



(deel 5)

# Systematisch review van diagnostisch onderzoek

*dr. Hester Vermeulen*  
projectleider afdeling Kwaliteit en Proces Innovatie van het AMC Amsterdam, docent EBP en lid lectoraat EBP bij de Amsterdam School of Nursing, Hogeschool van Amsterdam

*dr. Dirk Ubbink*  
programmameerder afdeling Kwaliteit en Proces Innovatie van het AMC Amsterdam

*dr. Lotty Hoof*  
Dutch Cochrane Centre, AMC Amsterdam

Dit is het vijfde artikel uit een serie van zeven artikelen over het beoordelen van diverse vormen van medisch-wetenschappelijke publicaties.<sup>1-4)</sup> De reden voor deze serie is om verpleegkundigen te helpen bij het efficiënt en kritisch beoordelen van wetenschappelijke literatuur op hun vakge-

bied. Het bijhouden van vakliteratuur is voor velen lastig omdat hiervoor eigenlijk (te) weinig tijd is. Daarnaast vertrouwt men vaak op de eigen ervaring en routine. Toch is het goed om te realiseren dat we, ondanks onze opgebouwde expertise, vele malen per dag keuzes maken, zonder zeker te weten wat de juiste keuze is.

» Een voorbeeld: kun je met een korte vragenlijst in de eerstelijns situatie screenen op depressie?<sup>5)</sup> Zijn er accurate en praktische screeningstesten voor dementie in de eerstelijns situatie?<sup>6)</sup> Wat is de diagnostische waarde van een fluïdertest bij zelfgerapporteerd gehoorverlies?<sup>7)</sup> Wat is de diagnostische waarde van lichamelijk onderzoek bij het diagnosticeren van verschillende soorten incontinentie?<sup>8)</sup> Als u zich deze onzekerheden realiseert, kunt u deze proberen op te lossen door gebruik te maken van bewijsmateriaal uit de vakliteratuur, zoals het EBP-principe ook bepleit.

Het doel van deze 'toolbox' is om de waarde van systematische reviews van onderzoek naar de waarde van een diagnostische test te beoordelen. In dergelijke reviews worden een of meer inductieve (de testen waarvan de diagnostische waarde geëvalueerd wordt) vergeleken met een referentietest (ook wel gouden standaard genoemd).<sup>4)</sup> Vanwege de complexiteit van dit onderwerp is het aan te bevelen om iemand met expertise op het gebied van systematische reviews van diagnostisch onderzoek bij de beoordeling te betrekken. Een kritische beoordeling van een willekeurige studie bestaat altijd uit

drie belangrijke punten. Ten eerste of de informatie in een studie valide is (dat wil zeggen antwoord kan geven op de vraagstelling), ten tweede wat, hoe groot en hoe precies de uitkomsten van de studie zijn en ten derde of deze toepasbaar zijn in uw eigen situatie.

### Checklist

Een checklist om deze drie punten te beoordelen is onder andere te vinden op de website van het Dutch Cochrane Centre<sup>9)</sup>, onderdeel van de Cochrane Collaboration. Deze checklist is bestemd voor het beoordelen van systematische reviews van onderzoek naar de waarde van een diagnostische test waarin een of meer indextesten worden vergeleken met een referentietest. Net als de eerder behandelde checklists<sup>1-4)</sup> bestaat deze uit een aantal vragen om de validiteit van de review te kunnen beoordelen. Deze vragen komen gedeeltelijk overeen met de vragen die gesteld worden over een systematisch review van interventiestudies (zie tabel 1).<sup>2)</sup> Indien het algemeen oordeel over de validiteit voldoende is, biedt het formulier de mogelijkheid om kritisch naar de resultaten te kijken. In dit soort studies worden resultaten voornamelijk weergegeven in sensitiviteit, specificiteit en voorspellende waarden.<sup>4)</sup> De lijst sluit af met vragen over de toepasbaarheid.

Ter illustratie van deze checklist gebruiken wij een systematisch review van Holsinger over het screenen van patiënten op dementie.<sup>6)</sup> Dit artikel is verschenen in het gezaghebbende JAMA. Dementie is ook in Nederland een groot probleem. Het is van belang om een goede, valide en liefst korte test te vinden om antwoord te krijgen op de volgende vraag: lijdt deze patiënt aan dementie?

### Validiteit

#### 1 Vraagstelling

Een systematisch review dient uit te gaan van een expliciete klinische vraagstelling. De vraag bevat bij voorkeur de volgende aspecten: de doelpopulatie, de indextest, de referentietest en het beoogde resultaat (bijvoorbeeld het aantonen of uitsluiten van een aandoening, een hoge voorspellende waarde van een testuitslag of een laag percentage foutnegatieven). Een leidraad kan het Engelse acroniem PICO zijn: Patient, Indextest(en), Comparison (gouden standaard of referentietest, waarvoor soms ook het beloop van de aandoening genomen wordt) en Outcome.

*In het voorbeeld is het een beetje zoeken naar de vraagstelling. Uit het artikel blijkt dat het om een oudere patiëntenpopulatie in de eerste lijn gaat. De auteurs zijn bij deze patiënten op zoek naar een accuraat en praktisch (lees kort) meetinstrument om cognitie te testen om daarmee op dementie te screenen. Deze korte test dient vergeleken te zijn met een betrouwbare referentietest die in de vraagstelling niet nader wordt omschreven.*

#### 2 Zoekactie

Vanuit de expliciete vraagstelling dienen duidelijk zoek-

Tabel 1. De checklist

Checklist voor de beoordeling van een systematisch review van diagnostisch onderzoek	Ja	Nee	?
1 Is de vraagstelling adequaat geformuleerd?			X
2 Is de zoekactie adequaat uitgevoerd?	X		
3 Is de selectieprocedure van artikelen adequaat uitgevoerd?		X	
4 Is de kwaliteitsbeoordeling adequaat uitgevoerd?	X		
5 Is adequaat beschreven hoe data-extractie heeft plaatsgevonden?	X		
6 Zijn de belangrijkste kenmerken van de oorspronkelijke onderzoeken beschreven?		X	
7 Is meta-analyse op een correcte manier uitgevoerd?			X

termen naar voren te komen. De zoektermen moeten zo gekozen zijn dat naar verwachting bijna alle relevante artikelen gevonden zullen worden voor de vraagstelling (sensitieve zoekactie). Hiermee moet op een gestructureerde wijze de internationale literatuur (via databases) grondig doorzocht zijn. Een minimumvereiste voor de databases is een zoekactie met PubMed in Medline en de referenties van de verzamelde artikelen. Indien niet in Medline gezocht is, hoeft u met de beoordeling niet verder te gaan. Het is nog onduidelijk of een aanvullende zoekactie in Embase of specialistische databases, zoals Cinahl en Psycit, meerwaarde heeft bij het vinden van onderzoeken over de waarde van een diagnostische test. Ook kan men nog zoeken door gedrukte indexen, het aanschrijven van experts en fabrikanten en het handmatig doorzoeken van tijdschriften. De eventuele beperkingen in de zoekactie moeten wel beschreven zijn. Hierbij kan gedacht worden aan een beperking voor wat betreft jaartal van publicatie, publicatiestatus (wel of niet gepubliceerd) of taal. Een beperking tot één taal (bijvoorbeeld alleen Engels) kan leiden tot minder precieze, en mogelijk ook vertekende, schattingen. Vanzelfsprekend moet de zoekactie aansluiten op de vraagstelling zoals beoordeeld in vraag 1.

*Voor deze review hebben de auteurs gezocht in Medline (de minimaal vereiste database) en daarnaast hebben ze in een passende specialistische database gezocht namelijk Psycinfo. Ze hebben een beperking toegepast door alleen te zoeken vanaf 2000 tot april 2006. Dit verantwoordt zij wel door op te merken dat deze review voortbouwt op een eerdere review.<sup>10)</sup>*

#### 3 Selectie

Een uitgebreide zoekactie (vraag 2) zal resulteren in een groot aantal hits van artikelen. Veel van deze hits zullen niet relevant zijn voor de review en erbuiten gelaten moeten worden (exclusie). De selectie van artikelen die wel geïncludeerd kunnen worden, dient bij voorkeur door twee beoordelaars onafhankelijk van elkaar te gebeuren aan de hand van van tevoren duidelijk beschreven in- en exclusiecriteria. Een minimum van twee beoordelaars voor de selectie is bedoeld om de kans op selectiebias te verkleinen. Onder in- en exclusiecri-

ria vallen bijvoorbeeld een specifieke onderzoekspopulatie (bijvoorbeeld patiënten in de eerste lijn), gebruik van een bepaalde referentietest enzovoort. De in- en exclusiecriteria dienen aan te sluiten op de vraagstelling en dienen expliciet beschreven te zijn om de reproduceerbaarheid van de review te verhogen. Tevens moet de selectie vanaf de eerste zoekactie tot de uiteindelijke inclusie in de systematische review duidelijk beschreven zijn.

*In het artikel worden expliciete in- en exclusiecriteria beschreven: (i) Engelstalige studies, (ii) screeninginstrumenten voor dementie die bruikbaar zijn voor huisartsen, (iii) toegepast bij mensen ouder dan 60 jaar zonder overduidelijke symptomen van dementie. In het artikel wordt niet aangegeven dat dit door twee onafhankelijke beoordelaars is gedaan.*

#### 4 Kwaliteitsbeoordeling

De geïncludeerde artikelen dienen op methodologische kwaliteit beoordeeld te zijn. Indien dit niet heeft plaatsgevonden, hoeft u niet met de beoordeling verder te gaan. Ook de kwaliteitsbeoordeling dient bij voorkeur door twee beoordelaars onafhankelijk van elkaar te gebeuren. Dit aan de hand van een expliciete lijst met criteria. Deze onafhankelijke beoordeling verkleint de kans op informatiebias. De kwaliteit van de geïncludeerde onderzoeken is mede bepalend voor de validiteit van de conclusies van en aanbevelingen in de systematische review. Er zijn verschillende criterialijsten om de kwaliteit van diagnostisch onderzoek te beoordelen. Van een aantal criteria is aangetoond dat, indien hieraan niet wordt voldaan, dit tot een overschatting of vertekening van de testkwaliteiten leidt. Het betreft de criteria over 'case control design' (een design waarin de niet-zieken niet representatief zijn voor de doelgroep, bijvoorbeeld gezonde vrijwilligers in plaats van de gangbare patiëntenpopulatie), het toepassen van verschillende referentietesten voor verschillende deelnemers, onvoldoende beschrijving van de indextest en onvoldoende beschrijving van de onderzochte populatie. De gebruikte criterialijst zal daarom ten minste deze criteria moeten bevatten. Het resultaat van de kwaliteitsbeoordeling per individuele beoordelaar zal in het artikel terug te vinden moeten zijn. De wijze waarop tot consensus gekomen wordt en de mate van overeenstemming tussen de beoordelaars dienen eveneens beschreven zijn.

*De auteurs geven duidelijk aan dat door twee onafhankelijke beoordelaars de kwaliteit beoordeeld is. Beoordeeld werden de studiegrootte, selectie van de patiënten, gebruik van een betrouwbare referentiestandaard en of deze blind en onafhankelijk van de indextest werden toegepast. Dit zijn vragen die in deel 4 in deze serie ook behandeld zijn.<sup>4)</sup>*

*De hier gebruikte set kwaliteitscriteria komt niet overeen met de minimale vragen die gescoord zouden moeten zijn. Wel wordt beschreven dat bij tegenstrijdig oordeel over een kwaliteitsitem een derde beoordelaar werd betrokken in het definitieve oordeel.*

#### 5 Data-extractie

Data-extractie is het verzamelen van de resultaten uit de oorspronkelijke onderzoeken. Gewoonlijk wordt dit in diagnostisch onderzoek samengevat in 2x2-tabellen. Indien de auteurs van de oorspronkelijke onderzoeken deze niet gemaakt hebben, dienen de beoordelaars deze 2x2-tabellen zelf te maken. Daarom dient het data-extractieproces duidelijk beschreven te zijn en weer bij voorkeur door twee beoordelaars onafhankelijk van elkaar plaatsgevonden te hebben.

*Ook dit punt is door twee beoordelaars onafhankelijk van elkaar gedaan. Uitgebreid wordt beschreven hoe de data werden geëxtraheerd en dat voor de berekening van sensitiviteit, specificiteit, voorspellende waarde en likelihood ratio's 2x2-tabellen zijn aangemaakt.*

#### 6 Beschrijving oorspronkelijke onderzoeken

Van de oorspronkelijke onderzoeken moeten in een systematische review de belangrijkste kenmerken beschreven zijn. Wat deze belangrijkste kenmerken zijn, hangt af van het onderwerp. Ten minste beschreven moeten zijn:

- De onderzochte populatie: per onderzoek ten minste de leeftijds- en geslachtsverdeling, aard, ernst en duur van de klachten, spectrum van de zieken en niet-zieken.
- De index- en referentietest: aard, wijze van uitvoering en criteria voor het duiden van het testresultaat; indien van toepassing dienen de gehanteerde afkappunten voor het benoemen van een negatieve en positieve test te zijn vermeld.
- De karakteristieken van de indextest: per indextest een samenvattende 2x2-tabel (aan de hand waarvan alle gewenste parameters berekend kunnen worden) of aantal onderzochte personen, prevalentie van de onderzochte aandoening, sensitiviteit, specificiteit, likelihood ratio van een positieve en negatieve uitslag van de indextest.

*In het artikel worden in tabel 2 de karakteristieken van de onderliggende studies beschreven. In deze tabel zijn niet alle bovengenoemde items te beoordelen, zoals leeftijd van de onderzochte populatie of de wijze van uitvoering van de test beschreven. Aangezien het een screeningtest betreft, is een goede beschrijving van de populatie cruciaal. Was het een homogene populatie? Was het een echte screeningpopulatie of is een groep geselecteerd met al bekende klachten? Beschrijving van de test is noodzakelijk voor de reproduceerbaarheid (met name zeer belangrijk bij fysische testen of beeldvormende technieken). Hier gaat het waarschijnlijk om het afnemen van een vragenlijst, dus minder relevant.*

#### 7 Meta-analyse

Omdat verschillen in resultaten (heterogeniteit) tussen de afzonderlijke onderzoeken vaak voorkomen, is het niet eenvoudig om bij diagnostische reviews tot samenvattende, preciezere schattingen van de verschillende

parameters te komen. Een diagnostisch review zal dan ook in de eerste plaats de bronnen van eventuele verschillen die de variatie kunnen verklaren moeten onderzoeken. In een aantal gevallen is het wel mogelijk om, bij voldoende homogeniteit, tot samenvattende schattingen te komen. Momenteel is er een geaccepteerde methode voor meta-analyse in een diagnostisch review; een paar jaar geleden was er nog geen overeenstemming over de ideale wijze om een diagnostische meta-analyse uit te voeren. Voor zo'n meta-analyse verwijzen we u graag naar de informatie bij de checklist van het Dutch Cochrane Centre.

*De auteurs hebben terecht besloten om geen meta-analyse uit te voeren en aldus de resultaten in een tabel weer te geven. Zij vonden dat er te grote verschillen waren in de diverse testen die onderzocht werden (er werden immers 25 verschillende meetinstrumenten onderzocht) en de verschillende onderzoeksontwerpen (niet elke studie gebruikte dezelfde testscore waarboven men sprak van dementie).*

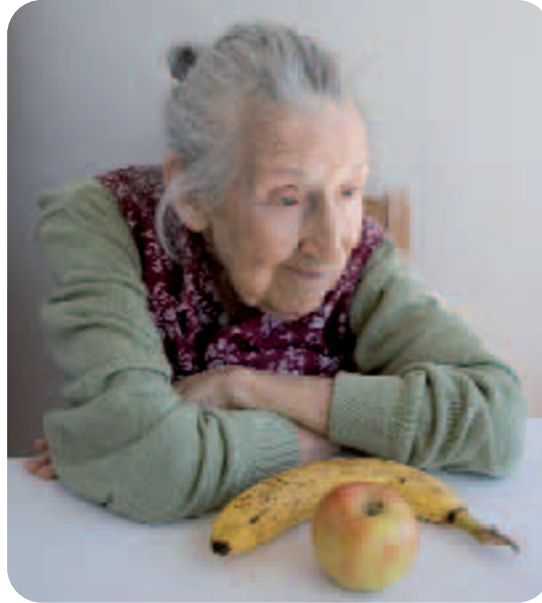
### 8 Algemeen oordeel

Hier wordt een inschatting gevraagd van de validiteit ('Is dit een goed uitgevoerde systematisch review?') en toepasbaarheid ('Gaat deze review over mijn PICO?'). Let hierbij ook op eventuele fouten in de uitvoering van de review die fataal zijn voor de validiteit ervan (zogenoemde red flags of fatal flaws). Er zijn geen regels te geven voor welke of hoeveel items positief gescoord moeten worden. Dit is deels afhankelijk van de state-of-the-art met betrekking tot het betreffende onderwerp. Het gaat hier om het samenvattende oordeel dat de beoordelaars de werkgroep zouden willen mededelen over de bruikbaarheid van het artikel voor de besluitvorming.

*Deze diagnostische review is goed uitgevoerd. Het belangrijkste minpunt is het feit dat niet geheel inzichtelijk is in wat voor soort populaties de onderliggende studies gedaan zijn. Dit is wellicht een praktische keuze van de auteurs om dit fantastische en veelomvattende werk toch op 'slechts' veertien pagina's weer te kunnen geven. Omdat het hier gaat om het afnemen van een vragenlijst, lijkt dit minpunt dus minder relevant. Echter, voor diagnostische studies moet benadrukt worden dat de beschrijving van de populatie nog vele malen belangrijker is dan voor interventiestudies. Afhankelijk van de onderzochte populatie (bijvoorbeeld patiënten die zeker of in een verder gevorderd stadium zijn) kan de test immers andere uitslagen opleveren.*

Door het beantwoorden van bovengenoemde punten krijgt u een indruk hoe degelijk ('valide') de studie is uitgevoerd. Als er minpunten zijn gevonden, moet u zich afvragen of de uitvoering van het onderzoek beter had gekund.

*Neem bijvoorbeeld de selectie van de artikelen. De onderzoekers hebben alleen Engelstalige artikelen meegenomen. Dit had beter gekund en ook de vraagstelling had scherper geformuleerd kunnen worden.*



Zijn er ernstige tekortkomingen gevonden in de validiteit die wel beter hadden gekund en gemoeten, zijn de resultaten vals of vertekend en dus niet betrouwbaar? In dat geval hoeft men niet verder te lezen. Het is dan beter om een andere publicatie te zoeken die misschien niet evenveel bewijskracht heeft, maar wel goed en dus betrouwbaar is uitgevoerd. Indien er geen (ernstige) tekortkomingen zijn gevonden in de validiteit, dan kunnen de resultaten veilig worden geïnterpreteerd.

### Resultaten

Indien er sprake is van homogene resultaten (maar dit is meer uitzondering dan regel!), dan kunnen de berekende diagnostische parameters (zoals sensitiviteit, specificiteit) uit de systematische review worden overgenomen. Neem de gepoolde schatters en hun 95%-betrouwbaarheidsinterval over (NB. In deel 4 wordt de betekenis van de verschillende diagnostische parameters toegelicht). Indien er sprake is van heterogene resultaten, dan moeten de resultaten beoordeeld worden door iemand met expertise op het gebied van systematische reviews van diagnostisch onderzoek. Zoals aangegeven in de inleiding is het verstandig zo iemand bij de beoordeling te betrekken.

*In deze review is terecht afgezien van een meta-analyse. In tabel 2 staan van alle studies ook de likelihood ratio's weergegeven. In deel 4 is vanwege de complexiteit nog niet ingegaan op deze samengestelde uitkomstmaat. Eenvoudig gezegd is het voordeel van deze maat dat het de sensitiviteit en de specificiteit van de test combineert. Hierdoor wordt extra zekerheid verkregen over het aanwezig of afwezig zijn van de aandoening als de test wordt uitgevoerd. De likelihood ratio geeft dus aan wat de diagnostische winst is van een test in de eigen patiëntenpopulatie waarvan de voorafkansen op een aandoening (in dit voorbeeld: dementie) bekend is.*

### Toepasbaarheid

De laatste stap is het beoordelen of de gevonden resultaten ook toepasbaar zijn in de eigen situatie of de toepas-

Tabel 2.

## Sceningtesten voor dementie in oudere patiënten in de eerstelijns gezondheidszorg

Type test	Tijdsinvestering	Aantal studies	Mediane +LR (range)	Mediane -LR (range)
Mini-Mental State	7-10 min	11	6.3 (3.4 to 47)	0.19 (0.06 to 0.37)
Korte instrumenten	1-10 min	15	6.9 (1.2 to 47)	0.16 (0.02 to 0.63)

baarheid in een andere populatie ('generaliseerbaarheid'). Dit hangt in de eerste plaats af van de beschikbaarheid en bruikbaarheid van de test (bijvoorbeeld als het gaat om nieuwe apparatuur of arbeidsintensiviteit), de aanwezige ervaring (bijvoorbeeld als het een nieuw aan te leren techniek of test betreft), en de kosten van de nieuwe diagnostische test in de eigen setting.

*Gezien de vergrijzing zal dementie een steeds grotere kwestie worden in de gezondheidszorg. Het is belangrijk om een praktisch en betrouwbaar instrument te hebben om dementie aan te tonen of uit te sluiten. De onderzoekers tonen aan dat het gebruik van de MMSE hiervoor een geschikt instrument is als eerste stap. Deze lijst is ook in Nederland beschikbaar en wordt toegepast. Gezien het feit dat het een korte test is zijn daardoor de personele kosten ook niet hoog. Vervolgens bepaalt de beschikbare tijd mede de keuze voor een eventuele (kortdurende) vervolgstap, zoals de Memory Impairment Screen. Deze studie richt zich op het gebruik bij patiënten boven de 60 jaar in de eerste lijn ofwel de huisartsenpraktijk. De kans op voorkomen van dementie in deze groep is anders dan de kans op voorkomen in een verpleeg- of ziekenhuispopulatie. Vooral voor verpleegkundigen in verpleeg-, verzorg- en ziekenhuizen is het van belang onderscheid te kunnen maken tussen dementie, depressie of een delier. De auteurs geven aan dat in de eerstelijns gezondheidszorg het afnemen van de MMSE de eerste keus is. Als de patiënt dan inderdaad van dementie verdacht wordt, kan een van de uitgebreidere lijsten afgenomen worden om patiënten die op dementie gelijkende symptomen hebben (door bijvoorbeeld bijwerkingen van medicijnen of infectie) niet onterecht als dement te bestempelen.*

### De likelihood ratio

Likelihood ratio van een positieve test (LR+) = de verhouding tussen de kans op een positieve uitslag van de indextest bij personen met de ziekte en de kans op een positieve testuitslag bij personen zonder de ziekte. Een diagnostische test is informatiever naarmate de LR+ hoger wordt (naar oneindig gaat). Bij een positieve uitslag van een indextest met een hoge LR+ is het 'aannemelijker' dat we met een zieke te maken hebben dan met een niet-zieke.

Likelihood ratio van een negatieve test (LR-) = de verhouding tussen de kans op een negatieve uitslag op de indextest bij personen met de ziekte en de kans op een negatieve testuitslag bij personen zonder de ziekte. Een diagnostische test is informatiever naarmate de LR- lager wordt (dichter tot 0 nadert). Bij een negatieve uitslag van een indextest met een lage LR- is het veel 'minder aannemelijk' dat we met een zieke te maken hebben dan met een niet-zieke.

### Samenvatting

Systematische reviews van diagnostisch onderzoek vereisen specifieke expertise en komen nog maar weinig voor. Systematische reviews van interventiestudies zijn in groten getale te vinden in de Cochrane Library. Binnenkort zullen via de Cochrane Collaboration ook steeds meer reviews van diagnostisch onderzoek beschikbaar komen. Net als bij de beoordeling van de onderliggende diagnostische studies moet ook hier gemeld worden dat voorspellende waarden van de test sterk afhankelijk zijn van de context en de patiënten waarin de test gebruikt wordt. Houd dus bij het lezen of kiezen van een studie in het achterhoofd hoe vaak in uw eigen populatie de aandoening voorkomt. Vervolgens kan met een checklist beoordeeld worden of de studie goed is uitgevoerd en of de resultaten inderdaad in de eigen praktijk toepasbaar zijn. «

### Literatuur

- 1) Ubbink D, Vermeulen H, Scholten R. Toolbox deel 1: Gerandomiseerd klinisch onderzoek. Nederlands Tijdschrift voor Evidence Based Practice 2007;6:8-12.
- 2) Vermeulen H, Ubbink DT, Hooft L. Toolbox deel 2: Systematisch literatuuronderzoek. Nederlands Tijdschrift voor Evidence Based Practice 2008;1:4-8.
- 3) Goossens A, Ubbink D, Vermeulen H. Evidence Based Richtlijnen. Nederlands Tijdschrift voor Evidence Based Practice 2008;2:18-23.
- 4) Leeflang M, Vermeulen H, Hooft L. Toolbox deel 4: Diagnostisch onderzoek. Nederlands Tijdschrift voor Evidence Based Practice 2008;3:18-23.
- 5) Mitchell AJ, Coyne JC. Do ultra-short screening instruments accurately detect depression in primary care? A pooled analysis and meta-analysis of 22 studies. Br J Gen Pract 2007;57:144-51.
- 6) Holsinger T, Deveau J, Boustani M, Williams WJ Jr. Does This Patient Have Dementia? JAMA 2007;297:2391-2404.
- 7) Pirozzo S, Papinczak T, Glasziou P. Whispered voice test for screening for hearing impairment in adults and children: systematic review. BMJ 2003;327:967.
- 8) van Gerwen, M, Lagro-Janssen ALM. De diagnostische waarde van anamnese en lichamelijk onderzoek bij ouderen met urine-incontinentie; een overzicht van de literatuur. Ned Tijdschr Geneesk 2006;150:1771-1775.
- 9) [www.cochrane.nl](http://www.cochrane.nl)
- 10) Boustani M, Peterson B, Hanson L, Harris R, Lohr KN. U.S. Preventive Services Task Force. Screening for dementia in primary care: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 2003;138(11):927-37.