

FORMULIER I

**voor het beoordelen van een artikel
over de waarde van een**

DIAGNOSTISCHE TEST



Evidence-Based RichtlijnOntwikkeling

Formulier I: beoordeling artikel over de waarde van een diagnostische test

Er zijn verschillende vormen van onderzoek naar de waarde van een diagnostische test. Dit formulier is bestemd voor het meest gangbare type diagnostisch onderzoek: een indextest (de test die geëvalueerd wordt) wordt vergeleken met een referentietest (een zogenoemde gouden standaard). Dit type onderzoek is een zogenoemd cross-sectioneel onderzoek (dwarsdoorsnedeonderzoek). Het formulier voldoet niet voor andere vormen van diagnostisch onderzoek, zoals onderzoeken waarbij de uitkomsten van verschillende indextesten met elkaar worden vergeleken, of onderzoeken waarin de toepassing van een diagnostische test of diagnostische strategie wordt geëvalueerd op basis van uitkomsten op patiëntniveau (bijvoorbeeld overleving of functionele status) (de zogenoemde managementstudies).

De waarde van een diagnostische test wordt ook wel door middel van een randomised controlled trial (RCT) onderzocht; in dat geval dient u formulier II voor het beoordelen van RCT's te gebruiken.

Dit formulier is ontwikkeld door een werkgroep bestaande uit vertegenwoordigers van het Dutch Cochrane Centre, het Kwaliteitsinstituut voor de Gezondheidszorg CBO, het Nederlands Huisartsen Genootschap, het institute for Medical Technology Assessment, de Werkgroep Onderzoek Kwaliteit, het College voor Zorgverzekeringen, Zorgonderzoek Nederland (ZonMw) en de Orde van Medisch Specialisten en wordt ondersteund door het Nederlands Paramedisch Instituut, de Vereniging voor Integrale Kankercentra en de Werkgroep Infectieziektenpreventie.

Voor het beoordelen van de kwaliteit van andere typen onderzoek zijn eveneens formulieren ontwikkeld. Deze staan samengevat in onderstaande tabel.

| Type onderzoek | Formulier |
|--|-----------|
| Dwarsdoorsnedeonderzoek (waarde diagnostische test) | I |
| Randomised controlled trial (RCT) | II |
| Cohortonderzoek | III |
| Patiënt-controleonderzoek | IV |
| Systematische review van RCT's (therapie en preventie) | Va |
| diagnostisch onderzoek | Vb |
| observatieonderzoek (etiologie/"harm"/prognose) | Vc |
| Economische evaluatie | VI |
| Richtlijn | AGREE |

Instructie beoordeling

- De bruikbaarheid van een publicatie voor een richtlijn wordt in de formulieren op drie facetten beoordeeld: validiteit, toepasbaarheid in de praktijk en toepasbaarheid in de Nederlandse gezondheidszorg
- Daarnaast wordt gevraagd om de belangrijkste kwantitatieve gegevens te extraheren en op een uniforme wijze te presenteren.
- De opmaak van de beoordelingsformulieren maakt het u makkelijk:
 - a) op diverse plaatsen is een beslismoment ingebouwd: indien een publicatie op dat moment niet aan de vereisten van validiteit of toepasbaarheid voldoet hoeft u met de beoordeling niet verder te gaan.
 - b) de criteria en manier van data-extractie worden telkens op de tegenoverliggende pagina kort toegelicht.

Zend opmerkingen en suggesties aangaande dit formulier naar cochrane@amc.uva.nl.

Dit formulier is geschikt voor het meest gangbare type diagnostisch onderzoek: een indextest (de test die geëvalueerd wordt, zie hieronder) wordt vergeleken met een referentietest (een zogenoemde gouden standaard).

Indextest. Vaak een nieuwe test die in het artikel wordt geëvalueerd. De indextest wordt dan met de referentietest vergeleken.

Referentietest. Een referentietest wordt ook wel gouden standaard genoemd. Bij afwezigheid van een gouden standaard wordt een indextest vaak vergeleken met de best voorhanden zijnde test of wordt het lange-termijn beloop van de aandoening als referentietest gebruikt. Uitkomsten van diagnostische onderzoeken kunnen onderling verschillen omdat verschillende referentietesten zijn gebruikt.

Vraag 1. *Referentietest.* Zie hierboven. Met "valide referentietest" wordt een referentietest bedoeld die de werkelijkheid zo goed mogelijk benadert. Toepassing van een niet-valide referentietest kan leiden tot misclassificatie van zieken en niet-zieken, waardoor het "effect" van de indextest "verdund" wordt.

Vraag 2. *Blinde beoordeling.* Testuitslagen kunnen worden beïnvloed wanneer de beoordelaar van een test reeds op de hoogte is van de uitslag van de test waarmee vergeleken wordt. Daarom dient de beoordelaar van de referentietest niet op de hoogte te zijn van de uitslag van de indextest. Hetzelfde geldt voor de beoordelaar van de indextest: deze mag niet weten wat de uitslag was van de referentietest.

Het niet toepassen van een onafhankelijke, blinde beoordeling kan leiden tot een overschatting van zowel de sensitiviteit als de specificiteit van de indextest.

Vraag 3. *Onafhankelijke beoordeling indextest.* Indien de beoordelaar van de indextest beschikt over andere relevante (klinische) informatie die zelf sterk gerelateerd is met de ziektestatus, dan kan dit leiden tot een (vertekende) beoordeling van de indextest en daardoor tot een kunstmatig hogere overeenstemming met de referentietest (zie ook vraag 2).

Vraag 4: *Work-up of verification bias.* De indextest en de referentietest moeten beide bij *alle* patiënten zijn uitgevoerd. Indien dit niet het geval is, is het mogelijk dat de keuze om de referentietest uit te voeren samenhangt met de uitkomst van de indextest. Zo kunnen bijvoorbeeld bij dure of belastende (invasieve) referentietesten alle indextestpositieven verwezen zijn voor het ondergaan van de referentietest, maar niet alle indextestnegatieven. Indien patiënten aldus op basis van de uitslag van de indextest verwezen zijn voor bevestiging met de referentietest worden de verhoudingen tussen de verschillende categorieën verstoord. De steekproef is dan "opgewerkt": de sensitiviteit van de indextest wordt kunstmatig hoger en de specificiteit lager.

Beantwoord vraag 4 derhalve met "Ja" als alle patiënten zowel de indextest als de referentietest hebben ondergaan. Het design is ook valide, als de referentietest is uitgevoerd bij alle indextest-positieven en bij een aselechte steekproef van de indextestnegatieven. Een andere, valide oplossing is het (mede) toepassen van een andere, algemeen geaccepteerde referentietest voor de indextest-negatieven, bijvoorbeeld het al dan niet optreden van de onderzochte ziekte tijdens een voldoende lange follow-up periode.

Wees op uw hoede, als er nauwelijks patiënten met een negatieve indextest in het onderzoek zijn opgenomen. Waarschijnlijk heeft er dan al vóór het onderzoek een selectie plaatsgevonden (zie ook vraag 6). Dit kan een kunstmatig hoge sensitiviteit opleveren van bijna 100%.

Beoordeling van de kwaliteit van een onderzoek naar de waarde van een diagnostische test

Naam beoordelaar:..... Datum:.....

Titel:

.....

Auteurs:

Bron:

Beoordeling van de validiteit

Korte beschrijving indextest:

Korte beschrijving referentietest:

VALIDITEIT

1. Is een valide referentietest toegepast?

Ja

Nee

Te weinig informatie in het artikel om dit te beantwoorden

2. Werden de referentietest en de indextest onafhankelijk (blind) van elkaar beoordeeld?

Ja

Nee

Te weinig informatie in het artikel om dit te beantwoorden

3. Werd de indextest afgenomen onafhankelijk van andere relevante informatie over de werkelijke ziektestatus van de patiënt?

Ja

Nee

Te weinig informatie in het artikel om dit te beantwoorden

4. Was de beslissing om de referentietest uit te voeren onafhankelijk van de uitslag van de indextest?

Ja

Nee

Te weinig informatie in het artikel om dit te beantwoorden

Vraag 5. *Behandelingsparadox*. Indien reeds met een behandeling is gestart voordat de referentietest is uitgevoerd zal een aantal echt-ziekten een negatieve referentietest krijgen. Hierdoor wordt de waarde van de referentietest ondermijnd. Zowel de sensitiviteit als de specificiteit nemen af.

Vraag 6. *Selectiebias bij inclusie in het onderzoek*. Het is van belang dat er vóór inclusie in het onderzoek geen selectie heeft kunnen optreden. Indien er niet met een aselechte steekproef of met opeenvolgende patiënten is gewerkt kan het zijn dat er bewust of onbewust geselecteerd is op klinische kenmerken die het moeilijker of juist eenvoudiger (meestal eenvoudiger) maken om met de index- of referentietest een juiste diagnose te stellen.

Vraag 7. *Spectrumbias*. De waarde van de indextest dient onderzocht te zijn in een relevante patiëntengroep (= "indicated population") zoals men die in de praktijk ook zou tegenkomen. Personen met de ziekte kunnen verschillen in ernst, stadium of duur van de ziekte. Als in een bepaalde setting normaliter sprake is van een breed ziektespectrum en de indextest wordt onderzocht bij louter personen met relatief ernstiger aandoeningen, dan zullen de eigenschappen van de indextest (sensitiviteit, specificiteit, likelihoodratio) doorgaans gunstiger zijn (overschat worden). De groep niet-ziekten moet bij voorkeur bestaan uit personen met aandoeningen die in de praktijk makkelijk tot verwarring met de onderzochte aandoening leiden. Hierdoor zal de proportie fout-positieve indextestresultaten toenemen (hetgeen in de dagelijkse praktijk ook het geval zou zijn), waardoor de specificiteit afneemt en een realistischer beeld geeft van de werkelijkheid.

Aspecten waarop bij de beoordeling kan worden gelet zijn:

- a) Spectrum van ziekte (bijvoorbeeld stadium tumor zoals vastgesteld met de referentietest)
- b) Spectrum onder niet-ziekten (de case-mix van degenen die als niet-ziek zijn gecategoriseerd)
- c) Setting (eerstelijns, tweedelijns, polikliniek, opgenomen patiënten)
- d) Duur van de ziekte voorafgaand aan de testen
- e) Voorafgaande testen / verwijfsfilters
- f) Co-morbiditeit, medicatiegebruik
- g) Demografische gegevens, zoals leeftijd en geslacht

Vraag 8. *Beschrijving test*. Aspecten waarop bij de beoordeling kan worden gelet:

- a) Uitvoering van de test
- b) Analyse en interpretatie van de test (afkappunt)
- c) Voorbereiding patiënt

Vraag 9. *Algemeen oordeel*. Hier wordt een inschatting van de validiteit en toepasbaarheid gevraagd. Let hierbij ook op eventuele fouten in het onderzoek die funest zijn voor de validiteit ervan (*red flags*, *fatal flaws*). Er zijn geen regels te geven voor welke items in ieder geval positief gescoord moeten worden of het minimum aantal items dat positief gescoord moet worden. Dit is deels afhankelijk van de "state-of-the-art" met betrekking tot het betreffende onderwerp. Probeer een inschatting te maken van de mate en richting van de vertekening, en laat dit in uw oordeel meewegen. Het gaat hier om het samenvattend oordeel van wat de beoordelaar de werkgroep zou willen mededelen over de bruikbaarheid van het artikel voor de besluitvorming.

5. Is de referentietest toegepast voordat op basis van de resultaten van de indextest reeds een behandeling is gestart?
- Ja / niet relevant
 - Nee
 - Te weinig informatie in het artikel om dit te beantwoorden
6. Was de selectie van patiënten voor het onderzoek valide?
- Toelating van opeenvolgende patiënten uit de bronpopulatie (valide)
 - Toelating op basis van een aselechte steekproef uit de bronpopulatie van patiënten (valide)
 - Toelating is mogelijk op selectieve wijze gebeurd (niet valide)
 - Te weinig informatie in het artikel om dit te beantwoorden

TOEPASBAARHEID IN DE PRAKTIJK

7. Komen de patiënten in het onderzoek overeen met de patiënten waarbij de indextest in de praktijk zal worden toegepast?
- Ja
 - Nee
 - Te weinig informatie in het artikel om dit te beantwoorden
8. Zijn de methoden om de indextest uit te voeren in voldoende detail beschreven om de test te kunnen reproduceren?
- Ja
 - Nee
 - Te weinig informatie in het artikel om dit te beantwoorden

TUSSENOORDEEL

9. Zijn de resultaten van het onderzoek valide en toepasbaar?
- Voldoende valide en toepasbaar ⇐ ga verder naar 10
 - Twijfelachtig ⇐ ga verder naar 10
 - Onvoldoende valide en toepasbaar ⇐ U kunt stoppen met het invullen van de checklist, tenzij er geen betere artikelen op dit gebied zijn (terugkoppelen naar de werkgroep)

Vraag 10. Toelichting resultaten diagnostisch onderzoek

In de tabellen dienen de meest relevante resultaten weergegeven te worden. Niet alle parameters zullen echter in het artikel vermeld staan. Deze zijn echter vaak zelf uit te rekenen met gegevens uit het artikel en onderstaande formules.

Relatie tussen de uitslag van een indextest en de uitslag van de referentietest.

| | Referentietest | | Totaal |
|---------------------|----------------|----------|---------------|
| | Ziekte + | Ziekte - | |
| Uitslag indextest + | a | b | a + b |
| Uitslag indextest - | c | d | c + d |
| Totaal | a + c | b + d | a + b + c + d |

Afkappunt(en) indextest en referentietest: geef hier de afkappunten aan waarvoor een test als positief gescoord wordt.

Dit schema kan alleen worden ingevuld, wanneer met zekerheid bekend is wie van de onderzochte personen de ziekte echt heeft. Dit betekent dat er een geaccepteerde gouden standaard beschikbaar moet zijn en dat de ziektestatus van iedere persoon zowel met de indextest als met de referentietest geëvalueerd is.

Op Internet zijn diverse sites beschikbaar voor het berekenen van de verschillende parameters (bijvoorbeeld op <http://www.cebm.utoronto.ca/practise/ca/statscal/> of <http://minerva.minervation.com/cebm/>). U kunt deze ook zelf berekenen aan de hand van de formules in de volgende tabel.

| Parameter | Berekening |
|---|---|
| Sensitiviteit | $a / (a + c)$ |
| Specificiteit | $d / (b + d)$ |
| Prevalentie (prior kans) | $(a + c) / (a + b + c + d)$ |
| Voorspellende waarde positieve test (VW+) | $a / (a + b)$ |
| Voorspellende waarde negatieve test (VW-) | $d / (c + d)$ |
| Likelihoodratio positieve test (LR+) | $\text{Sens} / (1 - \text{Spec}) = (a / (a + c)) / (b / (b + d))$ |
| Likelihoodratio negatieve test (LR-) | $(1 - \text{Sens}) / \text{Spec} = (c / (a + c)) / (d / (b + d))$ |

Sensitiviteit = de fractie terechtpositieven onder de zieken, ofwel de fractie van een groep personen met de onderzochte ziekte die met de indextest terecht als ziek geïdentificeerd wordt. Als alle zieke personen door de test als "ziek" worden aangeduid, d.w.z. een positieve testuitslag krijgen, is de sensitiviteit maximaal: 1 (ofwel 100%).

Specificiteit = de fractie terechtnegatieven onder de niet-zieken, ofwel de fractie van een groep personen zonder de onderzochte ziekte die met de indextest terecht als niet-ziek geïdentificeerd wordt.

Prevalentie (prior kans) = de prevalentie van (of de kans op) de ziekte in de onderzochte patiëntengroep.

Voorspellende waarde positieve test = de fractie van de onderzochte personen met een positieve uitslag op de indextest die de ziekte (vastgesteld met de referentietest) inderdaad heeft (= posterior kans op de ziekte bij een positieve indextestuitslag).

Voorspellende waarde negatieve test = de fractie van de onderzochte personen met een negatieve uitslag op de indextest die de ziekte (vastgesteld met de referentietest) inderdaad niet heeft (= posterior kans op het niet-hebben van de ziekte bij een negatieve indextestuitslag).

Likelihoodratio van een positieve test (LR+) = de verhouding tussen de kans op een positieve uitslag op 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.

Likelihoodratio 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.

10. Resultaten

Relatie tussen de uitslag van een indextest en de uitslag van de referentietest.

| | Referentietest | | Totaal |
|---------------------|----------------|----------|--------|
| | Ziekte + | Ziekte – | |
| Uitslag indextest + | | | |
| Uitslag indextest – | | | |
| Totaal | | | |

Afkappunt(en) indextest:

Afkappunt(en) referentietest:

Parameters voor de indextest

| Parameter | Waarde |
|---|--------|
| Sensitiviteit | |
| Specificiteit | |
| Prevalentie (prior kans) | |
| Voorspellende waarde positieve test (VW+) | |
| Voorspellende waarde negatieve test (VW–) | |
| Likelihoodratio positieve test (LR+) | |
| Likelihoodratio negatieve test (LR–) | |

Resultaten in relevante subgroep:

Relatie tussen de uitslag van een indextest en de uitslag van de referentietest.

| | Referentietest | | Totaal |
|---------------------|----------------|----------|--------|
| | Ziekte + | Ziekte – | |
| Uitslag indextest + | | | |
| Uitslag indextest – | | | |
| Totaal | | | |

Parameters voor de indextest

| Parameter | Waarde |
|---|--------|
| Sensitiviteit | |
| Specificiteit | |
| Prevalentie (prior kans) | |
| Voorspellende waarde positieve test (VW+) | |
| Voorspellende waarde negatieve test (VW–) | |
| Likelihoodratio positieve test (LR+) | |
| Likelihoodratio negatieve test (LR–) | |

Vraag 11 en 12. *Toepasbaarheid in de Nederlandse gezondheidszorg*. Beide vragen zijn een belangrijk onderdeel van richtlijnontwikkeling en dienen daarom in de werkgroep bediscussieerd te worden.

Vraag 13. *Conclusie met betrekking tot het artikel en de waarde van de diagnostische test*

Geef hier een globale samenvatting van het eindoordeel over het artikel. Probeer, indien aanwijzingen bestaan voor vertekening van de resultaten, tenminste een inschatting te maken van de richting van de vertekening (overschatting of onderschatting van de testkarakteristieken) en zo mogelijk ook over de grootte van de vertekening. Eventuele aanwijzingen voor mogelijke belangenverstremming van de auteurs met belanghebbende opdrachtgevers, kunt u hier ook rapporteren. Ook is het verstandig ingezonden brieven en/of redactionele commentaren op het hier door u beoordeelde onderzoek te raadplegen bij het formuleren van uw conclusie.

Voorbeeld: "Eindoordeel twijfelachtig. Goed opgezet artikel. De CDT lijkt een goede test om problematisch alcoholgebruik op te sporen. Door het ontbreken van een goede gouden standaard (referentietest) dienen de resultaten echter met voorzichtigheid te worden geïnterpreteerd".

Hier blijkt een enkel item (aard referentietest, vraag 1) voor de beoordelaar aanleiding te zijn om de resultaten van het onderzoek met voorzichtigheid te interpreteren.

TOEPASBAARHEID IN DE NEDERLANDSE GEZONDHEIDSZORG

11. Kan het gevonden resultaat worden toegepast op de Nederlandse situatie?
(hierbij valt bijvoorbeeld te denken aan de beschikbare diagnostische faciliteiten)

- Ja
- Nee
- Te weinig informatie in het artikel om dit te beantwoorden

12. Op welk(e) echelon(s) kan het resultaat worden toegepast?
(meerdere opties tegelijk mogelijk)

- algemene bevolking (bijvoorbeeld populatiescreening)
- eerste lijn
- tweede lijn
 - academische ziekenhuizen
 - perifere ziekenhuizen
- derde lijn (revalidatie instellingen, verpleeghuizen)

CONCLUSIE

13. Conclusie met betrekking tot het artikel en de waarde van de diagnostische test