

Tool GRADE voor diagnostiek

Ton Kuijpers*

Miranda Langendam*

Mariska Tuut*

Hans de Beer*

Philip van der Wees

* namens Dutch GRADE Network

In opdracht van Zorginstituut Nederland

Juni 2018

Inhoudsopgave

A. Inleiding	3
B. Doel van deze tool	4
C. Inhoud van de tool	5
Stappenplan voor een richtlijntekst voor een diagnostische uitgangsvraag	5
Toelichting op stappenplan richtlijntekst voor een diagnostische uitgangsvraag	8
D. Toelichting op de ontwikkeling en evaluatie	16
E. Referenties	18

A. Inleiding

In 2015 verscheen het rapport 'GRADE-methodiek voor diagnostische tests en teststrategieën'. Doel van dit rapport is om in een samenwerkingsverband tussen de verschillende richtlijnorganisaties in Nederland een voorstel te maken voor een stramien voor een richtlijntekst voor uitgangsvragen betreffende diagnostische test of teststrategieën, gebaseerd op de GRADE-methode. GRADE staat voor Grading of Recommendations Assessment, Development and Evaluation.

De GRADE-benadering van diagnostische vragen is een andere en meer complexe benadering dan die waaraan richtlijnontwikkelaars en klinici gewend zijn. Traditioneel kijken klinici en richtlijnontwikkelaars meestal naar diagnostische testeigenschappen als sensitiviteit en specificiteit wanneer zij een besluit nemen om een diagnostische test al dan niet te gebruiken. De sensitiviteit geeft aan hoe goed een test in staat is om mensen die de ziekte daadwerkelijk hebben ook op te sporen, en dus geen zieken over het hoofd ziet. De specificiteit geeft aan hoe goed een test in staat is om mensen die de ziekte niet hebben op te sporen, en dus geen niet-zieken het etiket "ziek" geeft. Maar het nut van een test hangt niet af alleen van hoe goed deze zieken en niet-zieken onderscheidt, maar ook van nadelen van de test (complicaties of belasting), het al dan niet beschikbaar zijn van een behandeling, de effectiviteit van de behandeling, en van de mate waarin het uitvoeren van de test iets aan het te volgen klinisch beleid verandert[1, 2] .

Om de meerwaarde van de nieuwe test te onderzoeken moet de nieuwe test vergeleken worden met de bestaande situatie. Er zijn verschillende studieontwerpen mogelijk. De meest directe en volledige is een gepaarde opzet: alle patiënten ondergaan zowel de nieuwe als de bestaande test, en de referentie test. Afhankelijk van de rol van de test zijn er nog andere mogelijkheden. Een bespreking daarvan is te vinden in het artikel van Bossuyt et al[3].

In de GRADE-benadering staan patiënt-relevante uitkomsten centraal, ook als het diagnostische vragen betreft. Het gaat het om de effectiviteit van de combinatie van test en behandelstrategie. Een gerandomiseerde klinische trial van de test en behandelstrategie als geheel is het ideale onderzoeksontwerp voor het evalueren van de effecten van verschillende tests op patiënt-relevante uitkomsten, maar deze studies zijn zeldzaam. Studies die rapporteren over diagnostische vragen beperken zich vaak tot de diagnostische accuratesse van een test. Wie de GRADE-benadering toepast, zal dan ook vaak op basis van het bewijs met betrekking tot testeigenschappen een inschatting moeten maken van de patiënt-relevante consequenties. Dit rapport beschrijft de daarvoor benodigde stappen.

B. Doel van deze tool

Om deze tool goed toe te kunnen passen is kennis van de methodologie van evidence-based richtlijnontwikkeling van belang voor richtlijnmethodologen, overige werkgroepleden en zorgverleners. Het beoogde doel van deze tool is het bieden van een stramien voor een richtlijntekst over diagnostische tests en teststrategieën aan werkgroepen (methodologen en overige leden). Hiermee kan de GRADE-methode op een stapsgewijze en transparante manier worden toegepast zodat een aanbeveling op een gestructureerde manier tot stand komt.

Deze tool is primair bedoeld voor richtlijnmethodologen. Overige werkgroepleden kunnen met deze tool de toegepaste methodiek volgen.

Deze tool is gebaseerd op het rapport '[GRADE-methodiek voor diagnostische tests en teststrategieën](#)' en er is gebruik gemaakt van nieuwe inzichten (waaronder aansluiting bij nieuwe ontwikkeling op het gebied van *Evidence to decision*) opgedaan tijdens bijeenkomsten van de GRADE Working Group in 2015 en 2016.

Voor algemene informatie over GRADE wordt verwezen naar de website van de internationale GRADE working group: www.gradeworkinggroup.org.

C. Inhoud van de tool

Raamwerk voor een richtlijntekst voor een diagnostische uitgangsvraag

De tool GRADE voor diagnostiek bevat zeven onderdelen die in onderstaand schema zijn samengevat. Een toelichting op de verschillende onderdelen volgt onder het schema.

Titel paragraaf

1. ACHTERGROND

- Korte beschrijving van de achtergrond van de uitgangsvraag (knelpunten huidige situatie, status van de huidige test).
- Beschrijving van de nieuwe test(s), rol van de nieuwe test(s) (vervanging, triage of add-on) en het vermeende voordeel.
- Beschrijving relevante setting.

1.1 Uitgangsvragen (opgesteld volgens vast format (PICO)).

- 1.1.1 T.a.v. de test(s): wat zijn de diagnostische testeigenschappen en wat zijn de directe gevolgen van de (combinatie van) test(s): hoe belastend is de test en wat is de kans op complicaties ten gevolge van de test?
- 1.1.2 T.a.v. de testuitslag: wat zijn de gevolgen van de testuitslag op de volgende stap in het zorgtraject (behandeling, verder testen), wat is de effectiviteit van behandeling op patiënt relevante uitkomstmaten? Presenteer in een tabel de mogelijke patiënt relevante consequenties van terecht/onterecht positieven en terecht/onterecht negatieven en de relatieve importantie daarvan (eventueel in bijlage) (zie voorbeeld in Tabel 1)

2. METHODE

2.1 Zoeken en selecteren van literatuur

- Geef aan of is uitgegaan van bestaande systematische review(s) (SR) of dat zelf een SR van de literatuur is uitgevoerd. Doe dit voor 1) diagnostische testeigenschappen, 2) directe gevolgen van de test (zoals complicaties of belasting), 3) effectiviteit behandeling en 4) connectie tussen test en behandeling.
- Beperkte beschrijving zoekstrategie en selectie van de literatuur (inclusief tabel met selectiecriteria: o.a. patiënten, interventies, uitkomstmaten, exclusiecriteria) (met uitgebreide toelichting in bijlagen).

3. RESULTATEN

3.1 Diagnostische testeigenschappen

3.1.1 Beschrijving studies

- Aantal gevonden studies, studieontwerp, totale n.
- Beschrijving van de patiëntengroepen, interventies en uitkomstmaten.

3.1.2 Kwaliteit van bewijs

- Beschrijving van de kwaliteit van bewijs op basis van de GRADE factoren om af te waarderen (zie GRADE handboek in de GRADEpro/Guideline Development Tool (GDT) en Tabel 2) (met verwijzing naar bijlage van GRADE profielen en/of summary of findings (SoF) tabellen).

3.1.3 Resultaten

- Presenteer terecht/onterecht positieven en terecht/onterecht negatieven en accuratesseparameters zoals sensitiviteit en specificiteit met de 95%-betrouwbaarheidsintervallen bij voorkeur in een SoF-tabel (met verwijzing naar bijlage van GRADE profielen en/of SoF tabellen). Zie Tabel 3.

3.2 Directe gevolgen van de test

- Beschrijf directe gevolgen door toepassen van de test, b.v. complicaties. Kwantificeer indien mogelijk en bepaal de kwaliteit van het bewijs met GRADE voor interventies.

3.3 (Optioneel: Connectie test – behandeling)

- Beschrijf hoe zeker het is dat een positieve testuitslag leidt tot behandeling. Kwantificeer indien mogelijk.)

3.4 Consequenties van testuitslag

- Geef hier een korte samenvatting van de mogelijke consequentie(s) en onderliggend bewijs (beschrijving studies, kwaliteit van bewijs en grootte effect) . Beschrijf daarbij:
 - Opties voor behandeling
 - Effectiviteit van de behandeling(en) op de patiënt-relevante uitkomsten. Verwijs zo mogelijk naar andere hoofdstukken in dezelfde richtlijn (zie de tool '[GRADE voor interventies](#)').
 - Natuurlijk beloop van de te diagnosticeren aandoening

4. Conclusies

4.1 Diagnostische testeigenschappen

- Format conclusie: Er is (veel / matig / geringe / zeer geringe)¹ zekerheid over de diagnostische testeigenschappen van (indextest) versus (vergelijkende test) als [rol test] voor het diagnosticeren van [aandoening] bij mensen met [gezondheidsprobleem/bepaalde kenmerken.

4.2 Directe gevolgen

- Vat hier kort de mogelijke consequentie(s), inclusief mate van zekerheid samen.

4.3 (Optioneel: Connectie test – behandeling)

4.4 Consequenties van testuitslag

Vat hier kort de mogelijke consequentie(s), inclusief mate van zekerheid samen (verwijs daarbij zo mogelijk naar een de uitwerking van de consequenties op een andere plek in dezelfde richtlijn).

4.5 (Optioneel: Algehele kwaliteit van bewijs)

- De algehele kwaliteit van bewijs is zo hoog als de zwakste schakel in de keten (de laagste kwaliteit van bewijs van de vier bovenstaande componenten.)

5. Van bewijs naar aanbeveling

Toelichting op welke van de volgende factoren in welke mate de (sterkte van) aanbevelingen hebben bepaald (eventueel met formulier in bijlage):

5.1 Kwaliteit van bewijs

¹Alternatief: de kwaliteit van bewijs is hoog /redelijk / laag / zeer laag.

- 5.2 Balans tussen gunstige en ongunstige effecten van de test
- 5.3 Waarden en voorkeuren van patiënten en professionals over de patiëntrelevante uitkomsten en de directe gevolgen van de test
Beschrijf eventueel de surveys, vignetten studies, en kwalitatief onderzoek zoals focusgroepen of diepte-interviews ter onderbouwing van waarden en voorkeuren een optie
- 5.4 Middelenbeslag (kosten)
Beschrijf eventueel economische evaluatiestudies ter onderbouwing van de kosten die gepaard gaan met de toepassing van een optie; verwijst zo nodig naar een evidence profile
- 5.5 (On-)gelijkheid in gezondheidsuitkomsten
- 5.6 Aanvaardbaarheid²
Beschrijf eventueel de surveys, kwalitatief onderzoek zoals focusgroepen of diepte-interviews ter onderbouwing van de aanvaardbaarheid van een optie.
- 5.7 Haalbaarheid³
Beschrijf eventueel de surveys, kwalitatief onderzoek zoals focusgroepen of diepte-interviews ter onderbouwing van de aanvaardbaarheid van een optie

Zie ook GRADEpro/GDT.

6. Rationale

Geef hier een korte toelichting op de aanbeveling en aan welke factoren de werkgroep de meeste waarde heeft toegedicht bij het formuleren van de aanbeveling.

7. Aanbevelingen

Het wordt geadviseerd om bij sterke aanbevelingen voor of tegen de test formuleringen als 'we bevelen aan' of 'het wordt aanbevolen' te gebruiken en bij zwakke (conditionele) aanbevelingen de formulering 'overweeg' en/of de verschillende opties aan te geven.

Referentielijst

Bijlagen

- Bepaling belang patiëntrelevante uitkomstmaten
- Literatuursearch
- Literatuurselectie
- Tabellen met studiekarakteristieken
- GRADE Profielen / Summary of Findings (SoF) tabellen
- Forest plots (optioneel)
- Uitwerking van procedure "Van bewijs naar aanbeveling" (eventueel met evidence-to-decision framework) (optioneel)

² Het betreft hier de vraag of de aan te bieden interventie aanvaardbaar is voor de voornaamste stakeholders, gegeven het relatief belang dat zij hechten aan de gewenste en ongewenste effecten van de aan te bieden interventie; gegeven de termijn waarop en bij wie zich de baten, nadelen en kosten voordoen, en gegeven hun ethische en morele waarden.

³ Het betreft hier de vraag of de aan te bieden optie daadwerkelijk implementeerbaar is gegeven de aanwezigheid van belemmerende en bevorderende factoren.

Toelichting op raamwerk richtlijntekst voor een diagnostische uitgangsvraag

1. ACHTERGROND

Beschrijf in deze paragraaf de test, de te diagnosticeren aandoening, de setting waarin patiënten worden gezien en de rol van de test. De te onderzoeken test kan drie rollen hebben, waarbij steeds vergeleken wordt met de bestaande situatie:

- a. *Vervanging* van een bestaande test omdat de nieuwe test nauwkeuriger, minder invasief, comfortabeler, sneller, goedkoper, eenvoudiger of gemakkelijker te interpreteren is (bijvoorbeeld MRI in plaats van mammografie voor de diagnostiek van mamma carcinoom)
- b. *Triage*: voorafgaand aan een bestaande test, om patiënten voor verder (of niet) testen te selecteren (bijvoorbeeld de faeces test voor de screening op coloncarcinoom; voorafgaand aan de bestaande test is colonoscopie). De nieuwe test kan minder accuraat zijn dan de bestaande test maar heeft andere voordelen zoals een eenvoudiger toepassing of geringere kosten
- c. *Add-on*: na een bestaande test om het aantal fout-positieven of fout-negatieven van de bestaande test te verminderen (bijvoorbeeld een moleculair diagnostische test [Xpert MTB/RIF assay] voor het bevestigen van tuberculose na een positieve speekseltest)

1.1 Uitgangsvraag

Voor het formuleren van de uitgangsvraag sluiten we aan bij GRADEpro/GDT. Hierin worden de uitgangsvragen opgesteld in PICO-format. In aanvulling daarop vindt de werkgroep het van belang de patiënt-relevante uitkomsten te benoemen.

Wordt [<test>*] als [*<rol van de test: vervanging, triage of add-on>*] versus [*vergelijkende test*] aanbevolen voor het diagnosticeren van [*<aandoening>*] bij mensen met [*<gezondheidsprobleem/bepaalde kenmerken>*]⁴?*

P ...

I *<test>*

C *<vergelijkende test>*

O *Voor de patiënt relevante uitkomstmaten:*

1.1.2 Beredeneren van patiënt relevante consequenties

In deze stap gaat het erom de voor de patiënt relevante consequenties van de verschillende testuitslagen (TP (true positives, terecht positieven, sensitiviteit), FP (false positives, fout-positieven), TN (true negatives, terecht negatieven, specificiteit), FN (false negatives, fout-negatieven)) en de directe gevolgen van de test (o.a. complicaties / belasting) te beredeneren. Dit is nodig om

⁴ De Engelstalige formulering in de Guideline Development Tool is: Should [test] versus [comparison] be used to diagnose [target condition] in [health problem and/or population]? In het ideale geval worden alle elementen van de PICO benoemd; in de praktijk wordt hier soms wat vrijer mee om gegaan om de leesbaarheid te behouden.

de voor- en nadelen van de test, met het oog op potentiële gezondheidswinst voor de patiënt, af te wegen.

Indien twee testen worden vergeleken kunnen de voor de patiënt relevante consequenties verschillend zijn, bijvoorbeeld vanwege de kans op complicaties door het toepassen van de test.

Stappen

- Ga uit van TP, TN, FP, FN⁵
- Beredeneer voor iedere categorie de gevolgen voor de patiënt: de voor de patiënt relevante consequenties van de verschillende testuitslagen (zie ook tabel 1 voor een ingevuld voorbeeld).
- Beschrijf de gevolgen van niet eenduidig te interpreteren testuitslagen, de belasting van de test en het beslag op middelen.
- Geef de importantie van de uitkomsten aan op een schaal van 1-9, conform de GRADE systematiek voor interventies. Hieruit blijkt of meer waarde wordt gehecht aan het vermijden van fout-positieve, of juist aan fout-negatieve uitslagen.

Neem de verantwoordig voor het beredeneren van voor de patiënt relevante consequenties op in in een tabel de lopende tekst of als bijlage (zie voorbeeld tabel 1).

2. METHODEN

Indien directe evidence beschikbaar is die de gehele test-behandelingsstrategie evalueert, wordt aanbevolen de tool '[GRADE voor interventies](#)' te volgen.

Indien geen directe evidence voorhanden is, zoek naar bestaande systematische reviews van DTA (diagnostische test accuratesse) studies (en actualiseer deze zo nodig) of maak zelf een systematische review indien goed uitgevoerde systematische reviews ontbreken.[4]^{6,7,8} Indien het niet mogelijk is een review te maken (vanuit methodologisch dan wel praktisch oogpunt), kan eventueel volstaan worden met het beschrijven van de individuele studies en het benoemen van een range van uitkomsten voor sensitiviteit en specificiteit.

⁵ Deze absolute aantallen volgen uit de prevalentie en gepoolde sensitiviteit en specificiteit (en worden berekend in de GDT wanneer gepoolde data worden ingevoerd). Een meta-analyse van diagnostische accuratessestudies zal meestal gebaseerd zijn op bivariate (simultane) analyse van de sensitiviteit en specificiteit. Sensitiviteit en specificiteit zijn eventueel om te rekenen naar andere accuratesse maten.

⁶ Bij bestaande systematische reviews van DTA studies is het belangrijk dat de vraagstelling van de review expliciet aansluit bij de door de richtlijnwerkgroep vastgestelde uitgangsvraag/PICO. De kwaliteit van een review van diagnostische accuratessestudies kan beoordeeld worden met de AMSTAR checklist. Voor een uitgebreide handleiding voor het maken van een review van diagnostische accuratessestudies verwijzen we naar het Cochrane handboek (<http://srdta.cochrane.org/handbook-dta-reviews>) (in ontwikkeling).

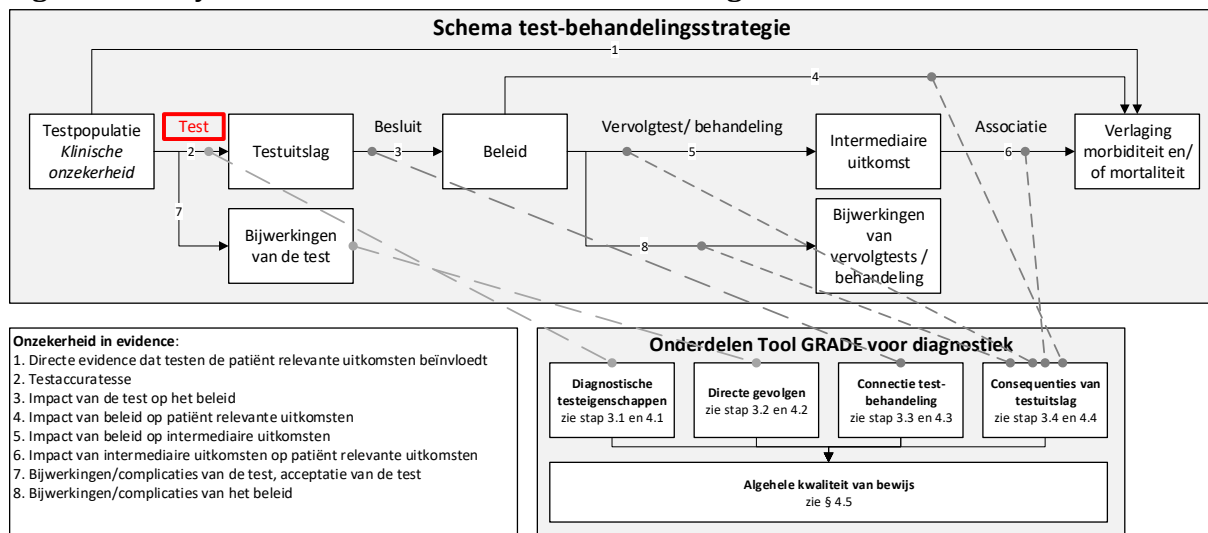
⁸ Voor het maken van een goede meta-analyse kan gebruik worden gemaakt van gratis software, zoals Open Meta Analyst en het softwarepakket R.

Daarnaast wordt naar bewijs gezocht voor:

- directe gevolgen van de test (belasting, complicaties);
- behandelconsequenties, zoals de effectiviteit van de behandeling die volgt op een positieve testuitslag of het natuurlijk beloop bij een negatieve testuitslag. Ook verder testen kan een vervolg stap zijn (bijvoorbeeld invasieve test na triage test);
- Aanname dat alle geteste personen ook daadwerkelijk het veronderstelde vervolgtraject doorlopen.

In figuur 1 is een overzicht gegeven van het analytisch raamwerk voor evaluatie van diagnostische testen, en relatie met de tool. De figuur is afgeleid van de figuur uit de Methods Guide for Medical Test Reviews van de Agency for Healthcare Research and Quality (AHRQ).⁹ De oorspronkelijke figuur van AHRQ is achterin deze tool in figuur 2 weergegeven.

Figuur 1. Analytisch raamwerk voor evaluatie van diagnostische testen



Bewijs voor de directe gevolgen van de test of het natuurlijk beloop bij een negatieve testuitslag kan mogelijk verkregen worden uit observationeel onderzoek. Bewijs voor de gewenste en ongewenste effecten van een behandeling kan via een (eventueel bestaand) evidence profile worden verkregen (zie daarvoor de tool '[GRADE voor interventies](#)').

3. RESULTATEN

3.1 Diagnostische testeigenschappen

3.1.1 Beschrijving studies

⁹ Aangepast naar voorbeeld van: Methods Guide for Medical Test Reviews. AHRQ Publication No. 12-EC017. Rockville, MD: Agency for Healthcare Research and Quality; June 2012. https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/methods-guidance-tests_overview-2012.pdf.

Beschrijf kort de gevonden studies, studieontwerp, totale n en geef daarnaast een beschrijving van de patiëntengroepen, interventies en uitkomstmaten.

3.1.2 Kwaliteit van bewijs

Beschrijving van de kwaliteit van bewijs op basis van de GRADE factoren om af te waarderen (zie GRADE handboek in de Guideline Development Tool) (met verwijzing naar bijlage van GRADE profielen en/of summary of findings tabellen). (zie tabel 2)

3.1.3 Resultaten

Presenteer terecht/onterecht positieven en terecht/onterecht negatieven en accuratesseparameters zoals sensitiviteit en specificiteit met de 95%-betrouwbaarheidsintervallen bij voorkeur in een (summary of findings (SoF)-tabel (met verwijzing naar bijlage van GRADE profielen en/of summary of findings tabellen).

3.2 Directe gevolgen van de test

Beschrijf de directe gevolgen van de test, hanteer dezelfde subparagrafen als bij 3.1. Bepaal de kwaliteit van bewijs voor de directe gevolgen van de test met GRADE voor interventies en beschrijf de resultaten.

3.3 (Optioneel: Connectie test - behandeling)

Beschrijf hoe zeker het is dat een positieve testuitslag leidt tot behandeling. Kwantificeer indien mogelijk. Hanteer dezelfde subparagrafen als bij 3.1.)

3.4 Consequenties van de testuitslag

Geef hier een korte samenvatting van de mogelijke consequentie(s) en onderliggend bewijs. Hierbij wordt bedoeld op de effectiviteit van de behandeling en het natuurlijk beloop. Hanteer dezelfde subparagrafen als bij 3.1. Verwijs zo mogelijk naar de uitwerking van de consequentie op een andere plek in dezelfde richtlijn.

4. CONCLUSIES

Formuleer conclusies over de diagnostische accuratesse, de directe gevolgen en de consequenties van de testuitslag. Noem daarbij ook de rol van de test.

(Optioneel: formuleer conclusies over de connectie tussen testuitslag en behandelconsequenties en over de algehele kwaliteit van bewijs).

5. VAN BEWIJS NAAR AANBEVELING

De stap van bewijs naar aanbeveling volgt de algemene GRADE-systematiek. Hierbij worden de volgende punten in kaart gebracht:

5.1 Kwaliteit van bewijs (zie voorgaand)

- 5.2 Waarden en voorkeuren van patiënten over de patiëntrelevante uitkomsten en de directe gevolgen van de test
- 5.3 Balans van voor- en nadelen van de test-behandel strategie ('netto' gezondheidswinst)
- 5.4 Kosten
- 5.5 Gelijkheid in gezondheidsuitkomsten
- 5.6 Aanvaardbaarheid en haalbaarheid
- 5.7 Haalbaarheid

6. RATIONALE

In de rationale beschrijft de richtlijnwerkgroep de beoordeling en weging van bovenstaande factoren. De werkgroep geeft hier aan of zij bijvoorbeeld meer gewicht heeft toegekend aan de gevolgen van fout-positieve dan aan die van fout-negatieve testuitslagen.

7. FORMULEREN VAN EEN AANBEVELING

De aanbeveling kan een sterke of zwakke (conditionele) aanbeveling voor of tegen het toepassen van een test of teststrategie zijn.

Voorbeelden van de uitwerking van een uitgangsvraag

Er zijn twee voorbeelden van een uitgangsvraag uitgewerkt:

- Echografie bij schouderklachten
- Allergische rhinitis

Beide voorbeelden zijn als aparte bijlagen bijgevoegd.

Tabel 1. Algemeen raamwerk voor het beredeneren van voor de patiënt relevante consequenties (voorbeeld: diagnostiek coeliakie)

Uitkomstmaat	Gevolgen	Voor de patiënt relevante consequenties (gebaseerd op gevolgen)	Importantie¹⁰
TP	Het kind wordt terecht gediagnosticeerd met coeliakie. Het krijgt terecht een glutenvrij dieet. Dit dieet is ingrijpend.	De ziektecontrole zal verbeteren en het kind zal minder pijn en overige gastro-intestinale symptomen hebben. De groeiachterstand zal normaliseren. De kwaliteit van leven zal mogelijk wijzigen; deze zal verhogen doordat er een verbeterde ziektecontrole is, maar verminderen doordat er sprake is van een strikt, ingrijpend dieet. De kosten voor het dieet zullen stijgen.	8
TN	Het kind wordt terecht niet gediagnosticeerd met coeliakie. Er volgt terecht geen glutenvrij dieet. Er vindt vervolgonderzoek plaats naar de oorzaak van de symptomen	De symptomen en overige gastro-intestinale symptomen zullen niet wijzigen op de korte termijn (totdat een andere diagnose is gesteld). De kwaliteit van leven neemt niet toe. De groeiachterstand neemt niet af. Er kan sprake zijn van ingrijpend en mogelijk risicovol vervolgonderzoek, maar met de kans dat een juiste diagnose gesteld wordt en de juiste behandeling wordt ingezet. De kosten voor de patiënt zullen niet wijzigen. Er is mogelijk sprake van geruststelling, omdat het kind niet de diagnose coeliakie heeft.	8
FP	Het kind wordt onterecht gediagnosticeerd met coeliakie. Het krijgt onterecht een glutenvrij dieet. Dit dieet is ingrijpend. De klachten persisteren (immers, het kind heeft geen coeliakie, maar waarschijnlijk wat anders)	De ziektecontrole, pijn en overige gastro-intestinale symptomen zullen niet verbeteren. De kwaliteit van leven neemt af, doordat de ziektecontrole niet verbeterd en er wel een ingrijpend dieet wordt gevolgd. De kosten voor het dieet stijgen. De groeiachterstand neemt niet af. Bij persisteren van klachten zal er na verloop van tijd mogelijk ingrijpend en risicovol vervolgonderzoek plaatsvinden.	9
FN	Het kind wordt onterecht niet gediagnosticeerd met coeliakie. Het krijgt onterecht geen glutenvrij dieet. Er vindt vervolgonderzoek plaats naar de oorzaak van de symptomen.	De symptomen en overige gastro-intestinale symptomen zullen niet wijzigen; de groeiachterstand neemt niet af. De kwaliteit van leven neemt niet toe. Er is mogelijk sprake van geruststelling, omdat de diagnose coeliakie niet gesteld is. De uiteindelijke diagnose coeliakie zal met vertraging gesteld worden via vervolgonderzoek (mogelijk ingrijpend en risicovol).	7
Niet eenduidig te interpreteren testuitslagen	Serologie: Er volgt vervolgonderzoek (biopt?). Biopt: niet mogelijk?	Vertraging van de uiteindelijke diagnose.	5
Belasting van de index- en referentietest	Serologie: dit is bloedonderzoek. De belasting voor de patiënt is te verwaarlozen. Biopt: dit is een invasief onderzoek, dat onder algehele narcose wordt		5

¹⁰ Optioneel

Uitkomst- maat	Gevolgen	Voor de patiënt relevante consequenties (gebaseerd op gevolgen)	Importan- tie ¹⁰
	uitgevoerd. Dit is belas- tend en er zijn risico's aan verbonden.		
Beslag op mid- delen (kosten)	Serologie: Deze zijn zeer laag in verhouding tot het alternatief (bi- opt) Biopt: specialistische in- greep, met narcose (hoge kosten)		5

Tabel 2. Beoordeling kwaliteit van bewijs verkregen uit diagnostische accuratessestudies

Diagnostische accuratessestudies, d.w.z. studies waarin patiënten met verdenking op de aandoening zowel de indextest als referentietest ondergaan starten als hoge kwaliteit van bewijs.

Factor	Beoordeling
Zijn er beperkingen in opzet en uitvoering van studies? (risk of bias)	Van alle individuele studies wordt eerst de risk of bias bepaald, bijvoorbeeld aan de hand van het QUADAS 2 instrument. ^{11,12} Het totaal aan methodologische beperkingen, dus van alle studies samen, bepaalt vervolgens of de kwaliteit van bewijs met een of twee niveaus verlaagd moet worden. Het QUADAS 2 domein Risk of bias biedt hiervoor een goed handvat.
Is er sprake van indirect bewijs? (indirectness)	Overweeg het verlagen van de kwaliteit van bewijs met een of twee niveaus als het gevonden bewijs niet overeenkomt met de patiëntenpopulatie (bijvoorbeeld ten aanzien van voorafgaande testen, ziektespectrum of co-morbiditeit), setting, indextest (bijvoorbeeld de expertise van de degene die de test uitvoert) of referentietest uit de uitgangsvraag. Het QUADAS 2 domein Applicability biedt hiervoor een goed handvat. Wanneer twee testen worden vergeleken en dit niet in dezelfde studie is gedaan, is reden om de kwaliteit te verlagen (vgl indirecte vergelijking van twee interventies). Het feit dat diagnostische accuratesse een tussenstap is (d.w.z. indirect bewijs is) voor het bepalen van de voor de patiënt relevante consequenties wordt meegenomen bij het doorredeneren van de accuratesse-uitkomsten naar de voor de patiënt relevante consequenties.
Onnauwkeurigheid van de uitkomsten? (imprecision)	Er zijn geen 'harde' afkappunten voor het beoordelen van imprecision. Beoordeel de betrouwbaarheidsintervallen rond sensitiviteit en specificiteit en rond het aantal TP, FP, TN en FN (per 1.000 geteste personen). Algemeen: als de ondergrens van het betrouwbaarheidsinterval tot een andere conclusie leidt dan de bovengrens, overweeg dan het verlagen van de kwaliteit.
Zijn de uitkomsten tussen studies niet consistent? (inconsistency)	De statistische methoden voor het bepalen van heterogeniteit zoals we die kennen bij interventie onderzoek zijn niet van toepassing op diagnostische accuratessestudies. Er is sprake van inconsistentie als de betrouwbaarheidsintervallen niet of nauwelijks overlappen. Dit kan door visuele inspectie van het forest plot worden vastgesteld. Beoordeel of de inconsistentie verklaard kan worden door verschillen tussen subgroepen. Als dat niet zo is, overweeg verlaging van de kwaliteit van bewijs. Houd er wel rekening mee dat niet alle sterke heterogeniteit onzekerheid met betrekking tot een aanbeveling met zich mee hoeft te brengen. Wanneer sterke heterogeniteit alleen verschil tussen sterk-positieve en matig-positieve effecten betreft en een matig-positief effect voldoende wordt gevonden om een positieve aanbeveling te formuleren, hoeft niet te worden afgewaardeerd.
Is er sprake van publicatiebias?	Over de statistische methoden voor het bepalen van publicatiebias bij diagnostische studies bestaat discussie. Totdat hierover duidelijkheid is verkregen, wordt geadviseerd dit criterium niet te beoordelen (en dit met een voetnoot in de GDT aan te geven). Indien de beoordelaar desondanks sterke aanwijzingen heeft voor het bestaan van publicatiebias (bijv. louter kleine studies die gefinancierd zijn door instanties die belang hebben bij de uitkomsten van het onderzoek), kan dit criterium toch gescoord worden onder de voorwaarde dat dit wordt toegelicht.

¹¹ QUADAS 2 wordt aanbevolen door de Cochrane Collaboration. Het instrument bestaat uit vier domeinen: patiënten, index test, referentie test en flow en timing van de studie. De eerste drie domeinen hebben een Risk of bias deel en Applicability (generaliseerbaarheid) deel, Voor het domein flow en timing is alleen Risk of bias relevant.

¹² Wanneer een bestaande systematische review als uitgangspunt wordt genomen is de beoordeling van de methodologische kwaliteit vaak al door de review auteurs gedaan.

D. Toelichting op de ontwikkeling en evaluatie

De tool is grotendeels gebaseerd op het rapport '[GRADE-methodiek voor diagnostische tests en teststrategieën](#)' en er is gebruik gemaakt van nieuwe inzichten opgedaan tijdens bijeenkomsten van de GRADE Working Group in 2015 en 2016. Bij de ontwikkeling en de evaluatie van de tool waren de volgende personen en organisaties betrokken:

	Naam	Organisatie
Voorzitter	Philip van der Wees	IQ Healthcare
Werkgroepleden	Ton Kuijpers	NHG
	Miranda Langendam	AMC
	Mariska Tuut	PROVA
Klankbordgroep	Leden van het Dutch GRADE netwerk	

De werkgroep heeft vier keer vergaderd in de periode april 2016 tot oktober 2017.

Evaluatie

De tool is geëvalueerd in een twee richtlijntrajecten. In de tabel hieronder staat een overzicht van de richtlijnen en de betrokken personen in deze evaluatie.

Richtlijn	Prospectief/ Retrospectief	Betrokken personen	Organisatie
NHG-Standaard Allergische Rhinitis	prospectief	Mariska Tuut	PROVA
		Miranda Langendam	AMC
NHG-Standaard Schouderklachten	prospectief	Maureen van den Donk	NHG
		Ton Kuijpers	NHG

Ervaringen met gebruik van de tool

De tool is bruikbaar en zinvol bleek uit interviews met drie respondenten (leden van de richtlijnwerkgroepen, niet ervaren in het gebruik van GRADE) die betrokken waren bij het testen van de tool. Het geeft een totaalplaatje en zorgt ervoor dat er niet alleen naar de diagnostische accurate van de test wordt gekeken. Het zet de voor- en nadelen van het gebruik van een diagnostische test goed op een rij en geeft structuur. De grootste meerwaarde van de tool zit in het op gang brengen van het denkproces: de tool helpt om bewust te zijn van het belang van de relevantie van de diagnostische test voor de patiënt en de te kiezen behandeling. Een van de respondenten geeft dat dat het wel een groot instrument is voor een relatief klein onderwerp, vooral voor diagnostische tests die al bekend en ingevoerd zijn. De meerwaarde zit met name bij nieuwe tests waar alle stappen van de tool relevant zijn om tot een aanbeveling te komen.

De stap waarin de behandelconsequenties en natuurlijk beloop bepaald worden werd door een van de respondenten aangemerkt als een cruciale stap in het proces omdat hier de relatie wordt gelegd met de consequenties van de test. Het wordt daarom gezien als belangrijkste stap in de tool.

Het geïnterviewde werkgroep lid van de betreffende NHG-standaard gaf aan dat de ondersteuning vanuit het NHG belangrijk was bij de toepassing van de tool. De presentatie over de toepassing van de tool was belangrijk voor het inzicht van de werkgroep. De uitwerking van de uitgangsvraag was gedaan door de richtlijnmethodoloog. Dat maakte het voor de werkgroep makkelijk om de tool te gebruiken.

Suggesties voor verbetering van de tool¹³

De volgende suggesties ter verbetering werden gedaan:

- Zorg voor een uitgebreidere toelichting over gebruik van de PICO (stap 1), diagnostische accuratesse (stap 3) en evidence to decision framework (stap 6). Pas de lay-out aan van de tabel voor invullen van de getallen over de diagnostische accuratesse.
- Maak duidelijk dat voor toepassing van de tool ervaring nodig is met de GRADE systematiek in het algemeen en specifiek met diagnostische accuratesse.
- Voeg een concreet en uitgewerkt voorbeeld toe.

Evaluatie en bijstelling

De projectgroep heeft de resultaten van de interviews besproken tijdens een gezamenlijke bijeenkomst en daarin ook de eigen ervaringen met het gebruik van de tool meegenomen. Op basis van de bespreking en bovenstaande suggesties zijn aanpassingen aan de tool gedaan (de twee uitgewerkte uitgangsvragen voor de NHG-standaarden Schouderklachten en Allergische Rhinitis zijn als aparte bijlagen bij deze bijgevoegd).

¹³ Hier zijn algemene suggesties en discussiepunten genoemd; tekstuele en beperkte inhoudelijke wijzigingen zijn direct in de meest recente versie van de tool verwerkt.

E. Referenties

1. Schunemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ (Clinical research ed)*. 2008;336(7653):1106-10. Epub 2008/05/17. doi: 10.1136/bmj.39500.677199.AE. PubMed PMID: 18483053; PubMed Central PMCID: PMCPMC2386626.
2. Brozek JL, Akl EA, Jaeschke R, Lang DM, Bossuyt P, Glasziou P, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines: Part 2 of 3. The GRADE approach to grading quality of evidence about diagnostic tests and strategies. *Allergy*. 2009;64(8):1109-16. Epub 2009/06/06. doi: 10.1111/j.1398-9995.2009.02083.x. PubMed PMID: 19489757.
3. Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: assessing new tests against existing diagnostic pathways. *BMJ (Clinical research ed)*. 2006;332(7549):1089-92. Epub 2006/05/06. doi: 10.1136/bmj.332.7549.1089. PubMed PMID: 16675820; PubMed Central PMCID: PMCPMC1458557.
4. Leeflang MM. Systematic reviews and meta-analyses of diagnostic test accuracy. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*. 2014;20(2):105-13. Epub 2013/11/28. doi: 10.1111/1469-0691.12474. PubMed PMID: 24274632.

Aanvullende bronnen

AHRQ. Methods Guide for Medical Test Reviews. AHRQ Publication No. 12-EC017. Rockville, MD: Agency for Healthcare Research and Quality; June 2012. https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/methods-guidance-tests_overview-2012.pdf.

AQUA. Leidraad voor kwaliteitsstandaarden. Diemen: Zorginstituut Nederland;2016.

de Beer JJA, Kuijpers T. Toepassen GRADE in Nederland. Den Haag: Regieraad Kwaliteit van Zorg;2012.

Everdingen, J.J.E. van, D.H.H. Dreesens, J.S. Burgers, J.A. Swinkels, T.A. van Barneveld; G. D. E. M. van der Weijden (2014). Handboek evidence-based richtlijnontwikkeling: een leidraad voor de praktijk. Houten: Bohn Stafleu van Loghum, Springer Media.

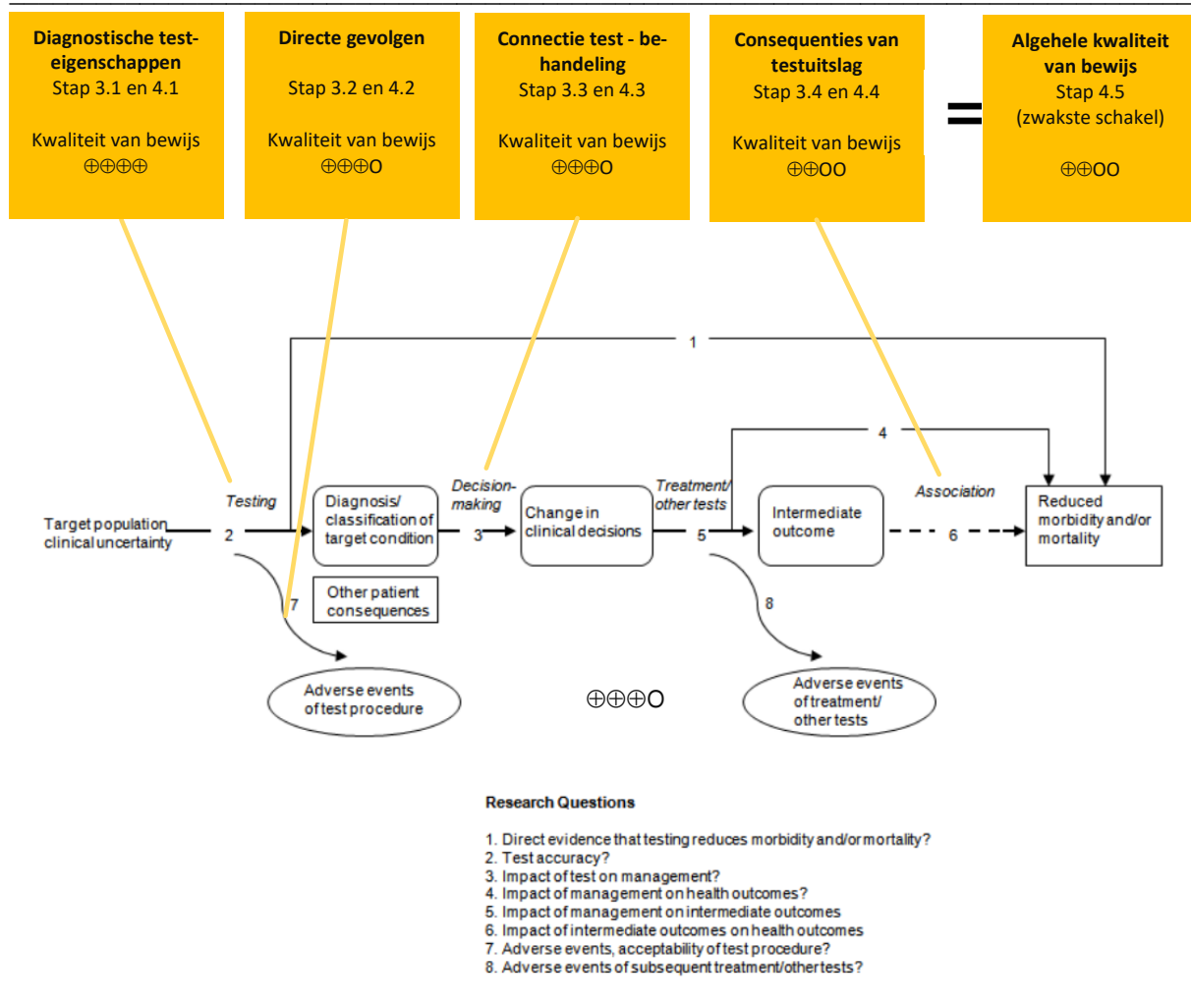
Kuijpers T, de Beer JJA. Tool GRADE voor interventies. Diemen: Zorginstituut Nederland;2014.

Tuut MK, Langendam MW, de Beer JJA, Kuijpers T. GRADE-methodiek voor diagnostische tests en teststrategieën. Diemen: Zorginstituut Nederland; namens het Dutch GRADE Network;2015.

Bijlage

Figuur 2 Analytisch raamwerk voor evaluatie van diagnostische testen

Toelichting: in de vierkante gekleurde vlakken staat de relatie met de tool aangegeven en hoe de algehele kwaliteit van bewijs wordt bepaald.



Uit: Methods Guide for Medical Test Reviews. AHRQ Publication No. 12-EC017. Rockville, MD: Agency for Healthcare Research and Quality; June 2012. www.effectivehealthcare.ahrq.gov/reports/final.cfm.