

3x2 tabellen?

Tom Poelman, Vakgroep Huisartsgeneeskunde en Eerstelijnsgezondheidszorg, Universiteit Gent

Wanneer de arts een medische beslissing neemt baseert hij zich vaak op informatie uit diagnostische testen. Net zoals onderzoek naar de effectiviteit van een behandeling is onderzoek naar de accuraatheid van diagnostische testen onontbeerlijk. Studies die een te evalueren test (=indextest) vergelijken met een test waarvan de accuraatheid voldoende is aangetoond (=referentietest) vormen hierbij een belangrijke stap. De resultaten van een diagnostisch onderzoek worden gewoonlijk uitgezet in een vierveldentabel die toelaat om sensitiviteit, specificiteit, positieve en negatieve voorspellende waarde, positieve en negatieve likelihoodratio, aantonnende en ontkenkende kracht te berekenen.

Maar, zoals onderzoek naar de effectiviteit van behandelingen, is ook diagnostisch onderzoek onderhevig aan verschillende vormen van bias^{1,2}. Zowel de **Standards for the Reporting of Diagnostic Accuracy (STARD)**- als de **Quality Assessment of Diagnostic Accuracy Studies (QUADAS)**-vragenlijst³, die respectievelijk de rapportering en de kwaliteit van diagnostisch onderzoek evalueren, verwijzen naar verschillende vormen van bias op het vlak van selectie van patiënten en wijze en tijdstip van uitvoering van index- en referentietest.

Een recente meta-analyse vergeleek de diagnostische accuraatheid van niet-invasieve coronaire CT-angiografie (de indextest) met conventionele coronarografie (de referentietest)⁴. In 109 van de 120 studies werden de onderzoekers geconfronteerd met niet-evalueerbare resultaten. Hierbij ging het meestal om vernauwde bloedvaten op CT-angiografie waarvan de onderzoekers de klinische relevantie (minstens 50% stenose) voor de patiënt niet konden inschatten. Daardoor was het dus onmogelijk om de indextest als ondubbelzinnig positief of negatief te klasseren. Hoewel zowel STARD als QUADAS de vraag stellen naar de rapportering van 'uninterpretable' en 'indeterminate' resultaten van de indextest was het voor 24 van de 109 studies niet duidelijk of en hoe ze niet-evalueerbare resultaten naar patiëntniveau vertaalden. Naast de rapportering heeft de STARD-vragenlijst ook aandacht voor de manier waarop met deze niet-evalueerbare resultaten is omgegaan.

| | | Referentietest | |
|-----------|-------------------|--|--|
| | | Positief | Negatief |
| Indextest | Positief | Terecht positief | Vals positief |
| | Niet-evalueerbaar | Niet-evalueerbaar met positieve referentietest | Niet-evalueerbaar met negatieve referentietest |
| | Negatief | Vals negatief | Terecht negatief |

Bron: Schuetz GM, Schlattmann P, Dewey M. Use of 3x2 tables with an intention to diagnose approach to assess clinical performance of diagnostic tests: meta-analytical evaluation of coronary CT angiography studies. *BMJ* 2012;345:e6717.

Uit de systematische review bleek dat 26 studies geen rekening hielden met niet-evalueerbare resultaten, dat 23 studies patiënten met niet-evalueerbare resultaten uitsloten en dat 26 en 7 studies alle patiënten met niet-evalueerbare resultaten respectievelijk als positief of negatief beschouwden. Vanuit klinisch

standpunt is het natuurlijk correct om alle niet-evalueerbare resultaten als positief te beschouwen en de patiënt voor verder onderzoek door te sturen. Deze aanpak zal de diagnostische accuraatheid van een indextest echter vals verhogen. Slechts 3 studies gebruikten een 3x2 tabel⁵ (zie tabel) om niet-evalueerbare resultaten een eigen plaats te geven. Voor 23 studies konden de auteurs van de systematische review zelf een 3x2 tabel tekenen omdat ook voor de niet-evalueerbare resultaten het resultaat op invasieve coronarografie (de referentietest) bekend was. Meteen bleek dat de diagnostische accuraatheid significant ($p < 0,05$) lager was voor de 3x2 tabellen dan voor de 2x2 tabellen. Als we de sensitiviteit en de specificiteit berekenen met een 3x2 tabel kunnen we, naar analogie met een intention to treat analyse bij interventioneel onderzoek, een analyse volgens 'intention to diagnose' uitvoeren. Wanneer men in de 3x2 tabel alle niet-evalueerbare resultaten als positief beschouwt, zullen het aantal terecht positieven en het aantal vals-positieven toenemen met als gevolg dat de sensitiviteit zal toenemen en de specificiteit zal dalen. Omgekeerd, worden alle niet-evalueerbare resultaten als negatief beschouwd, dan zullen het aantal terecht negatieven en het aantal vals-negatieven toenemen met als gevolg dat de sensitiviteit daalt en de specificiteit toeneemt. De auteurs van de meta-analyse kwamen hierdoor tot een sensitiviteit die varieerde van 98,3 (95% BI 96,9 tot 99,0) tot 92,9 (95% BI 88,8 tot 95,5) en een specificiteit die varieerde van 78,4 (95% BI 71,6 tot 84,0) tot 90,5 (95% BI 86,8 tot 93,2).

Om een reëel beeld (zonder overschatting) van de diagnostische accuraatheid te krijgen, moeten we de niet-evalueerbare resultaten bij patiënten met een positieve referentietest dus als vals-negatief en de niet-evalueerbare resultaten bij patiënten met een negatieve referentietest als vals-positief beschouwen. Het klinische belang van een intention to diagnose analyse met een 3x2 tabel versus een analyse met een 2x2 tabel wordt zeer goed geïllustreerd bij de berekening van de negatieve likelihoodratio. Deze is respectievelijk 0,09 (95% BI 0,06 tot 0,15) en 0,02 (95% BI 0,01 tot 0,04). De ontkenkende kracht daalt dus van 50 naar 11 door intention to diagnose analyse met een 3x2 tabel, m.a.w. een CT-angiografie zal minder goed coronair lijden uitsluiten als we correct rekening houden met niet-evalueerbare resultaten die zich in de praktijk voordoen.

Bij de bespreking van diagnostische onderzoeken in Minerva hebben we nooit rekening gehouden met de manier waarop de onderzoekers met niet-evalueerbare resultaten zijn omgegaan. Zeker voor diagnostische testen waarbij de uitvoerders van de indextest een zekere expertise moeten hebben⁶ om deze juist te interpreteren is het van belang om na te kijken of de verwerking van niet-evalueerbare resultaten correct gebeurde.

Besluit

Transparante rapportering van de resultaten van diagnostisch onderzoek in een 3x2 tabel met inclusie van niet-evalueerbare resultaten en het toepassen van een intention to diagnose analyse, geeft een realistischer beeld van de accuraatheid van een diagnostische test.

Referenties: zie website www.minerva-ebm.be