

# Testen en hertesten

## Over variaties in diagnostische testuitslagen

Iedere huisarts weet dat de bloeddruk van een patiënt in de ochtenduren wat hoger is dan later op de dag en dat het bij veel patiënten verschil maakt of deze thuis wordt gemeten of in de praktijk. Daarnaast is van invloed wie de meting verricht: de dokter of de praktijkassistente.

Bijna iedere (instrumentele) meting staat bloot aan drie bronnen van variatie:

- *biologische variaties*: intra-individuele verschillen, dagritmes en invloed van stress;
- *analytische variaties*: de voorbereiding van de patiënt, de betrouwbaarheid van het meetinstrument, de nauwkeurigheid waarmee de meting wordt uitgevoerd;
- het *onderscheidend vermogen* van de test: in welke mate voorspelt een gemeten hoge bloeddruk een werkelijk bestaande hypertensie.

Over het onderscheidend vermogen ten aanzien van aanvullend diagnostisch onderzoek door de huisarts is vorig jaar uitvoerig bericht in *Huisarts en Wetenschap*.<sup>1-4</sup> Beschreven werd hoe een schatting kan worden gemaakt van het optimale referentiewaardengebied en hoe de voorspellende waarde van testuitslagen te beïnvloeden is door het aanpassen van dat gebied. Dit overzicht is beperkt tot de achtergrond van de analytische en biologische variaties in diagnostisch testuitslagen en het klinische belang daarvan, geïllustreerd aan de hand van onderzoeksverslagen en praktische voorbeelden.

### Analytische variatie

In een recent onderzoek werd de betrouwbaarheid van de BSE-bepaling in de huisartspraktijk onderzocht door een serie uit het ziekenhuislaboratorium afkomstige bloedmonsters te verspreiden onder vijf huisartspraktijken met het verzoek om in elk monster op de gebruikelijke wijze de BSE te bepalen.<sup>5</sup> De onderzoekers vonden een verontrustend grote variatie in de uitslagen; in één geval bedroeg het verschil 29 mm. Op grond hiervan werd geconcludeerd dat de variatie klinisch relevant was. Een BSE van 4 mm zal in alle gevallen als normaal worden afgegeven, terwijl een BSE van 33 mm in principe kan leiden tot verder onderzoek, behandeling of zelfs verwijzing van de patiënt.

Voor zover men kon nagaan, werd de gevonden variatie niet veroorzaakt door een verkeerde verhouding van bloed en citraat, of een te groot tijdsinterval tussen bloedafname en bepaling. De pre-analytische fase van de bepalingen was dus niet debet aan het onbetrouwbare resultaat. Dit gold niet voor de meer instrumentele, analytische fase. Het gebruik van onvoldoende gereinigde bepalingsbuizen was waarschijnlijk de boosdoener, en deze bevinding leidde tot het praktische advies om in het vervolg slechts *disposable* buizen te gebruiken.

In het BSE-onderzoek werden ook de in de praktijk gevonden waarden vergeleken met een lege artis verrichte

**Dinant GJ. Testen en hertesten. Over variaties in diagnostische testuitslagen [Commentaar]. Huisarts Wet 1993; 36(1): 4-6.**

Correspondentie: Dr. G.J. Dinant, Vakgroep Huisartsgeneeskunde, Rijksuniversiteit Limburg, Postbus 616, 6200 MD Maastricht.

controle in het laboratorium: de *referentie* voor de bepaling. Door deze vergelijking werd een *systematisch verschil* in de uitslagen opgespoord: de uitkomsten in het laboratorium waren in bijna alle gevallen hoger dan de door de praktijken gemeten waarden. Dit verschil kon niet afdoende worden verklaard.

De grootte van een systematisch verschil is een maat voor de *accuracy* of *validiteit* van een test.<sup>6</sup> Daarbij kan de referentiebepaling soms worden beschouwd als een *gouden standaard* voor de onderzochte test. Zo werden door *Hiscoke et al.* de resultaten van urine-onderzoek met teststrips vergeleken met de uitslag van urinekweken als gouden standaard.<sup>7</sup> Vaak is er echter geen goede ('gouden') referentie of is deze te ingrijpend voor de patiënt. Zo lijkt het bezwaarlijk om de betrouwbaarheid van bloeddrukmeters in de huisartspraktijk te onderzoeken met invasieve referentiebepalingen.<sup>8</sup> In die gevallen blijft het betrouwbaarheidsonderzoek noodgedwongen beperkt tot herhaalde, zoveel mogelijk onder dezelfde omstandigheden uitgevoerde metingen en wordt in feite de *reproduceerbaarheid* van de testresultaten geschat. Op die manier onderzochten *Stoffers et al.* de betrouwbaarheid van de enkelarmindex, gemeten met behulp van Doppler-apparatuur bij patiënten met een mogelijk perifere arteriële vaatobstructie.<sup>9</sup>

In de praktijk worden de begrippen reproduceerbaarheid en betrouwbaarheid vaak als synoniem gebruikt en wordt daarbij verwezen naar studies met en zonder referentiebepalingen. Bij reproduceerbaarheidsonderzoek zonder referentiebepalingen kunnen de resultaten op nette en inzichtelijke wijze worden weergegeven door het *gemiddelde* van een aantal numerieke metingen (bijvoorbeeld de BSE- of de Hb-bepaling) uit te zetten tegen de *spreiding* van de resultaten rondom dit gemiddelde.<sup>10</sup> Omdat de grootte van de spreiding bij veel bepalingen a priori afhankelijk is van de hoogte van de gemiddelde waarde, wordt vervolgens vaak de variatiecoëfficiënt bepaald; dit is het quotiënt van de standaarddeviatie en het gemiddelde van een reeks uitslagen. Zo werd in het boven aangehaalde voorbeeld van een gemiddelde BSE van 19 mm de te grote spreiding bevestigd door een grote variatiecoëfficiënt.

## Biologische variatie

Indien kan worden aangenomen dat bepalingen op betrouwbare wijze en lege artis zijn uitgevoerd, zal nog steeds een zekere spreiding in de testuitslagen worden gevonden. Deze spreiding, die per definitie berust op toevallige fouten, is een maat voor de *precisie* (nauwkeurigheid) van de test. Aan toevallige fouten kunnen onder meer intra-individuele, dus biologische variaties ten grondslag liggen.<sup>11</sup> Dat geldt voor diverse bepalingen die huisartsen regelmatig verrichten, zoals de bloeddrukmeting en de cholesterbepaling. Mede om die reden wordt in de betreffende NHG-standaarden geadviseerd een diagnostische conclusie te baseren op een reeks met bepaalde tussenpozen verrichte metingen. We spreken dan van een *longitudinale* vergelijking van testuitslagen. Sommigen gaan nog verder en adviseren om bij iedere (nieuwe) patiënt een soort bestand met basisgegevens aan te leggen, verzameld tijdens klachtenvrije periodes, onafhankelijk van de aanwezigheid van (latere) risicofactoren. Vergelijking van deze basisgegevens met latere testuitslagen zou een optimale beoordeling van de biologische variatie mogelijk maken. Alleen al het medicaliserende en kostenverhogende effect van deze interventie is in deze tijd echter onacceptabel.

## Relevant of niet?

De variatie tussen twee testuitslagen is te groot – en daarmee is de testuitslag ‘onbetrouwbaar’ – indien de ene uitkomst tot een andere diagnostische, therapeutische, of prognostische conclusie leidt dan de andere. Zo had de door *Hendriksen et al.* gevonden variatie ten aanzien van de cholesterbepaling consequenties voor het indelen van personen in de verschillende risicocategorieën.<sup>12</sup> Een patiënt zou ten onrechte wel of juist geen medicamenteus advies hebben ontvangen. Een vergelijkbaar resultaat werd door *Dols et al.* en door *Hamer-van Lange et al.* gevonden bij in de huisartspraktijk uitgevoerde studies naar de reproduceerbaarheid van de cholesterbepaling met behulp van de Reflotron.<sup>13,14</sup> Beide onderzoekers concludeerden, onafhankelijk van elkaar, dat de gevonden spreiding (klinisch) relevant was, dat derhalve de introductie van de Reflotron in de huisartspraktijk voor de cholesterbepaling voorsnog geen aanbeveling verdient, en dat in ieder geval een nauwkeurige kwaliteitscontrole nodig is.

Uit deze studies blijkt dat ook de modernere en voor de huisartspraktijk beschikbare laboratoriumapparatuur minder betrouwbaar kan zijn. In het licht van de toenemende decentralisatie van laboratoriumonderzoek is dit een belangrijk gegeven. De huisarts kan niet alleen afgaan op door de fabrikant gepresenteerde informatie, hij dient zich

ook af te vragen of nieuwe apparatuur betrouwbaarder is dan de bestaande. Zo vonden *Ros et al.* dat moderne Hb-meters niet nauwkeuriger zijn dan de aloude Sicca of de Spencer.<sup>15</sup> Overigens geldt deze conclusie ook voor bepalingen die worden uitgevoerd in professionele (ziekenhuis)laboratoria, zoals *Kahn et al.* aantoonde ten aanzien van cholesterbepalingen in serum.<sup>16</sup>

Wellicht dient de huisarts extra op zijn hoede te zijn bij ‘doe het zelf’-tests. *Dujardin et al.* vonden bij reflectometers voor diabetische zelfcontrole een matige tot slechte overeenkomst met laboratoriumuitslagen en adviseerden de thuis gemeten waarden altijd te vergelijken met in de praktijk uitgevoerde bepalingen.<sup>17</sup> Het klinische belang van een betrouwbare glucosebepaling is immers evident, zoals *Marks* nog eens benadrukte naar aanleiding van een patiënt die overleed in een diabetisch coma in aansluiting op een glucosebepaling die abusievelijk was verricht met behulp van een teststrip die bij een ander instrument hoorde.<sup>18</sup>

Dit voorbeeld is natuurlijk een uitzondering. In de dagelijkse praktijk is de vraag hoe betrouwbaar testuitslagen zijn, vaak moeilijk te beantwoorden en ook niet altijd opportuun. Het diagnostisch oordeel van de huisarts is tenslotte niet gebaseerd op één testuitslag, maar op het geheel van anamnese en onderzoeksgegevens. Daarbij is het voor de huisarts van minder belang om te weten of een patiënt in werkelijkheid bijvoorbeeld een BSE van 40 of 50 mm heeft, dan van 15 of 25 mm.<sup>3</sup> Met andere woorden: het belang van betrouwbare testuitslagen is bij uitstek aan de orde in het grensgebied tussen normaal en afwijkend. Betrouwbaarheidsonderzoek dient derhalve te geschieden in een populatie waarin voor de test geïndiceerde patiënten zijn vertegenwoordigd: in de huisartspraktijk. Daar kunnen ook de effecten van een druk spreekuur, een relatief beperkte laboratoriumervaring en het grotendeels ontbreken van een traditie van kwaliteitscontroles worden gemeten.

Dat een betrouwbaarheidsstudie de kwaliteit van laboratoriumwerk ten goede kan komen, blijkt uit het aangehaalde BSE-onderzoek, waarin een grondige instructie over het uitvoeren van de test leidde tot een sterke reductie van de aanvankelijk gevonden variatie.<sup>5</sup> Met dit onderzoek werd geïllustreerd hoe huisartsen relatief eenvoudig kwaliteitscontroles in de eigen praktijk kunnen uitvoeren en hoe daarmee het inzicht in de variatie van testuitslagen kan worden vergroot. Het is een relatief kleine moeite om af en toe een dubbelgroot bloedmonster af te nemen voor twee BSE-bepalingen in plaats van één. Op vergelijkbare wijze kan een door de huisarts in de eigen praktijk gevonden verlaagd Hb-gehalte eenvoudig worden getoetst aan de laboratoriumuitslag, indien de patiënt voor een verdere analyse van zijn anemie (bijvoorbeeld ter bepaling van het MCV) naar het laboratorium wordt verwezen. Het verdient

aanbeveling het controleren van in eigen beheer uitgevoerd laboratoriumonderzoek tot een gewoonte te maken, zoals het ijken van de bloeddrukmeter dat al is. Daarnaast moeten we met het (jaarlijks) vervangen van verlopen ampullen wellicht ook de urine- en glucoseteststrips, de BSE-bepalingsbuizen en de Hb-meter controleren. Pas dan is de praktijk 'APK-goedgekeurd'.

Geert-Jan Dinant

### Dankbetuiging

Met dank aan de collegae Phile Govaert, André Knottnerus en Raymond Leclercq voor hun kritische reactie op het manuscript.

- 1 De Maeseneer J, Muris JWM. Interpretatie van aanvullend diagnostisch onderzoek door de huisarts. *Huisarts Wet* 1992; 35: 181-2.
- 2 Van der Schouw YT, Verbeek ALM, Ruijs JHJ. De ROC-curve voor de eerste evaluatie van nieuwe diagnostische tests. *Huisarts Wet* 1992; 35: 204-8.
- 3 Dinant GJ, Knottnerus JA, Van Wersch JWJ. Het onderscheidend vermogen van de BSE-bepaling in de dagelijkse praktijk. *Huisarts Wet* 1992; 35: 197-203, 224.
- 4 Zaat JOM, Schellevis FG, Kluijdt I, et al. Laboratoriumonderzoek bij de klacht moeheid in de huisartspraktijk. *Huisarts Wet* 1992; 35: 183-7.
- 5 Dinant GJ, Van Wersch JWJ, Knottnerus JA. De kwaliteit van de BSE in de huisartspraktijk. *Huisarts Wet* 1989; 32: 412-7, 438, 546.
- 6 Fletcher RH, Fletcher SW, Wagner EH. *Clinical epidemiology: the essentials*. Baltimore: Williams & Wilkins, 1988.
- 7 Hiscoke C, Yoxall H, Greig D, Lightfoot NF. Validation of a method for the rapid diagnosis of urinary tract infection suitable for use in general practice. *Br J Gen Pract* 1990; 40: 403-5.
- 8 Van Doorn BA, Van der Does E, Lubsen J, Rijsterborgh H. Betrouwbaarheid van de bloeddrukmeting; vergelijking van een elektronische meter en kwikmanometer in de huisartspraktijk. *Ned Tijdschr Geneesk* 1990; 134: 1646-50.
- 9 Stoffers J, Kaiser V, Schouten H, Knottnerus JA. Peripheral arterial occlusive disease in general practice: the reproducibility of the ankle-arm systolic pressure ratio. *Scand J Prim Health Care* 1991; 9: 109-14.
- 10 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1991; 337: 215-8.
- 11 Anonymous. Measurement imprecision: ignore or investigate? [Editorial]. *Lancet* 1992; 339: 587-8.
- 12 Hendriksen IJM, Bezemer PD, Boerma GJM, Zuiderveld U. Schommelingen van het serumcholesterolgehalte bij één persoon: wat is de werkelijke waarde en wanneer is er sprake van een significante verandering? *Ned Tijdschr Geneesk* 1992; 136: 1507-11.
- 13 Dols JLS, Sanders GTB, Dijksterhuis PH, et al. De betrouwbaarheid van droge-stofchemie in de huisartspraktijk. *Huisarts Wet* 1991; 34: 62-6.
- 14 Hamer-van Lange MAM, Dinant GJ, Van Wersch JWJ. De reproduceerbaarheid van de cholesterolbepaling met de Reflotron. *Huisarts Wet* 1992; 35: 243-5.
- 15 Van der Ros N, Nauta JJP. De nauwkeurigheid van kleine Hb-meters. *Huisarts Wet* 1988; 31: 285-7.
- 16 Kahn RF, McCord RS. Interlaboratory variability in serum lipid measurements. *Fam Pract Res J* 1991; 11: 415-20.
- 17 Dujardin BCG, Roijers AFM, Schoorl M. Laboratoriumonderzoek van reflectometers voor diabetische zelfcontrole. *Huisarts Wet* 1987; 30: 207-9.
- 18 Marks V. Stick testing; reliable, safe, and economical – but only if the instructions are followed. *Br Med J* 1991; 302: 482-3.

■