

## COMMENTAAR

## STARD 2015 voor de evaluatie van diagnostische tests

Daniël A. Korevaar en Patrick M.M. Bossuyt

Recent verscheen in de Cochrane Library een systematisch literatuuroverzicht van onderzoek naar de accuratesse van 2 radiologische tests voor het detecteren van neuroblastomen:  $^{123}\text{I}$ -MIBG-scintigrafie en  $^{18}\text{F}$ -FDG-PET.<sup>1</sup> De auteurs van het overzicht hadden geprobeerd de kwaliteit van elk van de 11 geïncludeerde studies in kaart te brengen. Ze deden dit aan de hand van een lijst van 16 vragen die vertekening of een beperkte toepasbaarheid van de resultaten kunnen opsporen. Voorbeelden van dit soort vragen zijn: 'Waren de beoordelaars van de tests geblindeerd voor de resultaten van de referentiestandaard?' en 'Was de referentiestandaard acceptabel?'. Het resultaat van deze exercitie was teleurstellend. In bijna de helft van de gevallen (78/176) konden de auteurs een vraag niet beantwoorden omdat de benodigde gegevens niet in het artikel te vinden waren.

Dit is geen uitzonderlijke casus. Maar al te vaak moeten lezers van de biomedische literatuur en ontwikkelaars van klinische richtlijnen concluderen dat ze de geloofwaardigheid van onderzoek naar diagnostische tests niet goed kunnen beoordelen, omdat de vereiste informatie ontbreekt in het onderzoeksverslag.<sup>2,3</sup> Kennelijk bestaat er een forse discrepantie tussen de informatie die hiervoor nodig is en wat onderzoekers denken te moeten vermelden in hun artikel.

#### ONDERZOEK NAAR DIAGNOSTISCHE ACCURATESSE

Jaarlijks verschijnen er veel nieuwe tests op de markt die betrouwbaarder, beter, sneller, goedkoper of minder invasief zouden zijn dan de reeds beschikbare tests. Onderzoek naar de diagnostische accuratesse van die tests kan helpen bij het beoordelen van dergelijke claims.<sup>4</sup> Het gaat dan om het vermogen van de test om een bepaalde aandoening te detecteren. De resultaten van zulk onderzoek worden meestal uitgedrukt als schattingen van de sensitiviteit en de specificiteit van de test.

Bij het interpreteren van diagnostische accuratessestudies is het echter oppassen geblazen: de prestaties kunnen vertekend zijn.<sup>5</sup> Zo schatten sommige studies de sensitiviteit van een test bij zeer zieke patiënten, terwijl ze de specificiteit baseren op de fout-positieve resultaten in een aparte groep supergezonde controlepersonen. In andere gevallen is de beoordelaar van de onderzochte test niet geblindeerd voor de resultaten van de referentiestandaard. Dit soort situaties kan leiden tot een overschatting van de accuratesse.

De sensitiviteit en specificiteit van een test liggen niet vast, maar kunnen variëren tussen populaties en situaties, afhankelijk van bijvoorbeeld behandelingen en tests die patiënten al hebben ondergaan, of van het doel van de test: gaat het om screening, diagnostiek, of monitoring?<sup>5</sup> We moeten dus in een onderzoeksverslag kunnen lezen waar, wanneer, bij wie en hoe de accuratesse is geschat, om te kijken of de geschatte accuratesse van toepassing is op de situatie waarin wij de test willen gebruiken.

#### STARD: EEN HULPMIDDEL OM VERSLAGLEGGING TE VERBETEREN

De Verklaring van Helsinki stelt onomwonden dat 'onderzoekers [...] verantwoordelijk (zijn) voor de compleetheid en accuraatheid van hun studierapporten'.<sup>6</sup> Maar wanneer is een diagnostische accuratessestudie compleet en accuraat gerapporteerd? Om auteurs te helpen besloot een internationale groep wetenschappers en andere belanghebbenden richtlijnen op te stellen voor de verslaglegging van dit type onderzoek. Dit leidde in 2003 tot STARD, dat staat voor 'STAndards for Reporting Diagnostic accuracy'. STARD werd gepubliceerd in veel tijdschriften, waaronder het *NTvG*,<sup>7</sup> en geadopteerd door nog veel meer andere tijdschriften. Recent heeft de STARD-groep, die inmiddels 85 personen bevat, een update uitgebracht.

#### STARD 2015: WAT IS ER VERANDERD?

STARD 2015 bestaat uit een lijst van 30 essentiële items (voorheen 25).<sup>8</sup> Het doel van de update was om het gebruiksgemak van STARD te bevorderen en tegelijk een aantal nieuwe, essentiële elementen op te nemen in de lijst. In de voorbije 10 jaar hebben meerdere onderzoekers gekeken of gepubliceerde artikelen wel voldoen aan STARD.<sup>2</sup> Daaruit bleek dat er soms verschil zit in hoe items geïnterpreteerd worden. De STARD-groep nam zich voor om dubbelzinnigheden te verwijderen en com-

Academisch Medisch Centrum-Universiteit van Amsterdam,  
afd. Klinische Epidemiologie, Biostatistiek en Bioinformatica,  
Amsterdam.

Drs. D.A. Korevaar, arts-onderzoeker; prof.dr. P.M.M. Bossuyt,  
klinisch epidemioloog.

Contactpersoon: prof.dr. P.M.M. Bossuyt  
(p.m.bossuyt@amc.uva.nl).

plexe items te verduidelijken en te vereenvoudigen. Daarnaast heeft de STARD-groep ernaar gestreefd om de gebruikte begrippen en bewoordingen zoveel mogelijk aan te laten sluiten bij die van andere rapportagerichtlijnen, zoals CONSORT voor gerandomiseerde trials (CONSORT staat voor 'Consolidated standards of reporting trials').<sup>9</sup>

STARD 2015 bevat ook geheel nieuwe items.<sup>8</sup> Een aantal hiervan hangt samen met de toenemende roep om meer transparantie en openbaarheid van onderzoeksgegevens. STARD verzoekt auteurs nu bijvoorbeeld om een registratienummer te vermelden, aan te geven waar het studieprotocol gevonden kan worden, en eventuele sponsors van het onderzoek expliciet te vermelden.

Andere nieuwe items hebben te maken met de doelgerichtheid van onderzoek. Op dat vlak lopen diagnostische accuratessestudies gemiddeld genomen namelijk achter op trials naar de effectiviteit van therapeutische interventies. Zo is het in onderzoeksverslagen meestal onduidelijk wáár de auteurs de onderzochte test zouden willen inpassen in het bestaande klinische pad, wat hun hypothese is over de vereiste of gewenste accuratesse van de onderzochte test, en hoeveel proefpersonen ze wensten te includeren om deze hypothese te toetsen.<sup>10</sup> De STARD-groep verzoekt auteurs nu om deze informatie te vermelden.

Ook vindt de STARD-groep dat een stroomdiagram een vast onderdeel hoort te zijn in de rapportage van elke diagnostische accuratessestudie. Dit stroomdiagram maakt het voor lezers makkelijker om na te gaan hoeveel patiënten oorspronkelijk in aanmerking kwamen voor

deelname aan het onderzoek, wat de redenen voor exclusie waren, hoeveel patiënten de verschillende tests ondergingen, en wat de testuitslagen waren.

#### EN NU?

10 jaar geleden bevatten de meeste onderzoeksverslagen nog niet de helft van de vereiste gegevens. In de voorbije jaren is de verslaglegging van diagnostische accuratessestudies geleidelijk verbeterd, maar deze is nog verre van ideaal.<sup>2,3</sup> De ontwerpers van STARD 2015 willen nieuwe instrumenten ontwikkelen, gebaseerd op de lijst van essentiële items. Naast deze afvinklijst komen er een mal voor het schrijven van artikelen en hulpmiddelen voor tijdschriftredacties.<sup>8</sup> Daarnaast zullen er ook aanvullende instructies komen voor specifieke vormen van tests, zoals beeldvormend onderzoek of laboratoriumdiagnostiek. Bij elkaar moeten die initiatieven ervoor zorgen dat de lezers van onderzoeksverslagen kunnen vinden wat ze nodig hebben om de evaluatie van een test goed te kunnen beoordelen, en desgewenst te repliceren.

Belangenconflict en financiële ondersteuning: geen gemeld./Belangenconflict en financiële ondersteuning: ICMJE-formulieren zijn online beschikbaar bij dit artikel.

Aanvaard op 4 februari 2016

Citeer als: Ned Tijdschr Geneeskd. 2016;160:D113

 **KIJK OOK OP [WWW.NTVG.NL/D113](http://WWW.NTVG.NL/D113)**

## LITERATUUR

- Bleeker G, Tytgat GA, Adam JA, et al. 123I-MIBG scintigraphy and 18F-FDG-PET imaging for diagnosing neuroblastoma. *Cochrane Database Syst Rev.* 2015;9:CD009263.
- Korevaar DA, van Enst WA, Spijker R, Bossuyt PM, Hooft L. Reporting quality of diagnostic accuracy studies: a systematic review and meta-analysis of investigations on adherence to STARD. *Evid Based Med.* 2014;19:47-54.
- Korevaar DA, Wang J, van Enst WA, et al. Reporting diagnostic accuracy studies: some improvements after 10 years of STARD. *Radiology.* 2015;274:781-9.
- Linnet K, Bossuyt PM, Moons KG, Reitsma JB. Quantifying the accuracy of a diagnostic test or marker. *Clin Chem.* 2012;58:1292-301.
- Whiting PF, Rutjes AW, Westwood ME, Mallett S; QUADAS-2 Steering Group. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *J Clin Epidemiol.* 2013;66:1093-104.
- World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA.* 2013;310:2191-4.
- Bossuyt PM, Reitsma JB, Bruns DE, et al. Verslaglegging van diagnostisch evaluatieonderzoek volgens een standaardmethode; de 'Standards for reporting of diagnostic accuracy' (STARD). *Ned Tijdschr Geneeskd.* 2003;147:336-40.
- Bossuyt PM, Reitsma JB, Bruns DE, et al; STARD Group. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ.* 2015;351:h5527.
- Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ.* 2010;340:c332.
- Ochodo EA, de Haan MC, Reitsma JB, Hooft L, Bossuyt PM, Leeflang MM. Overinterpretation and misreporting of diagnostic accuracy studies: evidence of 'spin'. *Radiology.* 2013;267:581-8.