



Huisarts en glazen bol: voorspellen op het spreekuur

Geert-Jan Geersing is *keynote speaker* op de NHG-Wetenschapsdag op 13 juni 2014 in Groningen. Thema van deze dag is: 'Geneeskunde-op-maat... wie wil dat niet?' In deze beschouwing licht de auteur alvast een tipje van de sluier op van wat hij daar zal vertellen.

Samenvatting

Geersing GJ. Huisarts en glazen bol: voorspellen op het spreekuur. *Huisarts Wet* 2014;57(6):298-301.

Een belangrijk deel van het werk van de huisarts op het spreekuur bestaat uit het maken van schattingen: hoe groot is de kans op (ernstige) ziekte? Predictieonderzoek – meer bepaald: diagnostisch predictieonderzoek – biedt de huisarts de mogelijkheid die kans zo nauwkeurig mogelijk te bepalen. Niet ieder predictieonderzoek is echter geschikt voor gebruik in de huisartsenpraktijk. Van belang zijn de populatie waarin het is uitgevoerd (homogeen of heterogeen?), het aantal diagnostische variabelen dat wordt meegewogen (één enkele testuitslag of een beslisregel?) en de referentietest (hebben alle onderzochte patiënten er een gehad?).

Een belangrijk deel van het werk van de huisarts op het spreekuur bestaat uit het maken van schattingen: hoe groot is de kans op een (ernstige) ziekte? Die kansschattingen zijn natuurlijk gebaseerd op een combinatie van symptomen en bevindingen uit lichamelijk onderzoek en soms op aanvullend bloed- of beeldvormend onderzoek, maar vaak voor een deel ook intuïtief. In het consult met mevrouw Willemse [casus] zullen de volgende zaken door uw hoofd gaan: is hier sprake van hartfalen? Of COPD? Of allicht beide? Mevrouw is een straffe rookster en dat al heel lang, en ook die bloeddrukbehandeling kwam lastig van de grond. Welke diagnostiek zou nu meerwaarde hebben? Hoe groot is bij dit risicoprofiel eigenlijk de kans dat ze in de toekomst een acuut coronair syndroom krijgt? Moet ik haar misschien verwijzen? Maar naar wie dan? De cardioloog? De longarts? Of misschien wel de vasculair internist?

Dergelijke vragen vormen de kern van de huisartsgeneeskunde en vullen een belangrijk deel van het spreekuur. Met dit soort vragen worstelt de huisarts veel vaker dan met vragen over de meest zinvolle behandeling. Immers, zodra het antwoord gevonden is en de diagnose of prognose bekend is, volgt het te voeren beleid vrijwel altijd vanzelf. Het onderzoek dat diagnostische of prognostische vragen tracht te beantwoorden, heet predictieonderzoek.^{1,2}

Predictieonderzoek gaat over voorspellen: het voorspellen van iets dat al heeft plaatsgevonden maar voor dokter en patiënt nog onbekend is – een diagnose – of van iets dat in de toekomst waarschijnlijk gaat gebeuren – een prognose. Predictieonderzoek is dus voor de huisarts wat de magische glazen bol is voor de waarzegger. In deze beschouwing bespreek ik de achtergronden en methoden van predictieonderzoek. Daarna ga ik in op enkele specifieke kenmerken waaraan predictieonderzoek moet voldoen, wil het bruikbaar zijn voor de huisarts. Tot slot geef ik een korte checklist op basis waarvan huisartsen kunnen nagaan of een bepaald diagnostisch model, zoals dat gepresenteerd wordt tijdens nascholingen of in de literatuur, ook toepasbaar is in hun praktijk. Ik zal me daarbij vooral richten op diagnostisch predictieonderzoek. Prognostisch predictieonderzoek volgt in essentie dezelfde methoden en kent vergelijkbare problemen, maar een gedetailleerde uiteenzetting zou in dit bestek te ver voeren.

DIAGNOSTISCH PREDICTIEONDERZOEK

Anders dan bij de evaluatie van interventies is de opzet van predictieonderzoek bijna altijd observationeel. Bij diagnostisch predictieonderzoek verzamelen de deelnemende huisartsen, bij voorkeur prospectief, gegevens van patiënten bij wie zij een specifieke aandoening vermoeden. Een belangrijk uitgangspunt is dat de selectie van patiënten plaatsvindt op basis van de verdenking en niet op basis van aanvullend onderzoek dat de diagnose al dan niet bevestigt. De selectie van patiënten zal dus bijna altijd plaatsvinden tijdens het spreekuur. Daarnaast is de opzet van predictieonderzoek pragmatisch: de deelnemende huisartsen scoren de relevante bevindingen uit anamnese en lichamelijk onderzoek, en eventueel de uitslag van een bloedtest (bijvoorbeeld een snel-

De kern

- Diagnostisch predictieonderzoek biedt de mogelijkheid de voorafkans op een bepaalde aandoening in te schatten.
- Niet alle met predictieonderzoek ontwikkelde instrumenten zijn geschikt voor gebruik in de huisartsenpraktijk; nog steeds worden echter ongeschikte instrumenten aanbevolen.
- Predictie-instrumenten moeten worden gebruikt in soortgelijke populaties als die waarin ze zijn ontwikkeld. Predictieonderzoek in geselecteerde (tweedelijs)populaties is meestal niet toepasbaar in de huisartsenpraktijk.
- Predictie-instrumenten staan nooit op zichzelf maar moeten worden geïnterpreteerd in samenhang met de bevindingen uit anamnese en lichamelijk onderzoek.
- Niet alle patiënten in een predictieonderzoek hoeven dezelfde referentietest ondergaan te hebben om de uiteindelijke diagnose vast te stellen, maar ze moeten wel allemaal een referentietest hebben gehad.

Casus

Mevrouw Willemse

Mevrouw Willemse, 72 jaar, komt op het spreekuur: 'Dokter, ik kom nu toch maar eens bij u. Ik kom al een hele tijd moeilijk de trap op; bovenaan ben ik helemaal buiten adem. Mijn adem giert dan echt zo, dokter. En ik hou er juist zo van om even dat rondje nog te maken op de Dappermarkt. Maar nu laat ik het maar; iedereen kijkt me aan, door dat hijgen. Laatst nog, toen sprak mijn oude buurvrouw me aan en zei ze: 'Mens, wat word jij oud!' Nou, leuk is anders! Deze week werd ik 's nachts wakker van benauwdheid, ik heb wel een uur voor het raam staan hijgen en hoesten... Ik wil toch echt nog even door. Een trap oplopen kan ik niet meer. Dat is toch niet normaal dokter?'

test). Daarna bepaalt men de referentietest: het (aanvullende) onderzoek dat kan vaststellen of de patiënt ook daadwerkelijk de betreffende aandoening heeft. Idealiter is de referentietest voor alle patiënten hetzelfde, maar dit is niet altijd haalbaar. Om een voorbeeld te noemen: in een recent onderzoek naar de diagnostiek van longembolie door de huisarts zou het de voorkeur hebben gehad alle patiënten een CT-scan te laten ondergaan maar dit bleek niet haalbaar, simpelweg omdat het in de klinische praktijk niet bij alle patiënten gebeurt. Een deel van de verwezen patiënten kreeg, al dan niet terecht, géén CT-scan na het maken van een ECG en/of bloedgasanalyse; bij die patiënten verving een follow-upperiode de referentiestandaard. Weer andere patiënten kregen, conform het lokale protocol, een ventilatie-perfusiescan.³ Niet iedereen kreeg dus dezelfde referentietest. Dat leidt inherent tot enige vertroebeling van de resultaten, maar volgt juist wel erg goed de klinische praktijk.⁴ Bovendien kan een al te sterke selectie van patiënten op basis van één bepaalde referentietest juist leiden tot bias: alleen patiënten met grote kans op de betreffende ziekte worden ingesloten in het onderzoek.⁵

In de analysefase van het predictieonderzoek bepaalt men de voorspellende waarde van de gescoorde items. In zijn simpelste vorm leidt dit tot de bekende 2 x 2-tabel die de diagnostische accuratesse van een bepaalde test aangeeft in termen van sensitiviteit en specificiteit. In de praktijk worden voorspellingen echter zelden gedaan op basis van één variabele; bijna altijd zijn het combinaties van verschillende variabelen. Huisartsen zullen de voorspellende waarde van iedere diagnostische test – met inbegrip van anamnese en lichamelijk onderzoek – beoordelen in de context van eerdere bevindingen. Met andere woorden, zij kijken naar de onafhankelijke, toegevoegde waarde van elk diagnostisch onderzoek. Deze onafhankelijke waarde is met behulp van logistische regressieanalyse in een formule te vatten. Ter wille van de praktische bruikbaarheid wordt deze formule vaak omgezet naar een nomogram of klinische beslisleiding. In [tabel 1] staat een voorbeeld van diepe veneuze trombose.⁶

In de dagelijkse praktijk kunnen beslisleidingen op twee manieren worden gebruikt. De eerste manier, die het meest gebruikt wordt, is een bepaalde score aan te wijzen als afkappunt om het beleid direct te sturen. Bij de beslisleiding in [tabel 1] is dat bijvoorbeeld: wel verwijzen voor een echo van het been bij een score ≥ 4, niet verwijzen bij een score ≤ 3 (inclusief D-dimeerwaarde). De tweede methode is de procentuele kans op

Tabel 1 Klinische beslisleiding bij verdenking op diepe veneuze trombose

Diagnostische variabele	OR (95%-BI)	Regressiecoëfficiënt	Puntenscore
Constante		-5,47	
Mannelijk geslacht	1,80 (1,36-2,16)	0,59	1
Pilgebruik	2,12 (1,32-3,35)	0,75	1
Maligniteit	1,52 (1,05-2,44)	0,42	1
Recente operatie	1,46 (1,02-2,09)	0,38	1
Afwezigheid van trauma	1,82 (1,25-2,66)	0,60	1
Uitgezette venen	1,62 (1,19-2,20)	0,48	1
Kuitomvang ≥ 3 cm	3,10 (2,36-4,06)	1,13	2
Afwijkende D-dimeerwaarde	20,3 (8,25-49,9)	3,01	6

BI = betrouwbaarheidsinterval; OR = oddsratio.

De kans op diepe veneuze trombose kan per patiënt worden berekend met behulp van de regressiecoëfficiënten uit het onderliggende model. Deze kans is $1 / (1 + \exp(-5,47 + 0,59 \times \text{mannelijk geslacht} + 0,75 \times \text{pilgebruik} + 0,42 \times \text{maligniteit} + 0,38 \times \text{recente operatie} + 0,60 \times \text{afwezigheid van trauma} + 0,48 \times \text{uitgezette venen} + 1,13 \times \text{toename kuitomvang} \geq 3 \text{ cm} + 3,01 \times \text{afwijkende D-dimeerwaarde}))$. Voor de dagelijkse praktijk kan men dit omrekenen naar een afkappunt: de patiënt moet verwezen worden bij een score ≥ 4 punten.

diepe veneuze trombose te berekenen aan de hand van de afzonderlijke regressiecoëfficiënten. In de legenda van [tabel 1] wordt daarvoor de formule gegeven. Bijvoorbeeld: de kans op diepe veneuze trombose bij een man wiens kuitomvang meer dan 3 cm is toegenomen en die een afwijkende D-dimeerwaarde heeft, is $1 / (1 + \exp(-5,47 + 0,59 + 1,13 + 3,01)) = 32\%$. Zonder afwijkende D-dimeerwaarde zou die kans $1 / (1 + \exp(-5,47 + 0,59 + 1,13)) = 2\%$ zijn. In het eerste geval is de puntenscore op de beslisleiding 1 + 2 + 6 = 9 en is diepe veneuze trombose niet uit te sluiten; de huisarts moet de patiënt verwijzen. In het tweede geval is de puntenscore 1 + 2 = 3, diepe veneuze trombose is dan onwaarschijnlijk en de patiënt hoeft niet verwezen te worden.

In de toekomst wordt het misschien mogelijk dit soort goed gevalideerde beslismodellen te integreren in huisartsinformatiesystemen. De huisarts zou dan de kans op (in dit geval) diepe veneuze trombose preciezer kunnen berekenen en zelf kunnen bepalen welk beleid daarbij het beste past.

Daarbij past de kanttekening dat iedere test, dus ook elke vraag bij de anamnese en het lichamelijk onderzoek, informatie toevoegt. Dat wil niet zeggen dat elke test altijd gedaan wordt, er is een soort natuurlijke hiërarchie. Huisartsen beginnen altijd met anamnesevragen, waarna het lichamelijk onderzoek de bevindingen uit de anamnese bijstuurt. Zij gaan vervolgens alleen over tot aanvullend onderzoek als dat zinvolle informatie zou kunnen opleveren. Een goede beslisleiding, en dus ook goed predictieonderzoek, kijkt allereerst naar de

Abstract

Geersing GJ. The GP and his crystal ball: prediction in general practice. *Huisarts Wet* 2014;57(6):298-301.

A major part of a general practitioner's consultation involves making estimates – what is the likelihood of a serious disorder or disease? Diagnostic prediction rules provide GPs with the opportunity to estimate this likelihood as accurately as possible. However, not all prediction rules are suitable for use in general practice, and it is important to pay attention to the variables on which the rule is based, such as the population studied (heterogeneous or homogeneous?), the number of diagnostic variables considered (a single test result or a decision algorithm?), and the reference test (is it carried out in all study participants?).

Tabel 2 Checklist om te beoordelen of diagnostisch onderzoek bruikbaar is voor de huisarts²³

Vraag	Niet of minder valide	Valide en bruikbaar
1 Populatie: wie selecteerde de patiënten en welke patiënten zitten in het onderzoek?	<ul style="list-style-type: none"> ■ patiënten niet geselecteerd op basis van verdenking maar op basis van referentietest ■ onverwacht hoge prevalentie; risicoprofiel past niet bij patiënten zoals de huisarts die ziet 	<ul style="list-style-type: none"> ■ patiënten geselecteerd op basis van verdenking en tijdens het spreekuur van de huisarts ■ type patiënten goed gelijkend op wat op het spreekuur wordt gezien
2 Diagnostische variabelen: is gekeken naar de toegevoegde waarde van verschillende tests?	<ul style="list-style-type: none"> ■ alleen de resultaten van één diagnostische test worden gegeven, bijvoorbeeld één specifieke bloedmarker ■ anamnese en lichamelijk onderzoek lijken overgeslagen 	<ul style="list-style-type: none"> ■ de onafhankelijke waarde van verschillende tests wordt gecombineerd, bijvoorbeeld in een beslisregel ■ natuurlijke hiërarchie van tests: eerst anamnese, dan lichamelijk onderzoek, daarna pas aanvullend onderzoek
3 Referentietest: welke test is gebruikt om de ziektestatus te definiëren?	<ul style="list-style-type: none"> ■ referentietest is niet bij alle patiënten uitgevoerd; deze zijn geëxcludeerd uit het onderzoek 	<ul style="list-style-type: none"> ■ bij alle van de aandoening verdachte patiënten is een referentietest uitgevoerd, bij voorkeur dezelfde

waarde van anamnese en lichamelijk onderzoek. Predictieonderzoek dat alleen kijkt naar een bepaalde beeldvormende test of bloedbepaling is voor de huisarts bijna per definitie niet bruikbaar.^{7,8}

BRUIKBAARHEID IN DE HUISARTSENPRAKTIJK

Het werk van de huisarts onderscheidt zich op enkele cruciale aspecten van het werk van de ziekenhuisspecialist. Over het geheel genomen is de kans op een ernstige ziekte doorgaans laag,⁹ maar de huisarts wil ernstige aandoeningen zeker niet missen. Het diagnostisch redeneren zal er dus vaak in eerste opzet op gericht zijn zulke aandoeningen uit te sluiten. Een bijkomend probleem is dat de huisarts ernstige pathologie vaak in een vroeg stadium ziet, als die nog moeilijk te onderscheiden is van onschuldige aandoeningen: vaak ontbreken de klassieke verschijnselen en ook de stijging van bepaalde biomarkers is meestal nog beperkt. Voor de huisarts zal de toegevoegde waarde (en dus ook het optimale afkappunt) van bijvoorbeeld een BNP-bepaling bij een verdenking op hartfalen anders zijn dan voor de cardioloog, want patiënten met hartfalen die de huisarts bezoeken, zitten in een relatief vroege fase van hun symptomatologie.¹⁰ Dit verschil in selectie ten opzichte van de patiëntenpopulatie in het ziekenhuis heeft invloed op de waarde van diagnostische tests in de huisartsenpraktijk. Dit is een sterk argument om predictieonderzoek dat voor huisartsen bruikbaar wil zijn, ook uit te voeren in de huisartsensetting. Predictieonderzoek dat plaatsvindt in het ziekenhuis is vaak niet goed bruikbaar voor de huisarts: de symptomen zijn anders en vaak ook duidelijker, terwijl de huisarts juist behoefte heeft aan diagnostische markers en tests die in een vroeg stadium van de aandoening houvast bieden.

Een ander kenmerkend aspect van de huisartsenpraktijk is dat de patiënten per definitie zeer heterogeen zijn. Heterogeen wat betreft leeftijd en ook wat betreft risicoprofiel en comorbiditeit.¹¹⁻¹³ De patiëntenpopulatie die een ziekenhuisspecialist (een longarts bijvoorbeeld) ziet, is veel homogener, want voorgeselecteerd – vaak door de huisarts – op basis van risicoprofiel en comorbiditeit (bijvoorbeeld ernstiger COPD). De heterogeniteit van de patiëntenpopulatie in de huisartsenpraktijk heeft belangrijke consequenties voor het diagnostisch

redeneren. Ten eerste beïnvloedt zij, door subgroepeffecten, de waarde van diagnostische tests. Subgroepeffecten zijn vooral bekend uit therapeutisch onderzoek, een bekend voorbeeld is dat antibiotica bij otitis doorgaans alleen zinvol zijn bij kinderen onder de 2 jaar met dubbelzijdige otitis of die al bij eerste presentatie een loopoor hebben.¹⁴ Subgroepeffecten zijn echter ook bekend binnen diagnostisch onderzoek; een recent voorbeeld is dat de waarde van een klinische beslisregel voor diepe veneuze trombose beperkter is bij patiënten die al eens een trombose hebben doorgemaakt.¹⁵ Een tweede consequentie van de heterogeniteit van de patiëntenpopulatie is dat de huisarts zelden slechts één enkele diagnose zal overwegen. Het voorbeeld van mevrouw Willemse illustreert dat. Op het moment dat een patiënt naar de huisarts gaat vanwege benauwdheid, is het aantal mogelijke aandoeningen nog zeer uitgebreid en bovendien komen deze aandoeningen vaak gezamenlijk voor.¹⁶⁻¹⁸ Waar de longarts vooral patiënten ziet met een verhoogd risico op longziekten en de cardioloog vooral patiënten met een verhoogd risico op hartziekten, moet de huisarts in eerste instantie met beide rekening houden. Huisartsen hebben dus vooral behoefte aan diagnostische modellen die de waarschijnlijkheid van verschillende aandoeningen tegelijk weergeven.¹⁹ Onderzoek naar dergelijke heterogeniteitseffecten binnen het diagnostisch redeneren is nog relatief schaars, maar moderne analysemethoden waarmee de patiëntgegevens van verschillende onderzoeken kunnen worden gecombineerd op het niveau van individuele patiënten, zullen naar verwachting meer van dit soort onderzoek mogelijk maken.²⁰⁻²²

CHECKLIST

Predictieonderzoek naar diagnostische modellen en tests gaat over de kern van de huisartsgeneeskunde: het spreekuur. Niet elk predictieonderzoek is echter zonder meer geschikt voor gebruik in de huisartsensetting. Dit type onderzoek is de afgelopen jaren steeds meer in de belangstelling gekomen, en het aantal goede onderzoeken dat ook bruikbaar is op het spreekuur van de huisarts is sterk toegenomen. Toch worden tijdens nascholingen of in wetenschappelijke artikelen voor huisartsen nog altijd minder goede voorbeelden gepresenteerd. Door een paar simpele vragen te stellen, kan de huisarts

snel bepalen of het betreffende onderzoek bruikbaar is voor zijn spreekuur.²³ Deze heb ik samengevat in [tabel 2]. Het is mijn stellige overtuiging dat goed diagnostisch predictieonderzoek niet alleen bijdraagt aan de zorg voor onze patiënten, maar ook aan het werkplezier van huisartsen. Het is spannend en uitdagend onderzoek, waarvan de klinische importantie vaak snel zijn weg vindt naar de dagelijkse praktijk. ■

LITERATUUR

- Grobee DE, Hoes A. *Clinical epidemiology: Principles, methods and applications for clinical research*. London: Jones & Bartlett, 2009.
- Hendriksen JM, Geersing CJ, Moons KG, De Groot JA. Diagnostic and prognostic prediction models. *J Thromb Haemost* 2013;11 Suppl 1:129-41.
- Geersing CJ, Erkens PM, Lucassen WA, Büller HR, Cate HT, Hoes AW, et al. Safe exclusion of pulmonary embolism using the Wells rule and qualitative D-dimer testing in primary care: prospective cohort study. *BMJ* 2012;345:e6564.
- Naaktgeboren CA, Bertens LC, van Smeden M, de Groot JA, Moons KG, Reitsma JB. Value of composite reference standards in diagnostic research. *BMJ* 2013;347:f5605.
- Lijmer JC, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;282:1061-6.
- Oudega R, Moons KG, Hoes AW. Ruling out deep venous thrombosis in primary care: A simple diagnostic algorithm including D-dimer testing. *Thromb Haemost* 2005;94:200-5.
- Moons KG, Biesheuvel CJ, Grobbee DE. Test research versus diagnostic research. *Clin Chem* 2004;50:473-6.
- Moons KG, Grobbee DE. Diagnostic studies as multivariable, prediction research. *J Epidemiol Community Health* 2002;56:337-8.
- Knottnerus JA. Between iatrotropic stimulus and interiatric referral: the domain of primary care research. *J Clin Epidemiol*, 2002;55:1201-6.
- Kelder JC, Cowie MR, McDonagh TA, Hardman SM, Grobbee DE, Cost B, et al. Quantifying the added value of BNP in suspected heart failure in general practice: an individual patient data meta-analysis. *Heart* 2011;97:959-63.
- Knottnerus JA, Tugwell P. Heterogeneity and clinical reality. *J Clin Epidemiol* 2013;66:809-11.
- Starfield B, Lemke KW, Herbert R, Pavlovich WD, Anderson G. Comorbidity and the use of primary care and specialist care in the elderly. *Ann Fam Med* 2005;3:215-22.
- Van den Akker M, Buntinx F, Metsemakers JF, Roos S, Knottnerus JA. Multimorbidity in general practice: prevalence, incidence, and determinants of co-occurring chronic and recurrent diseases. *J Clin Epidemiol* 1998;51:367-75.
- Rovers MM, Glasziou P, Appelman CL, Burke P, McCormick DP, Damoiseaux RA, et al. Antibiotics for acute otitis media: a meta-analysis with individual patient data. *Lancet* 2006;368:1429-35.
- Geersing CJ, Zuithoff NP, Kearon C, Anderson DR, Ten Cate-Hoek AJ, Elf JL, et al. Exclusion of deep vein thrombosis using the Wells rule in clinically important subgroups: individual patient data meta-analysis. *BMJ* 2014;348:g1340.
- Landahl S, Steen B, Svanborg A. Dyspnea in 70-year-old people. *Acta Med Scand* 1980;207:225-30.
- Rutten FH, Cramer MJ, Lammers JW, Grobbee DE, Hoes AW. Heart failure and chronic obstructive pulmonary disease: An ignored combination? *Eur J Heart Fail* 2006;8:706-11.
- Rutten FH, Moons KG, Cramer MJ, Grobbee DE, Zuithoff NP, Lammers JW, et al. Recognising heart failure in elderly patients with stable chronic obstructive pulmonary disease in primary care: cross sectional diagnostic study. *BMJ* 2005;331:1379.
- Biesheuvel CJ, Vergouwe Y, Steyerberg EW, Grobbee DE, Moons KG. Polytomous logistic regression analysis could be applied more often in diagnostic research. *J Clin Epidemiol* 2008;61:125-34.
- Debray TP, Koffijberg H, Vergouwe Y, Moons KG, Steyerberg EW. Aggregating published prediction models with individual participant data: a comparison of different approaches. *Stat Med* 2012;31:2697-712.
- Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010;340:c221.
- Tugwell P, Knottnerus JA. Advantages of individual patient data analysis in systematic reviews. *J Clin Epidemiol* 2010;63:233-4.
- Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529-36.

