

Vaststellen van symptoomvaliditeit in neuropsychologisch expertiseonderzoek

*Prof. dr. B.A. Schmand en dr. J.F.M. de Jonghe**

1. Inleiding

Rechtbanken, Openbaar Ministerie, advocaten, verzekeringsartsen en (andere) medisch deskundigen kunnen om een neuropsychologisch onderzoek (NPO) vragen wanneer zij een oordeel moeten vormen over letselschade, invaliditeit, toerekeningsvatbaarheid of wilsbekwaamheid van een persoon die neuropsychologische functiestoornissen heeft of zou kunnen hebben. De neuropsycholoog bestudeert dan eerst het dossier en gaat vervolgens met de persoon zelf aan de slag. Hij neemt een anamnese en zo nodig een heteroanamnese af, en legt de persoon een aantal cognitieve tests en klachtenvragenlijsten voor, en eventueel ook persoonlijkheidstests. Hierbij is het van belang dat de onderzochte persoon zijn klachten en problemen waarheidsgetrouw weergeeft, en dat hij zich in redelijke mate inspant bij het maken van de cognitieve tests. Bij veel expertiseonderzoeken spreekt dit niet vanzelf. De belangen die op het spel staan, kunnen de onderzochte persoon in de verleiding brengen zijn klachten te overdrijven en de tests minder goed te maken dan waartoe hij in staat is. We noemen dit 'aggraveren', respectievelijk 'onderpresteren'. Uiteraard is het ook mogelijk dat de klachten en problemen helemaal niet bestaan, maar dat ze worden gesimuleerd.

In de neuropsychologie worden deze kwesties aangeduid met de term 'symptoomvaliditeit': zijn de klachten en symptomen van de onderzochte persoon valide?

Vendrig heeft enkele jaren geleden in dit tijdschrift een uitgebreide beschouwing gegeven over symptoomvaliditeit bij het NPO.¹ Hij besprak niet alleen aggraveren, onderpresteren en simuleren en hoe dit kan worden vastgesteld, maar ook het voorkomen ervan, mogelijke oorzakelijke mechanismen, verbanden met verwante fenomenen, zoals somatiseren en medisch onverklaarde klachten, en de manier waarop men er in juridische kring mee omgaat. Zijn beschouwing is nog steeds actueel, en we verwijzen er dan ook graag naar.

* Prof. dr. B.A. Schmand is emeritus hoogleraar klinische neuropsychologie aan de Universiteit van Amsterdam. Dr. J.F.M. de Jonghe is klinisch neuropsycholoog in het Noordwest Ziekenhuis in Alkmaar.

1. A.A. Vendrig. Het meten van onderpresteren bij whiplashslachtoffers. Veelbelovend maar juridisch lastig te vertalen, TVP 2011, afl. 3, p. 73-81.

In deze bijdrage willen we enkele meer technische aspecten van het vaststellen van symptoomvaliditeit behandelen. Ons doel is de juridische en medische 'eindgebruikers' van neuropsychologische expertiseonderzoeken een beter inzicht te geven in de gehanteerde methoden. We hopen dat dit hen in staat stelt de conclusies die de neuropsycholoog trekt beter op hun waarde te schatten. Achtereenvolgens bespreken we de belangrijkste methoden waarmee symptoomvaliditeit wordt onderzocht, de manier waarop specifieke detectiemethoden worden geconstrueerd, de sensitiviteit en specificiteit ervan, en de manier waarop ze worden toegepast in het NPO. We zullen ook stilstaan bij de vraag of we kunnen aantonen dat de onderzochte persoon moedwillig slecht presteert. Ten slotte zullen we een aantal kwaliteitskenmerken van het NPO noemen waar de eindgebruiker op zou moeten letten.

2. Prestatievaliditeitstests en symptoomvaliditeitstests

Er zijn grofweg twee categorieën klachten en symptomen waar het NPO zich op richt:

1. cognitieve klachten of cognitieve functiestoornissen; en
2. emotionele of gedragsmatige klachten en stoornissen.

Bij beide categorieën is symptoomvaliditeit van belang. Voor de eerste categorie kan de symptoomvaliditeit worden gecontroleerd met cognitieve tests, waarvan leken denken dat ze niet goed gemaakt kunnen worden door patiënten met hersenletsel of hersenziekten, terwijl de tests in feite betrekkelijk gemakkelijk zijn voor die patiënten. Dergelijke tests noemen we tegenwoordig 'prestatievaliditeitstests' (PVT's). Een lage score op een PVT wijst op onderpresteren, dat wil zeggen presteren onder het niveau waartoe men feitelijk in staat is.

De term 'symptoomvaliditeitstest' (SVT) blijft gereserveerd voor de tweede categorie, en heeft betrekking op vragenlijsten die naar emotionele en gedragsmatige klachten vragen. In die lijsten staan onwaarschijnlijke klachten en symptomen, die leken niet als onwaarschijnlijk herkennen. Bijvoorbeeld, in een lijst die naar stemmingsklachten vraagt, kunnen beweringen staan die vrijwel altijd ontkennend beantwoord worden door depressieve patiënten, zoals 'Meestal voel ik me 's avonds somberder dan 's ochtends.' Wanneer veel van dergelijke onwaar-

schijnlijke klachten worden gemeld (hoge SVT-score), wijst dit op aggraveren, dat wil zeggen het overdrijven van klachten. Behalve met vragenlijsten kan de symtroomvaliditeit ook worden gecontroleerd door middel van een gestandaardiseerd interview. ‘Gestandaardiseerd’ wil zeggen dat altijd bepaalde vragen worden gesteld op een standaardmanier, en dat de antwoorden op een standaardmanier worden verwerkt tot kwantitatieve scores.

2.1 Hoe worden die tests geconstrueerd en gevalideerd?

Voorbeeld van een fictieve test

Aan de hand van een fictief voorbeeld zullen we uitleggen hoe dergelijke tests worden gemaakt en op hun kwaliteit worden getoetst.

De fictieve test is de Mondiale Geheugen Test (MGT). De test bestaat uit 20 toeristische ansichtkaarten van allerlei plaatsen in de wereld. De onderzochte persoon krijgt de kaarten een voor een te zien en moet proberen ze te onthouden. Na een interval van 10 minuten krijgt hij de kaarten opnieuw te zien met kriskras daartussendoor 20 nieuwe ansichtkaarten. Bij elke kaart moet hij zeggen of hij de kaart eerder heeft gezien. Dus hij moet 40 keer een tweekeuzevraag beantwoorden. Patiënten met geheugenstoornissen kunnen dit soort herkenningstaken meestal nog tamelijk goed uitvoeren, iets wat leken over het algemeen niet vermoeden.

We moeten nu allereerst toetsen of de MGT een goede PVT is. Verwarrenderwijs heet een dergelijke kwaliteitstoets in de psychometrie de ‘validering’ van de test. Is de test valide? Meet hij wat we willen dat hij meet? De eerste stap om dit te bepalen is de MGT uit te proberen bij vrijwilligers die de test zo goed mogelijk maken (controlepersonen) dan wel doen alsof ze een geheugenstoornis hebben (‘experimentele simulanten’). Als de test geen onderscheid kan maken tussen deze twee groepen, dan is hij waardeloos als PVT.

In tabel 1 (zie p. 17) is het (fictieve) resultaat van deze valideringsstap te zien. Bijna alle controles scoren maximaal, negen maken één fout, en één controlepersoon maakt twee fouten. Bij de simulanten maakt één persoon twee fouten en de rest maakt er meer, vaak zelfs veel meer. Als we nu alle scores van 37 en lager afwijkend noemen (afkappunt), dan classificeert de test alle controles en 99% van de simulanten correct. De test heeft – in technische termen – een sensitiviteit van 99% (gevoeligheid voor onderpresteren) en een specificiteit van 100% (niemand van de controles wordt ten onrechte beschuldigd van onderpresteren).

De volgende, veel belangrijker stap is de test te valideren bij een groep patiënten die werkelijk cognitieve functiestoornissen hebben. Het gaat er immers om of de test mensen die onderpresteren kan onderscheiden van patiënten met werkelijke stoornissen. Het resultaat van deze stap is te zien in de middelste kolommen van tabel 1. Patiënten met geheugenstoornissen doen de test niet zo goed als gezonde vrijwilligers,

maar over het algemeen toch beter dan de experimentele simulanten kennelijk veronderstellen. En dat is precies wat de test een goede PVT maakt.

We moeten nu het afkappunt lager kiezen en bijvoorbeeld alle scores van 34 en lager als afwijkend beschouwen. Dan worden 7 patiënten ten onrechte geclassificeerd als onderpresteerder, en 18 experimentele simulanten worden gemist als onderpresteerder. Dus de sensitiviteit is 82% en de specificiteit is 93%.

De MGT moge fictief zijn, deze cijfers zijn realistisch. Bij PVT’s (en SVT’s) streven we naar een hoge specificiteit, omdat we niet ten onrechte willen concluderen dat er sprake is van aggraveren of onderpresteren.

Een derde validatiestap die testconstructeurs kunnen zetten, is de test te toetsen in de expertisepraktijk bij groepen mensen van wie bekend is dat ze vaak onderpresteren, zoals mensen met een chronisch whiplashsyndroom die in een letselschadezaak verwickeld zijn. In dergelijke groepen is het percentage afwijkende scores op een goede PVT relatief hoog.

Mutatis mutandis worden klachtenlijsten die als SVT moeten gaan fungeren op dezelfde wijze gevalideerd. Relevante categorieën patiënten die in het validatieonderzoek van SVT’s betrokken dienen te worden, zijn dan niet alleen hersenbeschadigde patiënten, maar ook psychiatrische patiënten met stemmings- en/of gedragsstoornissen.

2.2 Sensitiviteit en specificiteit zijn geen vaststaande kenmerken

Terloops hebben we de begrippen sensitiviteit en specificiteit ingevoerd. Dit zijn belangrijke kenmerken van elk diagnostisch instrument. Er is enige analogie met de strafrechtspleging. Politie, openbaar aanklager en rechters zijn in zekere zin diagnostici. Ze moeten beoordelen of een persoon al dan niet schuldig is aan een misdaad. De politie en de openbaar aanklager geven daarbij voorrang aan de sensitiviteit: ze willen de schuldige vooral voor de rechter krijgen. De rechter geeft daarentegen voorrang aan de specificiteit: hij veroordeelt alleen als de schuld onomstotelijk bewezen is, en bij twijfel volgt vrij spraak. De neuropsycholoog moet echter beide aspecten mee laten wegen. Deze afweging bepaalt waar hij het afkappunt legt. In tabel 1 is een algemeen principe te zien, namelijk dat de sensitiviteit afneemt als het afkappunt lager wordt gelegd, en dat de specificiteit dan toeneemt, en vice versa.

Wat niet in de tabel is te zien, is dat sensitiviteit en specificiteit geen harde kenmerken zijn van een instrument, maar dat ze afhangen van kenmerken van de groepen waarbij het validatieonderzoek is gedaan. Stel, bijvoorbeeld, dat we de experimentele simulanten van tevoren coachen, dat wil zeggen grondig voorlichten over geheugenstoornissen en over PVT’s, dan zal de verdeling van hun scores naar boven verschuiven. Met andere woorden, ze zullen gemiddeld genomen minder fouten maken en hun prestatie zal meer lijken op die van de patiën-

Tabel 1 Verdeling van scores op de MGT in drie groepen van elk 100 personen, te weten gezonde vrijwilligers (controles), patiënten met geheugenstoornissen en vrijwilligers die de instructie hadden gekregen de test te maken alsof ze een geheugenstoornis hadden (simulanten). Cum % = cumulatief percentage.

MGT-score	Controles	Cum %	Patiënten	Cum %	Simulanten	Cum %
40	90	90	1	1		0
39	9	99	15	16		0
38	1	100	28	44	1	1
37			31	75	3	4
36			10	85	3	7
35			7	92	11	18
34			5	97	21	39
33			2	99	12	51
32			1	100	23	74
31					5	79
30					4	83
29					5	88
28					1	89
27					2	91
Lager					9	100

ten. Daarmee neemt de sensitiviteit van de MGT voor onderpresteren af. En stel dat we een andere groep patiënten onderzoeken met veel ernstigere geheugenstoornissen dan deze groep, dan zal de verdeling van hun scores lager liggen, omdat ze meer fouten maken. Daardoor valt de specificiteit van de test lager uit. Kortom, bij het beoordelen van de diagnostische kwaliteit van een instrument moet men altijd in ogenschouw nemen op welke steekproeven het instrument is gevalideerd. Sommige validiteitstests zijn niet geschikt om af te nemen bij mensen bij wie ernstige stoornissen vermoed worden. Dergelijke tests zijn gevalideerd bij patiënten met betrekkelijk lichte stoornissen, terwijl patiënten met ernstige stoornissen die tests niet goed meer kunnen maken. Andere validiteitstests zijn 'robuuster'. Ze worden door vrij ernstig gestoorde patiënten nog goed gemaakt, maar ze zijn meestal wat minder gevoelig voor onderpresteren. Met andere woorden, de specificiteit is goed, maar de sensitiviteit is minder. Dit soort specificaties van SVT's en PVT's maken deel uit van de vakkennis van de neuropsycholoog. De beste SVT's en PVT's zijn gevalideerd bij bonafide patiënten met vrij ernstige stoornissen en bij experimentele simulanten die goed op de hoogte zijn van de betreffende stoornissen en van SVT's en PVT's.

3. Bij positieve validiteitstests is er geen conclusie meer te trekken over stoornissen

Wanneer de resultaten van SVT's en/of PVT's positief zijn, dat wil zeggen de onderzochte persoon heeft afwijkende scores behaald, terwijl deze afwijkingen niet aan de ernst van de cognitieve of emotionele problematiek te wijten kunnen zijn, dan kunnen de resultaten van de overige tests niet meer als valide worden beschouwd. Dus dan wijzen afwijkende scores op de overige tests niet noodzakelijk op stoornissen. Stoornissen kunnen in dat geval echter evenmin worden uitgesloten. Er valt niets meer te concluderen, behalve dat de persoon in kwestie onderpresteerde of zijn klachten overdreef, en dat de testresultaten dus niet valide zijn.

Soms lukt het de psycholoog te achterhalen wat de reden van het onderpresteren of overdrijven was. In de expertisecontext liggen de redenen uiteraard voor de hand. Maar het fenomeen doet zich ook vrij vaak voor in klinische setting.² Dan blijft de reden of het achterliggende mechanisme meestal een raadsel en is evenmin duidelijk of de onderzochte persoon de intentie had afwijkende scores te behalen.

2. B. Dandachi-Fitzgerald, B. van Twillert, P. van de Sande, Y. van Os & R.W. Ponds, Poor symptom and performance validity in regularly referred hospital outpatients: Link with standard clinical measures, and role of incentives, *Psychiatry Research* (239) 2016, p. 47-53.

4. Presteren onder kansniveau

Er is één situatie waarin de (bewuste) intentie wel duidelijk is, en dat is het geval waarin de behaalde score op een PVT onder kansniveau ligt. We nemen de MGT weer als voorbeeld. De taak bij de MGT bestaat uit 40 keer beslissen of een ansichtkaart al eerder is gezien. Stel dat die taak wordt gedaan door iemand die de kaarten nog helemaal niet gezien heeft. Of door een patiënt die zo'n ernstige geheugenstoornis heeft, dat hij geen enkele kaart kan herkennen. Dan is elk van de 40 beslissingen een gok. We verwachten onder deze omstandigheden a priori dat de score (ongeveer) 20 zal zijn. Dat is het kansniveau. Als de persoon in kwestie meer dan 20 fouten maakt, dan laadt hij de verdenking op zich dat hij wel degelijk weet wat de goede antwoorden zijn, en dat hij dus expres fouten maakt. Maar wanneer is deze verdenking gerechtvaardigd? Hoeveel meer dan 20 fouten moet hij maken, voordat we hem kunnen betichten van boze opzet?

Om die vraag te beantwoorden roepen we de kansrekening te hulp, in casu de binomiaaltoets. Deze wijst uit dat de score 14 (of minder) correcte antwoorden moet zijn om met 95% zekerheid te kunnen zeggen dat de persoon bewust fouten maakt. Het tolereren van 5% kans op het trekken van de onjuiste conclusie is heel gebruikelijk bij dit soort hypothesetoetsing. Naarmate de score lager is, neemt de kans verder af dat we de onjuiste conclusie trekken. Bijvoorbeeld, bij een score van 12 correcte antwoorden is die kans kleiner dan 1%.

Let wel, dit alles gaat uit van de aanname dat de onderzochte persoon geen enkele kennis heeft van de juiste kaarten, en dus puur aan het gokken is. In de praktijk is die aanname te sterk en dus onjuist, want hij heeft de 20 ansichtkaarten wel degelijk bekeken. Ook de aanname dat zijn geheugen wellicht dermate gestoord is dat hij geen enkele kaart kan herkennen, is in de praktijk meestal te sterk en dus onjuist. Zelfs alzheimerpatiënten scoren boven kansniveau op dit soort tests, zolang ze niet in het eindstadium van de ziekte zijn. Kortom, dergelijke lage scores zijn extreem onwaarschijnlijk. Betrokkene antwoordt dus aantoonbaar slechter dan iemand die puur gokt, en slechter dan patiënten met zeer ernstige cognitieve stoornissen. Daaruit kan worden afgeleid dat hij weet heeft van de foute en de juiste antwoorden, en dat er sprake moet zijn van de intentie om foute antwoorden te geven. Bovendien heeft hij waarschijnlijk laten zien dat zijn geheugen veel minder ernstig gestoord is dan een score op kansniveau suggereert. Hij is immers naar de afspraak voor het NPO gekomen, heeft zijn klachten uiteen kunnen zetten, en dergelijke. De kans dat een redelijk zelfstandig functionerende persoon per toeval zo laag scoort, is nihil. Dus de score onder kansniveau is inconsistent met de rest van zijn gedrag in de zitting. Op basis van soortgelijke argumenten menen sommige neuropsychologen dat al eerder, dus met meer dan 5% kans op een onjuiste conclusie, kan worden gesproken van een score onder kansniveau.³

3. L.M. Binder, G.J. Larrabee & S.R. Millis, Intent to fail: Significance testing of forced choice test results, *The Clinical Neuropsychologist* (28) 2014, afl. 8, p. 1366-1375.

Presteren onder kansniveau impliceert dus een (bewuste) intentie, maar dit betekent nog niet dat de persoon zich altijd bewust is van zijn motieven daartoe. Betrokkene heeft weet van zijn foute antwoorden, maar het is denkbaar dat dit voortkomt uit achterliggende psychopathologie, zoals een nagebootste stoornis. Hoe het ook zij, de reguliere testuitslagen kunnen in ieder geval niet meer betrouwbaar worden geïnterpreteerd.

5. Combineren van validiteitstests

Tot nu toe hebben we steeds gesproken over opzichzelfstaande SVT's of PVT's. In een deugdelijk NPO worden echter meestal meerdere validiteitstests gebruikt. Wanneer meerdere tests in dezelfde richting wijzen, kunnen conclusies met grotere zekerheid worden getrokken. Zijn de resultaten niet eensluidend, dan is voorzichtigheid geboden. Bovendien verschillen validiteitstests wat betreft de cognitieve vaardigheden of de klachten en symptomen waaraan ze appelleren. Om deze redenen bevelen professionele organisaties aan meerdere validiteitstests te gebruiken.⁴

In geval van niet-eensluidende resultaten rijst de vraag aan welke uitslag de meeste waarde moet worden gehecht. Hier wordt de vakkennis van de neuropsycholoog over de specificaties van de tests weer belangrijk. Daarnaast zijn er statistische methoden waarmee testresultaten kwantitatief kunnen worden gecombineerd.⁵

6. Kenmerken van een degelijk neuropsychologisch onderzoek

Het zal nu duidelijk zijn dat SVT's en PVT's in een neuropsychologisch expertisearchief onmisbaar zijn. Een NPO waarin deze instrumenten ontbreken, is ondeugdelijk in de context van een expertise. Een ander kwaliteitskenmerk is de vakkennis van de psycholoog die het NPO uitvoert of onder wiens verantwoordelijkheid dit gebeurt. De beste waarborg hiervoor biedt de BIG-registratie als klinisch neuropsycholoog. Een BIG-geregistreerde klinisch neuropsycholoog heeft na de academische studie (meestal met neuropsychologie als hoofdrichting) eerst de tweejarige opleiding tot gezondheidszorgpsycholoog gedaan en vervolgens de vierjarige specialistische opleiding tot klinisch neuropsycholoog. Deze opleiding biedt een grondige kennis van onder meer principes van de

4. S.S. Bush, R.M. Ruff, A.I. Tröster, J.T. Barth, S.P. Koffler, N.H. Pliskin, C.R. Reynolds & C.H. Silver, Symptom validity assessment: Practice issues and medical necessity NAN policy & planning committee, *Archives of Clinical Neuropsychology* (20) 2005, afl. 4, p. 419-426; R.L. Heilbronner, J.J. Sweet, J.E. Morgan, G.J. Larrabee & S.R. Millis, Conference participants. American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering, *The Clinical Neuropsychologist* (23) 2009, afl. 7, p. 1093-1129; NIP Sectie Neuropsychologie, Richtlijn voor neuropsychologische expertise, Utrecht 2016, zie www.psynip.nl.

5. Binder, Larrabee & Millis 2014; G.J. Larrabee, Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios, *The Clinical Neuropsychologist* (22) 2008, afl. 4, p. 666-679.

neurowetenschappen, de relaties tussen hersenen en gedrag, de normale en abnormale ontwikkeling van de hersenen, diagnostiek en behandeling van ziekten en beschadiging van de hersenen, en van psychopathologie en psychotherapie. Indien het NPO wordt uitgevoerd door een klinisch psycholoog of gezondheidszorgpsycholoog, is van belang dat deze voldoende kennis van en ervaring met hersenbeschadigde patiënten heeft. Ook is voldoende kennis van psychometrie noodzakelijk. Ten slotte dienen de internationale en nationale richtlijnen met betrekking tot neuropsychologische expertise te worden gevolgd.⁶ Indien hiervan wordt afgeweken, moet dit met redenen omkleed gebeuren.

7. Conclusies

De winst die is geboekt met het in maat en getal vaststellen van onderpresteren en aggraveren kan niet licht worden overschat. De tijd dat patiënten werden geloofd 'op hun mooie blauwe ogen' ligt achter ons.

Neuropsychologisch expertiseonderzoek in een context van letselschade of arbeidsongeschiktheid biedt, in tegenstelling tot vroeger, kwantificeerbare informatie over de betrouwbaarheid van gerapporteerde klachten. Onderpresteren impliceert een incongruente samenwerkingsrelatie, en dat heeft gevolgen voor diagnostiek en behandeling. Psychiatrische diagnostiek zal erdoor worden bemoeilijkt, en een ingezette behandeling kan eindeloos duren zonder resultaat.

Bij voorkeur bespreekt de neuropsycholoog het onderpresteren met betrokkene. Dit kan ertoe leiden dat opnieuw NPO wordt ingezet. Indien in tweede instantie geen sprake meer is van onderpresteren, dan zou dit kunnen wijzen op een goede prognose; betrokkene reageert immers constructief op de gegeven feedback.

Zoals geldt voor elke test, zijn er ook grenzen aan wat men kan meten met een PVT. Ernstig hersenletsel, zoals bij gevorderde dementie, kan leiden tot valspositieve uitslagen op PVT's. De reden daarvoor is dat in zo'n geval het cognitief functioneren dusdanig gecompromitteerd is, dat de PVT niet meer uitvoerbaar is voor de patiënt.

In neuropsychologische expertiseonderzoeken zijn formele tests van de symptoom- en prestatievaliditeit onmisbaar. Een NPO waarin dergelijke tests ontbreken, kan men als ondeugdelijk terzijde schuiven. Ook een NPO waarbij de resultaten van PVT's of SVT's door de psycholoog worden genegeerd, is ondeugdelijk. Helaas komt dit af en toe voor, meestal bij NPO's die door onvoldoende gekwalificeerde psychologen zijn uitgevoerd.

De beste kwaliteitsgarantie voor het NPO is wanneer het wordt uitgevoerd door of onder toezicht van een BIG-geregistreerde klinisch neuropsycholoog.

6. Bush, Ruff, Tröster, Barth, Koffler, Pliskin, Reynolds & Silver 2005; Heilbronner, Sweet, Morgan, Larrabee & Millis 2009; NIP Sectie Neuropsychologie 2016.