

Leitlinien zur Klassifikation und Interpretation neuropsychologischer Testergebnisse

1 Vorbemerkungen

Die Urteilsbildung im Rahmen einer neuropsychologischen Untersuchung stützt sich in der Regel auf mehrere Bestandteile. Nebst den testdiagnostischen Ergebnissen müssen zur Entscheidung, ob eine neurokognitive Störung vorliegt oder nicht, selbstredend auch medizinische und psychopathologische Befunde, anamnestische und z.T. fremdanamnestische Daten, die subjektiven Beschwerden sowie die Verhaltensbeobachtung und auch die Symptomvalidierung ausreichend berücksichtigt werden. Qualitative Untersuchungsansätze haben in der Neuropsychologie weiterhin ihre Berechtigung (z.B. bei Anosognosie oder Apraxie) und stehen keineswegs im Widerspruch zu dem vom Ansatz her quantifizierenden Zugang der Testpsychologie. Die vorliegenden Leitlinien stellen eine Empfehlung im Sinne eines aktuellen Standards dar, um eine grössere interindividuelle Übereinstimmung bei der Klassifikation und Interpretation von Testergebnissen zu erzielen. Die Leitlinien sind konsensusorientiert, breit abgestützt, pragmatisch und sollen der weiteren Verbesserung der Qualität in diesem Bereich dienen. Sie ergänzen die Leitlinien zur Berichterstattung (www.neuropsych.ch), deren Hauptanliegen eine Verbesserung der Qualität von neuropsychologischen Berichten ist. Zunächst soll kurz auf den einer neuropsychologischen Untersuchung zugrundeliegenden Normbegriff und die Verwendung normalisierter Kennwerte eingegangen werden.

2 Zum Normbegriff

Neuropsychologische Testergebnisse werden in Form von Rohwerten erhoben. Rohwerte per se haben aber keine Aussagekraft und benötigen ein Bezugssystem (Normwerte), innerhalb dessen die Testleistungen einer einzelnen Person eingeordnet werden können (z.B. als unter- oder überdurchschnittlich). Dafür werden Rohwerte in Prozentrangwerte und in normalisierte Kennwerte (Standardwerte wie z-Werte oder T-Werte) transformiert, die einerseits einen Vergleich mit der Referenzpopulation und andererseits einen Vergleich von Testwerten untereinander zulassen. Die gängigsten Normskalen sind Variabilitätsnormen, Äquivalenznormen und Prozentränge (Fisseni, 2004). Während in der neuropsychologischen Diagnostik erwachsener Personen meist soziale Bezugsnormen (Variabilitätsnormen) sinnvoll sind, werden in der Kinder- und Jugendneuropsychologie zur Bestimmung des Entwicklungsalters (im Vergleich zum tatsächlichen Alter) zum Teil auch Äquivalenznormen beigezogen. Soziale Bezugsnormen stossen im Bereich der Erwachsenenneuropsychologie dann an Grenzen, wenn z.B. bei älteren Patienten die Fahreignung beurteilt werden soll. Hier versagt der Vergleich mit Gleichaltrigen, da die statistische Norm (z.B. Referenzgruppe der 90-Jährigen) nicht entscheidend sein darf in der Frage, wie sicher eine ältere Person sich im Strassenverkehr bewegt. Stattdessen muss hier ein kriteriumsorientierter Ansatz verfolgt werden (Könnerschaft oder Mastery). Der klinisch tätige Neuropsychologe sollte sich bei seiner Arbeit dessen bewusst sein, welcher Normbegriff einer gegebenen Fragestellung am besten gerecht wird.

3 Testinstrumentarium und Normierung

Die Auswahl des geeigneten Testinstrumentariums zur Beantwortung der diversen neuropsychologischen Fragestellungen soll nicht Gegenstand dieser Richtlinien sein. Generell ist jedoch zu empfehlen, dass – wenn immer möglich – standardisierte und gut nor-

mierte Testverfahren zur Anwendung kommen, welche die methodischen Ansprüche hinsichtlich Testgütekriterien erfüllen. Zudem sollten die Verfahren dem Schweregrad und der Art der Störung angepasst sein und hypothesengeleitet eingesetzt werden. Sofern verfügbar sind sogenannte co-normierte Testverfahren zu bevorzugen (Normierung an ein und derselben Normstichprobe). Diese erlauben eine eindeutigere und einzelfallstatistisch abgesicherte Identifikation von (differenziellen) Defiziten bzw. von Leistungsdiskrepanzen zwischen verschiedenen Testleistungsbereichen (vgl. z.B. Unter- teilung in „classical dissociation“, „strong dissociation“, „trend dissociation“ nach Shallice, 1988). Leistungsunterschiede können dahingegen nicht eindeutig interpretiert werden, wenn sie mit Testverfahren erfasst wurden, die an unterschiedlichen Normstichproben mit verschiedener Größe und Qualität aus sich unterscheidenden Kulturkreisen normiert wurden.

3.1 Berücksichtigung von Alters-, Geschlechts- und Bildungseffekten

Bei neuropsychologischen Untersuchungsverfahren besteht meist eine Abhängigkeit der Testleistung von demographischen Variablen wie Alter, Geschlecht und Bildungsgrad. Sofern verschiedene Testformen für die Verlaufsuntersuchung zur Verfügung stehen, kann es auch vorkommen, dass diese sich in ihrem Schwierigkeitsgrad unterscheiden. Falls möglich und sinnvoll in Bezug auf die Fragestellung sollten jeweils diesbezüglich differenzierte Gruppennormen beigezogen bzw. bei Verwendung von Gesamtnormen eine regressionsbasierte Adjustierung der Rohwerte vorgenommen werden. Bei Fragestellungen im Zusammenhang mit der Intelligenzbestimmung im Erwachsenenalter und bei Fahreignungsabklärungen ist es dahingegen wichtig, dass der Stichprobenvergleich alters-, geschlechts- und bildungsunabhängig durchgeführt wird.

3.2 Berücksichtigung des Differenzierungsgrades von Testkennwerten

Nicht alle Testkennwerte eines Untersuchungsverfahrens oder einer Testsammlung weisen den gleichen Differenzierungsgrad auf. Die Bereiche der in der Normierung tatsächlich erreichten Testwerte können unterschiedlich umfangreich sein. Insbesondere bei Testkennwerten, die bestimmte Formen von Fehlern erfassen oder die von gesunden Probanden leicht zu lösen sind, ist dieser Wertebereich stark eingeschränkt. Für diese Testkennwerte stehen in der Regel nur Grobnormen oder Cutoff-Werte zur Verfügung, deren Über- bzw. Unterschreiten auf beeinträchtigte Leistungen hinweisen. Andere Testkennwerte, mit in der Normierung breiter gestreuten Wertebereichen, lassen differenziertere Normen (sogenannte Feinnormen) zu. Bei elaborierten Testverfahren wird der Anwender meist über die empfohlene Standardnorm (z-Werte, T-Werte, C-Werte, Cutoff) auf den Differenzierungsgrad der verschiedenen Testkennwerte hingewiesen.

4 Diagnostische Einordnung der Untersuchungsergebnisse

Nach einer der Fragestellung angepassten Auswahl der diagnostischen Verfahren und der Durchführung der Tests sollten die erhobenen Daten nach dem neuesten Stand der Normen, Standards und des professionellen Wissens analysiert und interpretiert werden. Beschreibende, vergleichende und interpretierende Elemente sollten dabei voneinander getrennt behandelt werden (vgl. „Richtlinien für den diagnostischen Prozess“ von Westhoff et al., 2003). Der weiter oben erwähnte qualitative Ansatz müsste dementsprechend eher als beschreibendes Element bezeichnet werden, das aber ebenso in die Interpretation der Ergebnisse eingehen soll wie der vergleichende Ansatz der Testdiagnostik. Bevor Untersuchungsergebnisse oder Elemente einer Verhaltensbeobachtung interpretiert werden, erfolgt meist eine Klassifikation der Ergebnisse. Auf diese wird im folgenden Unterabschnitt eingegangen.

4.1 Klassifikation neuropsychologischer Testergebnisse

Klassifikationen können in der Neuropsychologie einerseits basierend auf unterschiedlichen Störungsbildern vorgenommen werden und andererseits auf der Basis von Leistungsunterschieden (vgl. z.B. Goldstein & Shelly, 1987). Diese Leitlinien behandeln das Thema der Klassifikation neuropsychologischer Testergebnisse.

Bei dem der Variabilitätsnorm zugrundeliegenden Begriff der statistischen Norm besteht Einigkeit darüber, dass ein bestimmter Bereich um den Mittelwert einer Gausschen Normalverteilung per definitionem als „normal“ bezeichnet werden kann. Es besteht unter Neuropsychologen aber weder in der Schweiz noch international ein Konsens darüber, wie gross dieser Normbereich sein soll. Eine nicht repräsentative Umfrage der Autoren ergab, dass in der Westschweiz, in Frankreich und in Italien die Normgrenze mehrheitlich bei Prozentrang 5 gezogen wird, wie dies in der Literatur z.B. von Benton et al. vertreten wird (Benton, Van Allen, Hamsher, & Levin, 1978; Benton, Varney, & Hamsher, 1977). Im deutschsprachigen sowie im englischsprachigen Raum scheint sich dahingegen die Normgrenze bei Prozentrang 16, also eine Standardabweichung unter dem Populationsmittelwert, eingebürgert zu haben. Diese Normgrenze dürfte v.a. auf die weite Verbreitung der Halstead-Reitan Battery und derer Normen zurückzuführen sein (Heaton, Grant, & Matthews, 1991; Heaton, Miller, Taylor, & Grant, 2004). Darüber hinaus findet teilweise auch Prozentrang 25 oder Prozentrang 10 als Cutoff-Wert Verwendung. In den Wechsler-Testverfahren werden traditionellerweise Testkennwerte unter Prozentrang 10 als „borderline“ und Testkennwerte unter Prozentrang 2 als „extremely low“ klassifiziert (Wechsler, 1997a, 1997b). Der mangelnde Konsens hinsichtlich Normbereich hat zur Folge, dass der Leser eines Berichtes in Bezug auf die Einordnung von Testergebnissen mit verschiedenen regionalen oder teilweise sogar klinikinternen Konventionen vertraut sein muss, was der Transparenz eines Berichtes und der Grundanforderung an einen Bericht – nämlich der Kommunikation und der Dokumentation zu dienen – abträglich ist.

Es bestehen zahlreiche Vorschläge, wie man Testergebnisse in Form von Standardwerten in Kategorien wie z.B. „unterdurchschnittlich“, „durchschnittlich“ oder „überdurchschnittlich“ einteilen kann (Brooks et al., 2011). Diese variieren von einer geringen Differenzierung mit drei Kategorien bis hin zu einer sehr ausdifferenzierten Einteilung mit neun Kategorien. Die SVNP schlägt im Sinne eines praktikablen Lösungsvorschlages im Rahmen der vorliegenden Leitlinien eine sieben-kategorielle Einteilung vor (siehe Tabelle 1). Als untere Normgrenze wird dabei Prozentrang 16 verwendet, wobei Testergebnisse zwischen $PR > 5$ bis $PR < 16$ als „unterdurchschnittlich“ klassifiziert werden und solche zwischen $PR \geq 2$ bis $PR \leq 5$ als „deutlich unterdurchschnittlich“. Eine Leistung mit $PR < 2$ (bzw. mehr als zwei Standardabweichungen unter dem Mittelwert) wird als „weit unterdurchschnittlich“ klassifiziert. Als „durchschnittlich“ wird der Bereich von einer Standardabweichung über und unter dem Populationsmittelwert bezeichnet ($PR \geq 16$ bis $PR \leq 84$). Leistungen im Bereich $PR > 84$ bis $PR < 95$ werden als „überdurchschnittlich“ klassifiziert. Als „deutlich überdurchschnittlich“ sind Ergebnisse zu bezeichnen, die im Bereich $PR \geq 95$ bis $PR \leq 98$ liegen und als „weit überdurchschnittlich“ solche, die mehr als zwei Standardabweichungen über dem Mittelwert liegen ($PR > 98$).

Tabelle 1: Klassifikation und klinische Interpretation neuropsychologischer Testergebnisse

Prozentrang	T-Wert	z-Wert	IQ	Klassifikation	klinische Interpretation
>98 - 100	>70 - 80	>2.0 - 3.0	>130 - 145	weit überdurchschnittlich	
≥95 - ≤98	≥66.4 - ≤70	≥1.6 - ≤2.0	≥125 - ≤130	deutlich überdurchschnittlich	
>84 - <95	>60 - <66.4	>1.0 - <1.6	>115 - <125	überdurchschnittlich	
≥16 - ≤84	≥40 - ≤60	≥-1.0 - ≤1.0	≥85 - ≤115	durchschnittlich	unauffällig
>5 - <16	>33.6 - <40	>-1.6 - <-1.0	>75 - <85	unterdurchschnittlich	möglicherweise beeinträchtigt
≥2 - ≤5	≥30 - ≤33.6	≥-2.0 - ≤-1.6	≥70 - ≤75	deutlich unterdurchschnittlich	beeinträchtigt
0 - <2	20 - <30	-3.0 - <-2.0	55 - <70	weit unterdurchschnittlich	schwer beeinträchtigt

Wichtig scheint uns zu erwähnen, dass mit den Begriffen „unterdurchschnittlich“, „deutlich unterdurchschnittlich“ und „weit unterdurchschnittlich“ nicht impliziert werden darf, dass es sich zwangsläufig um pathologische Ergebnisse handelt oder um Ergebnisse, die das Vorliegen einer Hirndysfunktion belegen.

4.2 Interpretation von Testergebnissen

Bei der Interpretation von Testergebnissen müssen verschiedenste Aspekte berücksichtigt werden. An dieser Stelle sollen lediglich einzelne methodische Aspekte kurz behandelt werden. Zu diesen gehören die Sensitivität und die Spezifität einer neuropsychologischen Untersuchung. Um Patienten mit erworbener Hirndysfunktion auch tatsächlich identifizieren zu können, sollen einerseits die Grenzen für Auffälligkeit vergleichsweise eng gesetzt werden. Andererseits ist der Diagnostiker in der Pflicht, dass neuropsychologisch unauffällige Personen (bzw. unauffällige Testleistungen) auch tatsächlich als solche erfasst werden. Um sowohl der Forderung nach Sensitivität als auch nach Spezifität gerecht zu werden, müssen gewisse psychometrische Grundprinzipien beachtet werden, auf welche z.B. Binder, Iverson & Brooks (2009) in ihrem Artikel „To Err is Human: ‚Abnormal‘ Neuropsychological Scores and Variability are Common in Healthy Adults“ eindrücklich hinweisen. Sie geben in dem Artikel zu bedenken, dass die Wahrscheinlichkeit für unterdurchschnittliche Testleistungen steigt, je mehr Testverfahren mit verschiedensten daraus resultierenden Testkennwerten eingesetzt werden. Da der Einsatz einer Vielzahl von Testverfahren im Rahmen einer neuropsychologischen Untersuchung die Regel ist und die erwähnte Wahrscheinlichkeit dadurch meist recht hoch ausfällt, sollte solchen Testsammlungen der Vorrang gegeben werden, die anhand von Base Rate-Tabellen darüber Auskunft geben, wie viele unterdurchschnittliche Testergebnisse auch bei den gesunden Personen der Normstichprobe bei einer vorliegenden Testkonstellation zu erwarten sind. Alternativ kann diese Wahrscheinlichkeit auch mit statistischen Modellen (binomiale Wahrscheinlichkeitsverteilungen, Monte Carlo-Simulationsverfahren) eingeschätzt werden.

Bei der Interpretation von Testergebnissen ist generell auf mögliche Inkonsistenzen zu achten. Hilfreich für deren Aufdeckung sind Verfahren zur einzelfallanalytischen Profilanalyse. Auch der Einsatz von Verfahren zur Symptomvalidierung kann je nach Fragestellung nötig sein. Weiter muss dringend darauf geachtet werden, ob die Untersuchungsergebnisse mit einer bekannten oder mutmasslichen zerebralen Ätiologie oder

den Daten aus der neurodiagnostischen Bildgebung vereinbar sind. Zu beachten ist hierbei, dass ein positiver neuroradiologischer Befund, wie beispielweise ein rechts frontaler Parenchymdefekt, nicht notwendigerweise auch zu einer neurokognitiven Funktionsstörung führen muss, und umgekehrt das Fehlen eines neuroradiologischen Befundes nicht notwendigerweise eine neurokognitive Störung unplausibel macht (z.B. eine Amnesie nach einer hypoxischen Hirnschädigung). Unzulässig ist es, bei fehlenden oder mangelnden anamnestischen Angaben, allein aufgrund kognitiver Minderleistungen auf eine Hirndysfunktion zu schliessen bzw. mit einer solchen zu spekulieren.

5 Beurteilung

Zum Beurteilungsteil einer neuropsychologischen Untersuchung gehört, dass die gesammelten Informationen nach ihrer Bedeutung für den Fall gewichtet und integriert werden. Die Interpretation der Daten soll dabei nicht dem Leser überlassen werden. Schlussfolgerungen sind klar darzustellen und vorläufige Schlüsse als solche zu kennzeichnen. Es empfiehlt sich in der Beurteilung eine Aufteilung in verschiedene Funktionsbereiche wie Aufmerksamkeit, Gedächtnis, exekutive Funktionen, visuelle Wahrnehmung und visuo-konstruktive Funktionen vorzunehmen und diese separat zu bewerten.

Am Ende einer Beurteilung erfolgt in der Regel eine Einschätzung des Schweregrades der neurokognitiven Funktionsstörung. Für die Einordnung des Gesamtschweregrades wird auf die Leitlinien der SVNP verwiesen (Frei, Balzer, Gysi, Leros, Plohmann & Steiger, 2016). Eine Alternative steht seit Einführung des DSM-5 (American Psychiatric Association, 2013) in Form der Diagnosen Major Neurocognitive Disorder und Mild Neurocognitive Disorder zur Verfügung. Unter Major Neurocognitive Disorder gibt es zudem die Möglichkeit einer Spezifikation in die Schweregrade Mild, Moderate und Severe. Bei Major Neurocognitive Disorder liegen die Testleistungen typischerweise zwei Standardabweichungen unter dem Populationsmittelwert, bei Mild Neurocognitive Disorder zwischen einer und zwei Standardabweichungen. Die in der Tabelle 1 vorgeschlagenen Richtlinien stimmen damit überein. In den entsprechenden Kapiteln des DSM-5 wird mehrfach betont, dass die Beurteilung im Kontext des prämorbidem Leistungsniveaus zu erfolgen hat. Es muss also eine Leistungsminderung stattgefunden haben, um diese Diagnosen stellen zu können. Diese Forderung ist selbstverständlich nicht haltbar bei der Beurteilung von Entwicklungsstörungen im Bereich der Kinder- und Jugendneuropsychologie oder bei Vorliegen mehrerer Ereignisse in zeitlicher Abfolge. Die Leistungsdefizite haben dann eher im Kontext des Entwicklungsalters (Äquivalenznormen) bzw. der sozialen Bezugsnormen zu erfolgen. Dabei stellt die Beurteilung von Patienten mit sehr hohem und sehr niedrigem Bildungs-/Intelligenzniveau, ebenso wie die Beurteilung von Patienten aus einem anderen Kulturkreis bzw. Sprachraum, eine besondere Herausforderung dar.

Quellen

- Benton, A. L., Van Allen, M. W., Hamsher, K. S., & Levin, H. S. (1978). Test of facial recognition, form SL. Iowa City, IA: Department of Neurology, University of Iowa Hospitals and Clinics.
- Benton, A. L., Varney, N. R., & Hamsher, K. S. (1977). Judgment of line orientation manual. Iowa City, IA: Department of Neurology, University of Iowa Hospitals and Clinics.
- Binder, L. M., Iverson, G. L., & Brooks, B. L. (2009). To err is human: "Abnormal" neuropsychological scores and variability are common in healthy adults. *Archives of Clinical Neuropsychology*, 24, 31-46.
- Brooks, L. B., Sherman, E. M. S., Iverson, G. L., Slick, D. J. & Strauss E. (2011). Psychometric Foundations for the Interpretation of Neuropsychological Test Results. In: *The little Black Book of Neuropsychology*, Schoenberg, M. R. & Scott, J. G. (Eds). Springer, New York.
- DSM-5 (2013). *Diagnostic and Statistical Manual of Mental Disorders. Fifth Edition*. American Psychiatric Association, Washington, DC.
- Fisseni, H.-J. (2004). *Lehrbuch der psychologischen Diagnostik. Mit Hinweisen zur Intervention. Ausgabe 3*. Hogrefe Verlag GmbH, Göttingen.
- Frei, A., Balzer, Ch., Gysi, F., Leros, J., Plohmann, A. & Steiger, G. (2016). Kriterien zur Bestimmung des Schweregrades einer neuropsychologischen Störung sowie Zuordnungen zur Funktions- und Arbeitsfähigkeit. *Zeitschrift für Neuropsychologie*, 27[2], 107-119.
- Goldstein, G., & Shelly, C. (1987). The classification of neuropsychological deficits. *Journal of Psychopathological and Behavioral Assessment*, 9, 183-202.
- Heaton, R. K., Grant, I., & Matthews, C. G. (1991). *Comprehensive norms for an extended Halstead-Reitan Battery: Demographic corrections, research findings, and clinical applications*. Odessa, FL: Psychological Assessment Resources.
- Heaton, R. K., Miller, S. W., Taylor, M. J., & Grant, I. (2004). *Revised comprehensive norms for an expanded Halstead-Reitan Battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults professional manual*. Lutz, FL: Psychological Assessment Resources.
- Shallice, T. (1988). *From neuropsychology to mental structure*. New York, NY: Cambridge University Press
- Wechsler, D. (1997a). *Wechsler Adult Intelligence Scale: Third Edition*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1997b). *Wechsler Memory Scale: Third Edition*. San Antonio, TX: The Psychological Corporation.
- Westhoff, K., Hornke, L.F., Westmeyer, H. (2003). *Richtlinien für den diagnostischen Prozess - Zur Diskussion gestellt*. [Deutsche Adaptation von: Fernandez-Ballesteros, R, De Bruyn, E.E.J., Godoy, A., Hornke, L.F., Ter Laak, J., Vizcarro, C., Westhoff, K., Westmeyer, H., & Zaccagnini, J.L. (2001). *Guidelines for the Assessment Process (GAP): A Proposal for Discussion*. *European Journal of Psychological Assessment*, 17, 187-200.] *Report Psychologie*, 28 (9), 504-517.