

Geautomatiseerde spraakanalyse bij onderscheiden van frontotemporale dementie en depressie

W.A. Krudop, A.E. Voppel, J.N. de Boer, H. van Schothorst, I.E.C. Sommer, Y.A.L. Pijnenburg

- Achtergrond** Klinisch onderscheid maken tussen patiënten met de gedragsvariant van frontotemporale dementie (gvFTD) en patiënten met een depressie is uitdagend. Mogelijk biedt geautomatiseerde spraakanalyse een diagnostische kans.
- Doel** Onderzoeken of geautomatiseerde spraakanalyse kan helpen bij het onderscheiden van de gedragsvariant van frontotemporale dementie van een depressieve stoornis.
- Methode** Bij 15 patiënten met gvFTD, 15 patiënten met een depressie en 15 controlepersonen zonder stoornis werd een semi-gestructureerd interview opgenomen, getranscribeerd en geanalyseerd, waarna zowel akoestische als semantische kenmerken werden geëxtraheerd en geclassificeerd met machinelearning.
- Resultaten** Met akoestische kenmerken, zoals klankkenmerken of tempo, konden we gvFTD van depressie onderscheiden met een accuratesse van 80% en met semantische (taalinhoudelijke) kenmerken met een accuratesse van 70,8%.
- Conclusie** Zowel akoestische als semantische maten onderscheidde de twee groepen. Bij spraakonderzoek is het van belang aandacht te hebben voor privacy en potentiële vertekende factoren, zoals leeftijd, geslacht en etniciteit. Het onderzoek zou in een grotere onderzoeksgroep herhaald moeten worden.

Frontotemporale dementie (FTD) wordt vaak, tot bij de helft van de patiënten, aangezien voor een andere (neuro)psychiatrische stoornis.¹ Deze diagnostische verwarring en de daaruit volgende vertraging (*delay*) in het stellen van de correcte diagnose duurt gemiddeld meer dan zes jaar en zorgt voor zowel over- als onderbehandeling.²

FTD omvat een neurodegeneratief spectrum bestaande uit twee taalvarianten (semantische dementie en progressieve niet-vloeiende afasie) en een gedragsvariant (gedragsvariant frontotemporale dementie; gvFTD). Bij gvFTD is er vaak sprake van dementie op jonge leeftijd (< 65ste), van sociale-cognitiestoornissen en opvallende gedragsveranderingen, terwijl 'klassieke' dementie-symptomen zoals verdwalen of vergeetachtigheid vaak ontbreken.

Een specifieke biomarker voor gvFTD is er (nog) niet. De meest gestelde andere classificatie bij een patiënt die eigenlijk gvFTD heeft, is een depressieve stoornis of andere stemmingsstoornis.³ Hierdoor kan overbehandeling optreden met medicatie of bijvoorbeeld elektro-

convulsietherapie. Omgekeerd wordt ook regelmatig de classificatie gvFTD gesteld, waar er onderliggend sprake blijkt van bijvoorbeeld een stemmingsstoornis, waardoor onderbehandeling optreedt.³

In de vroege fase van FTD krijgt de patiënt vaak klinische verschijnselen passend bij één variant (gvFTD of één van de twee taalvarianten), maar in een gevorderd stadium ontstaan vaak aanvullende symptomen, officieel behorend bij één van de andere varianten.⁴ Een patiënt met gvFTD zal dus in een latere ziektefase ook taalstoornissen krijgen en een patiënt met semantische dementie vertoont enkele jaren later ook gedragsstoornissen.⁴ Het is dus aannemelijk dat er ook bij patiënten met gvFTD ergens in het ziektebeloop – in eerste instantie subtiele - veranderingen van de taal zullen optreden. Patiënten met de niet-vloeiende variant spreken langzaam, moeizaam en haperend. Er ontstaan spraakfouten (fonologische distorsies), de grammaticale structuur wordt steeds beperkter en de spraak bestaat steeds meer uit losse woorden. Ook wordt de spraak monotoner. De semantische variant wordt gekenmerkt door stoornis-

AUTEURS

Welmoed Krudop, psychiater, afd. Psychiatrie en Psychologie, St Antonius ziekenhuis, Utrecht en postdoctoraal onderzoeker, Alzheimercentrum Amsterdam UMC, locatie VUMC.

Alban Voppel, postdoctoraal onderzoeker, afd. Psychiatrie, Universiteit van Groningen, UMCG, Groningen.

Janna de Boer, arts in opleiding tot psychiater, afd. Psychiatrie, Universiteit van Groningen, UMCG, Groningen, en afd. Psychiatrie van het UMCU en Utrecht Brain Center.

Henrita van Schothorst, bachelorstudent, Universiteit van Amsterdam.

Iris Sommer, hoogleraar Cognitieve aspecten van neurologische en psychiatrische aandoeningen, afd. Psychiatrie, Universiteit van Groningen, UMCG, Groningen.

Yolande Pijnenburg, Hoogleraar Dementie op jonge leeftijd, Alzheimercentrum, Amsterdam UMC, locatie VUMC, Amsterdam Neuroscience, Amsterdam.

Correspondentie

Dr. W.A. Krudop (w.krudop@antoniusziekenhuis.nl).

Geen strijdige belangen meegedeeld.

Het artikel werd voor publicatie geaccepteerd op 18-11-2022.

Citeren

Tijdschr Psychiatr. 2023;65(3):193-197

sen in woordbegrip en in het benoemen van objecten. Patiënten kiezen vaak een globale aanduiding voor een specifiek woord of omschrijven dit woord. De spraak is vloeiend, maar leeg en patiënten kunnen moeilijk te onderbreken zijn.⁵ Eerder toonde geautomatiseerde analyse van spraak bij patiënten met gvFTD veelbelovende resultaten in vergelijking met een controlegroep (90% nauwkeurigheid), maar diagnostische groepen (gvFTD versus taalvarianten versus controlegroep zonder klachten) onderscheiden bleek lastiger (59,4% nauwkeurigheid).⁶

Sinds enkele jaren kunnen met geautomatiseerde taalanalyses van spontane spraak variabelen geïdentificeerd worden met goede diagnostische kenmerken voor psychosespectrumaandoeningen.^{7,8} Bij dit soort onderzoek worden talige kenmerken gebruikt in een zelflerend algoritme dat zo goed mogelijk de diagnostische groep probeert te voorspellen aan de hand van de beschikbare data. Ook bij depressieve stoornissen zijn aspecten van de spontane spraak gevonden die afwijken, zoals een lager spreektempo, minder toonhoogtevariabiliteit en meer referentie naar het 'zelf' (meer gebruik van het woord 'ik'), waarmee versus controlegroepen een diagnostische nauwkeurigheid van 91% kon worden bereikt.⁹ Het is dus denkbaar dat zowel akoestische (spreektempo en toonhoogte) als semantische (inhoudelijke betekenis) taalkenmerken nuttig kunnen zijn bij de (vroeg) diagnostiek naar gvFTD. Echter, we weten dat er grote overlap kan zijn met een depressieve episode, omdat bijvoorbeeld afvlakking, verminderde initiatiefname en sociale terugtrekking zowel bij een depressieve stoornis als bij gvFTD voor kunnen komen. Of deze overlap ook aanwezig is in semantische en akoestische variabelen is tot dusver onbekend.

DOEL

Verkennen of we met geautomatiseerde spraakanalyse taalkenmerken van patiënten met gvFTD en die met een depressieve stoornis kunnen karakteriseren, wat zou kunnen bijdragen aan nauwkeurigere differentiatie tussen de groepen.

METHODE

Deelnemers

Per groep (patiënten met gvFTD, patiënten met depressie en deelnemers zonder stoornis als controlegroep) werden 15 deelnemers geïncludeerd. De patiënten met gvFTD voldeden allen aan de classificatie van waarschijnlijke gvFTD (*probable bvFTD*) volgens de internationale consensuscriteria.¹⁰ Alle deelnemers waren ouder dan 18 jaar en hadden Nederlands als moedertaal. Exclusiecriteria waren niet-gecorrigeerde gehoorstoornissen of spraakstoornissen zoals stotteren. Schriftelijke toestemming werd verkregen van alle deelnemers. Binnen de groep patiënten met depressie werd de *Hamilton depressie-schaal* afgenomen, om een indruk te krijgen van de ernstmaat. De diagnose in deze groep was gesteld door de behandelend psychiater. Deelnemers uit de groep met een *major depressive disorder* (MDD) en de gematchte controlepersonen zonder stoornis kregen een kleine vergoeding (10 euro) vanwege losse deelname aan enkel deze studie (UMC Utrecht); deelnemers met gvFTD deden mee als onderdeel van een regulier onderzoekstraject (Alzheimercentrum, Amsterdam UMC).

Opgenomen interviews

Een semigestructureerd interview van 15 minuten werd gebruikt om spontane spraak op te nemen. Dit interview bestond uit neutrale vragen over ervaringen, gedachten en herinneringen uit het dagelijks leven. Expliciete vragen over gezondheid of welzijn werden vermeden. Voorbeelden van vragen waren ‘Kan je iets vertellen over je zwemlessen vroeger?’ en ‘Wat zou je doen als je een miljoen zou winnen?’ De spraak van zowel de interviewer als de deelnemer werd opgenomen met losse microfoons (AKG C544I), op het hoofd gedragen, aangesloten op een TASCAM DR40-recorder, zodat separate audiokanalen per persoon beschikbaar waren.

Taalverwerking

De opgenomen tweekanaals-audiobestanden werden voorberekt met de PRAAT-software om de spraak van de interviewer te verwijderen.¹¹ De resterende spraaksegmenten werden aaneengeschaald. De spraak van de deelnemer werd handmatig getranscribeerd met CLAN-software volgens de CHILDES-CHAT-richtlijn.¹²

Akoestische analyse

De methode van De Boer e.a.⁸ gebruikten we om het akoestische signaal van de spraak van deelnemers te analyseren, met OpenSMILE.¹³ Specifiek gebruikten we de *Extended Geneva Minimalistic Acoustic Parameter Set*.¹⁴ Deze parameterset kan worden onderverdeeld in vier verschillende soorten kenmerken: temporeel (bijv. spreeknelheid), frequentie (bijv. toonhoogte), spectraal (bijv. *mel-frequency cepstral* coëfficiënt (MFCC), een maat voor het spectrale profiel (het *powerspectrum*) van geluid) en energie/amplitude (bijv. luidheid). Voor elke deelnemer werden 88 parameters geëxtraheerd en gegroepeerd in deze 4 soorten kenmerken.

Transcriptie en vectorisatie

We gebruikten het semantische taalmodel word2vec om de coherentie van spraak te bepalen.⁷ Word2vec is een model dat gebruikmakend van het samen voorkomen van woorden in taal de samenhang in betekenis tussen woorden kan bepalen.¹⁵ Voor dit onderzoek trainden we word2vec op het Corpus Gesproken Nederlands.¹⁶ Met dit model werden de transcripten omgerekend tot vectoren in een multidimensionale ruimte. Woorden die frequent in vergelijkbare contexten voorkomen, zijn qua betekenis verwant. Woorden die niet vaak in dezelfde context voorkomen, zullen juist meer onderscheid

in betekenis laten zien, wat resulteert in een grotere afstand tussen de bijbehorende vectoren.

Als voorbeeld kan als context de zin ‘Ik zit op een ...’ gebruikt worden, waarbij zowel ‘stoel’ als ‘bank’ vaak voor zullen komen; hun vectoren zullen dicht bij elkaar liggen, terwijl het woord ‘pinguïn’ niet vaak in deze context voor zal komen. De vectoren van pinguïn en stoel zullen dus sterk afwijken van elkaar.

Vervolgens kan de cosinus van de hoek tussen de vectoren worden berekend als een maat voor coherentie. De berekening wordt uitgevoerd binnen venstergroottes van een bepaald aantal woorden.⁵⁻¹⁰ Door het venster steeds een woord verder te schuiven tot het einde van het transcript, verkregen we waarden voor coherentie van het hele interview.

Statistische analyse

We vergeleken de demografische gegevens tussen groepen statistisch met een χ^2 -test voor categorische uitkomstmaten en een variantieanalyse (ANOVA) voor continue uitkomstmaten. De groepen werden vergeleken op akoestische en semantische taalkenmerken door multivariate analyses van covariantie (MANCOVA's) uit te voeren, inclusief leeftijd als covariabele. Vervolgens gebruikten we ANCOVA's voor post-hoc tests.

Voor het zelflerende *random-forest* algoritme gebruikten we het R-package Caret.¹⁷ Dit algoritme trainden we op 90% van de data, waarna we met de overige 10% van de data keken hoe adequaat het algoritme kon voorspellen; dit herhaalden we 10 keer om steeds te trainen op een ander deel van de data. De algoritmes werden getraind met de 88 spraakparameters met als doel het onderscheiden van de twee klinische groepen, FTD en MDD. We gebruikten de resulterende waarschijnlijkheidsschattingen om specificiteit, sensitiviteit, en een receiver-operatorcurve (ROC) en de oppervlakte onder de curve (AUC) te berekenen.

RESULTATEN

Demografische gegevens

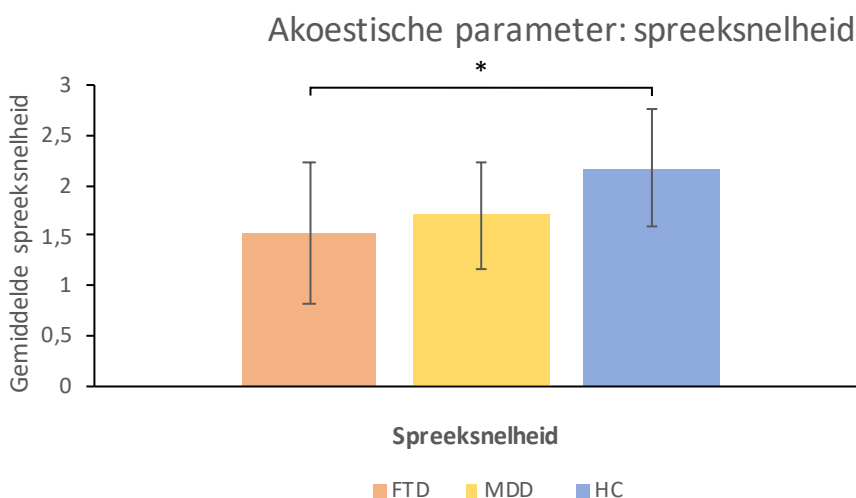
Demografische kenmerken van de deelnemersgroepen zijn weergegeven in **tabel 1**. De controlegroep en patiënten met FTD en depressie verschilden niet significant in leeftijd en geslacht. Patiënten met FTD hadden een significant kortere ziekteduur in vergelijking met patiënten met een depressie ($p < 0,001$).

Tabel 1. Kenmerken deelnemers

	gvFTD	Depressie	Controlepersonen	
Groepsgrootte	15	15	15	
Mannen, n	8	8	8	
Leeftijd, jaren (SD)	64,7 (9,9)	60,6 (7,2)	56,9 (8,0)	F = 3,204; p = 0,51
Ziekteduur, jaren (SD)	2,8 (2,9)	17,4 (12,4)		F = 13,687; p < 0,001
HDRS (SD)		17,3 (6,0)		

gvFTD: gedragsvariant frontotemporale dementie, SD: standaarddeviatie, HDRS: *Hamilton Depression Rating Scale*

Figuur 1. Spreeknelheid in woorden per seconde



FTD: patiënten met frontotemporale dementie; MDD: patiënten met major depressive disorder; HC: controlepersonen zonder stoornissen.

Akoestische resultaten

De uitgevoerde MANCOVA's op akoestische gegevens lieten een significant verschil zien tussen deelnemersgroepen op het gebied van temporele kenmerken ($F(2,12) = 2,465$; $p = 0,009$). Er werden echter geen significante groepsverschillen gevonden op frequentie ($F(2,48) = 0,976$; $p = 0,354$), energie/amplitudekenmerken ($F(2,28) = 0,795$; $p = 0,157$) en spectraal gerelateerde akoestische kenmerken ($F(2,86) = 0,000$; $p = 1$). De ANCOVA voor groepsverschillen op de afzonderlijke parameters van de temporele kenmerken liet significante groepeffecten op snelheid van luidheidspieken zien ($F(2,48) = 3,156$; $p = 0,032$ post-hoc Tukey's HSD-test), wat een approximatie is van spreeknelheid, zie **figuur 1**.

Semantische resultaten

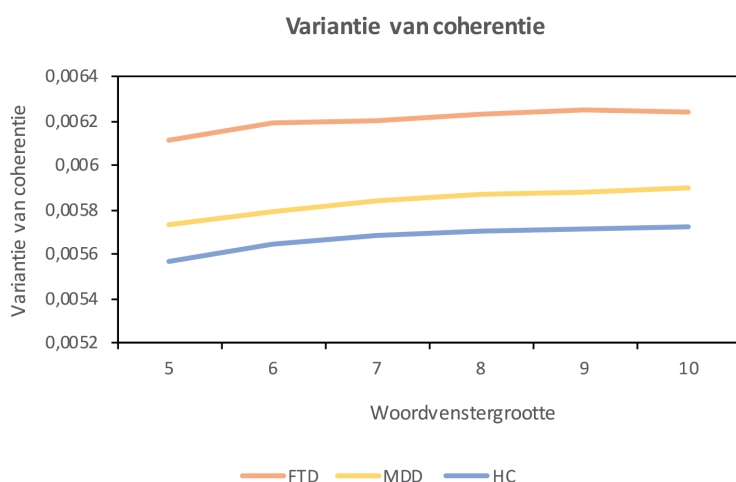
De verschillende MANCOVA-tests op de semantische gegevens lieten geen significante groepsverschillen zien

in gemiddelde ($F(2,12) = 1,080$; $p = 0,389$) en minimale coherentie ($F(2,12) = 0,838$; $p = 0,612$). Wel werd een significant groepsverschil in variantie van coherentie gevonden ($F(2,12) = 4,091$; $p = 0,044$). De daaropvolgende ANCOVA liet zien dat de groepen significant verschilden in de variantie van coherentie in alle vensterformaten (5-10). Post-hoc Tukey's HSD-tests toonden een significant hogere variantie van coherentie bij patiënten met gvFTD in vergelijking met de controlepersonen, zie **figuur 2**.

Classificatie

Met het random-forestalgoritme konden we aan de hand van akoestische kenmerken een onderscheid maken tussen de groep met gvFTD en de groep met depressie, met een accuratesse van 80%, met daarbij een specificiteit van 80% (95%-BI: 0,55-0,93) en een sensitiviteit van 80% (95%-BI: 0,55-0,93). De bijbehorende *area under the curve*

Figuur 2. Spreiding van de semantische coherentie



FTD: patiënten met frontotemporale dementie; MDD: patiënten met major depressive disorder; HC: controlepersonen zonder stoornissen.

(ROC-AUC) van dit model was 0,84. In de *classifier* was de MFCC de informatiefste variabele, gevolgd door het gemiddelde aantal geluidspieken (een maat voor spreek-snelheid).¹⁸

Gebruikmakend van semantische kenmerken konden we met het random-forest algoritme een onderscheid maken tussen gvFTD en depressie met een accuratesse van 70,8%, met daarbij een specificiteit van 77% (95%-BI: 0,48-0,89) en een sensitiviteit van 67% (95%-BI: 0,42-0,85). De ROC-AUC van dit model was 0,75. De informatiefste maat voor het algoritme om onderscheid te maken was de minimale coherentie bij een venster-grootte van 6 woorden; de op één-na hoogste was de variantie van coherentie bij een venster-grootte van 7. Dat de informatiefste waarden voor onderscheid taal-kenmerken waren, die niet significant tussen de groepen verschilden, kan verklaard worden doordat deze algoritmes complexe patronen in data kunnen herkennen, waarbij er bijvoorbeeld door een combinatie van niet-significante maten toch goed onderscheid gemaakt kan worden.

DISCUSSIE

Het gebruik van geautomatiseerde analyse van spontane spraak is veelbelovend voor het onderscheiden van gvFTD van depressieve stoornissen. Aangezien dit onderscheid in de klinische praktijk vaak op langdurige diagnostische twijfel en onduidelijkheid over behandel-mogelijkheden stuit, zou de geautomatiseerde spraak-analyse een praktisch toepasbare en patiëntvriendelijke bijdrage kunnen leveren in het voorkomen van deze diagnostische vertraging met onder- of overbehandeling tot gevolg.

Hoewel spraak nauwkeurig te kwantificeren is en gemakkelijk te verkrijgen is, weten we op dit moment niet in welke mate classificatiemodellen gedreven worden door vertekende factoren. In een recent artikel toonde men bijvoorbeeld aan dat taalanalyses meer gerelateerd waren aan sociale, culturele en demografische factoren (bijv. etnische achtergrond) dan aan de symptomen die ze probeerden vast te leggen (in dit geval formele denkstoornissen).¹⁹ Hoewel de invloed van sociale kenmerken op spraak de afgelopen jaren meer aandacht heeft gekregen, is er nog veel onbekend.²⁰ Wat wel al bekend is, is dat geslacht een duidelijk effect heeft op spraak, wat evident is in het verschil in toon-hoogte, maar het heeft ook invloed op andere maten, zoals spreektempo.²¹ Ten tweede is leeftijd een belangrijke confounder. Jongere mensen spreken bijvoorbeeld gemiddeld 5% sneller dan oudere.²¹ Ten derde laat recent onderzoek zien dat taalanalyses een systematische bias kunnen hebben, vermoedelijk gebaseerd op verschillende etnische kenmerken, die racisme en ongelijkheid jegens minderheidsgroepen zou kunnen vergroten.²²

Voor de huidige pilotstudie hebben we de groepen gematcht op leeftijd en geslacht en hebben we alleen patiënten met Nederlands als moedertaal geïncludeerd om dit soort invloeden op de uitkomsten te beperken, maar hiervoor is dus waakzaamheid geboden. Privacy is een belangrijke overweging bij onderzoek met audio-opnames, aangezien zowel de inhoud van een gesprek als het stemgeluid uniek voor een persoon zijn. Via een klankbordgroep is onderzocht hoe patiënten denken over het opnemen van interviews. 66% van ondervraagden gaf aan hier positief tegenover te staan gezien de potentie voor vroegdiagnostiek, maar privacy kwam wel als grootste bezwaar naar voren.²³ De pilotstudie toont dat zowel de akoestische maten als de semantische maten verschillen tussen de twee groepen tonen. Een belangrijke beperking echter is de kleine groepsgrootte. Met tienvoudige crossvalidatie hebben we geprobeerd om *overfitting* te voorkomen, maar desondanks zou het onderzoek in een grotere onderzoeksgroep herhaald moeten worden om betrouwbaardere conclusies te kunnen trekken. De bevindingen laten zien dat machinelearning aan de hand van talige kenmerken het onderscheiden van gvFTD en MDD zou kunnen verbeteren, leidend tot accuratere diagnostiek en behandeling.

LITERATUUR

- 1 Woolley JD, Khan BK, Murthy NK, e.a. The diagnostic challenge of psychiatric symptoms in neurodegenerative disease: rates of and risk factors for prior psychiatric diagnosis in patients with early neurodegenerative disease. *J Clin Psychiatry* 2011; 72: 126-33.
- 2 Van Vliet D, de Vugt ME, Bakker C, e.a. Time to diagnosis in young-onset dementia as compared with late-onset dementia. *Psychol Med* 2013; 43: 423-32.
- 3 Krudop WA, Dols A, Kerssens CJ, e.a. The pitfall of behavioral variant frontotemporal dementia mimics despite multidisciplinary application of the FTDC criteria. *J Alzheimer's Dis* 2017; 60: 959-75.
- 4 Ulugut H, Stek S, Wagemans LEE, e.a. The natural history of primary progressive aphasia : beyond aphasia. *J Neurol* 2022; 269: 1375-85.
- 5 Gorno-Tempini ML, Hillis AE, Weintraub S, e.a. Classification of primary progressive aphasia and its variants. *Neurology* 2011; 76: 1006-14.
- 6 Zimmerman VC, Hardy CJD, Eastman J, e.a. Automated profiling of spontaneous speech in primary progressive aphasia and behavioral-variant frontotemporal dementia: An approach based on usage-frequency. *Cortex* 2020; 133: 103-19.
- 7 Voppel AE, de Boer JN, Brederoo SG, e.a. Quantified language connectedness in schizophrenia-spectrum disorders. *Psychiatry Res* 2021; 304: 114130.
- 8 De Boer JN, Voppel AE, Brederoo SG, e.a. Acoustic speech markers for schizophrenia-spectrum disorders: A diagnostic and symptom-recognition tool. *Psychol Med* 2021; doi: 10.1017/S0033291721002804.
- 9 Koops S, Brederoo SG, de Boer JN, e.a. Speech as a biomarker for depression. *CNS Neurol Disord Drug Targets* 2023; 22: 152-60.
- 10 Rascovsky K, Hodges JR, Knopman D, e.a. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain* 2011; 134: 2456-77.

De overige literatuurverwijzingen zijn online te raadplegen.