

GRADE-methodiek

voor diagnostische tests en teststrategieën

- Rapport -

september 2015

Mariska Tuut
Miranda Langendam
Hans de Beer
Ton Kuijpers

Namens het Dutch GRADE Network

The application of the GRADE approach requires a shift in clinicians' thinking to clearly recognize that, whatever their accuracy, diagnostic tests are of value only if they result in improved outcomes for patients.

(BMJ 4 August 2008)

Project uitgevoerd in opdracht van Zorginstituut Nederland,
waarnemer Ilse Verstijnen

Inhoudsopgave

Woord vooraf	5
Samenstelling werkgroep	7
Kernpunten	9
1. Inleiding	9
2. Kenmerken van diagnostische tests	9
3. Evaluatie van diagnostische tests en teststrategieën	10
4. Verschillende typen bewijs	12
5. Van bewijs naar aanbeveling	12
6. Stappen in het proces	13
Stap 1. Formuleren van uitgangsvragen en voor de patiënt relevante uitkomstmaten	13
Stap 2. Beredeneren van de voor de patiënt relevante consequenties	16
Stap 3. Bewijs verzamelen en samenvatten	18
Stap 4. Bepalen van de kwaliteit van bewijs	19
Stap 5. Formuleren van een of meer conclusies	21
Stap 6. Van bewijs naar aanbeveling	21
Stap 7. Rationale	21
Stap 8. Formuleren van een aanbeveling	21
Referenties	22
Appendix verklarende woordenlijst	23
Bijlage 1. Raamwerk voor tekst over diagnostische uitgangsvraag	27
Bijlage 2. Taken bij beantwoorden uitgangsvraag diagnostiek met GRADE	31
Bijlage 3. Voorbeelden ter illustratie	33
Voorbeeld 1. Diagnostiek coeliakie	34
Voorbeeld 2. Diagnostiek urinestenen	41
Voorbeeld 3. Triage test voor dementie	51

Woord vooraf

In 2012 verscheen het rapport 'Toepassen van GRADE in Nederland'. Doel van dit rapport was om in een samenwerkingsverband tussen de verschillende richtlijnorganisaties in Nederland een voorstel te maken voor een raamwerk voor een richtlijntekst, gebaseerd op de GRADE-methodologie.

De directe aanleiding voor dit rapport was dat het adopteren van de GRADE-methodologie consequenties heeft voor het raamwerk van een richtlijntekst, zoals dat in Nederland in het kader van de EBRO-methodiek gangbaar was. Een van de aspecten van dit raamwerk is het specificeren van uitkomstmaten bij een interventie. Binnen de EBRO-methodiek werd geen onderscheid gemaakt naar aard en relatief belang van uitkomstmaten. In de GRADE-benadering echter staan uitkomstmaten centraal die voor de patiënt van belang zijn. Een uitkomstmaat als botdichtheid bijvoorbeeld is voor een persoon met osteoporose niet direct van belang, het optreden of voorkómen van een fractuur echter wel. Ook wordt in de GRADE-benadering onderscheid gemaakt in het belang van de uitkomstmaten voor de patiënt. Onder andere door meer en minder cruciale uitkomstmaten te benoemen. Aan een uitkomstmaat als sterfte of overleven zal bijvoorbeeld vaak een groter belang worden toegekend dan aan een fractuur.

Het rapport 'Toepassen van GRADE in Nederland' had betrekking op therapeutische uitgangsvragen. In het voorliggende rapport betreft het vragen van diagnostische aard. Lezers die nog niet eerder met GRADE gewerkt hebben doen er goed aan zich eerst te verdiepen in de algemene GRADE systematiek, bijvoorbeeld door het lezen van de GRADE publicatierreeksen in het British Medical Journal en Journal of Clinical Epidemiology (zie www.GRADEworkinggroup.org).

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 naar 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. De achterliggende redenering om te kijken naar testeigenschappen is dat een test aantrekkelijker is, en het gebruik ervan in de praktijk verantwoord, naarmate deze sensitiever en specifiek is. Maar het nut van een test hangt ook af van het al dan niet beschikbaar zijn van een effectieve behandeling, of van de mate waarin het uitvoeren van de test, iets aan het te volgen beleid verandert (Schünemann 2008; Brozek 2009).

In de GRADE-benadering staan patiënt-relevante uitkomsten centraal, ook als het diagnostische vragen betreft. In deze benadering gaat het feitelijk om de combinatie van test- en behandelstrategie, en is een gerandomiseerde klinische trial van de strategie als geheel het ideale onderzoeksontwerp voor het evalueren van de effecten van verschillende tests op patiënt-relevante uitkomsten. 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.

Kaders voor dit rapport

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

Er wordt vanuit gegaan dat voor het uitwerken van diagnostische uitgangsvragen met behulp van GRADE gebruik wordt gemaakt van de *Guideline Development Tool*, online beschikbare software ontwikkeld door de internationale GRADE Working Group, <http://www.guidelinedevelopment.org/>).

Dit rapport is gebaseerd op het diagnostiek artikel uit de GRADE serie in het British Medical Journal (Schünemann 2008), hieraan gerelateerde publicaties (Brozek 2009; Hsu 2011; Gopalakrishna, 2014), het GRADE handbook (onderdeel van de Guideline Development Tool) en de *Agency for Healthcare Research and Quality Methods Guide* (Samson 2012; Singh 2012). Tevens is gebruik gemaakt van nog niet gepubliceerde GRADE working group manuscripten over GRADE voor diagnostiek. Diverse figuren in dit rapport zijn aanpassingen van bestaande figuren. Voorts is gebruik gemaakt van het rapport 'Toepassen van GRADE in Nederland' (2012).

Dit rapport sluit nauw aan op andere initiatieven die ontplooid zijn op het gebied van evaluatie van bewijs met betrekking tot diagnostische accuratesse van test en patiënt-relevante consequenties van diagnostische uitkomsten:

- DECIDE (www.decide-collaboration.eu)
- GRADE Working Group (www.gradeworkinggroup.org)
- GRADE voor diagnostiek workshops (gegeven door Holger Schünemann)

Het beoogde doel van dit rapport is het geven van handvatten aan werkgroepen (methodologen en overige leden) waarbij diagnostiek een van de onderdelen van een richtlijn is. Met de aanwijzingen in dit rapport kunnen methodologen de systematiek toepassen. Overige werkgroepleden kunnen dit rapport gebruiken om de toegepaste methodiek te volgen. Dit rapport richt zich op de principes en uitvoering van GRADE en gaat niet diep in op de bouwstenen die daarvoor nodig zijn (bijv. (het maken van een) goede DTA-meta-analyse en processen en besluitvorming in richtlijnwerkgroepen). Deze bouwstenen zijn echter wel essentieel om tot een kwalitatief goede gradering te komen.

Wij hebben ervoor gekozen niet alle elementen uit de door de GRADE working group voorgestelde aanpak in detail op te nemen in dit rapport. Dit betreffen stappen die nog onvoldoende ontwikkeld en ongepubliceerd zijn. Dit wordt aangegeven in het rapport.

Dit rapport bestaat uit de volgende onderdelen:

- Een overzicht van het *proces* c.q. de stappen die vanuit GRADE-optiek nodig zijn om van een uitgangsvraag tot een aanbeveling te komen;
- Een voorstel voor een raamwerk voor het beschrijven van de resultaten van het hiervoor genoemde proces;
- Drie illustraties hoe een richtlijntekst op basis van dit proces en raamwerk eruit zou kunnen zien.

Samenstelling werkgroep

De werkgroepleden maken allemaal dele uit van het Dutch GRADE Network

Instanties	Personen
AMC:	Miranda Langendam
Dutch Cochrane Centre:	Rob Scholten, Lotty Hooft
Guide2Guidance:	Hans de Beer
IKNL:	Olga van der Hel
Kennisinstituut:	Koert Burger, Anneflood van Enst
NHG:	Ton Kuijpers, Miranda Kurver, Maureen van den Donk
NVK:	Leonie Venmans
PROVA:	Mariska Tuut
Trimbos Instituut:	Marleen Hermens

Vier werkgroepleden (Miranda Langendam, Mariska Tuut, Ton Kuijpers en Hans de Beer) die nauw betrokken waren bij de eerste fase van het uitwerken van de GRADE benadering voor diagnostische vragen, hebben de conceptversie van dit rapport geschreven en de voorbeelden uitgewerkt. De overige leden hebben in twee bijeenkomsten door het leveren van commentaar en het doen van inhoudelijke en tekstuele suggesties aanzienlijk bijgedragen aan het verbeteren van de conceptversie.

Kernpunten

- In aanbevelingen over diagnostische tests staat de te verwachten gezondheidswinst door het toepassen van de test of test-strategie centraal
- De accuratesse van een test is een surrogaat-uitkomstmaat voor de voor de patiënt relevante uitkomstmaten (gezondheidswinst)
- Verschillende typen bewijs worden in een aanbeveling over een diagnostische test in ogenschouw genomen: naast bewijs uit accuratessestudies bijvoorbeeld ook bewijs over de effecten van de behandeling die volgt op een positieve testuitslag.

1. Inleiding

Klinisch handelen begint met het toepassen van een diagnostische test, en op de testuitslag volgt een actie. Die kan bestaan uit verder testen, bevestiging (en eventueel behandeling) of uitsluiten van de aandoening. Het doel van het toepassen van een diagnostische test is dat het de patiënt gezondheidswinst oplevert, of dat nu door juiste en tijdige behandeling is of door geruststelling bij uitsluiting van een aandoening. In die zin is het doel hetzelfde als het toepassen van een behandeling. Deze zal pas aanbevolen worden als deze effectief is en de patiënt er baat bij heeft.

Het opstellen van aanbevelingen met betrekking tot diagnostische tests volgt dezelfde logica als een aanbeveling voor een behandeling, maar kent een aantal uitdagingen. De belangrijkste is om op basis van bewijs met betrekking tot de accuratesse van een test een inschatting te maken van de uiteindelijke consequenties die voor de patiënt relevant zijn. Bijvoorbeeld minder sterfte of beter functioneren door tijdige en adequate behandeling bij een positieve uitslag, het vermijden van onnodige invasieve tests bij een negatieve uitslag, en geruststelling als de aandoening uitgesloten kan worden.

De GRADE-methodiek voor diagnostische tests en teststrategieën (“GRADE voor diagnostiek”) biedt een kader om met deze uitdagingen om te gaan. In dit rapport is de methodiek verwerkt in een raamwerk voor een richtlijntekst over een diagnostische vraag.

2. Kenmerken van diagnostische tests

Voor het stellen van een diagnose – het aantonen of uitsluiten van een aandoening – is een accurate diagnostische test nodig. Alle soorten onderzoeken die bedoeld zijn om een diagnose te stellen vallen onder de term ‘diagnostische test’, zoals lichamelijk onderzoek (bijvoorbeeld de proef van Lasègue om hernia lumbalis vast te stellen), laboratoriumonderzoek (bijvoorbeeld HbA1c om diabetes vast te stellen) en beeldvormende diagnostiek (bijvoorbeeld een MRI om een hernia vast te stellen).

Het toepassen van een diagnostische test geeft, uitzonderingen daargelaten, geen absolute zekerheid dat de aandoening aan- of afwezig is. De testuitslag geeft informatie over de kans dat de patiënt de aandoening waarop getest wordt wel of niet heeft. Een diagnostische test is informatief als een positieve test uitslag betekent dat de kans op de aandoening groot is, en als bij een negatieve test uitslag de aandoening met grote zekerheid uitgesloten kan worden. Met andere woorden, als er weinig fout-positieve en fout-negatieve uitslagen zijn.

De accuratesse van een diagnostische test wordt onderzocht door een groep patiënten met verdenking op de aandoening zowel de te evalueren nieuwe test (indextest) als een referentietest te laten ondergaan. De referentietest is de test die met grote zekerheid de ziekte of aandoening kan aantonen dan wel uitsluiten. Idealiter is dit een volledig betrouwbare referentietest ofwel de gouden standaard. In de praktijk is de referentietest vaak de test die als standaard test wordt gezien. De uitkomsten van een accuratessestudie geven de verhouding tussen het aantal terecht-positieven (TP), terecht-negatieven (TN), fout-positieve (FP) en fout-negatieven (FN) weer. Veelgebruikte, hiervan afgeleide maten zijn sensitiviteit, specificiteit, positief en negatief voorspellende waarde en positieve en negatieve likelihood-ratio (zie appendix voor de definities hiervan).

Een accurate diagnostische test heeft een goede sensitiviteit en/of specificiteit, dus weinig fout-positieve en/of fout-negatieve uitslagen. Welke mate van sensitiviteit en specificiteit acceptabel is hangt onder andere af van het doel van de test. Een test bedoeld voor screening – bijvoorbeeld testen op de aanwezigheid van prostaat-specifiek antigeen in het bloed in het kader van screening op prostaatkanker – dient zo min mogelijk fout-negatieve uitslagen te geven en dus een hoge sensitiviteit te hebben. De hoogte van de specificiteit is minder belangrijk, fout-positieve uitslagen metodeloze ongerustheid als consequentie worden minder cruciaal gevonden. Echter bij bijvoorbeeld erfelijkheidsonderzoek, waar afzien van kinderwens of afbreking van de zwangerschap de consequentie van een positieve testuitslag is, moet de test zo min mogelijk fout-positieve uitslagen geven.

Een diagnostische test staat zelden op zichzelf. Vaak is een test onderdeel van een teststrategie of diagnostisch proces - een reeks van testen die uiteindelijk tot een diagnose leiden, zoals anamnese, lichamelijk onderzoek en bijvoorbeeld aanvullend onderzoek. Een belangrijk onderdeel van de vraagstelling is de rol van de nieuwe test in het diagnostisch proces: gaat het om vervanging van een bestaande test of om toevoeging aan een bestaande strategie? Daarnaast moet de test in relatie gezien worden tot de effectiviteit van het vervolgbeleid.

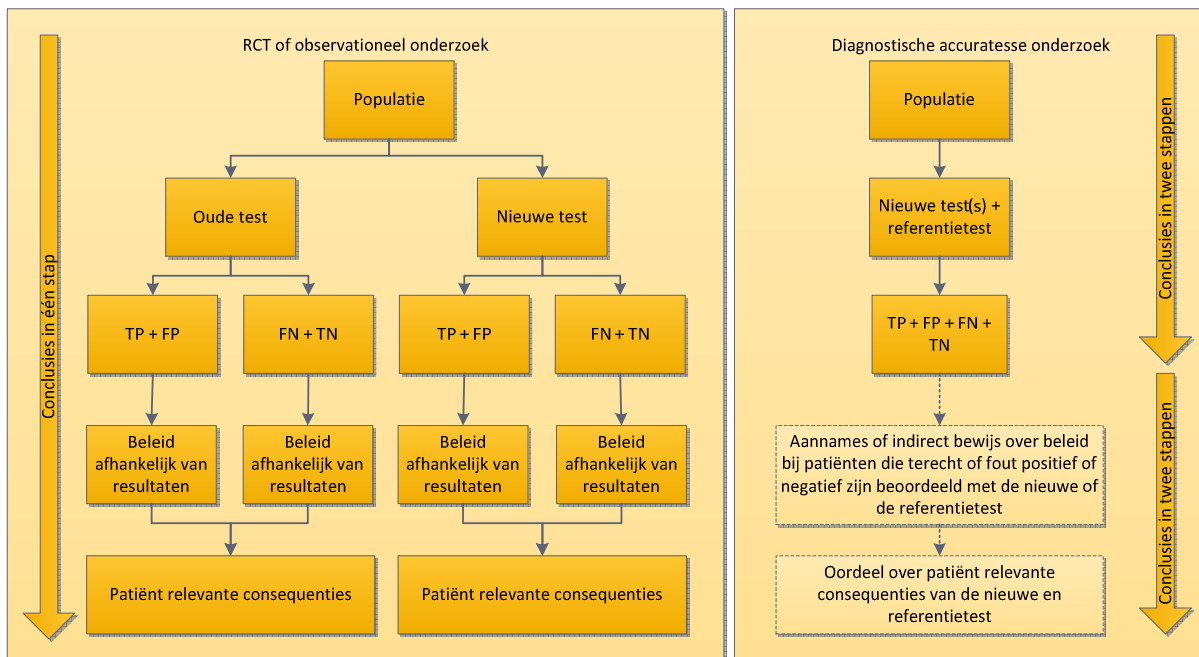
3. Evaluatie van diagnostische tests en teststrategieën

Voor het opstellen van een aanbeveling over een diagnostische test of teststrategie moet de verwachte gezondheidswinst van de test of teststrategie bepaald worden. Immers, op basis van de uitslag van de diagnostische test wordt een vervolgbeleid bepaald. Er wordt bijvoorbeeld een behandeling gestart bij degenen met een (al dan niet terechte) positieve testuitslag en geen behandeling gestart bij degenen met een (al dan niet terechte) negatieve testuitslag. In principe kan de waarde van een diagnostische test pas beoordeeld worden als duidelijk is of de gevolgen van die test (bijvoorbeeld behandeling of niet) klinisch relevante voordelen (bijvoorbeeld minder mortaliteit) voor de patiënt opleveren.

Er zijn twee mogelijkheden om de gezondheidswinst van diagnostische tests in de klinische praktijk te evalueren: de directe en indirecte methode. De meest ideale methode is de directe methode: evaluatie door middel van een 'randomised controlled trial' (RCT) (of eventueel vergelijkend onderzoek met observationele opzet). Patiënten worden 'at random' toegewezen aan de nieuwe of de oude teststrategie (of de te vergelijken testen) en vervolgens behandeld op geleide van de testuitslag. Voor alle patiënten worden de patiënt-relevante uitkomsten (gezondheidswinst) gemeten en vergeleken.

Wanneer er direct bewijs voorhanden is, kan de GRADE-systematiek voor interventies toegepast wor-

den om een aanbeveling te formuleren. RCTs die direct bewijs leveren, zijn echter zeldzaam. Ter illustratie: tussen 2004 en 2007 zijn slechts 37 van dergelijke RCTs gepubliceerd (Ferrante di Ruffano 2011). Een schematische weergave van beide methoden is opgenomen in figuur 1.



Figuur 1. Directe en indirecte manier om een diagnostische test(-strategie) te evalueren (Schünemann, 2008)¹

Omdat direct bewijs slechts sporadisch aanwezig is, richt dit rapport zich op de situatie dat direct bewijs niet voorhanden is.

Voor situaties dat er geen direct bewijs is kunnen de resultaten van accuratessestudies gebruikt worden om een schatting van de gezondheidswinst te maken. Voor dit doel is GRADE voor diagnostiek ontwikkeld. De indirecte methode heeft twee onderdelen:

- 1 samenvatten van bewijs met betrekking tot de accuratesse van de nieuwe test(s). De accuratesse wordt uitgedrukt in de verdeling van het aantal patiënten met een terecht-positieve, terecht-negatieve, fout-positieve en fout-negatieve uitslag,
- 2 doorredeneren naar mogelijke gevolgen van de testuitslag en de voor de patiënt relevante consequenties op basis van de testuitslagen. Bijvoorbeeld: patiënten met een positieve testuitslag worden behandeld. Patiënten met een terecht-positieve testuitslag verminderen door de behandeling hun kans op sterfte maar hebben ook kans op bijwerkingen. Patiënten met een fout-positieve uitslag ondervinden niet het gunstige effect van de behandeling: hun kans op sterfte vermindert niet omdat de werkelijke reden van hun klachten onbekend blijft, maar zij hebben wel kans op bijwerkingen.

¹ Let wel: de figuur links verschilt ook van de figuur rechts voor zover links sprake is van vervanging van een test door een andere test, terwijl rechts de plaats van de nieuwe test niet is weergegeven.

Op deze manier worden de testuitslagen gebruikt om de potentiële gezondheidswinst in kaart te brengen. Omdat accuratesse als tussenstap wordt gebruikt spreken we van indirect bewijs.

4. Verschillende typen bewijs

In de indirecte methode is er sprake van verschillende typen bewijs:

- bewijs verkregen uit accuratessestudies,
- bewijs over directe gevolgen van de test (zoals mogelijke complicaties en de belasting van de test),
- bewijs behorend bij de consequenties, zoals de gewenste en ongewenste effecten van de behandeling die volgt op een positieve testuitslag of het natuurlijk beloop bij een fout-negatieve testuitslag.

Voor elk van deze typen bewijs kan de kwaliteit (mate van zekerheid van de gevonden effecten) bepaald worden.

Bij het bepalen van de impact van de test op de voor de patiënt relevante uitkomsten is de aanname dat alle geteste personen het vervolgttraject passend bij de testuitslag doorlopen. Naast de drie hiervoor genoemde typen bewijs is het daarom van belang na te gaan hoe zeker deze aanname is. Bijvoorbeeld: worden alle personen met een positieve testuitslag daadwerkelijk behandeld of is er sprake van praktijkvariatie?

In de GRADE-benadering voor vragen over (therapeutische) interventies wordt de algehele kwaliteit van bewijs bepaald nadat voor iedere uitkomstmaat de kwaliteit van bewijs is bepaald, in samenhang met het relatief belang dat iedere uitkomstmaat voor de patiënt heeft. Deze redenering wordt ook in de GRADE-benadering voor diagnostische vragen gevolgd. De algehele kwaliteit van bewijs heeft betrekking op alle schakels in de keten van bewijs voor de geschatte gezondheidswinst van de test. De algehele kwaliteit van het bewijs is gelijk aan de laagste kwaliteit van bewijs van de schakel in de keten van bewijs voor het bepalen van de gunstige en ongunstige effecten van een test.

Neem als voorbeeld de volgende situatie: de richtlijnwerkgroep oordeelt dat het natuurlijk beloop en de effectiviteit van de behandeling cruciaal zijn voor bepalen van de gunstige en ongunstige effecten van een test. De kwaliteit van bewijs is laag voor deze schakels. De algehele kwaliteit van bewijs wordt daarmee laag, ondanks een eventueel hoge kwaliteit van bewijs voor de diagnostische accuratesse.

5. Van bewijs naar aanbeveling

De stap van bewijs naar aanbeveling volgt de algemene GRADE-systematiek. Naast kwaliteit van bewijs zijn dit de balans van gunstige en ongunstige effecten van de test(en) op de gezondheidstoestand van de patiënt, de waarden en voorkeuren van patiënten met betrekking tot de effecten, en het kostenaspect. Ook wordt aandacht besteed aan ongelijkheid in gezondheidsuitkomsten (inequity), aanvaardbaarheid van de test door belangrijke stakeholders en haalbaarheid van het toepassen van de test in de praktijk.

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

6. Stappen in het proces

Voor het beantwoorden van een diagnostische uitgangsvraag wordt navolgend schema (zie figuur 2) stapsgewijs doorgelopen.



Figuur 2. Schema stappen beantwoorden uitgangsvraag diagnostiek

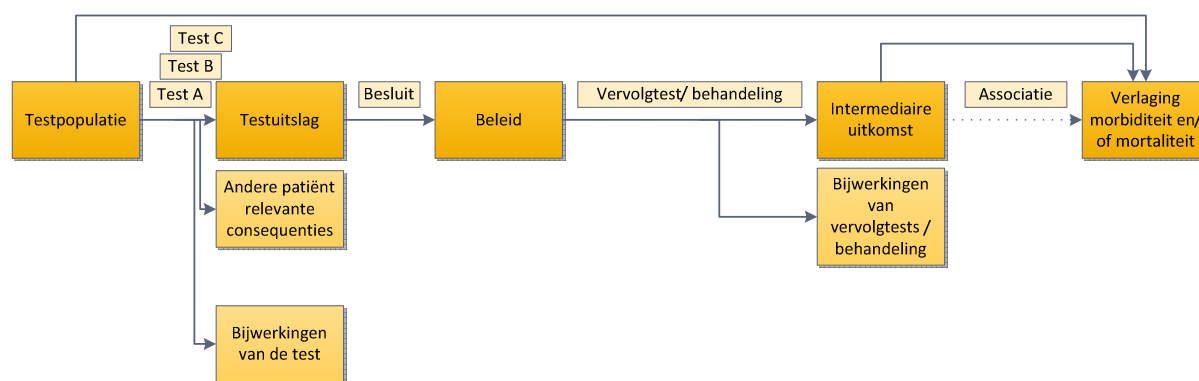
Stap 1. Formuleren van uitgangsvragen en voor de patiënt relevante uitkomstmaten

Zoals hiervoor al is besproken staat de potentiële gezondheidswinst voor de patiënt centraal bij het toepassen van een diagnostische test. Dit komt tot uitdrukking in de formulering van de uitgangsvragen. Omdat het traject van toepassen van de test in een bepaalde populatie tot en met het bepalen van de

voor de patiënt relevante consequenties vele stappen bevat kan het opstellen van een analytisch raamwerk nuttig zijn, zie figuur 3.²

De zorgverlener heeft een reden om een test uit te voeren en de uitkomst van een test heeft een gevolg (bijvoorbeeld inzetten van een behandeling of uitvoeren van een vervolgtest). De te onderzoeken test maakt daarom deel uit van een analytisch raamwerk dat een testuitslag koppelt aan klinisch beleid en aan de uitkomsten van dat beleid. De verwachte gezondheidswinst wordt uitgedrukt in voor de patiënt relevante uitkomstmaten. Deze zijn gerelateerd aan de aandoening waarop getest wordt.

Bij het in kaart brengen van het traject dat de patiënt doorloopt na het toepassen van de nieuwe test is het belangrijk om de rol van de nieuwe test ten opzichte van de bestaande testsituatie ('analytisch raamwerk') vast te stellen. De centrale vraag bij het opstellen van een aanbeveling is immers of het toepassen van een nieuwe test voldoende gezondheidswinst (voor de patiënt relevante voordelen) oplevert ten opzichte van de bestaande teststrategie.



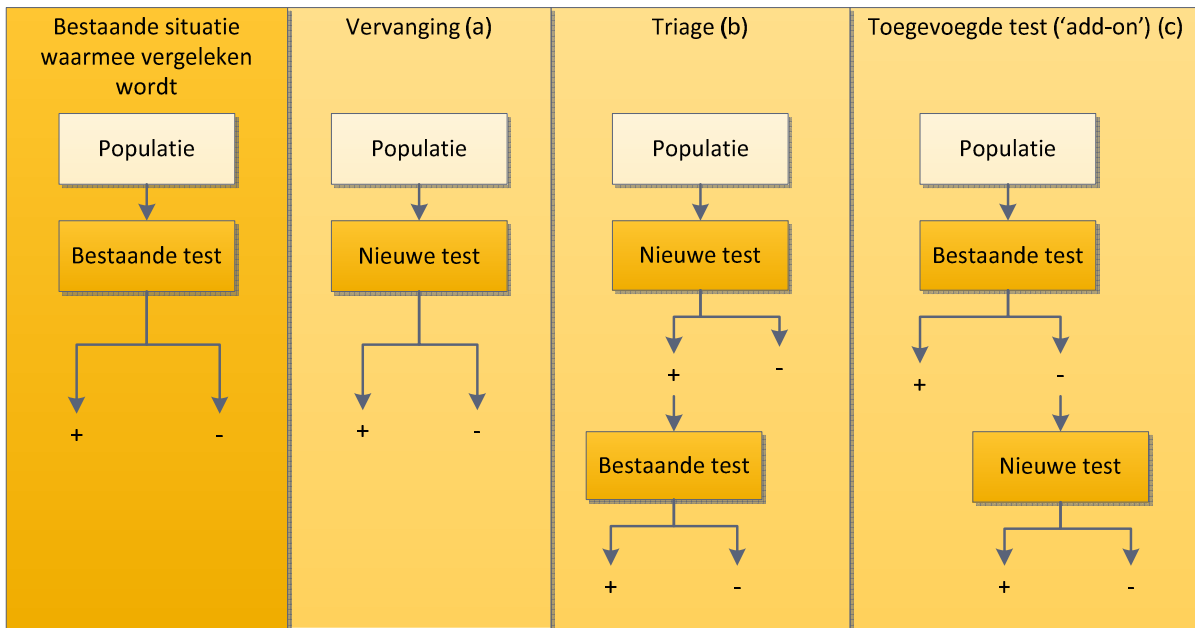
Figuur 3. Schematische weergave van een analytisch raamwerk (Samson 2012)

De te onderzoeken test kan drie rollen hebben, waarbij steeds vergeleken wordt met de bestaande situatie (zie figuur 4)³:

- Vervanging van een bestaande test omdat de nieuwe test nauwkeuriger, minder invasief, comfortabeler, sneller, goedkoper, eenvoudiger of gemakkelijker te interpreteren is (bijvoorbeeld MRI versus mammografie voor de diagnostiek van mammacarcinoom) (figuur 4, a)
- Triage: voorafgaand aan een bestaande test, om patiënten voor verder testen (of niet) te selecteren (bijvoorbeeld de faecetest voor de screening op coloncarcinoom; 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 (figuur 4, b)
- 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) (figuur 4, c)

² Samson (Chapter 2): "a specific form of graphical representation that specifies a path from the intervention or test of interest to all-important health outcomes, through intervening steps and intermediate outcomes".

³ Een uitgebreide bespreking van deze rollen, inclusief een groot aantal voorbeelden, is te vinden in het artikel van Bossuyt et al, 2006.



Figuur 4. Rol van de nieuwe test ten opzichte van de bestaande situatie (Bossuyt 2006)

Uitgangsvraag (PICO)

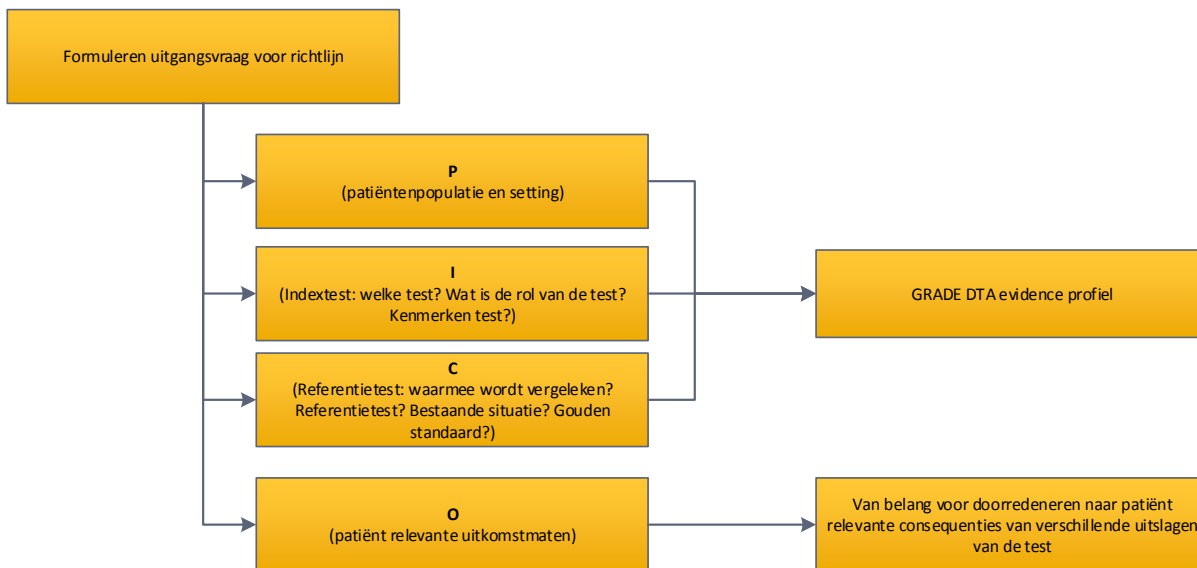
De uitgangsvragen worden opgesteld in PICO-format:

Wordt [*<test>*] als [*<rol van de test>*] versus [*vergelijkende test*] aanbevolen voor het diagnosticeren van [*<aandoening>*] bij mensen met [*<gezondheidsprobleem/bepaalde kenmerken>*]⁴?

- P ...
- I ...
- C ...
- O Voor de patiënt relevante uitkomstmaten:

Schematisch kan dit als onderstaand worden weergegeven:

⁴ De werkgroep heeft ervoor gekozen aan te sluiten bij de Guideline Development Tool; in aanvulling hierop vindt de werkgroep het van belang wel de voor de patiënt relevante consequenties te benoemen. 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.



Figuur 5. Formuleren uitgangsvraag

Stap 2. Beredeneren van de voor de patiënt relevante consequenties

In deze stap gaat het erom de voor de patiënt relevante consequenties van de verschillende testuitslagen (TP, FP, TN, FN) en de directe gevolgen van de test (o.a. complicaties / belasting) te beredeneren.⁵ Dit is nodig om 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 het aantal TP, TN, FP, FN per 1.000 geteste personen⁶
- 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 ten slotte de gevolgen van niet eenduidig te interpreteren testuitslagen, de belasting van de test en het beslag op middelen.

⁵ Een alternatief voor het beredeneren van de voor de patiënt relevante consequenties, is het kwantificeren van de consequenties d.m.v. het ontwikkelen van besliskundig model. Modelleren van de voor de patiënt relevante consequenties valt buiten de scope van dit rapport.

⁶ 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.

- 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.

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)	Importantie
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 test	Serologie: dit is bloedonderzoek. De belasting voor de patiënt is te verwaarlozen. Biopt: dit is een invasief onderzoek, dat onder algehele narcose wordt uitgevoerd. Dit is belastend en er zijn risico's aan verbonden.		5
Beslag op mid-delen (kosten)	Serologie: Deze zijn zeer laag in verhouding tot het alternatief (biopt) Biopt: specialistische ingreep, met narcose (hoge kosten)		5

Stap 3. Bewijs verzamelen en samenvatten

Het bewijs voor accuratesse van de diagnostische test wordt verzameld en samengevat in een GRADE evidence profile of Summary of Findings tabel ⁷ (zie stap 4). Zoek naar bestaande systematische reviews van DTA studies (en actualiseer deze zo nodig) of maak zelf een systematische review indien goed uitgevoerde systematische reviews ontbreken.^{8, 9, 10} 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.

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 vervolgstap zijn (b.v. invasieve test na triage test).
- Aanname dat alle geteste personen ook daadwerkelijk het veronderstelde vervolgetraject doorlopen.

⁷ Zie GRADE Handbook in GDT voor de verschillende formats (paragraaf 7.5.3)

⁸ 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://srdata.cochrane.org/handbook-dta-reviews>) (in ontwikkeling). 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 (2006).

⁹ Leeflang MM. Systematic reviews and meta-analyses of diagnostic test accuracy. Clin Microbiol Infect 2014 Feb;20(2):105-1

¹⁰ Voor het maken van een goede meta-analyse kan gebruik worden gemaakt van gratis software, zoals Open Meta Analyst en het softwarepakket R.

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 consequenties van een behandeling kan via een (eventueel bestaand) evidence profile worden verkregen (zie daarvoor het rapport "Toepassen GRADE in Nederland").

Stap 4. Bepalen van de kwaliteit van bewijs

a. Diagnostische accuratesse

Het samengevatte bewijs verkregen uit diagnostische accuratessestudies vormt de input voor het GRADE-profiel voor de diagnostische accuratesse. Voor het maken van een GRADE-profiel kan de Guideline Development Tool (GDT) gebruikt worden (<http://www.guidelinedevelopment.org/>). Op basis van de gecombineerde sensitiviteit en specificiteit van de nieuwe test en te vergelijken test en de prevalentie van de aandoening (voorafkans) wordt het aantal TP, FN, TN en FP per 1.000 patiënten berekend. Het is mogelijk verschillende prevalenties in te voeren. Vervolgens kan de kwaliteit van bewijs bepaald worden aan de hand van de bekende GRADE-factoren risk of bias, indirectness, inconsistency, imprecision en publication bias. De kwaliteit van bewijs wordt bepaald voor de sensitiviteit (TP en FN) en de specificiteit (TN en FP).

Tabel 2 geeft een kort overzicht van het beoordelen van de kwaliteit van bewijs verkregen uit diagnostische accuratessestudies. In grote lijnen gelden de criteria zoals die voor het bewijs met betrekking tot interventies, maar voor sommige factoren zijn nog onvoldoende 'harde' richtlijnen opgesteld. Deze zijn onderwerp van theoretisch en empirisch onderzoek, wat betekent dat de richtlijnen in onderstaande tabel als voorlopig gezien moeten worden. Onderstaande tabel is gebaseerd op het GRADE handboek (te vinden via de Help functie in GDT).

b. Bepaal de kwaliteit van bewijs voor de directe gevolgen van de test (GRADE voor interventies)

c. Bepaal de kwaliteit van bewijs voor de behandelconsequenties van de test (GRADE voor interventies)

d. Bepaal de mate van zekerheid voor de aanname dat behandelconsequenties één op één volgen op testuitslag

e. Bepaal de algehele kwaliteit van bewijs: dit is gelijk aan de laagste kwaliteit van bewijs van de schakels in de keten van bewijs die het belangrijkste zijn voor het bepalen van de gunstige en ongunstige effecten van een test.

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, indextest, referentietest 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.

Stap 5. Formuleren van een of meer conclusies

Formuleer een conclusie over de diagnostische accuratesse. Bij voorkeur wordt ook over de directe gevolgen en patiëntrelevante consequenties een conclusie geformuleerd.

Stap 6. Van bewijs naar aanbeveling

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

- Kwaliteit van bewijs (zie voorgaand)
- Waarden en voorkeuren van patiënten over de patiëntrelevante uitkomsten en de directe gevolgen van de test
- Balans van gunstige en ongunstige effecten van de test ('netto' gezondheidswinst)
- Kosten
- Gelijkheid in gezondheidsuitkomsten
- Aanvaardbaarheid en haalbaarheid.

Stap 7. 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.

Stap 8. Formuleren van een aanbeveling

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

Referenties

- Beer JJA de, Kuijpers T, namens de werkgroep GRADE_NL. Toepassen GRADE in Nederland. Utrecht, 2012.
- Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: assessing new tests against existing diagnostic pathways. *Brit Med J* 2006; 332: 1089-92.
- Brozek JL, Akl EA, Jaeschke R, Lang DM, Bossuyt P, Glasziou P, et al; GRADE Working Group. 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: 1109-16.
- Ferrante di Ruffano L, Davenport C, Eisinga A, Hyde C, Deeks JJ. A capture-recapture analysis demonstrated that randomized controlled trials evaluating the impact of diagnostic tests on patient outcomes are rare. *J Clin Epidemiol* 2012; 65: 282-7.
- Gopalakrishna G, Mustafa RA, Davenport C, Scholten RJ, Hyde C, Brozek J, et al. Applying Grading of Recommendations Assessment, Development and Evaluation (GRADE) tot diagnostic tests was challenging but doable. *J Clin Epidemiol* 2014 67: 760-8.
- Hsu J, Brozek JL, Terraciano L, Kreis J, Compalati E, Stein AT, et al. Application of GRADE: Making evidence-based recommendations about diagnostic tests in clinical practice guidelines. *Implementation Science* 2011; 6: 62.
- Samson D, Schoelles KM. Developing the topic and structuring systematic reviews of medical tests: utility of PICOTS, analytic frameworks, decision trees, and other frameworks. AHRQ Publication No. 12-EHC073-EF. Chapter 2 of *Methods Guide for Medical Test Reviews* (AHRQ Publication No. 12-EHC017). Rockville, MD: Agency for Healthcare Research and Quality; June 2012. www.effectivehealthcare.ahrq.gov/reports/final.cfm. Also published in a special supplement to the *Journal of General Internal Medicine*, July 2012.
- Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, et al; GRADE Working Group. GRADE: grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *Brit Med J* 2008; 336: 1106-10.
- Singh S, Chang S, Matchar DB, Bass EB. Grading a body of evidence on diagnostic tests. AHRQ Publication No. 12-EHC079-EF. Chapter 7 of *Methods Guide for Medical Test Reviews* (AHRQ Publication No. 12-EHC017). Rockville, MD: Agency for Healthcare Research and Quality; June 2012. www.effectivehealthcare.ahrq.gov/reports/final.cfm. Also published as a special supplement to the *Journal of General Internal Medicine*, July 2012.

Appendix verklarende woordenlijst

A-priori kans	Zie voorafkans.
A-posteriori kans	Zie achterafkans.
Achterafkans	De kans op de vermoede ziekte of aandoening op basis van een diagnostische testuitslag.
Beperkingen in de studie-opzet (ook wel risk of bias genoemd)	In GRADE: Eén van de vijf factoren die kunnen leiden tot het verlagen van het niveau van de kwaliteit van bewijs. Eerst wordt de methodologische kwaliteit van de afzonderlijke studie bepaald en daarna volgt een oordeel over de beperkingen in de studieopzet van alle studies samen (per uitkomst). Bijvoorbeeld: kwaliteit van bewijs wordt verlaagd omdat in drie van de vijf studies meer dan 20% van de patiënten uitvielen ("attrition bias").
Diagnostische uitkomstmaten	Uitkomstmaten met betrekking tot de diagnostische accuratesse van een test, zoals sensitiviteit, specificiteit, voorspellende waarde, positieve en negatieve likelihood-ratio, terecht negatieven, terecht positieven, fout-negatieven en fout-positieven.
Diagnostische accuratesse	Vermogen van een test om zieken van niet-zieken te onderscheiden.
Fout-negatieve uitslag	Op grond van de uitslag van een diagnostische test (indextest) worden personen ten onrechte als niet-ziek geduid.
Fout-positieve uitslag	Op grond van de uitslag van een diagnostische test (indextest) worden personen ten onrechte als ziek geduid.
Gouden standaard	De test waarvan algemeen geaccepteerd wordt dat deze de werkelijke situatie (ziek of niet-ziek) het beste weergeeft, ook wel referentietest genoemd. Bij afwezigheid van een gouden standaard wordt de te bestuderen indextest vaak vergeleken met de beste voorhanden zijnde referentietest of wordt het langetermijnbehoefte van de aandoening als referentietest gebruikt. Uitkomsten van onderzoeken kunnen onderling verschillen omdat verschillende referentietests zijn gebruikt.
GRADE	Grading Recommendations, Assessment and Evaluation. Zie www.gradeworkinggroup.nl
Guideline Development Tool	Online software voor het maken van GRADE profielen en Summary of Findings tabellen. Nieuwe versie van GRADEPro.
Onnauwkeurigheid	In GRADE: Eén van de vijf factoren die kunnen leiden tot het verlagen van het niveau van de kwaliteit van bewijs. Onnauwkeurigheid treedt op indien het geschatte effect berust op een kleine onderzoeksgroep en/of weinig events, met als gevolg brede betrouwbaarheidsintervallen.

Inconsistentie	In GRADE: Eén van de vijf factoren die kunnen leiden tot het verlagen van het niveau van de kwaliteit van bewijs. Er is sprake van inconsistentie als de resultaten van de afzonderlijke studies in een systematische review uiteenlopen in grootte en/of richting en deze heterogeniteit niet kan worden verklaard worden door verschillen in bijvoorbeeld de dosering van in de studies gebruikte medicatie; de betrouwbaarheidsintervallen overlappen niet of nauwelijks.
Index test	In diagnostisch onderzoek: de test waarvan de eigenschappen onderzocht worden.
Indirect bewijs	In GRADE: Eén van de vijf factoren die kunnen leiden tot het verlagen van het niveau van de kwaliteit van bewijs. Indirectheid treedt op als het gevonden bewijs niet (geheel) aansluit bij een of meer PICO-elementen (bijvoorbeeld seizoensgriep in plaats van vogelgriep, vergelijking A versus placebo en B versus placebo in plaats van A versus B).
Likelihood-ratio negatieve test (LR-)	De verhouding tussen het voorkomen van een negatieve uitslag op de indextest bij personen met de ziekte en het voorkomen van een negatieve testuitslag bij personen zonder de ziekte. Een diagnostische test is informatiever naarmate de LR- dichter tot 0 nadert.
Likelihood-ratio positieve test (LR+)	De verhouding tussen het voorkomen van een positieve uitslag op de indextest bij personen met de ziekte en het voorkomen van een positieve testuitslag bij personen zonder de ziekte. Een diagnostische test is informatiever naarmate de LR+ dichter tot oneindig nadert.
Meta-analyse	Onderdeel van een systematische review waarin kwantificering van de resultaten plaatsvindt. In een meta-analyse worden de afzonderlijke resultaten gecombineerd tot één overall schatting van het effect van de bestudeerde interventie ('poolen'). Meta-analyse kan daarnaast tevens analyse van bronnen van heterogeniteit bevatten.
Negatief voorspellende waarde (VW-)	Achterafkans op de afwezigheid van ziekte bij een negatieve uitslag van de indextest, ofwel de proportie niet-zieken (vastgesteld met de referentietest of gouden standaard) onder de personen met een negatieve uitslag op de indextest. Wordt ook voorspellende waarde negatieve testuitslag genoemd.
Positief voorspellende waarde (VW+)	Achterafkans op de aanwezigheid van ziekte bij een positieve uitslag van de indextest, ofwel de proportie zieken (vastgesteld met de referentietest of gouden standaard) onder de personen met een positieve uitslag op de indextest. Wordt ook voorspellende waarde positieve testuitslag genoemd.

Voor de patiënt relevante consequenties	Uitkomsten die van belang zijn voor de patiënt, zoals sterfte, morbiditeit, functioneren en kwaliteit van leven, complicaties.
Publicatiebias	Vertekening in meta-analyse van gepubliceerde onderzoeken die wordt veroorzaakt door het feit dat onderzoeken met positieve resultaten meer kans hebben om gepubliceerd te worden dan onderzoeken met negatieve resultaten. Het gevolg is dat in overzichten van gepubliceerde literatuur de behandeling positievere resultaten lijkt op te leveren dan in werkelijkheid het geval is. In GRADE: Eén van de vijf factoren die kunnen leiden tot het verlagen van het niveau van de kwaliteit van bewijs.
Referentietest	Zie gouden standaard.
Sensitiviteit	De proportie terecht-positieven onder de zieken, ofwel de proportie van de groep personen met de onderzochte ziekte die met de indextest terecht als ziek geclassificeerd wordt.
Specificiteit	De proportie terecht-negatieven onder de nietzieken, ofwel de proportie van een groep personen zonder de onderzochte ziekte die met de indextest terecht als niet-ziek geclassificeerd wordt.
Terecht-negatieve uitslag	Op grond van de uitslag van een diagnostische test (indextest) worden personen terecht als niet-ziek geïdentificeerd.
Terecht-positieve uitslag	Op grond van de uitslag van een diagnostische test (indextest) worden personen terecht als ziek geïdentificeerd.
Voorafkans	De kans op de vermoede ziekte of aandoening in de onderzochte groep personen. Deze is gelijk aan de prevalentie van de ziekte of aandoening.
Voorspellende waarde negatieve testuitslag	Zie negatief voorspellende waarde.
Voorspellende waarde positieve testuitslag	Zie positief voorspellende waarde.

Referentie

Scholten RJPM, Offringa M, Assendelft WJJ (red.). Inleiding in Evidence-Based Medicine. Klinisch handelen gebaseerd op bewijsmateriaal. Vierde druk. Houten: Bohn, Stafleu, Van Loghum, 2013. ISBN 9789031399031.

Bijlage 1. Raamwerk voor tekst over diagnostische uitgangsvraag

Titel paragraaf

1. ACHTERGROND

- Korte introductie op de achtergrond van de uitgangsvraag (knelpunten huidige situatie, status van de referentietest¹³).
- Beschrijving rol van de test(s) en het vermeende voordeel.
- Beschrijving relevante setting.

1.1 Uitgangsvraag (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 test(s): hoe belastend is de test en wat is de kans op complicaties?
- 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 en, indien relevant, de belasting van de vervolgttest? Wat is het natuurlijk beloop?
- 1.2 Presenteer in een tabel de mogelijke gevolgen en de voor de patiënt relevante consequenties van terecht/onterecht positieven en terecht/onterecht negatieven en de relatieve importantie daarvan (eventueel in bijlage)

2. METHODE

2.1 Zoeken en selecteren van literatuur

- Geef aan of is uitgegaan van bestaande SR(s) of dat zelf een SR van de literatuur is uitgevoerd. Voor 1) diagnostische testeigenschappen, 2) directe gevolgen van de test, 3) effectiviteit behandeling en andere gevolgen testuitslag, 4) natuurlijk beloop 5) waarden en voorkeuren patiënt, 6) aanvaardbaarheid en haalbaarheid.
- 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 Guideline Development Tool) (met verwijzing naar bijlage van GRADE profielen en/of summary of findings tabellen).
- 3.1.3 Diagnostische testeigenschappen
 - 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).

¹³ Voor het onderscheid tussen een referentietest en de gouden standaard, zie: GRADE Handbook in GDT, paragraaf 7.2.

3.2 Directe gevolgen van de test

- Beschrijving idem als in 3.1.

3.3 Consequenties

- Geef hier een korte samenvatting van de mogelijke consequentie(s) en onderliggend bewijs (beschrijving als in 3.1) (verwijs daarbij zo mogelijk naar de uitwerking van de consequentie op een andere plek in dezelfde richtlijn).

4. Conclusies

4.1 Diagnostische testeigenschappen

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

4.2 Directe gevolgen en consequenties

- 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).

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

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, vignette 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; verwijs zo nodig naar een evidence profile

5.5 (On-)gelijkheid in gezondheidsuitkomsten

5.6 Aanvaardbaarheid¹⁴

Beschrijf eventueel de surveys, kwalitatief onderzoek zoals focusgroepen of diepte-interviews ter onderbouwing van de aanvaardbaarheid van een optie.

5.7 Haalbaarheid¹⁵

Beschrijf eventueel de surveys, kwalitatief onderzoek zoals focusgroepen of diepte-interviews ter onderbouwing van de aanvaardbaarheid van een optie

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.

¹⁴ 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.

¹⁵ Het betreft hier de vraag of de aan te bieden optie daadwerkelijk implementeerbaar is gegeven de aanwezigheid van belemmerende en bevorderende factoren.

7. Aanbevelingen

Het wordt geadviseerd om bij sterke aanbevelingen voor of tegen de test gebruik te maken van termen zoals 'dienen of behoren' en bij zwakke (conditionele) aanbevelingen de verschillende opties aan te geven.

Referentielijst

Bijlagen

- Bepaling belang patiëntrelevante uitkomstmaten
- Literatuursearch
- Literatuurselectie
- Tabellen met studiekarakteristieken
- GRADE Profielen
- Forest plots (optioneel)
- Uitwerking van procedure "Van bewijs naar aanbeveling" (optioneel)

Bijlage 2. Taken bij beantwoorden uitgangsvraag diagnostiek met GRADE

Stap	Inhouds- en ervaringsdeskundigen richtlijnwerkgroep	Methodologisch expert richtlijnwerkgroep
1. Formuleren uitgangsvragen en patiëntrelevante uitkomstmaten	<ul style="list-style-type: none"> - Aangeven wat precies het knelpunt is dat opgelost moet worden - Specifiek benoemen van de gewenste en de huidige situatie, incl. rol van de test(s) die onderzocht wordt / worden 	<ul style="list-style-type: none"> - Formuleren PICO
2. Beredeneren van de voor de patiënt relevante consequenties	<ul style="list-style-type: none"> - Aangeven welke uitkomstmaten voor patiënten relevant zijn, en het relatief belang hiervanweergeven 	<ul style="list-style-type: none"> - Faciliteren van het prioriteren van de voor de patiënt relevante consequenties
3. Bewijs verzamelen en samenvatten	<ul style="list-style-type: none"> - Inhoudelijke expertise leveren aan methodoloog voor literatuuronderzoek, 	<ul style="list-style-type: none"> - Zoeken, selecteren en beoordelen van bewijs volgens tevoren bepaalde strategie - Evt. verrichten van eigen systematische review van DTA studies - Zoeken van bewijs voor directe gevolgen test, behandelconsequenties en gekoppeld bewijs - Opstellen SoF tabel
4. Bepaal de kwaliteit van bewijs		<ul style="list-style-type: none"> - Bepalen kwaliteit van bewijs voor diagnostische accuratesse, directe gevolgen, behandelconsequenties en relatie tussen testuitslag en behandelconsequenties aan de hand van GRADE factoren
5. Formuleer een of meerdere conclusies		<ul style="list-style-type: none"> - Formuleren conclusies over diagnostische accuratesse, directe gevolgen test, behandelconsequenties en relatie tussen testuitslag en behandelconsequenties per uitkomstmaat en gradeer deze
6. Van bewijs naar aanbeveling	<ul style="list-style-type: none"> - Geven van informatie over waarden en voorkeuren van patiënten, balans van gunstige en ongunstige effecten van de test, kosten, gelijkheid in gezondheidsuitkomsten, aanvaardbaarheid en haalbaarheid 	<ul style="list-style-type: none"> - Zoeken naar informatie over waarden en voorkeuren van patiënten, balans van gunstige en ongunstige effecten van de test, kosten, gelijkheid in gezondheidsuitkomsten, aanvaardbaarheid en haalbaarheid
7. Rationale	<ul style="list-style-type: none"> - Beoordeling en weging van alle factoren, met nadruk op het geven van gewicht aan de gevolgen van fout-positieve en fout-negatieve testuitslagen 	<ul style="list-style-type: none"> - Faciliteren weging
8. Formuleer aanbeveling	<ul style="list-style-type: none"> - Formuleren aanbevelingen en gradering daarvan 	<ul style="list-style-type: none"> - Faciliteren formuleren aanbevelingen

Bijlage 3. Voorbeelden ter illustratie¹⁶

¹⁶ Deels gesimplificeerd voorbeeld – niet getoetst op klinische relevantie.

Voorbeeld 1. Diagnostiek coeliakie

1. Achtergrond

Het betreft de introductie van serologisch onderzoek (eenvoudige, weinig belastende bloedtest) om coeliakie bij kinderen aan te tonen dan wel uit te sluiten, ter vervanging van het invasieve en belastende darmbiopt, bij kinderen onder behandeling van de kinderarts. Het darmbiopt wordt gezien als gouden standaard. Bij een positief darmbiopt wordt de diagnose coeliakie gesteld en wordt een glutenvrij dieet aanbevolen. Dit is een effectieve (ten aanzien van 'signs and symptoms'), maar voor de kwaliteit van leven ingrijpende, en mogelijk kostbare behandeling.

1.1 Uitgangsvraag

Dient serologisch onderzoek (I) ter vervanging van een darmbiopt (C) te worden toegepast voor het diagnosticeren van coeliakie bij kinderen met symptomen die passen bij coeliakie (P)?

Tabel 1.1 Selectiecriteria

Type patiënten	Kinderen (1-18 jaar) met symptomen zonder IgA-deficiëntie die een glutenvattend dieet volgen. Symptomen: - Gastro-intestinaal: chronische diarree, chronische constipatie, abdominale pijn, misselijkheid, braken, opgezwollen abdomen - Extra-intestinaal: 'failure to thrive', achtergebleven groei, verlate puberteit, chronische anemie, verminderde botmineralisatie (osteopenie/osteoporose), tandglazuurdefecten, prikkelbaarheid, chronische vermoeidheid, neuropathie, artritis/artralgie, amenorroe, verhoogde leverenzymen
Type indextest	- Serologisch onderzoek: Anti-TG2 IgA bloedtest
Type referentietest	- Darmbiopt: oesophagogastroduodenoscopie met biopten
Type uitkomstmaten	- Patiëntrelevante uitkomsten (zie onderstaand; patiëntrelevante consequenties)

Patiëntrelevante uitkomsten

- Ziektecontrole (antiTG2 na 5 maanden)
- Pijn (VAS)
- Kwaliteit van leven (CD-QOL; ziektespecifieke kwaliteit van leven)
- Overige gastro-intestinale symptomen (m.u.v. pijn)
- Groeiachterstand
- Kosten

Tabel 1.2 Gevolgen en consequenties van diagnostische testeigenschappen

Uitkomstmaat	Gevolgen	Patiëntrelevante consequenties (gebaseerd op gevolgen)	Im- por- tan- tie
Terecht positieven	Het kind wordt terecht gediagnostiseerd 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
Terecht negatieven	Het kind wordt terecht niet gediagnostiseerd 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. De kosten voor de patiënt zullen niet wijzigen. Er is mogelijk sprake van geruststelling, omdat het kind niet de diagnose coeliakie heeft.	8
Fout positieven	Het kind wordt onterecht gediagnostiseerd 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
Fout negatieven	Het kind wordt onterecht niet gediagnostiseerd 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 test	<p>Serologie: dit is bloedonderzoek. De belasting voor de patiënt is te verwaarlozen.</p> <p>Biopt: dit is een invasief onderzoek, dat onder algehele narcose wordt uitgevoerd. Dit is belastend en er zijn risico's aan verbonden.</p>	5
Beslag op middelen (kosten)	<p>Serologie: Deze zijn zeer laag in verhouding tot het alternatief (biopt)</p> <p>Biopt: specialistische ingreep, met narcose (hoge kosten)</p>	5

2 Methode

2.1 Zoeken en selecteren van literatuur

Voor het beantwoorden van de uitgangsvraag is voor het bepalen van de diagnostische accuratesse in PubMed gezocht met de volgende searchstring: (((coeliac disease[Title] OR celiac disease [Title]) AND diagnos*[Title]) AND children[Title]) AND systematic review. Literatuur werd geselecteerd als het systematisch onderzoek naar de diagnostiek van coeliakie bij kinderen betrof, in het Engels, Nederlands, Frans of Duits was gepubliceerd en indien de publicatiedatum na 2000 was. De kwaliteit van de review werd bepaald met AMSTAR. Verder werd de GRADE methodiek gevolgd. Een beperkte search naar de directe gevolgen van het onderzoek in PubMed werd uitgevoerd. In PubMed werd gezocht naar het natuurlijk beloop van coeliakie met de volgende search string: natural course (celiac OR coeliac) disease. Informatie over de effectiviteit van de behandeling werd verkregen uit de richtlijn coeliakie [NVMDL, 2008].

3 Resultaten

3.2 Diagnostische testeigenschappen

3.2.1 Beschrijving studie

In de review van Giersiepen et al werden cohortonderzoeken geïncludeerd onder kinderen met symptomen, bij wie nog diagnostische onzekerheid bestond [Giersiepen, 2012]. Ook werden patiënt-controleonderzoeken geïncludeerd waarbij patiënten met coeliakie vergeleken werden met kinderen met een andere aandoening of met bekenden van patiënten met coeliakie. In de review werden studies geïncludeerd over IgA, IgG-AGA, EmA, anti-TG2 (ook wel anti-tTG), anti-DGP en point-of-care tests voor IgA-anti-TG2. Alleen onderzoeken bij wie bij ten minste 90% van de deelnemers de resultaten van bioptonderzoek bekend waren (met Marsh of vergelijkbare criteria), werden meegenomen. Ten minste 80% van de deelnemers met coeliakie moest minimaal Marsh II laesies hebben. Er werd een minimum van 10 patiënten met coeliakie vastgesteld als inclusie criterium en de mediane leeftijd van de deelnemers moest < 18 jaar zijn. Alle kinderen moesten een glutenbevattend dieet volgen. Na search en uitgebreide selectie werden 16 publicaties geïncludeerd in de review (11 cohortonderzoeken en 6 patiënt-controleonderzoeken). In totaal werden 1876 patiënten met coeliakie (geverifieerd m.b.v. biopsie) vergeleken met 1234 patiënten zonder coeliakie.

3.2.2 Kwaliteit van bewijs

De kwaliteit van de geïncludeerde studies is bepaald met de QUADAS tool. Hierbij werd een behoorlijke risk of bias vastgesteld. Om deze reden is afgewaardeerd bij zowel de bepaling van de kwaliteit van bewijs voor de sensitiviteit als de specificiteit. Er zijn geen aanwijzingen dat de kwaliteit van bewijs afgewaardeerd zou moeten worden voor indirect bewijs, inconsistentie van de uitkomsten, onnauwkeurigheid van de uitkomsten of publicatiebias. De kwaliteit van bewijs is matig.

3.2.3 Diagnostische testeigenschappen

Data zijn gebaseerd op 5 studies met in totaal 470 patiënten met coeliakie en 399 patiënten zonder coeliakie. De gepoolde sensitiviteit is 0,964 (95%BI: 0,943-0,979). De gepoolde specificiteit is 0,977 (95%BI: 0,958-0,990). Zie ook de *summary of findings* in de bijlage.

3.2 Directe gevolgen van de test

In PubMed werd geen literatuur gevonden die de belasting van de onderzoeken beschrijft. Complicaties van gastroduodenoscopie worden in casuïstiek beschreven.

3.3 Consequenties

3.3.1 Beschrijving studies

Voor het doorredeneren naar consequenties van de test is gebruik gemaakt van de bestaande richtlijn voor de behandeling van coeliakie.

3.3.2 Kwaliteit van bewijs

De kwaliteit van bewijs voor de effectiviteit van de behandeling is hoog.

3.3.3 Behandelconsequenties

Een levenslang glutenvrij dieet is de enige wetenschappelijk onderbouwde behandeling die beschikbaar is voor symptomatische individuen met coeliakie. Het glutenvrije dieet houdt in dat de granen tarwe, rogge, gerst, spelt en kamut en alle voedingsmiddelen die daarmee bereid zijn het hele leven strikt vermeden moeten worden. Voedingsmiddelen die wel geschikt zijn in het glutenvrije dieet zijn de van nature glutenvrije (onbewerkte) voedingsmiddelen, zoals aardappelen, maïs, rijst, peulvruchten, groente, fruit, melkproducten, kaas, vlees, olie en roomboter. Naast deze voedingsmiddelen bestaan er ook voedingsmiddelen die speciaal gemaakt zijn voor het glutenvrije dieet. Die zijn gemaakt op basis van de van nature glutenvrije voedingsmiddelen of op basis van glutenvrij gemaakt tarwe(zet)meel, of een combinatie van beiden. Het aanbod en de verkrijgbaarheid van deze voedingsmiddelen wordt steeds groter. Het volgen van een glutenvrij dieet is zeer ingrijpend. Het wordt derhalve aanbevolen dat de behandeling alleen wordt gestart nadat de diagnose is bevestigd door een darmbiopsie volgens het diagnostisch algoritme, zoals elders in deze richtlijn aangegeven.

4 Conclusies

4.1 Diagnostische testeigenschappen

Er is matige zekerheid over de hoge sensitiviteit en specificiteit van IgA-TG2 POC voor het aantonen dan wel uitsluiten van coeliakie bij kinderen met symptomen passend bij coeliakie.

4.2 Directe gevolgen en consequenties

Er is geen bewijs over de directe gevolgen van IgA-TG2 POC.

De behandeling van coeliakie met een glutenvrij dieet is effectief.

5 Van bewijs naar aanbevelingen

5.1 Kwaliteit van bewijs

Er is bewijs van matige kwaliteit voor de diagnostische accuratesse en bewijs van hoge kwaliteit voor de effectiviteit van de behandeling.

5.2 Balans tussen voor- en nadelen van de test

De diagnostische accuratesse van het serologisch onderzoek benadert die van het darmbiopt en wordt acceptabel geacht. De belasting van het serologisch onderzoek wordt beperkter geacht dan van het darmbiopt en de kosten zijn lager. De positieve effecten van de test overtreffen hiermee de ongunstige effecten van de test.

5.3 Waarden en voorkeuren van patiënten over de patiëntrelevante uitkomsten en de directe gevolgen van de test

Ouders van patiënten hebben een sterke voorkeur voor niet-invasief onderzoek indien de betrouwbaarheid hiervan acceptabel is. Invasief onderzoek en de daarmee gepaard gaande narcose hebben grote nadelen en een verhoogde kans op complicaties volgens ouders. Een beperkt aandeel fout-positieve en fout-negatieve waarnemingen weegt volgens ouders van patiënten op tegen de voordelen van het niet-invasieve serumonderzoek. Immers, het gevolg van een fout-negatieve waarneming is dat de diagnose coeliakie uiteindelijk later gesteld wordt. En, bij een fout-positieve waarneming, zullen de klachten persisteren ondanks het ingezette glutenvrije dieet. Ook dan zal een uiteindelijke andere diagnose met vertraging gesteld worden. Ouders van patiënten achten het niet aannemelijk dat zeer ernstige, en levensbedreigende situaties, door deze uitgestelde uiteindelijke diagnostiek zullen ontstaan.

5.4 Kosten

De kosten van serologisch onderzoek zijn geringer dan die van het afnemen en histopathologisch beoordelen van een darmbiopt.

5.5 Ongelijkheid in zorg en zorguitkomsten

Aangenomen is dat het toepassen van de serologische test in ieder geval sociale verschillen in gezondheidsstatus niet vergroot. Mogelijk is er zelfs een positief effect te verwachten, omdat de kosten van het serologisch onderzoek veel lager zijn dan die van het biopt, en het eigen risico van de patiënt daarom minder wordt belast.

5.6 en 5.7 Aanvaardbaarheid en haalbaarheid

Serologisch onderzoek is een acceptabele en haalbare interventie. Verschuiving in budgetten (van ok naar lab) kan een knelpunt zijn in de implementatie.

6 Rationale

Professionals en patiënten zijn het erover eens dat het relatief beperkt aantal fout-positieve en fout-negatieve uitslagen ruimschoots opweegt tegen de nadelen van het invasieve darmbiopt (risico's biopt, pijn, kosten). De gevolgen van een fout-positieve, als ook van een fout-negatieve test zijn dat de definitieve diagnose (al dan niet coeliakie) met vertraging wordt gesteld.

7 Aanbeveling

Serologisch onderzoek wordt aanbevolen voor de diagnostiek van coeliakie bij patiënten zonder IgA-deficiëntie met symptomen die passen bij coeliakie. Serologisch onderzoek komt hiermee in de plaats van invasieve bipten, die daarmee een beperktere plaats krijgen. [*sterke aanbeveling*]

Referenties

- Biagi F, Marchese A, Ferretti F, Ciccocioppo R, Schiapatti A, Volta U, et al. A multicentre case control study on complicated coeliac disease: two different patterns of natural history, two different prognoses.. BMC Gastroenterol 2014; 14: 139.
- Giersiepen K1, Lelgemann M, Stuhldreher N, Ronfani L, Husby S, Koletzko S, Korponay-Szabó IR; ESPGHAN Working Group on Coeliac Disease Diagnosis. Accuracy of diagnostic antibody tests for coeliac disease in children: summary of an evidence report. J Pediatr Gastroenterol Nutr. 2012 Feb;54(2):229-41.
- Nederlandse Vereniging van Maag-Darm-Leverartsen, et al. Richtlijn Coeliakie en Dermatitis Herpetiformis. Haarlem, NVMDL, 2008.

Bijlage 1. GRADE-profiel

Question: Should anti-TG2 IgA POC be used to diagnose celiac disease in symptomatic children suspected of having celiac disease, without IgA-deficiency, and consuming a gluten containing diet?

Pooled sensitivity	0.964 (95% CI: 0.943 to 0.979)
Pooled specificity	0.977 (95% CI: 0.958 to 0.99)

Prevalences	4.6%	17%
-------------	------	-----

Outcome	N° of studies (N° of patients)	Study design	Factors that may decrease quality of evidence					Effect per 1000 patients/year		Test accuracy QoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability 4.6% ^{2 3}	pre-test probability 17%	
True positives (patients with celiac disease)	5 studies 470 patients	cross-sectional (cohort type accuracy study)	serious ¹	not serious ²	not serious	not serious	not assessed	44 (43 to 45)	164 (160 to 166)	⊕⊕⊕○ MODERATE
False negatives (patients incorrectly classified as not having celiac disease)								2 (3 to 1)	6 (10 to 4)	
True negatives (patients without celiac disease)	5 studies 399 patients	cross-sectional (cohort type accuracy study)	serious ¹	not serious ²	not serious	not serious	not assessed	932 (914 to 944)	811 (795 to 822)	⊕⊕⊕○ MODERATE
False positives (patients incorrectly classified as having celiac disease)								22 (40 to 10)	19 (35 to 8)	

1. Downgrade op RoB, vanwege relatief lage QUADAS scores; niet transparant in review

1. Geschatte prevalenties gebaseerd op richtlijn coeliakie; betreft kinderen met hoog risico op coeliakie. Dit is de populatie waarbinnen de test gebruikt zal worden

Voorbeeld 2. Diagnostiek urinestenen

1. Achtergrond

5 De huisarts zal in eerste instantie zijn beleid bepalen zonder daarbij gebruik te maken van beeldvormende diagnostiek. Als tijdens controle na 5 tot 7 dagen blijkt dat de klachten persistenten of recidiveren is mogelijk beeldvormende diagnostiek aangewezen om vast te stellen of er een steen zichtbaar is en of het pyleum of de ureter gedilateerd zijn. De vraag is welke beeldvormende diagnostiek is aangewezen.¹⁷

1.1 Uitgangsvraag

10 Wordt bij patiënten met flank of buikpijn (met een verdenking op urinestenen) (P) echosopie (I) *ter vervanging* van een CT-scan (C) aanbevolen voor het diagnosticeren van urinestenen?

Tabel 1.1 Selectiecriteria

Type studies	- SRs - RCTs - Diagnostisch accuratesseonderzoek (cross-sectioneel, prospectief of retrospectief cohort onderzoek, patiëntcontrole onderzoek)
Type patiënten	- patiënten flank of buikpijn (met een verdenking op urinestenen)
Type inductest	- 1. Echoscopie - 2. CT-scan
Type referentietest	- Bevestigde steendiagnose ¹⁸
Type uitkomstmaten	- Patiëntrelevante uitkomsten (zie onderstaand; patiëntrelevante consequenties)
Type setting	- Spoedeisende hulp / eerste lijn

15 *Patiëntrelevante uitkomsten (O)*

- gemiddelde pijnscore
- steenlozing
- hoog-risico diagnoses met complicaties
- cumulatieve stralingsbelasting
- 20 - (serieuze) bijwerkingen
- ziekenhuisopnames
- kosten

1.2 Gevolgen en consequenties van diagnostische testeigenschappen

25 In Tabel 1.2 worden mogelijke gevolgen en consequenties van diagnostische testeigenschappen gepresenteerd. De werkgroep heeft vooraf bepaald het meeste belang te hechten aan zo min mogelijk fout positieve testuitslagen. In het geval van een fout-positieve testuitslag wordt (onrecht) geen onderzoek om de werkelijke oorzaak van de klachten te achterhalen en bestaat de

¹⁷ Het betreft hier een deels hypothetisch en gesimplificeerd voorbeeld.

¹⁸ Bevestiging van passage van de steen door observatie van de patiënt of door rapportage dat de steen chirurgisch is verwijderd.

kans op het missen van hoog-risico complicaties. Er wordt minder waarde gehecht aan fout-negatieve testuitslagen, ondanks het uitblijven van een behandeling met tamsulosine kan een steen ook spontaan worden geloosd (gemiddeld is de tijd tot steenlozing dan wel langer).

Tabel 1.2 Gevolgen en consequenties van diagnostische testeigenschappen

Knelpunt: Er is behoefte aan een minder belastend alternatief voor het maken van een CT-scan bij patiënten met persistente of recidiverende klachten of met erythrocyturie.		
Uitgangsvraag: Wordt bij patiënten met flank- of buikpijn (met een verdenking op urinestenen) (P) echoscopie (I) (ter vervanging van een CT-scan (C)) aanbevolen voor het diagnosticeren van urinestenen?		
Populatie (P)	patiënten flank of buikpijn (met een verdenking op urinestenen)	
Interventie (I)	Echoscopie	
Vergelijking (C)	CT-scan	
Uitkomsten (O)	Gevolgen	Voor de patiënt relevante consequenties (gebaseerd op gevolgen)
TP	Starten van de behandeling met tamsulosine: het bevorderen van steenlozing. Geruststelling van de patiënt door bevestiging van de verdenking op urinesteenlijden.	Steenlozing
TN	De diagnose urinestenen wordt verlaten, zoeken naar ander oorzaken van de koliekpijn.	Geen kans op bijwerkingen tgv tamsulosine. Eerder op zoek naar andere oorzaak van de koliekpijn en daardoor mogelijk tijdige adequate behandeling.
FP	Onterechte behandeling met tamsulosine. Het onterecht niet verder zoeken naar hoog-risico complicaties.	Kans op bijwerkingen tgv tamsulosine. Mogelijke schade door hoog-risicocomplicaties.
FN	Het niet behandelen van de urinestenen waardoor een behandel-effect van tamsulosine achterwege blijft. Onterecht ondergaan van onnodige diagnostiek. Ongerustheid over de diagnose.	Ontbreken van een effect van tamsulosine. Stralingsbelasting ten gevolge van onnodige diagnostiek.
Niet eenduidig te interpreteren testuitslagen	Echoscopie: ondergaan van vervolgonderzoek. Ongerustheid over de diagnose. CT-scan: ondergaan van vervolgonderzoek. Ongerustheid over de diagnose.	Stralingsbelasting.
Belasting van de test	Echoscopie: echoscopie geeft geen stralingsbelasting. CT-scan: stralingsbelasting.	Stralingsbelasting
Beslag op middelen (kosten)	Echoscopie: bezoek aan ziekenhuis of diagnostisch centrum voor het ondergaan van echoscopie. Ct-scan: CT-scan vergt dezelfde tijdsinvestering, maar genereert meer kosten.	Kosten

5

2 Methode

2.1 Zoeken en selecteren literatuur

Er is een search (november 2014) uitgevoerd in Pubmed (zie bijlage 1). Er is gezocht naar SRs, RCTs en observationeel onderzoek, zie bijlage 1 voor de zoekstrategie. In tabel 1 zijn de selectiecriteria gepresenteerd voor de selectie van de literatuur.

10

3 Resultaten

3.1 Diagnostische testeigenschappen

Er werden 296 potentieel relevante abstracts gevonden, hiervan werden er 291 geëxcludeerd op basis van titel of abstract. Op basis van de volledige tekst werden er nog 4 onderzoeken geëxcludeerd. Er werd één studie ingesloten naar de diagnostische waarde van echoscopie in vergelijking met een CT-scan. In deze studie werd een randomised test comparison gemaakt en was geen sprake van het koppelen van behandelstrategieën aan de verschillende testuitslagen.

3.1.1 Beschrijving studies

In het onderzoek (n=2759) [Smith-Bindman, 2014] werd een vergelijking gemaakt tussen echoscopie uitgevoerd door een SEH-arts (n = 908)¹⁹, echoscopie uitgevoerd door een radioloog (n = 893) en een CT-scan (n = 958). Deze opties werden vergeleken met de 'gouden standaard' observatie van steenlozing door de patiënt of chirurgische verwijdering van de steen.

3.1.2 Kwaliteit van bewijs

De RCT scoorde goed op beide domeinen (risk of bias en toepasbaarheid) van het QUADAS-2 instrument (zie bijlage 2), er werd niet afgewaardeerd vanwege risk of bias en indirect bewijs. Voor de andere categorieën inconsistentie, onnauwkeurigheid en publicatie bias werd ook niet afgewaardeerd. Er was sprake van zeer veel zekerheid (kwaliteit van bewijs is hoog) van de diagnostische accuratesse parameters voor echoscopie en CT-scan (zie bijlage 3).

3.1.3 Diagnostische testeigenschappen

In tabel 3 worden de sensitiviteit en specificiteit van echoscopie en CT-scan gepresenteerd. Zie bijlage 4 voor de SoF (summary of findings)-tabel.

Tabel 3 Sensitiviteit en specificiteit echoscopie en CT-scan

	Sensitiviteit % (95% BI)	Specificiteit % (95% BI)
Echoscopie	54 (48-60)	71 (67-75)
CT-scan	88 (84-92)	58 (55-62)

3.2 Directe gevolgen van de test (belasting voor de patiënt en kosten)

Een nadeel van een CT-scan is de stralingsbelasting. Bij echoscopie is er geen sprake van stralingsbelasting. Beide testen vergen doorgaans een bezoek aan het ziekenhuis of diagnostisch centrum voor het ondergaan van echoscopie waarbij een CT-scan meer kosten genereert.

3.3 Behandelconsequenties

Indien een kleine steen (<5 mm) wordt vastgesteld zonder dilatatie kan spontane lozing worden afgewacht eventueel in combinatie met tamsulosine (kwaliteit van bewijs is laag voor een het bestaan van een klinisch relevant voordeel van tamsulosine; zie uitgangsvraag hierover in de NHG-Standaard Urinesteenlijnden 2014). Indien er wel dilatatie wordt vastgesteld, wordt er

¹⁹ In deze voorbeeldtekst is omwille van de overzichtelijkheid gekozen om alleen de vergelijking tussen echoscopie uitgevoerd door de SEH-arts (at the point-of-care) uit te werken (resultaten zijn overigens vergelijkbaar met echoscopie uitgevoerd door een radioloog).

vanuit de eerste lijn verwezen naar een uroloog voor verdere behandeling (opties variëren van conservatieve behandeling tot open chirurgie; kwaliteit van bewijs is zeer laag tot laag; zie richtlijn Nierstenen NVU 2014).

4. Conclusies

4.1 Diagnostische testeigenschappen

Er is sprake van veel zeer zekerheid (bewijs van hoge kwaliteit) over de diagnostische testeigenschappen van echoscopie (Se 54%; 95% BI 48-60 en Sp 71%; 67-75) en CT-scan (Se 88%; 95% BI 84-92 en Sp 58%; 95% BI 55-62) voor het diagnosticeren van urinestenen bij patiënten met flank of buikpijn (met een verdenking op urinestenen).

4.2 Behandelconsequenties

Er is sprake van veel onzekerheid (bewijs van lage kwaliteit) over de behandelconsequenties afwachtend beleid (eventueel ondersteund met tamsulosine) bij een positieve testuitslagen voor het vaststellen van een urinesteen.

5. Van bewijs naar aanbeveling

5.1 Kwaliteit van bewijs

- De kwaliteit van bewijs voor de diagnostische testeigenschappen is hoog.
- De kwaliteit van bewijs voor het directe gevolg *stralingsbelasting* is laag.
- De kwaliteit van bewijs voor het natuurlijk beloop (FN) is laag.
- De kwaliteit van bewijs van de behandelconsequenties 'afwachten beleid' eventueel gecombineerd met tamsulosine is laag (deze behandelbeslissing volgt op positieve testuitslagen²⁰).
- De kwaliteit van bewijs is laag over het verschil in het aantal hoog-risico-complicaties dat gemist wordt door een lager aantal fout positieve testuitslagen bij echoscopie vergeleken met CT-scan.
- De kwaliteit van bewijs over de link tussen de testuitslag en de behandelbeslissingen is laag.

Zie voor meer informatie bijlage 4.

5.2 Balans tussen voor- en nadelen van de interventie

De werkgroep hecht belang aan het lager aantal fout positieve testuitslagen van echoscopie in vergelijking met een CT-scan. Hierdoor is de verwachting is dat door het gebruik van echoscopie er minder kans is op het missen van hoog-risico complicaties doordat verder geen onderzoek wordt uitgevoerd om de werkelijke oorzaak van de klachten te achterhalen. Of dit ook werkelijk het geval is onzeker vanwege gebrek aan gegevens over een verschil in het optreden van hoog-risico complicaties tussen echoscopie en CT-scan.

5.3 Waarden en voorkeuren van patiënten en professionals

Patiënten mogelijk een voorkeur hebben voor een echoscopie omdat een minder belastende test is (aannee van de werkgroep)

²⁰ Hieronder wordt verstaan een test die een steen aantoon waarbij geen sprake is van dilatatie. De behandelconsequentie is afwachtend beleid eventueel gecombineerd met tamsulosine.

5.4 Middelenbeslag (kosten)

Beide testen vergen doorgaans een bezoek aan het ziekenhuis of diagnostisch centrum voor het ondergaan van echoscopie waarbij een CT-scan meer kosten genereert (aannee van de werkgroep).

5 5.5(On-)gelijkheid in gezondheidsuitkomsten

Niet besproken.

5.6 Aanvaardbaarheid

Zowel echoscopie als CT-scan worden gezien als aanvaardbare interventies voor zowel patiënten als professionals (aannee van de werkgroep).

10 5.7 Haalbaarheid

Het verrichten van een echoscopie vereist voldoende ervaring. Mogelijk is er onvoldoende vertrouwen van huisartsen of urologen in echoscopisten waardoor zij alsnog een CT-scan aanvragen. Dit staat de implementatie van deze aanbeveling mogelijk in de weg (aannee van de werkgroep).

15

6. Rationale

De werkgroep hecht belang aan het lager aantal fout positieve testuitslagen van echoscopie in vergelijking met een CT-scan. Hierdoor is de verwachting dat er minder kans is op het missen van hoog-risico complicaties doordat verder geen onderzoek wordt uitgevoerd om de werkelijke oorzaak van de klachten te achterhalen. Daarnaast is er geen stralingsbelasting ten gevolge van echoscopie en zijn de kosten lager dan een CT-scan. De werkgroep suggereert daarom echoscopie als de initiële test voor het diagnosticeren van urinestenen. Bij persisteren van de klachten of bij twijfel aan de juiste diagnose kan eventueel in het verdere diagnostische traject alsnog een CT-scan worden uitgevoerd om meer diagnostische zekerheid te verkrijgen (zie bijlage).

25 7. Aanbevelingen

Echoscopie wordt aanbevolen als de initiële test voor het diagnosticeren van urinestenen bij patiënten met flank- of buikpijn die zich melden bij de huisarts of op de spoedeisende hulp.

Referenties

- 30 - Smith-Bindman R, Aubin C, Bailitz J, Bengiamin RN, Camargo CA Jr, Corbo J, et al. Ultrasonography versus computed tomography for suspected nephrolithiasis. N Engl J Med. 2014;371(12):1100-10.

Bijlage 1 Zoekstrategie

Database searched	Search terms
PUBMED	((ureteral calculi[tw] OR ureteral colic[tiab] OR urolithasis[tiab] OR acute flank pain[tiab] OR erythrocyturia[tiab]) AND (computed tomography[tiab] OR "tomography, x-ray computed"[MeSH Terms] OR ultrasound[tiab] OR nonenhanced CT[tiab] OR unenhanced CT[tiab] OR nonenhanced computed tomography[tiab] OR non-contrast[tiab] OR "radiography"[Subheading] OR "ultrasonography"[Subheading]) AND ("sensitivity and specificity"[MeSH Terms] OR "prospective studies"[MeSH Terms] OR "predictive value of tests"[MeSH Terms] OR observational[tiab] OR cohort[tiab] OR cross-sectional[tiab] OR "retrospective studies"[MeSH Terms] OR "follow-up studies"[MeSH Terms]) AND ("2007/01/01"[PDAT] : "3000/12/31"[PDAT])) OR (("Urinary Calculi/radiography"[Majr] OR "Urinary Calculi/ultrasonography"[Majr] OR "colic/radiography"[MAJR] OR "colic/ultrasonography"[MAJR]) AND ("Sensitivity and Specificity"[Mesh] OR predictive value[tiab]) AND ("2007/01/01"[PDAT] : "3000/12/31"[PDAT])) AND ("2007/01/01"[PDAT] : "3000/12/31"[PDAT]))

Bijlage 2 QUADAS-2 score

Tabel 1 QUADAS-2 score Echoscopie versus CT-scan

Study	Risk Of Bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow And Timing	Patient Selection	Index Test	Reference Standard
Smith-Bindman, 2014	😊	😊	?	😊	😊	😊	😊

😊 Low Risk
 ☹️ High Risk
 ? Unclear Risk

Bijlage 3 GRADE evidence profile / SoF-table

Question: Should ultrasonography vs. computed tomography be used to diagnose urinary stones in patients with suspected nephrolithiasis?

Sensitivity ultrasonography	0.54 (95% CI: 0.48 to 0.6)	Sensitivity computed tomography	0.71 (95% CI: 0.67 to 0.75)	Prevalences	33%					
Specificity ultrasonography	0.88 (95% CI: 0.84 to 0.92)	Specificity computed tomography	0.58 (95% CI: 0.55 to 0.62)							
Outcome	N° of studies (N° of patients)	Study design	Factors that may decrease quality of evidence					Effect per 1000 patients/year		
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	ultrasonography	computed tomography	DTA QoE
True positives (patients with urinary stones)	1 Studies ²¹ Echo: 299 CT: 316	observational studies	not serious	not serious	not serious	not serious	not assessed	178 (158 to 198)	234 (221 to 248)	⊕⊕⊕⊕ HIGH
False negatives (patients incorrectly classified as not having urinary stones)								152 (172 to 132)	96 (109 to 82)	
								FN absolute difference: 56 more		
True negatives (patients without urinary stones)	1 Studies ²² Echo: 609 CT: 642	observational studies	not serious	not serious	not serious	not serious	not assessed	590 (563 to 616)	389 (369 to 415)	⊕⊕⊕⊕ HIGH
False positives (patients incorrectly classified as having urinary stones)								80 (107 to 54)	281 (301 to 255)	
								FP absolute difference: 201 fewer		

²¹ TP + FN = 299 (or 316) patients

²² TN + FP = 609 (or 642) patients

Bijlage 4 Evidence to decision-framework

Question: Should ultrasonography vs. computed tomography be used to diagnose urinary stones in patients with suspected nephrolithiasis?

Evidence and judgements								
Domain	Criteria	Judgements					Evidence	Comments
Burden of illness or problem	Is it frequent?	Very infrequent ○	Infrequent ●	Moderately frequent ○	Frequent ○	Very frequent ○		
	Is it severe?	Very mild ○	Mild ○	Moderate ●	Severe ○	Very severe ○		
Confidence in diagnostic test accuracy	What is the confidence in the diagnostic test accuracy information?	Very low ○	Low ○	Moderate ○	High ●		True Positives: high quality evidence True Negatives: high quality evidence False positives: high quality evidence False negatives: high quality evidence	
Benefits & harms of the options	What is the overall certainty of this evidence?	Very low ○	Low ●	Moderate ○	High ○		Effect of tamsulosine versus placebo on: Expulsion rate: RR 1.51 (1.34 to 1.69)	
	Is there important uncertainty about how much people value the main outcomes?		No ●		Yes ○		Stone expulsion time 3.4 days earlier (2.29 to 4.50)	
	Are the desirable anticipated effects large?	Very uncertain ○	Uncertain ○	Moderately certain ●	Certain ○	Very certain ○		

Confidence in benefits & harms	What is the overall confidence in the estimates of effect for benefits and harms?	Very low ○	Low ●	Moderate ○	High ○	
Values	What is the confidence in the values that patients place on the benefits and harms?	Very low ○	Low ●	Moderate ○	High ○	Very high ○
Resources	Is the cost small relative to the net benefits?	No ○	Uncertain ○	Yes ●		

Voorbeeld 3. Triage test voor dementie

1. Achtergrond

Dementie is een syndroom gekarakteriseerd door cognitieve en neuropsychologische achteruitgang met gevolgen voor het dagelijks functioneren. De neurodegeneratie en klinische manifestatie zijn progressief en tot op heden is er geen genezing. Uit onderzoek blijkt dat verschillende interventies de cognitieve achteruitgang kunnen vertragen. Prevalentie schattingen lopen uiteen van 5% bij mensen ouder dan 60 jaar tot 50 bij mensen ouder dan 85 jaar. De toenemende vergrijzing maakt dat het aantal mensen met dementie zal stijgen, met grote gevolgen voor de gezondheidszorg en de toegankelijkheid daarvan.

In de bestaande situatie vindt uitgebreid, multidisciplinair gespecialiseerd onderzoek plaats bij ouderen met klachten die mogelijk passen bij dementie, zoals geheugenstoornissen. Triage met een initiële test door zou kunnen bijdragen aan toegankelijkheid van de zorg, tijdige herkenning van dementie en het organiseren van passende zorg.

Een veelgebruikte test voor het diagnosticeren van mogelijke dementie is de Informant Questionnaire on Cognitive Decline in the Elderly, een vragenlijst bestaande uit 26 of 16 items. Deze test wordt gebruikt om via een 'proxy', een relevante naaste van de oudere, de verandering in cognitie over de afgelopen 10 jaar te meten. De IQCODE is ontwikkeld als 26-item lijst, de kortere 16-item lijst wordt veel gebruikt in de klinische praktijk en geldt als aanbevolen format. De IQCODE wordt ingevuld door de relevante naaste. Voor ieder item geeft de respondent aan in welke mate verandering is opgetreden in de afgelopen 10 jaar: 1 – is veel beter geworden tot 5 – is veel slechter geworden. Voor de 26-item lijst geeft dit een totale score met een range van 26 tot 130, waaruit een gemiddelde score per item berekend kan worden. Deze gemiddelde score heeft een range van 1 tot 5, hogere scores wijzen op een grotere daling van cognitief functioneren.

Een totale score groter dan 93 of een gemiddelde score van groter dan 3.31 geldt als een positieve test. Er is geen consensus over de optimale afkapwaarde.

1.1 Uitgangsvragen en patiëntrelevante uitkomstmaten

Dient bij zelfstandig wonende ouderen (P) de Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) (I) te worden toegepast als triage test voor het diagnosticeren van dementie?

Tabel 1.1 Selectiecriteria

Type patiënten	- Zelfstandig wonende ouderen
Type inductietest	- IQCODE
Type referentietest	- Klinisch multidisciplinair onderzoek
Type uitkomstmaten	- Patiëntrelevante uitkomsten (zie onderstaand; patiëntrelevante consequenties)

*Patiëntrelevante uitkomsten (O)*²³

- Cognitief functioneren
- Neuropsychologisch functioneren
- Dagelijks functioneren (ADL en iADL)
- Kwaliteit van leven
- Belasting mantelzorg
- Morbiditeit
- Mortaliteit

Welke scenario's worden vergeleken?

Bestaande situatie (geen triage): mensen met verdenking op dementie (klachten van vergeetachtigheid e.d.) worden onderzocht in gespecialiseerde klinische setting.

Nieuwe situatie (triage met IQCODE): initiële test bij mensen zonder klachten; alleen mensen met een positieve testuitslag worden onderzocht in gespecialiseerde klinische setting.

Wat is de gezondheidswinst voor de oudere van een initieel, laagdrempelig assessment, in vergelijking met assessment in specialistische setting wanneer er al een verdenking op dementie is?

Tabel 1.2 Voor de patiënt relevante consequenties

Uitkomstmaat	Gevolgen	Patiëntrelevante consequenties (gebaseerd op gevolgen)	Importantie
Terecht positieven (dementie waarschijnlijk)	Vervolg onderzoek in specialistische klinische setting. Wanneer diagnose bevestigd wordt kan passende zorg (waaronder medicatie, ondersteunende of activerende begeleiding, beschermde woonomgeving) ingezet worden.	Verbetering uitkomstmaten, afhankelijk van effectiviteit van behandeling bij in een vroeg stadium gediagnosticeerde dementie; ongerustheid en stress door mogelijke dementie diagnose	8
Fout positieven	Onterecht vervolg onderzoek in specialistische klinische setting	Onnodige investering van tijd en middelen, mogelijk stress bij patiënt en naasten door ongerustheid	9
Terecht negatieven	Geen vervolg onderzoek		8
Fout negatieven	Geen vervolg onderzoek. Uitgestelde klinische diagnose.	Verslechtering uitkomsten door latere diagnose en mogelijk minder voordeel van behandeling, grotere belasting van mantelzorg(er)s)	7
Niet eenduidig te interpreteren testuitslagen	N.v.t.		
Belasting van de test	Test is niet/weinig belastend; geen nadelige gevolgen		3
Beslag op middelen (kosten)	Tijd om vragenlijst af te nemen, training van interviewers		6

²³ Uitkomsten op gezondheidszorg niveau worden buiten beschouwing gelaten

2. Methode

2.1 Zoeken en selecteren van literatuur

Diagnostische testeigenschappen

De zoekactie was gericht op systematische reviews over de diagnostische testaccuratesse van de IQCODE uitgevoerd bij zelfstandig wonende ouderen (algemene bevolking). De kwaliteit van de systematische reviews werd beoordeeld met het AMSTAR instrument. De meest recente review van goede kwaliteit is als uitgangspunt genomen.

Directe gevolgen van de test

Hiervoor baseren wij ons op de systematische reviews over de diagnostische testeigenschappen. Er is geen separate uitputtende search voor deze vraag gedaan.

Indirecte gevolgen van de test

Voor een schatting van de effecten van interventies voor dementie in de verschillende stadia van de aandoening, de psychosociale consequenties van het krijgen van de diagnose dementie en de belasting van mantelzorgers is gezocht naar bestaande multidisciplinaire richtlijnen en recente overviews van reviews en systematische reviews.

3. Resultaten

3.1 Diagnostische testaccuratesse

3.1.1 Beschrijving studies

De Cochrane review van Quinn et al (2014) was de meest recente systematische review van goede kwaliteit (Quinn 2014). De review auteurs includeerden 'community-based' cohort studies, waarbij 'community-based' gedefinieerd was als een setting waarin deelnemers niet geselecteerd waren op basis van cognitieve scores of symptomen of verwijzing naar gespecialiseerde zorg. Deelnemers moesten tenminste 18 jaar zijn, maar verwacht werd dat de meeste deelnemers 65 jaar of ouder zouden zijn. Diagnostische patiënt controle onderzoeken, case studies en onderzoeken met 10 of minder personen werden niet geïncludeerd.

De onderzochte index test moest de IQCODE zijn, de 'target condition' klinisch gediagnosticeerde dementie (alle typen), gediagnosticeerd volgens erkende criteria zoals ICD-10 of DSM-IV classificatie.

De zoekactie werd uitgevoerd op 28 januari 2013 in de volgende databases: ALOIS, MEDLINE, EMBASE, PsycINFO, BIOSIS, Web of Science, LILACS, MEDION, DARE en ARIF. De auteurs gebruikten een sensitieve zoekstrategie gebaseerd op onder andere MeSH termen.

De methodologische kwaliteit van de studies werd beoordeeld met QUADAS-2.

Van de 16.144 referenties beschreven 71 artikelen de testaccuratesse van IQCODE. Tien studies voldeden aan de inclusie criteria. De tien studies rapporteerden de resultaten van 11 datasets met in totaal 2.644 personen. De prevalentie van dementie was 14%.

Geen enkele studie had een laag risico op vertekening in alle categorieën van QUADAS-2. Belangrijkste beperkingen waren samenstelling van de cohorten (geen inclusie van opeenvolgende deelnemers) en onvoldoende informatie over de wijze van toepassen van de IQCODE. Ook waren er zorgen over de generaliseerbaarheid, vooral met betrekking tot het selecteren van deelnemers (cohort niet representatief voor zelfstandig wonende ouderen).

3.1.2. Kwaliteit van het bewijs

De kwaliteit van bewijs was laag²⁴ (zie GRADE profiel in bijlage 1).

3.1.3 Diagnostische testeigenschappen

Bij een afkapwaarde van 3,3 was de gepoolde sensitiviteit is 0,80 (95% BI 0,75-0,85) en de gepoolde specificiteit 0,74 (95% BI 0,78-0,90). De testaccuratesse van de 16-item en 26-item lijst was vergelijkbaar.

Bij een prevalentie van 20% worden met de IQCODE 40 per 1000 (95% BI 30-50) geteste mensen met dementie gemist en 128 per 1000 (95% BI 80-176) geteste mensen ten onrechte met mogelijke dementie gediagnosticeerd. De test lijkt beter in het uitsluiten dan aantonen van dementie.

3.2 Directe gevolgen van de test

Eventuele belasting van de test werd niet besproken in de systematische review.

3.2.2 Kwaliteit van bewijs: geen bewijs gevonden

3.3 Gevolgen van de test op de patiëntrelevante uitkomsten; natuurlijk beloop en effectiviteit van behandeling

3.3.1 Beschrijving studies

Zie Zorgstandaard Dementie 2013 en Richtlijn diagnostiek en behandeling van dementie 2014.

3.3.2 Kwaliteit van het bewijs: laag- matig

3.3.3 Consequenties

Het ziekteproces van dementie verschilt sterk per persoon, maar is altijd een proces van achteruitgang. Uiteindelijk zal iemand met dementie overlijden aan de gevolgen daarvan. De interventies voor (in een vroeg stadium gediagnosticeerde) dementie zijn beperkt tot matig effectief. (Zorgstandaard Dementie 2013, Richtlijn diagnostiek en behandeling van dementie, 2014).

4. Conclusies

4.1 Diagnostische testeigenschappen

Er is geringe zekerheid over de diagnostische testeigenschappen van IQCODE (vergeleken met klinische evaluatie) voor het diagnosticeren van dementie bij ouderen in de algemene bevolking. Bij een dementie prevalentie van 20% worden met de IQCODE 40 per 1000 (95% BI 30-50) geteste mensen met dementie gemist en 128 per 1000 (95% BI 80-176) geteste mensen ten onrechte met mogelijke dementie gediagnosticeerd. De test lijkt beter in het uitsluiten dan aantonen van dementie.

4.2 Directe gevolgen van de test en consequenties

Er is geringe tot matige zekerheid over de effectiviteit van behandeling.

²⁴ De kwaliteit van bewijs was matig voor de sensitiviteit en laag voor de specificiteit vanwege methodologische beperkingen in het samenstellen van de onderzoekspopulatie (sensitiviteit en specificiteit), en vanwege een grote onzekerheidsmarge van de fout-positieven (specificiteit).

5. Van bewijs naar aanbevelingen

5.1 Kwaliteit van bewijs

De algehele kwaliteit van bewijs is laag, met name vanwege de geringe zekerheid over de diagnostische testeigenschappen en het ontbreken van een effectieve behandeling.

5.2 Waarden en voorkeuren van patiënten over patiëntrelevante uitkomsten en de directe gevolgen van de test

Door het dynamische ziekteproces is er geen standaard zorgpad te omschrijven en vergt de zorg continue afstemming met de persoon met dementie en zijn naasten (Zorgstandaard Dementie 2013). De waarden en voorkeuren van patiënten zullen daarom variëren.

5.3 Balans tussen gunstige en ongunstige effecten van de test

Ten aanzien van medicamenteuze therapie is het van belang om de bijwerkingen van medicatie die het dagelijks functioneren beïnvloeden (bijv. gastro-intestinale bijwerkingen, sufheid) zwaar te laten wegen. Gegeven de onzekerheid over en de geringe klinische relevantie van de grootte van de effecten van medicatie op cognitie en ADL is er geen duidelijk overwicht van gunstige of gunstige effecten van de test.

5.4 Kosten

De kosten van de IQCODE zijn beperkt in vergelijking met het gespecialiseerde klinische onderzoek (inschatting werkgroep). Op de test berust geen copyright.

5.5 (On-)gelijkheid in gezondheidsuitkomsten

Niet besproken.

5.6 en 5.7 Aanvaardbaarheid en haalbaarheid

Wat de haalbaarheid betreft kan tijdgebrek in de eerstelijnssetting een probleem zijn. Eventuele kennistekorten bij het toepassen en interpreteren van de triagetest kunnen relatief gemakkelijk worden opgelost (inschatting werkgroep).

6. Rationale

Het gebruik van de IQCODE als op zichzelf staande triage test zal leiden tot een aanzienlijk percentage fout-positieve uitslagen. Gezien de publieke opinie over dementie is het waarschijnlijk dat de ongerustheid en stress door de diagnose dementie zwaarder weegt dan het in eerste instantie missen van dementie (Quinn 2014). Daarnaast is er variatie in de waarden en voorkeuren van patiënten en onzekerheid over de effectiviteit van behandeling.

7. Aanbevelingen

De werkgroep doet de suggestie om de IQCODE van 16 items als triagetest toe te passen in een eerstelijnssetting, gevolgd door klinische evaluatie in een in dementie gespecialiseerde instelling indien de test positief is, en door een afwachtend beleid indien de test negatief is (*conditionele aanbeveling gebaseerd op lage kwaliteit van bewijs*).

Referenties

- Quinn TJ, Fearon P, Noel-Storr AH, Young C, McShane R, Stott DJ. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the diagnosis of dementia within community dwelling populations. Cochrane Database of Systematic Reviews 2014, Issue 4. Art. No.: CD010079. DOI: 10.1002/14651858.CD010079.pub2.
- Alzheimer Nederland & Vilans. Zorgstandaard Dementie, 2013 (http://www.vilans.nl/docs/vilans/publicaties/Zorgstandaard_Dementie.pdf)
- Nederlandse Vereniging voor Klinische Geriatrie. Richtlijn Diagnostiek en Behandeling van dementie, 2014

Bijlage 1. GRADE evidence profile

Question: Should IQCODE be used to diagnose dementia in community dwelling populations?

Pooled sensitivity	0.80 (95% CI: 0.75 to 0.85)
Pooled specificity	0.84 (95% CI: 0.78 to 0.90)

Prevalences	5%	20%	50%
-------------	----	-----	-----

Outcome	N ^o of studies (N ^o of patients)	Study design	Factors that may decrease quality of evidence					Effect per 1000 patients/year			DTA QoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Publication bias	pre-test probability of 5%	pre-test probability of 20%	pre-test probability of 50%	
True positives (patients with dementia)	11 Studies 2644 Patients	observational studies	serious ¹	not serious ²	not serious	not serious	not assessed	40 (38 to 43)	160 (150 to 170)	400 (375 to 425)	MODERATE
False negatives (patients incorrectly classified as not having dementia)								10 (12 to 7)	40 (50 to 30)	100 (125 to 75)	
True negatives (patients without dementia)	11 Studies 2644 Patients	observational studies	serious ¹	not serious ²	not serious ³	serious ⁴	not assessed	798 (741 to 855)	672 (624 to 720)	420 (390 to 450)	LOW
False positives (patients incorrectly classified as having dementia)								152 (209 to 95)	128 (176 to 80)	80 (110 to 50)	

1. Areas of particular concern for bias: participant sampling procedures (n = 2 papers graded low risk, with few papers using a true consecutive sampling frame) and application of index test (n = 1 paper graded low risk of bias, with most papers giving insufficient detail on how the IQCODE was actually applied in practice).
2. Concerns around applicability, particularly concerning patient selection procedures (n = 1 paper graded no concern, with few studies recruiting a cohort representative of community-dwelling older adults) - downgraded under risk of bias.
3. Some heterogeneity, no downgrading
4. Wide CI around false positive rate

