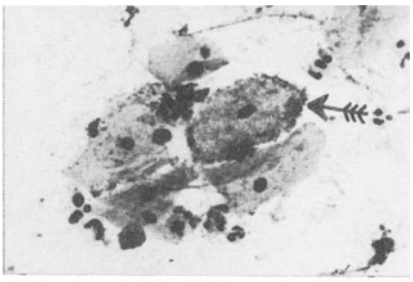
Interdokter-betrouwbaarheid bij de diagnostiek van vaginitis

JOAN BOEKE EN ARIE FAAS*

Twee huisartsen beoordeelden onafhankelijk van elkaar: 1) een aantal microscopische fluorpreparaten op de aanwezigheid van *clue cells*, Candida en Trichomonas; 2) de geur van fluorpreparaten na toevoeging van KOH 10%; 3) een aantal dia's van vaginaslijmvlies op de aanwezigheid van Candida-vaginitis. Ondanks voorafgaande scholing bleek de interdokter-betrouwbaarheid tijdens een eerste onderzoekssessie volstrekt onvoldoende. Een tweede sessie, na hernieuwde scholing en aanscherping van de diagnostische criteria, leverde echter een aanzienlijk grotere overeenstemming tussen de beoordelingen op.



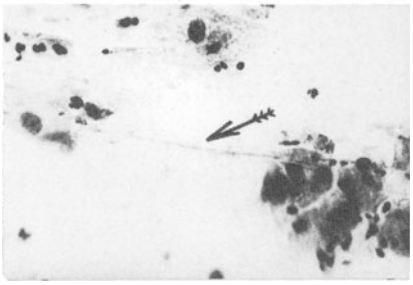


Foto boven. Een *clue cell.*
Foto onder. Een Candida-hyfe.

Inleiding

De huisarts wordt vaak geconfronteerd met „vaginitisklachten": klachten over jeuk en irritatie in en rond de schede, en als hinderlijk ervaren afscheiding. *Lamberts e.a.* komen alleen al voor de klacht vaginale afscheiding tot vier per 1000 contactredenen.[1]
Uit onderzoek blijkt dat bij deze klachten drie organismen de boventoon te voeren. In volgorde van voorkomen zijn dat:
- de aan de *niet-specifieke vaginitis* geassocieerde Gardnerella vaginalis (vroeger ook wel Haemophilus vaginalis en Corynebacterium vaginale genoemd);
- Candida;
- Trichomonas vaginalis.[2 3]

De huisarts maakt bij de diagnostiek van vaginitisklachten over het algemeen gebruik van anamnese, onderzoek van de vagina (inspectie), microscopisch onderzoek van een nat preparaat (fysiologische zoutoplossing) en een KOH-preparaat, en tegenwoordig ook van de beoordeling van de geur van de fluor. Langs deze weg zouden de drie meest voorkomende micro-organisme gemakkelijk zijn op te sporen, aan de hand van de volgende criteria:
- De Gardnerella openbaart zich in het natte preparaat onder de microscoop door de gespikkelde epitheelcellen (*clue cells*); verder is door toevoeging van KOH 10% aan de vers afgenomen fluor een rotte vis-lucht waarneembaar in geval van een Gardnerella-infectie;
- Een korrelige witte fluor (gestremde melk) op een rood geïrriteerd slijmvlies heet kenmerkend voor Candida, terwijl het KOH 10% preparaat onder de microscoop op de Candida hyfen aan het licht brengt;
- De Trichomonas vaginalis is in het natte preparaat onder de microscoop gemakkelijk te herkennen door zijn beweeglijkheid.[4]

Uit de resultaten van een onderzoek in een Amerikaanse eerstelijns gezondheidscentrum blijkt echter dat de huisarts met deze methode slechts bij een derde van de vrouwen met vaginitisklachten een diagnose kon stellen. Maakte de huisarts gebruik van de duurdere onderzoekmethoden, zoals kweken, dan vond hij in tweederde van de gevallen een diagnose.[2] Kennelijk is de huisartsendiagnostiek niet bijzonder valide en brengt zij niet goed aan het licht, wat er werkelijk aan de hand is. Bovendien wordt, zelfs met gebruikmaking van de beste diagnostiek, in een derde van de gevallen geen somatische oorzaak gevonden van de klachten.

Het lijkt ons tegen deze achtergrond belangrijk dat er onderzoek wordt gedaan naar het vóórkomen van vaginitis en de belangrijkste oorzaken daarvan, en naar de validiteit van de diagnostiek van de huisarts. Omdat de validiteit van een onderzoekmethode afhankelijk is van de betrouwbaarheid van die methode, hebben wij besloten eerst de *interdokter-betrouwbaarheid* te bepalen. In dit artikel doen wij verslag van een onderzoek naar de betrouwbaarheid van de diagnostiek bij vaginitisklachten, met de volgende vraagstelling:
- Komen twee onderzoekers onafhankelijk van elkaar tot dezelfde conclusie bij het beoordelen van een fluorpreparaat en inspectie van de vagina?
- Is het mogelijk – in geval van onvoldoende overeenstemming – de mate van overeenstemming op te voeren door scholing en aanscherpen van de diagnostische criteria?

Methoden

De onderzoekers (A en B) zijn begonnen met zichzelf te scholen door:
1) samen foto's te bekijken van fluorpreparaten;
2) samen foto's te beoordelen van de vaginawand op de aanwezigheid van Candida-vaginitis.
Daarbij werd afgesproken dat het vinden van één *clue cell*, één Candida-hyfen of één Trichomonas-flagellaat in het microscopisch preparaat een positieve bevinding betekende.

Beoordeling fluorpreparaten

Het eerste deel van het onderzoek vond plaats in een geslachtsziektenpoliklinniek in Amsterdam, tijdens een speciaal spreekuur voor aan heroïne verslaafde prostituées. A en B zaten in verschillen-

* Respectievelijk huisarts in Amsterdam en huisarts in Hengelo; beide auteurs zijn lid van de Commissie Wetenschappelijk onderzoek van het NHG.

de ruimten en mochten niet met elkaar praten. De verpleegkundige nam fluor af met een spatel en maakte hiervan vier preparaten. Eerst ontvingen A en B elk één preparaat; na toevoeging van KOH 10% beoordeelden zij de geur. Vervolgens beoordeelden zij na elkaar de overige twee preparaten; één met KOH 10% en één met een fysiologische zoutoplossing (steeds gedurende drie minuten).
Afgesproken werd dat bij een onvoldoende overeenstemming een tweede onderzoekssessie zou plaatsvinden. Daaraan voorafgaand zouden A en B bespreken welke preparaten zij verschillend hadden beoordeeld. Hierdoor en door middel van discussie over afbeeldingen van microscopische preparaten zouden zij tot nauwkeuriger beoordelingscriteria moeten komen.
Tijdens de tweede sessie zaten A en B in dezelfde ruimte; zij beoordeelden steeds, schriftelijk en onafhankelijk van elkaar, de preparaten van één patiënt, en bespraken vervolgens dat oordeel.

Macroscopische beoordeling
A en B bekeken de dia's tezamen en bespraken, na vastlegging van hun beider oordeel, elke dia en het hoe en waarom van hun diagnose (wel of niet Candida). Ook nu was afgesproken dat bij onvoldoende overeenstemming een tweede sessie zou plaatsvinden. Daaraan voorafgaand zouden beide onderzoekers nogmaals de gehanteerde criteria bespreken en bovendien werd de door de gynaecoloog bij elke dia geleverde diagnose gebruikt om de criteria aan te scherpen.

Betrouwbaarheid
Als maat voor de betrouwbaarheid is de π van Scott gebruikt (*kader*). Deze maat geeft een niet vertekend beeld van de werkelijke betrouwbaarheid bij waarnemingen door twee onderzoekers, ook in situaties waarin door scheve verdelingen van de randtotalen gemakkelijk een zeer hoog percentage overeenstemming kan worden bereikt.
Hoewel wij uiteraard streefden naar een maximale betrouwbaarheid ($\pi=1$) spreken wij van voldoende overeenstemming bij een $\pi \geqslant .80$. Afgesproken werd dat bij een $\pi<80$ een tweede onderzoekssessie zou plaatsvinden.

Resultaten

Beoordeling fluorpreparaten
Uit *tabel 1* blijkt dat de beoordeling van de microscopische preparaten en van de geur van de fluor onvoldoende betrouwbaar was. Daarom werd besloten tot een tweede sessie, waarbij de volgende nadere afspraken werden gemaakt:
- om van een rotte vis-geur te kunnen spreken moest de geur direct na de toediening van de KOH 10% worden waargenomen;
- ten aanzien van *clue cells* zou alleen op het al dan niet aanwezig zijn van bacteriën aan de buitenkant van de cel worden gelet.

Het nader bespreken van de diagnostische criteria wierp vruchten af, want in de tweede sessie werd voldoende interdoktor-betrouwbaarheid bereikt voor de rotte vis-geur, de *clue cells* en Trichomonas (*tabel 2*).
In geen van beide sessies werd Candida vastgesteld, zodat een π niet kon worden berekend. Deze negatieve bevindingen zijn wellicht terug te voeren op de geringe incidentie van Candida vaginitis bij deze populatie. Daarom werd nog een derde sessie toegevoegd volgens het protocol van de eerste sessie. Hierbij was de betrouwbaarheid voldoende (*tabel 2*).

Macroscopische beoordeling
De beoordeling van dia's met afbeeldingen van de vaginawand op het bestaan van Candida vaginitis leverde eveneens een te lage betrouwbaarheid op. Voor

De π van Scott

		Huisarts A		
		+	−	Totaal
Huisarts B	+	$P_{1.1}$	$P_{1.2}$	P_1
	−	$P_{2.1}$	$P_{2.2}$	P_2
	Totaal	$P_{.1}$	$P_{.2}$	1

$$P = \frac{\text{het gevonden aantal}}{\text{totaal}}.$$

Formule π van Scott:

$$\frac{4\,(P_{1.1}\,P_{2.2} - P_{1.2}\,P_{2.1}) - (P_{1.2} - P_{2.1})^2}{(P_{1.} + P_{.1})\,(P_{2.} + P_{.2})}$$

Tabel 1 – Table 1. *De beoordeling van de preparaten tijdens de eerste sessie (n=13) – Assessment of the specimens during the first session (n=13).*

	Similarities		Differences		π
	A+/B+	A−/B−	A+/B−	A−/B+	
Odour of decaying fish	8	2	3	–	.41
Clue cells	7	2	4	–	.38
Trichomonas	1	11	1	–	.65
Candida	–	13	–	–	

Tabel 2 – Table 2. *De beoordeling van de preparaten tijdens de tweede sessie (n=21) c.q. de derde sessie (alleen Candida, n=18) – Assessment of the specimens during the second session (n=21) and during the third session (only Candida, n=18).*

	Similarities		Differences		π
	A+/B+	A−/B−	A+/B−	A−/B+	
Odour of decaying fish	10	10	1	–	.92
Clue cells	12	8	1	–	.92
Trichomonas	1	20	–	–	1
Candida, first session	–	21	–	–	
Candida, second session	5	12	1	–	.88

Tabel 3 – Table 3. *De beoordeling van foto's van de vaginawand op de aanwezigheid van Candida vaginitis – Asse ment of the slides of the vaginal wall for the presence of Candida vaginitis.*

	Similarities		Differences		π
	A+/B+	A−/B−	A+/B−	A−/B+	
First session (n=13)	6	4	2	1	.54
Second session (n=28)	12	11	2	3	.66

de tweede sessie werd overeengekomen dat alleen „gestremde melk"-fluor als criterium zou gelden. Deze sessie leverde een iets hogere π-waarde op, maar niet het door ons nagestreefde niveau (*tabel 3*).

Beschouwing

Bij het kritisch lezen van verslagen van wetenschappelijk onderzoek in medische vakbladen blijkt, dat weinig aandacht wordt besteed aan de betrouwbaarheid van diagnostische onderzoekmethoden. Ook is bekend dat artsen nogal eens tot een verschillende uitspraak komen bij de beoordeling van een patiënt. In ons onderzoek bleek een betrouwbaarheidsonderzoek enerzijds betrekkelijk eenvoudig uitvoerbaar te zijn en anderzijds interessante gegevens op te leveren.

De intradokter-betrouwbaarheid – de mate waarin één dokter op verschillende momenten bij een en hetzelfde preparaat tot dezelfde conclusie komt – hebben wij buiten beschouwing gelaten. Het bepalen van de intradokter-betrouwbaarheid heeft in deze fase nog niet zoveel zin, omdat bij onvoldoende interdokter-betrouwbaarheid minstens één arts zijn gedrag moet veranderen, met als onvermijdelijk gevolg een lage intradokter-betrouwbaarheid.

De twee onderzoekers bleken, ondanks hun belangstelling voor de diagnostiek van vaginitisklachten en ondanks de voorafgaande scholing, in de eerste sessie niet voldoende tot een zelfde oordeel te komen. Pas na verdere scholing en aanscherping van de diagnostische criteria kwamen zij meer op één lijn. Dit betekent dat de interdokter-betrouwbaarheid van de diagnostische methoden van vaginitisklachten in de huisartspraktijk zodanig is op te voeren, dat een onderzoek naar de validiteit zin heeft. Alleen de betrouwbaarheid van de macroscopische beoordeling van de vaginawand met fluor bereikte nog niet het door ons nagestreefde niveau.

Overigens is het heel wel mogelijk de hier beschreven procedure ook toe te passen bij andere diagnostische methoden, zoals het bepalen van het Hb en het beoordelen van een urinesediment.

Synopsis

Boeke J, Faas A. Inter-doctor concurrence in the diagnosis of vaginitis. Huisarts en Wetenschap 1985; 28 : 395-7.

Introduction. This report on a study of diagnostic concurrence in the case of vaginitis complaints proceeds from the following problem definition:

- Do two investigators independently reach the same conclusion after assessing a discharge specimen and inspecting the vagina?
- If concurrence is insufficient, can it be improved by training and more precise diagnostic criteria?

Methods. The investigators (A and B) started training themselves by jointly examining photographs and discussing:

– the possible presence of *clue cells* and Trichomonas on photographs of discharge specimens;

– the possible presence of Candida vaginitis on slides of the vaginal wall.

The first part of the study was carried out at an out-patient clinic for venereal diseases in Amsterdam, during a special surgery hour for heroin-addicted prostitutes. A and B were seated in different rooms and not allowed to communicate. The nurse obtained vaginal discharge with a spatula and prepared four specimens. First, A and B each received one specimen, of which they assessed the odour after addition of 10 percent KOH. Next, they assessed in succession the remaining two specimens: one with 10 percent KOH and one with a physiological saline solution.

It was agreed that a second session was to take place in the case of insufficient agreement. Prior to this session A and B were to discuss the specimens they had assessed differently.

During the second session A and B were seated in the same room; they always assessed the specimens from one patient, working independently and writing down their findings; next they discussed their assessments.

A and B jointly studied the slides, wrote down their findings independently and then discussed each slide and the how and why of their diagnosis (Candida or no Candida). Again it was agreed that a second session was to be held in the case of insufficient concurrence. Prior to that session the investigators were once more to discuss the criteria applied; moreover, the diagnosis provided by the gynaecologist with each slide was used to define the criteria more precisely.

Scott's π was used as a measure of concurrence. Although the aim was of course maximal concurrence ($\pi=1$), concurrence was regarded as sufficient if $\pi \geqslant 0.80$. It was agreed that a second session was to be held if $\pi < 0.80$.

Results. *Table 1* shows that there was insufficient concurrence in the assessment of the microscopic specimens and of the odour of the discharge. A second session was therefore held. Further discussion of the diagnostic criteria was apparently productive, for in the second session sufficient interdoctor concurrence was achieved for the odour of decaying fish, the *clue cells* and Trichomonas (*Table 2*).

In neither session was Candida identified, and consequently π could not be calculated. A third session was therefore held using the protocol of the first session. Concurrence then proved to be sufficient (*table 2*).

The assessment of slides of the vaginal wall in order to diagnose Candida vaginitis also showed insufficient concurrence. The second session yielded a slightly higher π-value, but not up to the level aimed at (*table 3*).

Keywords: Diagnosis; Family practice; Vaginitis.

Address of authors: Nederlands Huisartsen Genootschap, Postbus 14006, 3508 SB Utrecht, The Netherlands.

Dankbetuiging

Voor de totstandkoming van dit artikel was de hulp van de volgende personen onontbeerlijk en daarvoor danken wij hen hartelijk: mevrouw J. A. R. van den Hoek, arts, polikliniek geslachtziekten GG en GD te Amsterdam, verpleegkundigen van deze polikliniek, Dr. K. J. H. Verschoof, gynaecoloog te Enschede, die de foto's leverde, Annemieke Knippenborg en Ineke Ruhe, doktersassistentes, Annie de Reuver, secretaresse NHG-bureau, die het manuscript typte, de leden van de Commissie Wetenschappelijk Onderzoek, en in het bijzonder Jan Gubbels, die statistische adviezen gaf.

[1] Boeke J, Faas A. Vaginitis in de huisartspraktijk. Een literatuuronderzoek. Huisarts en Wetenschap 1985; 28: 392-4, 399.

[2] Berg AO, Heindrich FE, Fihn SD, et al. Establishing the cause of genitourinary symptoms in women in a family practice. JAMA 1984; 251: 620-5.

[3] Komaroff AL, Pass TM, Mc Cue JD, Cohen AB, Hendricks TM, Friedland G. Management strategies for urinary and vaginal infections. Arch Intern Med 1978; 138: 1069-73.

[4] Van der Meyden WI. Niet-specifieke vaginitis. Huisarts en Wetenschap 1983; 26: 54-7.

[5] Landis JR, Koch GC. A review of statistical methods in the analysis of data arising from observer reliability studies. Statistica Neerlandica 1975; 3: 101-23, 151-61.