

4. De invloed van telediagnostiek op de testkwaliteit

4.1. Algemene bevindingen

Het idee van telegeneeskunde gaat al sinds het begin gepaard met bezorgdheden. In de meeste ruime zin zou dergelijke toepassing van zorgverlening een bedreiging betekenen voor de therapeutische relatie, het zou gevolgen hebben voor kwaliteit, veiligheid (de clinicus beschikt over minder informatie dan in de fysische-klinische ontmoeting), aansprakelijkheid (ziekteverzekeraars vrezen dat telecommunicatie het aantal consultaties zal doen verdubbelen), enzovoort. Specifiek voor telediagnostiek stellen clinici zich, geheel terecht overigens, vragen over de impact die digitale afnames hebben op de (psychometrische en klinische) kwaliteit van de testafname.

In meerdere recente wetenschappelijke studies werd reeds de face-to-face afname van reguliere psychodiagnostische tests vergeleken met de afname van dezelfde tests via beeldbellen. De aanleiding voor deze studies heeft niets met COVID-19 te maken, maar komt veelal voort vanuit het gegeven dat patiënten die woonachtig zijn in buitengebieden van bijvoorbeeld Australië of de Verenigde Staten, minder toegang tot deze noodzakelijke vorm van gezondheidszorg hebben. Hoewel het opzet, de aantallen patiënten en het gebruikte testmateriaal verschillen tussen de studies, laten de resultaten duidelijke aanwijzingen zien dat psychodiagnostisch onderzoek via beeldbellen haalbaar en betrouwbaar is. [9]

Hierbij valt op dat het werkveld van de neuropsychologie bepaald niet vooroploopt bij de digitalisering van de materialen en methoden en dat we nu vooral reguliere tests via beeldbellen aanbieden die oorspronkelijk bedoeld zijn voor face-to-faceafname. Aan de hand van praktische overwegingen over triage, validiteit, betrouwbaarheid en ethiek laten de auteurs zien dat het in bepaalde situaties goed mogelijk is om psychodiagnostisch onderzoek bij volwassenen via beeldbellen te verrichten, zonder in te boeten aan de kwaliteit van de diagnostiek. [9]

Hoewel de huidige evidentie (cf. supra) dus hoopgevend is, benadrukt het Kwaliteitscentrum voor Diagnostiek (KCD) echter dat er nog een aanzienlijke behoefte is aan verdere wetenschappelijke onderbouwing. Het Kwaliteitscentrum stelt als volgt: **zowel vanuit klinisch als wetenschappelijk onderzoek kunnen er géén of onvoldoende argumenten aangehaald worden die erop duiden dat de kwaliteit van een diagnostisch proces kan gewaarborgd worden bij een uitvoering vanop afstand.** [5] Of anders gezegd: wat het Kwaliteitscentrum betreft is er onvoldoende evidentie op basis waarvan men kan stellen dat telediagnostiek geen impact heeft op de kwaliteit van psychodiagnostisch (test)onderzoek.

Verder wetenschappelijk onderzoek is dus noodzakelijk. Daarbij staat voorop dat er gestructureerd gegevens worden verzameld over de consequenties met betrekking tot de validiteit en betrouwbaarheid van instrumenten bij afname via beeldbellen, alsook over de standaardisering en normering ervan. Ook het ontwikkelen van digitale (versies van) taken kan het vak verder vooruithelpen. [9]

4.2. Betrouwbaarheid en validiteit

Een belangrijk criterium bij het samenstellen van een psychologische testbatterij voor is of de validiteit, betrouwbaarheid en normering voldoende goed zijn. Het is vanzelfsprekend dat men dit met de grootste zorg doet om de kwaliteit van testonderzoek te garanderen. Wat telediagnostiek betreft is het dan ook een belangrijke vereiste in de APA-richtlijn en AST dat gekeken moet worden naar de kwaliteit van de data, in het bijzonder wat betreft de betrouwbaarheid en validiteit, wanneer deze zijn vergaard op een manier die afwijkt van de standaard wijze. [9]

Bij voorkeur wordt gebruikgemaakt van **tests waarbij equivalentiestudies zijn uitgevoerd** om vast te stellen of scores verkregen via digitale aanbidding vergelijkbaar zijn met scores verkregen via standaard (papieren) afname en daarmee een vergelijkbare betrouwbaarheid en validiteit hebben. Uitgever Pearson heeft dergelijke equivalentiestudies uitgevoerd, waarbij over het geheel genomen goede equivalentie kon worden aangetoond tussen standaard en digitale afname (Daniel e.a., 2014). En hoewel een digitale afname van tests natuurlijk iets anders is dan een afname via beeldbellen, verschaffen deze equivalentiestudies wel inzicht in hoe geschikt een test met visuele stimuli is om via een beeldscherm aan te bieden. [9]

4.2.1. Betrouwbaarheid

De betrouwbaarheid van een instrument duidt op de mate waarin een instrument vrij is van meetfouten. Hoe waardevol en betrouwbaar is het resultaat van een specifieke test wanneer deze via technologische toepassingen wordt afgenomen?

Het literatuuroverzicht aan het begin van dit hoofdstuk (cf. supra) geeft hier een eerste indruk van. In aanvulling hierop is een duidelijk advies uit de APA-richtlijn (APA, 2020) om bij het maken van conclusies en klinische beslissingen de **betrouwbaarheidsintervallen te verbreden** bij het interpreteren van de resultaten. Dit om de invloed van testafname buiten de richtlijnen (eventueel grotere foutenmarge bij telediagnostiek) en mogelijke wijzigingen in het scoren van de testprestatie te ondervangen.

4.2.2. Validiteit

De validiteit, meer specifiek de constructvaliditeit, geeft aan in hoeverre de taak meet wat hij beoogt te meten. Welke invloed kan de tele-methodiek hebben op dit criterium?

We weten dat het nauwgezet volgen van de instructies met betrekking tot de afname in de testhandleiding van een instrument draagt bij aan de validiteit van de afname. Bij het inzetten van een taak door middel van beeldbellen is het dan ook van cruciaal belang om zo dicht mogelijk bij deze instructies te blijven. Des te meer men van de standaardprocedure afwijkt, des te groter de mogelijke impact op de validiteit.

Voor een goede beoordeling van het behoud van de validiteit is het belangrijk om te bepalen of de processen die ten grondslag liggen aan de prestatie op een taak, worden beïnvloed door de werkwijze en of de validiteit van de verzamelde gegevens is behouden. Voor goed psychodiagnostisch onderzoek moeten we dus ook terug naar de basis van waarom we doen wat we doen.

- Welk gedrag willen we uitlokken?
- Waar letten we op bij de afname van een taak?

De mate waarin je moet afwijken van de testhandleiding, waarmee de validiteit in het geding kan komen, verschilt bovendien per taak en type taak.

- De kwaliteit van verbale taken wordt bijvoorbeeld veel minder makkelijk beïnvloed door afname via beeldbellen.
- Bij de non-verbale (en vooral ook visuele) taken:
 - zijn er aandachtspunten wat betreft de zichtbaarheid en de kwaliteit van de getoonde afbeelding, maar bij het gebruik van het originele stimulusmateriaal is ook bij deze taken afname via beeldbellen mogelijk (Hildebrand e.a., 2004).
 - Het mogelijke effect van het aanbieden van visueel materiaal via een beeldscherm in plaats van via een stimulusboek is daarbij aanvullend een aandachtspunt.

4.3. Interpretatie van de testresultaten

4.3.1. Bevindingen vanuit onderzoek

In principe geldt dat, om de resultaten van een test betrouwbaar te kunnen interpreteren, een test op dezelfde gestandaardiseerde manier en in dezelfde modaliteit (e.g. pen-en-papier afname) moet worden afgenomen als tijdens het normeringsonderzoek (of moet equivalentie tussen beide versies worden aangetoond). Er bestaan echter nauwelijks testen waarbij een tele-afname genormeerd is. Bovendien is het onderzoek naar de equivalentie tussen face-to-face afnames en afnames op afstand erg schaars en staat dit onderzoek nog in zijn kinderschoenen¹.

Omwille van deze redenen stelt men dat de resultaten die verkregen zijn via telediagnostiek in de meerderheid van de gevallen niet betrouwbaar geïnterpreteerd kunnen worden. Wel kunnen deze resultaten een eerste indicatie geven (wanneer de nood hoog is en indien de diagnosticus kan beargumenteren dat bepaalde aspecten van het diagnostisch onderzoek zich lenen voor teleconsultaties).

Ander onderzoek, vooral uitgevoerd bij neuropsychologische tests, is meer hoopgevend. In 2017 werd een systematische review verricht waarin de resultaten van twaalf studies met in totaal 497 proefpersonen werden samengevat (Brearly e.a., 2017). De resultaten laten zien dat de verschillen in testcores tussen beide aanbiedingsprocedures overwegend klein zijn, namelijk ongeveer 10% van een standaarddeviatie lager voor de onderzoeken via beeldbellen in vergelijking met face-to-faceonderzoek. De oorzaak van dit kleine, maar consistente verschil is niet duidelijk, mogelijk speelt de kwaliteit van de videoverbinding en een grotere variatie in prestaties bij mensen ouder dan 75 jaar hierbij een rol (Brearly e.a., 2017).

¹ De American Psychological Association (APA) geeft een overzicht van het beperkt beschikbare onderzoek naar equivalentie. Het Inter Organisational Practice Committee (IOPC) verwijst naar 25 studies die onderzoek hebben verricht naar het uitvoeren van neuropsychologische testen op afstand. Er wordt echter geconcludeerd dat er momenteel nog te weinig evidentie beschikbaar is om equivalentie te garanderen.

Prestaties op veelgebruikte neuropsychologische tests zoals cijferreeksen, verbale vloeiendheid en verbale leertaken verschillen vrijwel niet. Op de kloktekening worden vaker verschillen gevonden die mogelijk samenhangen met de beperkte range van de scores op deze taak (range 0 tot 3). In veel studies werden verbaal aangeboden neuropsychologische tests gebruikt, maar ook bij tests waarbij stimuli visueel worden aangeboden zijn de resultaten tussen face-to-faceafname en afname via beeldbellen vergelijkbaar (bijvoorbeeld Hildebrand e.a., 2004). De effect size g gemeten over alle studies en alle instrumenten is $-0,03$ ($-0,08$ tot $0,02$), duidend op een te verwaarlozen verschil dat niet klinisch relevant geacht wordt. Onderzoekers bevelen wel aan om voorzichtig te zijn bij het interpreteren van lichte cognitieve tekorten (niet op stoornisniveau) en bij prestaties die dicht bij de grensscore van een test liggen (bijvoorbeeld MMSE=24). Over het meten van verandering bij herhaald testen via beeldbellen is nog weinig bekend (zie bijvoorbeeld Castanho e.a., 2017). Onderzoekers signaleren ook dat niet alleen lagere, maar ook hogere scores bij testafname via beeldbellen voorkomen. Dit hangt mogelijk samen met een betere concentratie bij patiënten en/of duidelijker spreken door de onderzoeker ter compensatie van de onlineverbinding (Brearly e.a., 2017).



Ondanks deze eerder positieve bevindingen blijft het belangrijk om te benadrukken dat het totaal aantal onderzoeken naar neuropsychologische testafname via beeldbellen tot op heden beperkt blijft. Hoewel bovenstaande bevindingen ondersteunend zijn voor de mogelijkheden van afname via beeldbellen, is meer onderzoek in dit veld nodig om verdere conclusies ten aanzien van de betrouwbaarheid en validiteit te bestendigen. [9]

4.3.2. Conclusie

Men dient steeds het feit in gedachte te houden dat de gelijkwaardigheid tussen in-person testen en tele-testen niet (noodzakelijk) gegarandeerd is. Elke samenvatting van testcores met referentiescores, percentielen of andere interpretatieve opmerkingen, moet melding maken van het feit dat de toediening niet standaard was en dat de niet-standaard toediening het risico in zich draagt te resulteren in meetfouten.

De uitvoerder dient de nodige voorzichtigheid en klinische deskundigheid aan de dag te leggen bij het interpreteren van de gegevens en het integreren ervan met andere informatie om conclusies te trekken en klinische beslissingen te nemen. Dit betekent dan ook dat de validiteit van de gegevens in het rapport openlijk aan de orde moet komen. [4]

Omdat er onvoldoende gegevens zijn om te suggereren dat er systematische aanpassingen van normen worden gebruikt om testresultaten anders te interpreteren, zijn de huidige aanbevelingen om te vertrouwen op normatieve en validiteitsgegevens verkregen met behulp van de standaard assessments, met duidelijke documentatie in het rapport.

4.4. Rapportage van de testresultaten

Zoals hierboven reeds is aangehaald dient de diagnosticus in de rapportage duidelijke verklaringen neer te schrijven over de algemene beperkingen en eventuele risico's ten gevolge van het niet-standaard uitvoeren van psychodiagnostisch onderzoek, en de mogelijke invloed/interactie die hiervan uit kan gaan op diagnostische conclusies en aanbevelingen voor behandeling. Daarnaast dient de diagnosticus expliciet melding te maken van de verschillende modaliteiten en aanpassingen die gepaard gaan met het niet-standaard uitvoeren van het psychodiagnostisch onderzoek (e.g. het gebruik van technologische toepassingen, de computervaardigheden van patiënt én testleider, de gebruikte platforms en hoe de tests zijn aangepast of gemodificeerd). [9]

In de verslaglegging is het belangrijk aan te geven op welke wijze de afname van het testonderzoek heeft plaatsgevonden en welke aanpassingen zijn doorgevoerd.

- Specifieke aanpassingen aan de testsituatie kunnen in een aparte alinea 'onderzoekscondities' worden weergegeven. Dit is vergelijkbaar met het beschrijven van het toestandsbeeld ten tijde van de afname in regulier diagnostiek onderzoek.
- De afwijkingen ten opzichte van het reguliere testprotocol kunnen worden beschreven, evenals eventuele incidenten die tijdens de testafname voorkwamen en mogelijk van invloed zijn geweest op het onderzoek.
- Ook informatie over het verbreden van de betrouwbaarheidsintervallen en de potentiële fouten die gemaakt kunnen worden tijdens de afname, interpretatie en bij de rapportage van de resultaten zijn belangrijk om in de verslaglegging te vermelden.

Bij het opstellen van het verslag, de conclusies en de aanbevelingen integreer je informatie uit diverse bronnen: dossieronderzoek, (hetero)anamnese, observatie en testafname. Gebruik van verschillende bronnen ter ondersteuning van de conclusies gaat de foutenmarge van het testonderzoek tegen.

De vraag is wel of en hoe deze overige gegevensverzameling, vooral observatie en (hetero)anamnese, beïnvloed wordt door beeldbellen. Beïnvloedt dit de betrouwbaarheid van deze gegevens? Mis je hierdoor mogelijk informatie? Er zijn voor zover bekend geen studies die onderzocht hebben of informatie verkregen uit deze bronnen even optimaal is wanneer ze verkregen is via beeldbellen. Er is dus zowel geen evidentie die wijst op suboptimale als optimale vergaring van deze gegevens via beeldbellen. [9]