

Functionele hersennetwerken voor diagnostiek en de voorspelling van behandeluitkomst

G.A. van Wingen

- Achtergrond** Functionele MRI biedt inzicht in het functioneren van hersennetwerken van patiënten met psychiatrische stoornissen. Met machinelearninganalyse kunnen modellen worden gemaakt voor diagnostiek en het voorspellen van behandeluitkomst.
- Doel** Een overzicht bieden van recente inzichten in diagnostische en voorspellende neuro-imagingbiomarkers.
- Methode** Narratieve review van recente literatuur.
- Resultaten** Grootschalige studies suggereren dat diagnostische modellen voor de meeste stoornissen een beperkte nauwkeurigheid hebben. Daarentegen suggereren meta-analyses van kleinschalige studies dat de behandeluitkomst voor depressie en psychotische stoornissen goed kan worden voorspeld.
- Conclusie** Dit creëert de mogelijkheid om voorspellingsmodellen te ontwikkelen die behandelaars kunnen helpen bij het maken van een behandelplan en daarmee de behandeluitkomsten te verbeteren.

In dit artikel geef ik een overzicht van de laatste inzichten in het gebruik van neuro-imaging en machinelearning voor het maken van diagnostische modellen en prognostische modellen die de behandeluitkomst kunnen voorspellen. Na een korte inleiding over functionele *magnetic resonance imaging* (fMRI) en machinelearning bespreek ik de toepassing voor het maken van diagnostische biomarkers op basis van resultaten van meta-analyses en grote multicenterstudies. Vervolgens ga ik in op het voorspellen van behandeluitkomst met een focus op functionele netwerken voor depressie en psychotische stoornissen, waarvoor op dit moment de meeste literatuur beschikbaar is. Tot besluit volgt een beschouwing over het benodigde onderzoek om de onderzoeksresultaten te vertalen naar klinische toepassingen.

Neuro-imaging

Functionele MRI heeft het mogelijk gemaakt om de communicatie tussen verschillende hersengebieden in beeld te brengen. De hersengebieden die het sterkst met elkaar verbonden zijn via grote witte stofbanen vertonen ook de grootste mate van synchrone activiteit, ofwel functionele connectiviteit.¹ Functionele MRI is daarom ook veelvuldig gebruikt om het functioneren van hersennetwerken bij patiënten met verschillende psychiatrische stoornissen te onderzoeken.

Er zijn verschillende benaderingen voor het onderzoek naar functionele hersennetwerken. De meest gebruikte methodes zijn *functional connectomics* en *onafhankelijke componentanalyse*.² Bij functional connectomics worden de hersenen opgedeeld in afzonderlijke hersengebieden volgens een anatomische (of functionele) atlas. Vervolgens wordt de verandering in hersenactiviteit gedurende de fMRI-scan van al deze hersengebieden met elkaar gecorreleerd. Dit levert een correlatiematrix op die laat zien welke hersengebieden samen een netwerk vormen. Bij onafhankelijke componentanalyse wordt de hersenactiviteit van alle hersengebieden tegelijk geanalyseerd om daar een aantal functionele netwerken uit te halen. Onderzoek naar functionele netwerken bij psychiatrische stoornissen heeft een enorme vlucht genomen door de oprichting van internationale consortia waarbij fMRI-data of afgeleiden daarvan worden gedeeld. Dit heeft geleid tot beter inzicht in de neurale basis van onder andere autismespectrumstoornissen, schizofrenie en depressie. In dergelijke studies gebruikt men doorgaans univariate statistische methodes om verschillen tussen patiënten en controlepersonen zonder klachten te onderzoeken. Daarbij wordt de gemiddelde connectiviteit van een patiëntengroep vergeleken met de gemiddelde connectiviteit van controlepersonen met variantieanalyses of t-testen voor alle hersenverbindingen

AUTEUR

Guido van Wingen, hoogleraar Neuro-imaging in de psychiatrie, afd. Psychiatrie, Amsterdam UMC, locatie AMC, Amsterdam.

Correspondentie

Prof. dr. Guido van Wingen
(g.a.vanwingen@amsterdamumc.nl).

Geen strijdige belangen meegegeeld.

Het artikel werd voor publicatie geaccepteerd op 3-11-2023.

Citeren

Tijdschr Psychiatr. 2023;65(10):605-608

afzonderlijk. Daarnaast kunnen multivariate methoden zoals machinelearning worden gebruikt om maximaal onderscheid tussen patiënten en controlepersonen te maken door alle verbindingen tegelijk te analyseren. Machinelearning biedt ook de mogelijkheid om uitspraken te doen voor individuele patiënten, wat noodzakelijk is voor klinische toepassing.

Machinelearning

Het voornaamste doel van machinelearning is om voorspellingsmodellen te maken op basis van alle beschikbare data. Hierbij wordt een deel van de data (meestal 80 of 90%) gebruikt om modellen te maken, ofwel te *trainen*. Het resterende deel van de data (10 of 20%) wordt gebruikt om de voorspellende nauwkeurigheid te *testen*. Dit wordt vaak gedaan met kruisvalidatie, waarbij de data die worden gebruikt voor het trainen en testen steeds afwisselen. Er wordt dan vooral gekeken naar hoe nauwkeurig (*balanced accuracy*, ofwel het gemiddelde van sensitiviteit en specificiteit) de getrainde modellen voorspellingen kunnen maken voor de testdata. Bij machinelearning worden modellen dus altijd gevalideerd door een model te testen met nieuwe deelnemers. Zolang de testdata onafhankelijk zijn van de trainingsdata geeft dit een betrouwbare schatting.

Maar helaas worden hier nog regelmatig fouten mee gemaakt waardoor informatie van de testdata doorlekt naar de getrainde modellen (*data leakage*). Dit heeft tot gevolg dat de resultaten beter lijken dan ze zullen zijn wanneer het model op werkelijk onafhankelijke data wordt getest (positieve *bias*). In een recente systematische review vonden wij dat dit het geval kan zijn bij de helft van de gerapporteerde studies,³ waardoor veel publicaties in de psychiatrie een overschatting geven van de nauwkeurigheid. Om dit in de toekomst te voorkomen hebben Poldrack en collega's een korte handleiding geschreven voor correct gebruik van machinelearningmethoden voor de psychiatrie.⁴ De benodigde effectgrootte om voorspellingen te doen voor individuele patiënten is aanzienlijk. Deze is ook substantieel groter dan wat gebruikelijk is om significante verschillen tussen groepen te detecteren en op

basis waarvan huidige behandelingen zijn gebaseerd. Een richtlijn voor de nauwkeurigheid van diagnostische biomarkers van de American Psychiatric Association (APA) is 80%.⁵ Dat betekent dat de biomarker voor 8 van de 10 patiënten een correcte uitslag geeft, wat neerkomt op een oddsratio van 16. Ter vergelijking: meta-analyses laten zien dat de proporties patiënten die een respons vertonen op antidepressiva en antipsychotica maximaal twee keer zo groot zijn als de proporties patiënten die een respons vertonen op placebo. Dit komt neer op oddsratio's van rond de 2.^{6,7} De eis aan de effectgrootte voor voorspellingsmodellen is dus vele malen hoger dan die voor medicatie.

Diagnostische biomarkers

Diagnostische biomarkers voor psychiatrische stoornissen zijn in eerste instantie onderzocht in kleine steekproeven. Meta-analyses van dergelijke studies laten zien dat functionele netwerken een (zeer) goed onderscheid kunnen maken tussen controlepersonen zonder psychiatrische aandoeningen en patiënten met onder andere autismespectrumstoornissen (*balanced accuracy*: 74%),⁸ schizofrenie (81%)⁹ en depressie (84%)¹⁰ en voldoen veelal aan de richtlijn van de APA. De nauwkeurigheid van de diagnostische modellen wordt echter lager wanneer kleine steekproeven worden gecombineerd tot grote datasets voor multicenteronderzoek.^{11,12} Dit is opmerkelijk aangezien machinelearningmodellen altijd beter presteren naarmate meer data beschikbaar zijn om de modellen te trainen.¹³ De lagere nauwkeurigheid van grote studies komt vermoedelijk doordat de klinische heterogeniteit bij grote studies ook groter is.¹⁴ Er is dus een balans tussen nauwkeurige modellen voor specifieke doelgroepen en minder nauwkeurige modellen die goed bruikbaar zijn voor de verscheidenheid aan patiënten die wel allemaal aan dezelfde diagnose voldoen.

Recente onderzoeken laten zien dat de nauwkeurigheid van diagnostische modellen getraind op functionele netwerkdata van grote consortia rond de 67% ligt voor autismespectrumstoornissen,¹⁵ 61% voor depressie,¹⁶ 81% voor schizofrenie¹⁷ en 63% voor obsessieve-compul-

sieve stoornis.¹⁸ Hieruit blijkt dat hersennetwerkfunctie wel degelijk verschilt tussen patiënten met verschillende psychiatrische stoornissen, maar dat de nauwkeurigheid om individuele patiënten te onderscheiden doorgaans laag is. De uitzondering lijkt schizofrenie, waarbij verschillen in hersennetwerkfunctie ten opzichte van controlepersonen mogelijk groter zijn en/of de verscheidenheid binnen de patiënten kleiner is dan bij andere psychiatrische stoornissen.

De klinische relevantie van diagnostische modellen is echter voornamelijk beperkt. Het wordt relevanter wanneer dergelijke modellen ook geschikt zijn voor differentiële diagnostiek, bijvoorbeeld om onderscheid te kunnen maken tussen bipolaire stoornis, unipolaire depressie en schizofrenie.¹⁹ Maar studies om het volledige spectrum van stoornissen te kunnen onderscheiden zijn nog niet voor handen.

Voorspellen van behandeluitkomst

Een toepassing die wel direct klinisch relevant is, is het voorspellen van behandeluitkomst. De effectiviteit van verschillende behandelingen is zoals in het voorgaande beschreven beter dan die van placebo. Maar bij elke fase van een behandeltraject herstelt een aanzienlijk deel van de patiënten niet voldoende. Zo laten grote naturalistische studies zien dat hooguit een derde van de patiënten met een depressie herstelt na elke behandeling met antidepressiva of psychotherapie.²⁰ En zo herstelt hooguit de helft van de patiënten met een psychotische stoornis na elke behandeling met antipsychotica.²¹ De effectiviteit van deze behandelingen zou kunnen verbeteren wanneer een behandeling kan worden afgestemd op de patiënt. Het is dan niet meer nodig om via *trial-and-error* een passende behandeling te vinden. Maar voor zulke zogenaamde gepersonaliseerde, precisie- of *high-performance*zorg²² is het nodig om vooraf de beste behandeling te kunnen selecteren. Behandelaars kunnen de behandeluitkomst niet goed voorspellen en hebben een positieve *bias* dat hun behandeling zal werken.²³ Daarom bestaat er al een lange geschiedenis om methoden te ontwikkelen die wel een nauwkeurige voorspelling kunnen geven. Zo ontwikkelde Hobson al in 1953 een scoringsmethode om de respons op elektroconvulsie therapie (ECT) te voorspellen.²⁴ Maar tot op heden zijn er nog geen instrumenten beschikbaar die bruikbaar zijn in de klinische praktijk.

De laatste jaren is machinelearning enorm in opkomst om de respons op een behandeling te voorspellen. Bij psychotische stoornissen voorspellen klinische variabelen, demografische gegevens en vooral de zorgbehoeften een goede uitkomst na 4 weken behandeling met haloperidol, amisulpride, olanzapine, quetiapine of ziprasidon met 75% nauwkeurigheid.²⁵ Bij depressie voorspellen demografische gegevens en vooral de symptomen of er wel of geen remissie zal zijn na 12 weken behandeling met citalopram met 65% nauwkeurigheid.²⁶ Deze scores werden behaald wanneer de nauwkeurigheid werd getest voor nieuwe patiënten uit dezelfde

steekproef, ofwel met *interne validatie*. Wanneer het voorspellingsmodel voor psychotische stoornissen werd getest op patiënten van nieuwe ziekenhuizen, daalde de nauwkeurigheid naar 72%. En wanneer het voorspellingsmodel voor depressie werd toegepast op een andere studie met escitalopram, daalde de nauwkeurigheid naar 60%, wat niet veel beter is dan het 50%-kansniveau.

Deze manier van *externe validatie* geeft de beste voorspelling van de verwachte nauwkeurigheid wanneer een model ook daadwerkelijk wordt toegepast in de klinische praktijk. Het gebruik van klinische variabelen is dus veelbelovend voor psychotische stoornissen. Maar de voorspellende waarde van klinische variabelen voor depressie is beperkt. Er is daarom ook een zoektocht gaande naar de beste biomarkers om de voorspelling te verbeteren. Daarvoor wordt onder andere onderzoek gedaan naar genetica, maar is vooral neuro-imaging gebruikt.^{27,28}

Voorspellen met neuro-imaging

Uit een eerste meta-analyse naar het gebruik van verschillende typen data om met machinelearning de behandeluitkomst bij depressie te voorspellen kwam neuro-imaging als meest belovend naar voren.²⁹ De 13 geïncludeerde neuro-imagingstudies hadden een gemiddelde nauwkeurigheid van 85%, terwijl de 5 studies met klinische variabelen een gemiddelde nauwkeurigheid van 69% hadden. De neuro-imagingstudies bestonden echter veelal uit kleine steekproeven, terwijl men in de studies met klinische variabelen grote steekproeven had gebruikt. De schatting voor de nauwkeurigheid van de neuro-imagingstudies is derhalve minder betrouwbaar dan voor de klinische studies en zou bij nieuwe studies zowel hoger als lager uit kunnen vallen. Daarnaast werden bij deze meta-analyse verschillende typen neuro-imaging, behandelingen en typen uitkomsten samengevoegd, waardoor men in de meta-analyse vooral heeft onderzocht *waarmee* kan worden voorspeld en het onduidelijk blijft *wat* er kan worden voorspeld.

Om te onderzoeken welke behandeluitkomsten goed voorspeld kunnen worden, hebben wij een nieuwe meta-analyse naar de beschikbare MRI-studies bij depressie uitgevoerd.³⁰ We hebben 27 studies geïncludeerd met een gemiddelde steekproefgrootte van 44 patiënten, variërend van 12 tot 124. Uit de resultaten bleek de gemiddelde nauwkeurigheid 82% voor ECT (9 onafhankelijke studies), 74% voor verschillende antidepressiva (8 onafhankelijke studies), 81% voor repetitieve transcraniële magnetische stimulatie (rTMS; 3 onafhankelijke studies) en 78% voor verschillende psychotherapieën (2 onafhankelijke studies). In de studies hebben onderzoekers allemaal verschillende neuro-imagingmethoden gebruikt, waarvan ongeveer een derde structurele MRI, een derde taakgerelateerde fMRI en een derde *resting-state*-fMRI. Meta-regressieanalyses lieten geen significant verschil in de nauwkeurigheid tussen de verschillende behandelingen en typen MRI zien.

Deze meta-analyse laat zien dat verschillende MRI-methoden de behandeluitkomst voor verschillende behandelingen bij depressie met hoge nauwkeurigheid kunnen voorspellen. Een meta-analyse van de beschikbare elektro-encefalogram(eeg)-studies toont vergelijkbare resultaten.³ Maar door het gebruik van verschillende methoden blijft zowel voor MRI als eeg onduidelijk met welke variabelen men die voorspelling kan doen. In een vergelijkbare meta-analyse van resting-state-fMRI-studies heeft men de respons op antipsychotica bij psychotische stoornissen onderzocht.³¹ In deze studie werden alleen studies geïnccludeerd waarbij men functionele netwerken had gebruikt. De onderzoekers hebben 9 datasets uit 7 verschillende studies geïnccludeerd met een gemiddelde steekproefgrootte van 50 patiënten (27 tot 90 patiënten) die de respons op verschillende antipsychotica hebben voorspeld. De gemiddelde nauwkeurigheid van deze studies was 79%. Meta-regressieanalyses toonden alleen een significante invloed van ziekte duur, waarbij de voorspelling nauwkeuriger is voor patiënten die al langer psychotisch zijn. Deze meta-analyse laat zien dat functionele netwerken ook de respons op verschillende antipsychotica bij psychotische stoornissen kunnen voorspellen. Maar ook hier is door het gebruik van verschillende definities van hersennetwerkwijkfunctie niet duidelijk wat nu precies de voorspellende netwerken zijn.

Explainable artificial intelligence

Machinelearningalgoritmen zoeken naar een combinatie van variabelen (*features*) om de beste voorspelling te maken. Door deze multivariate benadering wordt het niet direct inzichtelijk hoe een model de voorspelling maakt. Met aanvullende methoden kan de zogenaamde *feature importance* worden bepaald, maar dat betreft echter nog steeds de relatieve bijdrage van deze *features* in combinatie met andere variabelen. Er zijn ook methoden beschikbaar om de invloed van individuele hersengebieden te bepalen, maar daarvoor zijn grotere steekproeven nodig. Die zijn nog niet beschikbaar voor het voorspellen voor de behandeluitkomst, maar sinds kort wel voor diagnostische biomarkers. Zo hebben wij onlangs in twee grote consortia voor depressie de bijdrage van verschillende hersengebieden onderzocht met verschillende zogenaamde visualisatietechnieken. Daaruit blijkt hyperconnectiviteit van de thalamus in beide consortia het belangrijkste om patiënten met een depressieve stoornis te identificeren.¹⁶ Voor schizofrenie blijkt bij twee grote consortia vooral de connectiviteit tussen subcorticale hersenstructuren en de temporaalkwab van belang, maar de unieke bijdrage is daarbij niet onderzocht.¹⁷

Naar klinische toepassing

Uit de beschreven meta-analyses voor depressie en psychotische stoornissen blijkt dat functionele netwerken de behandeluitkomst goed kunnen voorspellen, met een gemiddelde nauwkeurigheid van rond de 80%.^{30,31}

De APA geeft 80% nauwkeurigheid als richtlijn voor klinische toepassing van *diagnostische* biomarkers, maar geeft geen richtlijn voor *prognostische* biomarkers.⁵ Met een nauwkeurigheid van 80% kan het behandelings succes toenemen tot 80% wanneer alleen patiënten met een goede prognose de betreffende behandeling krijgen. Op dit moment is het behandelings succes van antidepressiva en antipsychotica hooguit 50%.^{20,21} Prognostische modellen die minder nauwkeurig zijn dan 80% kunnen daarom ook voor betere behandelresultaten zorgen en zullen daarom waarschijnlijk ook kosteneffectief zijn. Zo ver is het echter nog niet. De steekproeven van de geïnccludeerde studies waren doorgaans klein, de *features* verschilden per studie en in elke studie onderzocht men slechts de uitkomst van één type behandeling. Het is daarom nog onduidelijk of de gerapporteerde resultaten ook in de praktijk kunnen worden behaald. In toekomstige studies zal men daarom eerst moeten onderzoeken of voorspellingsmodellen ook toegepast kunnen worden voor patiënten van andere ziekenhuizen. Dit lijkt in ieder geval mogelijk voor de voorspelling van herstel na ECT.³² Daarnaast is het voor de toepassing van voorspellingsmodellen vooral relevant om de respons voor verschillende behandelingen in kaart te brengen. Op die manier kunnen voorspellingsmodellen de behandelaar ondersteunen om voor elke patiënt de best passende behandeling te selecteren en daarmee de behandeling te personaliseren. Maar dergelijke studies zijn nog nauwelijks uitgevoerd.

Conclusie

Ik concludeer dat functionele netwerken de behandeluitkomst voor individuele patiënten met depressie en psychotische stoornissen met ongeveer 80% nauwkeurigheid kunnen voorspellen. Deze resultaten zijn echter behaald in kleinschalige studies, waarbij onduidelijk is of dezelfde resultaten ook in nieuwe steekproeven kunnen worden behaald. Daarnaast moet nog blijken of toepassing van voorspellingsmodellen het behandelings succes daadwerkelijk verhoogt.

LITERATUUR

- 1 Haggmann P, Cammoun L, Gigandet X, e.a. Mapping the structural core of human cerebral cortex. *PLoS Biol* 2008; 6: e159.
- 2 van den Heuvel MP, Hulshoff Pol HE. Exploring the brain network: a review on resting-state fMRI functional connectivity. *Eur Neuropsychopharmacol* 2010; 20: 519-34.
- 3 Cohen SE, Zantvoord JB, Wezenberg BN, e.a. Electroencephalography for predicting antidepressant treatment success: a systematic review and meta-analysis. *J Affect Disord* 2023; 321: 201-7.
- 4 Poldrack RA, Huckins G, Varoquaux G. Establishment of Best Practices for Evidence for Prediction: A Review. *JAMA Psychiatry* 2020; 77: 534-40.
- 5 First M, Botteron K, Carter C, e.a. Consensus report of the APA work group on neuroimaging markers of psychiatric disorders. Washington: American Psychiatric Association; 2012.

De overige literatuurverwijzingen zijn online te raadplegen.