oordeel hierop volgens de IGZ-norm van toepassing had moeten zijn. Bij het eventueel ontbreken van argumenten is de beslisregel gehanteerd 'niet aanwezig' te scoren.

Bij de onderbouwingen waarover geen overeenstemming was tussen beide beoordelaars, is door middel van discussie tussen de twee beoordelaars consensus bereikt (tabel 2). Hierbij waren de argumenten van de beoorde-

laar met werkervaring als inspecteur doorslaggevend. Tabel 2 laat zien dat er bij de criteria "passende hulp bij eten en drinken" en "adequaat toezicht" sprake was van een relatief hoge overeenstemming tussen de twee beoordelaars (85% en 93%) terwijl dat bij de criteria "zorgzwaartemeting" en "decubitus" in iets mindere mate het geval was (62% en 64%). We hebben onderzocht in hoeveel gevallen de voorgeschreven oordelen op grond van de IGZ-normen verschilden van de feitelijke inspecteursoordelen. Ook hebben we onderzocht of er een verband is tussen de wijze waarop inspecteurs van de

IGZ-norm afwijken en het gemiddelde oordeel van inspecteurs.

RESULTATEN

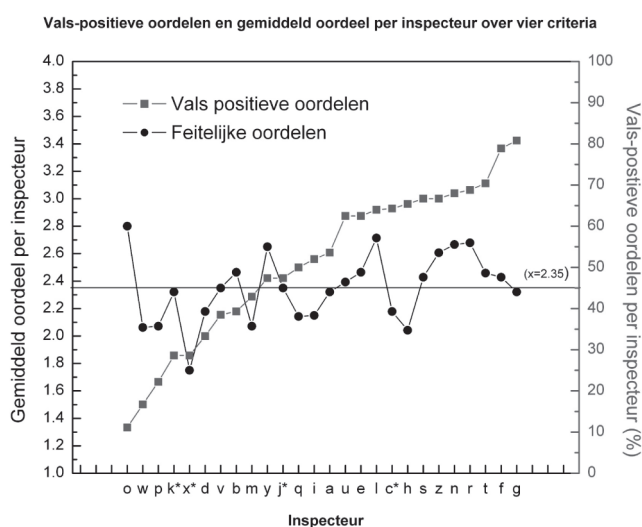
Uit de analyse blijkt dat de betekenis van de feitelijke oordelen zeer divers is. Dit is bijvoorbeeld het geval bij oordelen “geborgd” op het criterium “voldoende geïnstrueerde hulp bij eten en drinken”. Inspecteurs kennen aan dit oordeel onder andere de volgende betekenissen toe:

- “De personele inzet tijdens de maaltijden op de verpleegunits en afdeling meerzorg in het verzorgingshuis zijn niet expliciet vastgelegd.”
- “Het beleid rondom het zorgonderdeel eten en drinken, evenals de werkwijze en de registratie ervan in het zorgplan, zijn uitvoerig en overzichtelijk HKZ-protocolair vastgelegd. Het weegbeleid maakt daarvan onderdeel uit. Het protocol vermeldt de verantwoordelijkheden van de bij dit zorgonderdeel betrokken personeelsleden van de instelling. Het personeel is bekend met de procedures en werkwijze.”
- “Een richtlijn over de aanwezigheid van passende hulp bij eten en drinken van cliënten is aanwezig en is in de praktijk geborgd.”

Ook blijkt uit de analyse dat op basis van de aanwezige argumenten in de onderbouwingen, de feitelijke oordelen niet altijd aansluiten bij de gehanteerde IGZ-normen. Op grond van de geanalyseerde onderbouwingen blijkt in 52% van de gevallen dat het feitelijke oordeel hoger is dan het op grond van de IGZ-normen voorgeschreven oordeel had mogen zijn. In 1% van de gevallen is sprake van het omgekeerde en is het feitelijke oordeel lager dan het voorgeschreven oordeel (tabel 3). Deze oordelen worden vals-positieve, respectievelijk vals-negatieve oordelen genoemd. Vals-positieve oordelen geven ten onrechte een te positieve situatie weer terwijl er in feite sprake is van een negatievere situatie. Het omgekeerde geldt voor vals-negatieve oordelen.⁴ In tabel 3 zijn de vals-positieve oordelen en de vals-negatieve oordelen weergegeven. Van deze vals-positieve oordelen is 22% twee of meer categorieën hoger dan het voorgeschreven oordeel. Dit zijn

bijvoorbeeld oordelen die “operationeel” zijn gescoord, terwijl het voorgeschreven oordeel “afwezig” was. Van de feitelijke oordelen “operationeel” en “geborgd” is 34% vals-positief en had als “afwezig” of “aanwezig” gescoord moeten worden.

Uit de analyse blijkt dat het geven van vals-positieve oordelen niet persoonsgebonden is ($x=51,1$, mediaan=52,8, $Sd=19,3$). Het percentage vals-positieve oordelen verschilt tussen inspecteurs (minimum=11,1, maximum=80,0). In figuur 1 is zowel het percentage vals-positieve oordelen per inspecteur als het gemiddelde oordeel over de vier criteria weergegeven voor alle inspecteurs ($x=2,35$). Dit is het gemiddelde oordeel over de criteria “zorgzwaartemeting”, “decubitus”, “adequaat toezicht” en “hulp bij eten en drinken”. Zichtbaar is dat alle inspecteurs vals-positieve oordelen hebben gegeven op de vier onderzochte criteria. Ook inspecteurs die in opleiding waren, hebben regelmatig vals-positieve oor-



Figuur 1 Percentage vals-positieve oordelen per inspecteur op de criteria “zorgzwaartemeting”, “decubitus”, “adequaat toezicht” en “hulp bij eten en drinken” ($x=51,1$, mediaan=52,8, $Sd=19,3$) uitgezet tegen het gemiddelde oordeel over deze criteria per inspecteur. *= inspecteur in opleiding

Tabel 2 Overeenstemming tussen eerste en tweede beoordelaar in eerste beoordelingsronde bij beoordeling van de aanwezige argumenten in de onderbouwingen van vier criteria die in 2005/2006 gebruikt werden in het Gefaseerd Toezicht op de verpleeghuiszorg.

	Zorgzwaartemeting (n=168)	Decubitus (n=151)	Adequaat toezicht (n=145)	Passende hulp bij eten en drinken (n=151)
% overeenstemming na 1 ^e ronde	62%	64%	93%	85%

Tabel 3 Door de 2 onafhankelijk beoordelaars voorgeschreven oordelen op basis van gerapporteerde onderbouwingen en strikte hantering van de IGZ-normen, afgezet tegen de feitelijke inspecteursoordelen in 2005/2006.

Voorgeschreven oordeel 2008	Feitelijk inspecteursoordeel 2005/2006		Operationeel	Geborgd	Totaal
	Afwezig	Aanwezig			
Afwezig	176	107*	97*	19*	399
Aanwezig	4**	76	60*	23*	163
Operationeel	0**	1**	30	12*	43
Geborgd	0**	1**	2**	7	10
	180	185	189	61	615

**Vals negatieve oordelen (1%)

*Vals positieve oordelen (52%)

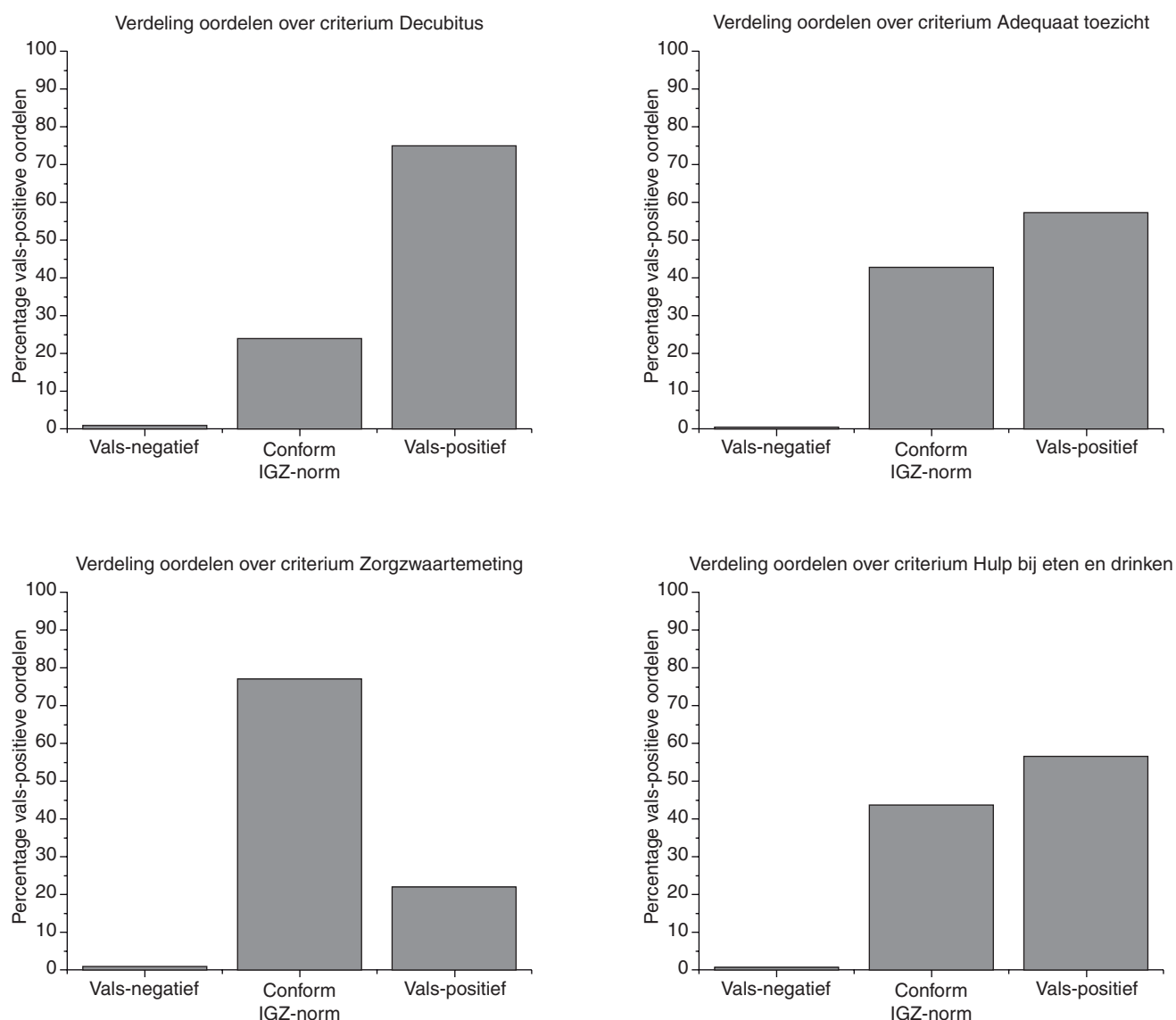
delen gegeven. Figuur 1 laat zien dat er geen verband is tussen de hoogte van het oordeel en de hoogte van het percentage vals-positieve oordelen: inspecteurs die lager dan gemiddeld oordelen, hebben zowel relatief hoge als lage percentages vals-positieve oordelen. Ook het omgekeerde is te zien: inspecteurs die hoger dan gemiddeld oordelen, hebben naast hoge, ook relatief lage percentages vals-positieve oordelen.

Wel verschillen de percentages vals-positieve oordelen per criterium (figuur 2). Bij de criteria “passend toezicht” en “hulp bij eten en drinken” is het percentage vals-positieve oordelen bijna gelijk (57% en 56%). Bij het criterium “decubitus” is het percentage vals-positieve oordelen hoger (75%) en bij het criterium “zorgzwaarte” is het percentage vals-positieve oordelen het laagst (22%).

BESCHOUWING

De onderzoeksresultaten laten zien dat er sprake is van een validiteitsprobleem bij de inspecteursoordelen. Gelijke oordelen hebben verschillende betekenissen en niet alle onderbouwingen van oordelen sluiten aan bij de

IGZ-normen. Hierdoor zijn de feitelijke inspecteursoordelen in meer dan de helft van de gevallen als ‘on nauwkeurig’ te beschouwen. Deze onnauwkeurigheid kenmerkt zich hoofdzakelijk door vals-positieve oordelen. Inspecteurs oordelen dus over het algemeen te positief. Er is geen verband tussen de hoogte van het oordeel en het percentage vals-positieve oordelen. Of inspecteurs tijdens de bezoeken al dan niet in opleiding waren, heeft geen invloed op het geven van vals-positieve oordelen. Omdat vals-positieve oordelen bij alle inspecteurs in meer of mindere mate voorkomen, maar bij sommige criteria meer dan bij andere, is het voorkomen van vals-positieve oordelen criteriumspecifiek en daarmee gerelateerd aan het toezichtsinstrument. Als het toezichtsinstrument minder eenduidig is, bestaat er voor inspecteurs de mogelijkheid eigen accenten in de oordeelsvorming te leggen. Omdat alle inspecteurs op hoofdzakelijk dezelfde wijze de ruimte tussen het instrument en de werkelijkheid invullen door voornamelijk te positieve oordelen, is dit gedrag een situatiegebonden kenmerk van inspecteurs.



Figuur 2 Overzicht van percentage oordelen conform de IGZ-norm, vals-negatieve oordelen en vals-positieve oordelen bij vier onderzochte criteria uit het Gefaseerd Toezicht op de verpleeghuiszorg in 2005/2006.

Instrument gerelateerde verklaringen

In het beoordelingsproces in de tweede fase van het GT gebruiken inspecteurs een officieel document met de te beoordelen criteria en de daarbij horende IGZ-normen. Dit inspectie-instrument dient als een richtlijn voor alle inspecteurs om tot gelijke oordelen te komen. Ook in het beste instrument zal altijd enige ruimte bestaan tussen het instrument en de praktijksituaties die inspecteurs in zorginstellingen tegenkomen. Inspecteurs hebben dan aanvullende argumenten nodig om tot oordelen te komen. Het is het vakmanschap van de inspecteur om ook in die gevallen tot een professioneel oordeel te komen. Als een instelling bijvoorbeeld geen decubitusregistratie heeft, moet dit volgens de IGZ-norm beoordeeld worden als “afwezig” en als de instelling decubitus registreert en over een decubitusprotocol beschikt, is het oordeel “aanwezig” van toepassing. Als een instelling alleen over een decubitusprotocol beschikt, dan voorziet het inspectie-instrument niet in een passend oordeel omdat de instelling noch aan “afwezig” noch aan “aanwezig” voldoet. De ene inspecteur vindt de betekenis van “aanwezig” het meest van toepassing terwijl de andere voor “afwezig” kiest. Doordat de gehanteerde IGZ-normen niet naadloos aansluiten op de werkelijkheid, ontstaat er ruimte om een instelling het voordeel van de twijfel te geven met als gevolg over het algemeen te positieve oordelen.

Niet alleen de mate waarin het inspectie-instrument aansluit op de praktijksituaties in zorginstellingen, maar ook de eenduidigheid van de IGZ-normen is een mogelijke verklaring voor het gevonden validiteitsprobleem. Daar waar operationalisaties van normen voor meerdere uitleg vatbaar zijn, is de kans op validiteitsproblemen groter omdat normen dan op verschillende manieren worden gehanteerd. Zowel de mate van beoordelaars-overeenstemming tussen de twee onafhankelijke beoordelaars over de aanwezige argumenten in onderbouwingen, als het percentage vals-positieve oordelen zijn indicaties voor de eenduidigheid van de IGZ-normen voor de inspecteurs: hoe hoger de overeenstemming en hoe lager het percentage vals-positieve oordelen, hoe eenduidiger de norm.

Bij de criteria “decubitus”, “adequaate toezicht” en “passende hulp bij eten en drinken” wordt dit verband tussen de mate van beoordelaarsovereenstemming en het percentage vals-positieve oordelen bevestigd. Bij twee van deze criteria (“adequaate toezicht” en “passende hulp bij eten en drinken”) is er sprake van een relatief hoog percentage overeenstemming tussen de beoordelaars (85% en 93%) en een relatief laag percentage vals-positieve oordelen (57% en 56%). Deze twee criteria lijken door beide beoordelaars op gelijke wijze toegepast te worden, waardoor zij in hoge mate met elkaar overeenstemmen. Bij deze twee criteria is het percentage vals-positieve oordelen relatief laag. Dit wijst erop dat de normen voor deze twee criteria relatief eenduidig zijn. Ook bij het criterium “decubitus” zijn de mate van beoordelaars-overeenstemming (64%) en het percentage vals-positieve oordelen (75%) een aanwijzing voor de eenduidigheid van de norm. In dit geval is er sprake van een relatief laag percentage beoordelaarsovereenstemming en een relatief

hoog percentage vals-positieve oordelen. Deze uitkomst bevestigt dat de mate van beoordelaarsovereenstemming en het daaraan gerelateerde percentage vals-positieve oordelen indicatief is voor de eenduidigheid van een norm.

Bij het criterium “zorgzwaartemeting” wordt dit verband niet bevestigd: daar is sprake van een relatief laag percentage beoordelaarsovereenstemming en een relatief laag percentage vals-positieve oordelen in plaats van een relatief hoog percentage vals-positieve oordelen. Deze uitkomst kan mogelijk verklaard worden door de manier waarop de beoordelaars tot consensus zijn gekomen over de verschillen in beoordeling van argumenten in onderbouwingen bij dit criterium. Als in de onderbouwing stond dat “er zorgzwaartemetingen worden uitgevoerd” is dit gescoord als “de zorgzwaarte wordt volgens een vaste methodiek geregistreerd”. Hierbij hebben de beoordelaars “volgens een vaste methodiek” gedefinieerd als “met een systeem voor zorgzwaartemeting”. Deze definitie is bepaald aan de hand van hoe inspecteurs in de dagelijkse toezichtpraktijk omgaan met de norm. Het percentage vals-positieve oordelen is met deze definitie relatief laag.

Inspecteur gerelateerde verklaringen

Naast de kwaliteit van het gehanteerde instrument, speelt compliance van inspecteurs met het instrument een belangrijke rol. Als de compliance laag is, is de kans groter dat inspecteurs het instrument op eigen wijze hanteren en de inspecteursoordelen een inspecteursafhankelijke betekenis krijgen. Daarnaast kan het eigen referentiekader van inspecteurs een belangrijke rol spelen. Aspecten die niet zijn opgenomen in de normen van de IGZ maar die volgens de inspecteur wel van belang zijn voor de kwaliteit van zorg, kunnen van invloed zijn op zijn of haar oordeelsvorming. Hygiëne is bijvoorbeeld geen criterium in de IGZ-normen, maar kan door de inspecteurs wel in hun feitelijke oordeel over de instelling betrokken zijn. Het oordeel over hygiëne heeft dan het feitelijke oordeel over een ander criterium dat wel in de IGZ-normen was opgenomen beïnvloed. Dat experts ook andere dan de voorgeschreven informatie over criteria in hun beoordeling betrekken, is een bekend verschijnsel dat kan leiden tot oordelen die afwijken van de officiële voorschriften.^{6,7} Als beoordelaars zich bij de beoordeling van afzonderlijke criteria niet houden aan specifieke voorschriften over die criteria maar zich laten leiden door hun globale indruk van het te beoordelen object, is er sprake van het halo-effect.⁷ Uit onderzoek naar het beoordelen van schriftelijk werk door leraren, blijkt dat een positieve verwachting of een gunstig beeld van een student, ertoe kan leiden dat goede prestaties meer in het oordeel worden betrokken dan minder goede prestaties.⁸

Naast bovengenoemde punten kunnen de vals positieve oordelen ook verklaard worden door het type handhavingstijl (coöperatief of repressief) dat de inspecteur gebruikt.^{9,10,11} Toezichhouders gebruiken lang niet altijd de aan hen toegekende formele bevoegdheden. Om de inspectie te beïnvloeden gebruiken inspecteurs bij voorkeur informele middelen zoals overleggen, advise-

ren, waarschuwen en dreigen. In elke toezichtsrelatie bestaat een tendens tot vermindering van de afstand tussen inspecteur en onder toezicht gestelde. De inspecteur wordt adviseur, partner, vraagbaak en/of vertrouwenspersoon. Harmonie is prettiger dan confrontatie en is mogelijk effectiever in het bereiken van de doelstelling van het toezicht.⁹ In tegenstelling tot de repressieve handhavingsstijl waarbij veel gecontroleerd en gesancioneerd wordt, kenmerkt de coöperatieve stijl zich door overleg om de betreffende instantie de regels na te laten leven.¹⁰ Een coöperatieve handhavingsstijl is gericht op het bewaken van de goede relatie tussen inspectee en inspecteur.¹⁰ Problemen van een coöperatieve handhavingsstijl zijn het “humaness-problem”,¹² “client intimacy”, “capturing”¹³ en “the shadow of the future”.¹⁴ De inspecteur wordt door zijn betrokkenheid bij de onder toezicht gestelde, ingekapseld en oordeelt daardoor niet meer objectief,^{13,15} of de inspecteur wil de relatie met de instelling niet schaden omdat ze ook in de toekomst nog samen verder moeten.¹⁴ In een coöperatieve handhavingsstijl oordeelt de inspecteur eerder te positief dan te negatief.¹²

Implicaties en oplossingen validiteitsprobleem

De IGZ gebruikt het GT om risicogericht toezicht te houden. Instellingen waar risico's in de zorg vermoed worden, bezoekt de IGZ om ter plaatse de kwaliteit van zorg in kaart te brengen en deze te stimuleren. De vals-positieve oordelen beperken de transparantie en controleerbaarheid van het toezicht. Daarnaast leidt het validiteitsprobleem mogelijk tot een effectiviteitsverlies van het toezicht. Als gevolg van de vals-positieve oordelen blijven verbetermaatregelen achterwege.

Ligt de oplossing voor het terugdringen van validiteits- en betrouwbaarheidsproblemen dan in het ontwikkelen van een instrument waarmee de inspecteur geen speelruimte meer heeft? Als het al mogelijk zou zijn om een instrument te ontwikkelen waarin inspecteurs geen discretionaire ruimte hebben, dan is het geen wenselijke oplossing voor bovengenoemde problemen: niet alleen wordt de complexe werkelijkheid van de zorg hiermee tekort gedaan, maar ook de professionaliteit van de inspecteur.¹⁶

Een werkelijkheid die meerdimensionaal is, past niet in een eendimensionaal instrument.¹⁶ Instrumenten waarbij een te eenzijdig accent ligt op het meetbare kunnen tot gevolg hebben dat politiek-normatieve afwegingen alleen worden bepaald door de mate waarin prestaties meetbaar zijn.¹⁷ Het accent komt dan te liggen op datgene wat gemeten is (bijvoorbeeld het percentage doorligwonden) in plaats van de kwaliteit of de prestaties van de zorg ter voorkoming van doorligwonden. Het risico hiervan is dat verantwoording vooral een manier wordt om instellingen te dwingen om meetbare prestaties te noteren, wat slechts een schrale afspiegeling kan zijn van de werkelijkheid.¹⁸ De werkelijkheid is complexer dan het instrument en de inspecteur heeft in deze situatie discretionaire ruimte nodig om de situatie te beoordelen. Als die ruimte er niet is, wordt de inspecteur een enquê-

teur en wordt de zorg teruggebracht naar een eendimensionale werkelijkheid: een onwenselijke situatie.^{16,19}

De oplossing lijkt te liggen in een combinatie van factoren. Een goed instrument dat recht doet aan de (complexiteit van de) zorg met een bijbehorend expliciet normenkader is van groot belang.¹⁶ Ook commitment van inspecteurs aan het instrument en scholing van inspecteurs in het gebruik van het instrument zijn essentieel.^{20,21} Niet alleen gaat het dan om scholing in het gebruik van het instrument zelf, maar (vooral) ook in de omgang met de ruimte die tussen het instrument en de werkelijkheid bestaat. Het is belangrijk dat inspecteurs leren om gemotiveerd af te wijken van de normen en het instrument als de situatie daarom vraagt.

Beperkingen van dit onderzoek

Bij de gebruikte onderzoeksmethode kunnen kritische kanttekeningen geplaatst worden. In dit onderzoek hebben we alleen de gerapporteerde argumenten van inspecteurs betrokken. Zo kunnen inspecteurs in het toezichtsbezoek wel gesproken hebben over de aanwezigheid van hulp bij de maaltijden van bewoners door geïnstrueerde vrijwilligers, maar dit niet hebben opgeschreven in de onderbouwing. Argumenten die wel een rol hebben gespeeld bij de oordeelsvorming maar die niet zijn gerapporteerd, kunnen ertoe geleid hebben dat de gerapporteerde onderbouwing en het feitelijke oordeel niet overeenkomen met de IGZ-normen. In dit onderzoek is bij het ontbreken van argumenten de beslisregel gehanteerd 'niet aanwezig' te scoren. Mogelijk heeft dit geleid tot een overschatting van het percentage van vals-positieve oordelen. Omdat zowel hoofdzakelijk op dezelfde manier onnauwkeurig geoordeeld wordt (vals-positief) en bijna een kwart van de vals-positieve oordelen twee of meer dan twee schaalpunten hoger is dan het voorgeschreven oordeel is aannemelijk dat het hier daadwerkelijk de nauwkeurigheid van oordelen betreft en niet de nauwkeurigheid van rapporteren.

Vervolg onderzoek

Onduidelijk is nog welke andere factoren mogelijk een rol spelen in de beoordelaarsverschillen van IGZ-inspecteurs. Eerder onderzoek laat zien dat ook handhavingsstijlen van inspecteurs een rol spelen bij de oordeelsvorming.²²⁻²⁴ Uit onderzoek naar interdoktervariatie bleken organisatieomstandigheden van invloed te zijn op beoordelaarsverschillen tussen artsen.²⁵ We onderzoeken welke factoren, naast handhavingsstijlen, inspecteursvariabelen en de kwaliteit van het inspectie-instrument, van invloed zijn op beoordelingsvariatie bij IGZ-inspecteurs. Door een systematische literatuurstudie onderzoeken we welke interventies effectief zijn om beoordelaarsverschillen te verkleinen. In het vervolg op deze studie onderzoeken we deze interventies in de toezichtspraktijk.

DANKBETUIGING

Mevr. K.N. Middendorp-Kolenbrander (Hoofd IGZ-Loket) werkte als tweede beoordelaar mee aan dit onderzoek.

Begeleidingscommissie: Mevr. Dr. C.A.J. Ketelaars

(Coördinerend inspecteur ketenzorg IGZ), Dhr. Drs. E. van Ankum (programmameider Specialistische Somatische Zorg IGZ) en Dhr. Drs. P. van Dyk (senior adviseur IGZ).

ABSTRACT

The relation between the employment of standards and judgment in supervision on healthcare

Objective: To evaluate the validity of judgments given by inspectors working with risk based supervision for the Dutch Health Care Inspectorate. **Design:** Descriptive, qualitative, retrospective study. **Method:** A total of 615 foundations of arguments from 26 inspectors of the Dutch Healthcare inspectorate were collected from 182 supervision reports. These judgments were given to 4 criteria which were specified for nursing homes in 2005/2006: "pressure ulcers", "adequate help during meals", "permanent supervision in living rooms" and "registration of severity of care". The precision of the foundations of arguments was analyzed by two independent observers. A comparison between those arguments and the conventions designed by the Dutch Health Care Inspectorate was made to determine whether the arguments correspond to those conventions. **Results.** Within the source of 615 foundations of arguments, 53% of the foundations were inaccurate. In 52% of the cases, the judgments inspectors gave were identified as false-positive and 1% as false-negative. No dependency has been found between the percentage of false-positive judgments and the average judgment of the inspector. The percentages of false-positive judgments differ between criteria. **Conclusion:** A problem in the validity of judgments has been demonstrated. Giving false-positive judgments is not a personal characteristic of an inspector: all inspectors show this type of behavior, even inspectors who were receiving training. The occurrence of false-positive judgments seems to depend on the criteria that are being evaluated. Therefore, giving false-positive judgments depends on the quality of the instrument.

Keywords: supervision interobserver disagreement, validity, evaluative survey

LITERATUUR

1. Tuijn SM, Janssens FJG, Bergh H van den, Robben PBM. Het ene oordeel is het andere niet: Interinspecteursvariatie bij inspecteurs van de IGZ: een kwantitatieve analyse. Ned Tijdschr Geneesk 2009;153:322-6.
2. Uebersax JS. Validity inferences from interobserver agreement, Psychol Bull 1988;104:405-16.
3. Nijveldt MJ. Validity in Teacher Assessment. An exploration of the judgement processes of assessors. Proefschrift. Enschede: Gildeprint, 2007.
4. Mehrens WA, Lehmann IJ. Measurement and Evaluation in Education and Psychology. New York: Holt, Rinehart and Winston, 1973.
5. Bergh H van den, Zwarts M, Peter-Sips M. Kwaliteit van het onderwijsleerproces, Tijdschr Onderwijsresearch 2000; 25:20-39.
6. Geath GJ, Shanteau J. Reducing the influence of irrelevant information on experienced decision makers, Organizat Behav Human Perform 1984;33:263-82.
7. Meuffels B, Ansink M, Donselaar J van. Een bijzonder halo-effect bij het beoordelen van opstellen, Tijdschr Taalbeheersing 1986;8:177-93.
8. Zillig M. Einstellung und Aussage. Zeitschr Psychol 1928;106: 54-106.
9. Ridder J de. Een goede raad voor toezicht. Den Haag: Boom Juridische Uitgevers, 2004.
10. Huisman W, Tussen winst en moraal. Proefschrift. Den Haag: Boom Juridische Uitgevers, 2001.
11. Day P, Klein R. The regulation of nursing homes: a comparative perspective, Milbank Q 1987;65:303-47.
12. Albanese MA. Challenges in using rater judgements in medical education, J Eval Clin Pract 2000;6:305-19.
13. Bruijn H de, Heuvelhof E ten. Handhaving. Het spel tussen inspecteur en inspectee. Utrecht: Lemma, 2005.
14. Leeuw FL. Gedragsmechanismen achter overheidsinterventies en rechtsregels. Oratie. Maastricht: Faculteit der Rechtsgeleerdheid van de Universiteit Maastricht, 2008.
15. Shover N. Zelfregulering door ondernemingen. Ontwikkeling, beoordeling en bange voorgevoelens, Tijdschr Criminologie 2008;50:169-81.
16. Janssens FJG. Van onderzoek naar evaluatie. De methodologie van de onderwijsinspectie. Den Haag: Delwel, 1997.
17. WRR. Bewijzen van goede dienstverlening, Den Haag: Wetenschappelijke Raad voor het Regeringsbeleid, 2004 p.198.
18. Tonkens EH. Mondige burgers, getemde professionals. Marktwerking, vraagsturing en professionaliteit in de publieke sector. Utrecht: Nizw, 2003.
19. Weiss DJ, Shanteau J. The vice of consensus and virtue of consistency. Psychological Investigations of Competence in Decision Making. Cambridge: Cambridge University Press, 2004.
20. Francke AL, Smit MC, Veer AJE de, Mistiaen P. Factors influencing the implementation of clinical guidelines for health care professionals: A systematic meta-review. BMC Med Informat Decis Making. 2008;38.
21. Ploeg J, Davids B, Edwards N, Gifford W, Miller PE. Factors Influencing Best-Practice Guideline Implementation: Lessons Learned from Administrators, Nursing Staff, and Project Leaders. Worldviews on Evidence-Based Nursing 2007;4:210-19.
22. Gray C, Gardner J. The impact of school inspections, Oxford Rev Educ 1999;25:455-69.
23. Ehren MCM. Toezicht en schoolverbetering. Proefschrift. Delft: Eburon, 2006.
24. Kist S, Hutschemaekers GJM. Beroep inspecteur in de gezondheidszorg, Den Haag: Inspectie voor de Gezondheidszorg, 2006.
25. Jong J de. Explaining medical practice variation. Proefschrift. Enschede: Print Partners Ipskamp BV, 2008.

CORRESPONDENTIEADRES

Mevr. drs. S.M. Tuijn, Inspectie voor de Gezondheidszorg, Postbus 2680, 3500 GR Utrecht, tel: 030-2338727, e-mail: s.tuijn@igz.nl