

Bedeutung neuropsychologischer Beschwerdvalidierung für die Beurteilung der funktionellen Leistungsfähigkeit bzw. Arbeitsfähigkeit in der versicherungsmedizinischen Begutachtung

Inauguraldissertation
zur
Erlangung der Würde
einer Doktorin der Philosophie
vorgelegt der
Fakultät für Psychologie
der Universität Basel

von

Andrea Maria Plohmann, Dipl.- Psych.

aus Lörrach (Deutschland)

Basel, April 2017



Genehmigt von der Fakultät für Psychologie

auf Antrag von

Prof. Dr. rer. nat. Dipl.-Psych. Rolf-Dieter Stieglitz
Prof. Dr. med. Ralph Mager

Basel, den 19.09.2017

Prof. Dr. phil. Roselind Lieb



ERKLÄRUNG ÜBER DIE SELBSTÄNDIGKEIT

Die zur Promotion eingereichten Zeitschriftenbeiträge wurden in Zusammenarbeit mit den jeweiligen Koautoren angefertigt. Es handelt sich um eine Originalarbeit, einen Übersichtartikel und eine Leitlinie, die zur Veröffentlichung an eine Zeitschrift eingereicht, bzw. bereits veröffentlicht wurden. Es wurden nur die angegebenen Quellen und Hilfsmittel verwendet und die entnommenen Gedanken und Formulierungen gekennzeichnet. Die empirische Studie wurde vom Forschungsfond der SUVA finanziell unterstützt; ein Interessenskonflikt bestand jedoch nicht.

Die folgenden Publikationen sind in die vorliegende Dissertation eingegangen und in der unten aufgeführten Reihenfolge im Anhang A-1 bis A-3 angefügt:

Publikation 1

Plohmann, A.M., Hurter M. (eingereicht). Prevalence of poor effort and malingered neurocognitive dysfunction in litigating patients in Switzerland. *Zeitschrift für Neuropsychologie*

Publikation 2

Keppler, C., Plohmann, A.M., Pflüger, M., Rabovsky, K., Langewitz, W., & Mager, R. (2017). Beschwerdenvalidierung in der versicherungsmedizinischen Begutachtung. *Fortschritte der Neurologie und Psychiatrie*, 85, 17-33.

Publikation 3

Frei, A., Balzer, C., Gysi, F., Leros, J., Plohmann, A. M., & Steiger-Bächler, G. (2016). Kriterien zur Bestimmung des Schweregrades einer neuropsychologischen Störung sowie Zuordnungen zur Funktions- und Arbeitsfähigkeit. *Zeitschrift für Neuropsychologie*, 27 (2), 107-119.

Basel, April 2017

Andrea Maria Plohmann

INHALTSVERZEICHNIS

ERKLÄRUNG ÜBER DIE SELBSTÄNDIGKEIT	1
ZUSAMMENFASSUNG	3
1 EINLEITUNG	4
1.1 AUSGEWÄHLTE PROBLEME IN DER BEGUTACHTUNG	5
1.1.1 Validität der Beschwerdenpräsentation	5
1.1.2 Validität von Expertenurteilen	6
1.2 LÖSUNGSANSÄTZE	7
1.2.1 Konsistenzprüfung	7
1.2.2 Plausibilitätsprüfung	8
1.2.3 Beschwerdenvalidierungsverfahren	9
1.2.4 Kriterienkataloge für vorgetäuschte Gesundheitsstörungen	10
1.3 STATISTISCHE KENNWERTE ZUR ABSICHERUNG VON NEGATIVER ANTWORTVERZERRUNG UND AGGRAVATION	16
1.4 DIE BEDEUTUNG DER GRUNDRATE / PRÄVALENZ	17
1.5 ERHÖHUNG DER DIAGNOSTISCHEN SICHERHEIT	19
2 UNTERSUCHUNGSANSATZ UND ZIELE DER EMPIRISCHEN ARBEIT	21
3 METHODIK	23
3.1 STICHPROBE	23
3.2 UNTERSUCHUNGSINSTRUMENTE	24
3.3 STATISTISCHE ANALYSEN	25
4 ERGEBNISSE	27
4.1 BEZUG ZU SOZIODEMOGRAPHISCHEN FAKTOREN	27
4.2 BEZUG ZU NEUROLOGISCHEN STÖRUNGSBILDERN	32
4.3 BEZUG ZU PSYCHIATRISCHEN DIAGNOSEN	33
5 DISKUSSION	35
5.1 PRÄVALENZ UNZUREICHENDER ANSTRENGUNGSBEREITSCHAFT UND AGGRAVATION	35
5.2 STÖRUNGSBILDER MIT ERHÖHTER WAHRSCHEINLICHKEIT NICHT-AUTHENTISCHER BEFUNDE	36
5.3 MÖGLICHE FALSCH POSITIVE KLASSIFIKATIONEN	37
5.4 SCHWÄCHEN DER STUDIE	38
6 SCHLUSSFOLGERUNGEN	40
7 LITERATUR	45
ANHÄNGE	52

ZUSAMMENFASSUNG

Im Rahmen der vorliegenden Dissertation werden die Notwendigkeit neuropsychologischer Beschwerdenvalidierung in der versicherungsmedizinischen Begutachtung dargestellt sowie diagnostische Ansätze hierfür vorgestellt und diskutiert. Zudem liefert diese Dissertation erstmals empirisch abgesicherte Prävalenzschätzungen für unzureichende Anstrengungsbereitschaft und Aggravation in der Schweiz.

In der **ersten Publikation** (hier Publikation 3) wird eine Systematik vorgestellt, die es erlaubt, unter Einbezug aller erhältlicher Informationen aus Aktenlage, Exploration, Fremdanamnese und Verhaltensbeobachtung von den psychometrischen Befunden auf den Schweregrad neuropsychologischer Einschränkungen und damit die Funktions- bzw. Arbeitsfähigkeit eines Exploranden zu schliessen. Dies bedingt jedoch, dass die erhobenen Befunde als auch die eigenanamnestischen Angaben des Exploranden valide sind.

Die **zweite Publikation** legt auf der Grundlage einer aktuellen Literaturübersicht wissenschaftliche Konzepte und Modelle zur Vortäuschung von Gesundheitsstörungen sowie Methoden, diese aufzudecken, dar. Die Autoren zeigen auf, dass die Validierung von beklagten Beschwerden und demonstrierter Leistungseinschränkung einen wesentlichen Beitrag zu einem wissenschaftlich abgesicherten Urteil eines Gutachters liefert. Die grösste und bislang unbefriedigend gelöste Herausforderung stellt dabei die Frage nach der Intention und Bewusstheit negativer Antwortverzerrung dar.

Um im Einzelfall eine möglichst präzise Aussage über die Wahrscheinlichkeit negativer Antwortverzerrung treffen zu können, ist es neben dem Einsatz möglichst sensitiver und hochspezifischer Verfahren notwendig, die Grundrate negativer Antwortverzerrung in einer, einem Exploranden möglichst gut entsprechenden, Vergleichspopulation zu kennen. Für Patientengruppen in der Schweiz existierten hierzu bislang keine Daten. Diese Lücke schliesst die hier vorgelegte **empirische Arbeit (Publikation 1)**. Zudem weist sie soziodemographische Faktoren und Diagnosegruppen aus, welche die Wahrscheinlichkeit unzureichender Anstrengungsbereitschaft und sicherer Aggravation signifikant erhöhen.

1 EINLEITUNG

Eine der zentralen Aufgaben von medizinischen und psychologischen Gutachtern im Bereich des Sozialversicherungs- und Haftpflichtrechts besteht in der Beurteilung der funktionellen Leistungsfähigkeit bzw. Arbeitsfähigkeit eines Exploranden.

Hierbei kommt der psychologischen Testdiagnostik eine entscheidende Bedeutung zu, konnten doch mit ihrer Hilfe eine Reihe von Konstrukten messbar gemacht werden. Zu diesen zählen die kognitive Leistungsfähigkeit, psychische Beschwerden, Persönlichkeitsmerkmale, Neigungen und Interessen, Leistungsmotivation, Stresstoleranz, Team- und Führungsfähigkeit u.a. mehr. Die Erfassung der kognitiven Leistungsfähigkeit stellt dabei bis heute eines der Kerngebiete der klinischen Neuropsychologie dar, wobei immer auch emotionale und Verhaltensänderungen in Folge einer neurologischen Erkrankung oder Hirnverletzung mit zu erheben sind. Wegen der besonderen diagnostischen Kompetenz klinischer Neuropsychologen hält das Fachgebiet auch zunehmend Einzug in die Psychiatrie, mit der Aufgabe die funktionelle Leistungsfähigkeit von Patienten mit psychiatrischen Störungsbildern möglichst präzise zu erheben. Die standardisierten, untersucherunabhängigen, auf normativen Vergleichswerten basierenden und in diesem Sinne objektiven Testbefunde stellen eine wichtige Ergänzung der Exploration und des aufgrund des klinischen Eindrucks erhobenen psychopathologischen Befundes dar. Die mit einer, in der Regel mehrstündigen, psychometrischen Diagnostik verbundene Verhaltensbeobachtung erbringt weitere nützliche Informationen als Ergänzung zum reinen Interview, welches wenig mehr als die Beurteilung des Gesprächsverhaltens erlaubt. Bei optimierter Gestaltung der testpsychologischen Untersuchung kann diese als eine Art „Arbeitsprobe“ unter den Bedingungen leichter körperlicher Arbeit und einer, in Abhängigkeit vom prämorbidem Niveau, unterschiedlicher gestaltbarer mentalen Belastung gelten (Frei, Balzer, Gysi, Leros, Plohmann et al., 2016). Gemäss der aktuellen Rechtsprechung ist es Sache des Exploranden, seine gesundheitlichen Einschränkungen nachzuweisen. Während letztere bei vordergründiger Betrachtung bei beklagten somatischen Beschwerden noch einfach zu objektivieren zu sein scheinen, ist der Gutachter im Falle

von chronischen Schmerzen, chronischer Fatigue, psychosomatischen und psychiatrischen Störungsbildern sowie bei beklagten kognitiven Störungen wie etwa nach Hirnverletzungen im Wesentlichen auf die subjektive Beschwerdenschilderung des Exploranden (im Rahmen der Exploration als auch unter Verwendung von Selbstbeurteilungsinstrumenten) und seine mittels psychometrischer Testverfahren erhobene Leistungsfähigkeit angewiesen. Als ergänzende Informationsquellen stehen dem Gutachter fremdanamnestic Angaben sowie die vom Auftraggeber überlassenen und ggf. zusätzlich eingeforderten medizinischen Unterlagen zur Verfügung. Auf die in seltenen Fällen zur Verfügung gestellten Materialien aus Observationen (z.B. Videoaufzeichnungen oder Protokolle von ausserhäuslichen Aktivitäten eines Versicherten) seitens des Auftraggebers, die zwischenzeitlich höchstrichterlich untersagt worden sind (EGMR-Urteil 61838/10 vom 18. Oktober 2016), soll in dieser Arbeit nicht näher eingegangen werden.

1.1 Ausgewählte Probleme in der Begutachtung

1.1.1 Validität der Beschwerdenpräsentation

Eigenanamnestische Angaben und testpsychologisch erhobene Befunde können nur dann zur Beurteilung der funktionellen Leistungsfähigkeit herangezogen werden, wenn deren Validität bzw. Authentizität sichergestellt ist (Frei et al., 2016; Plohmman, 2008). Mit anderen Worten muss überprüft werden, ob sich der Explorand kooperativ verhalten, seine Beschwerden in authentischer Weise geschildert und bei der Bearbeitung von Leistungstests eine ausreichende Leistungsbereitschaft gezeigt hat. Eine solche Kooperation stillschweigend vorauszusetzen, ist spätestens seit der zunehmenden Zahl an Veröffentlichungen, die in gutachterlichen Kontexten redundant und unter Verwendung verschiedener methodischer Ansätze bedeutsame Raten nicht-authentischer Beschwerdendarstellung und unzureichender Leistungsmotivation bis hin zur Aggravation/Simulation nachgewiesen haben, nicht mehr vertretbar (Bush, et al., 2005; Bush, Heilbronner und Ruff, 2014; Heilbronner, et al., 2009). Nicht nur die Beschwerdenschilderung im Interview, sondern auch die gezeigten Testleistungen sind von Laien ohne besondere Mühe, Sach-

kenntnis oder Vorbereitung verfälschbar. Gerade weil es dem Exploranden obliegt, seine vorhandenen gesundheitlichen Einschränkungen nachzuweisen, befindet er sich in der paradoxen Situation, dass er einerseits bestmöglich kooperieren, d.h. sich bei der Bearbeitung von Testverfahren maximal anstrengen soll, genau dadurch aber riskiert, dass insbesondere leichtere Einschränkungen übersehen werden könnten. Hier ist es also Aufgabe des Gutachters den Exploranden zu einer optimalen Mitarbeit zu überzeugen. Negative Antwortverzerrungen sind ebenso bei Selbstbeurteilungsinstrumenten zu erwarten. Diese zu erkennen stellt angesichts der Vielzahl verzerrender Antwortstile eine besondere Herausforderung dar. Schätzungen unzureichender Anstrengungsbereitschaft und übertriebener Beschwerdenschilderung reichen in Abhängigkeit von der untersuchten Population, dem Untersuchungskontext, den eingesetzten Verfahren und der Art der Klassifikation von 2.4% ausserhalb eines medico-legalen Kontexts untersuchten demennten Patienten (Mittenberg, Patton, Canyock, & Condit, 2002), 40 ± 10 % bei Personen mit leichter traumatischer Hirnverletzung (Larrabee, 2003) bis hin zu 89.5% bei kriminologisch-forensischen Begutachtung (Ardolf, Denney, & Houston, 2007).

1.1.2 Validität von Expertenurteilen

Unter vielen Ärzten und Psychologen im deutschen Sprachraum hält sich bis heute die längst als widerlegt zu betrachtende Annahme, dass eine Beurteilung der Glaubwürdigkeit der Angaben und des Verhaltens eines Exploranden alleine auf der Grundlage des klinischen Eindrucks und oder von Testbefunden sowie einer Prüfung der Konsistenz zwischen verschiedenen Datenquellen zu leisten sei (Plohmann, 2013a). Einige Autoren warnen sogar explizit vor „einem generellen und regelmäßigen Einsatz von BVT [Beschwerdenvalidierungsverfahren] im Bereich der psychiatrischen Begutachtung“ (Dressing, Widder, & Förster, 2010). Dabei mehren sich spätestens seit den 70er Jahren des vorigen Jahrhunderts empirische Belege, dass Gutachter ebenso wenig wie Laien in der Lage sind, die Glaubhaftigkeit gemachter Aussagen allein aufgrund eines klinischen Urteils mit der notwendigen Sicherheit zu beurteilen (Rosenhan, 1973; Bartlett, Littlewort, Frank, & Lee, 2014; Miller, 2005). Es konnte wiederholt gezeigt werden, dass die Treffsicherheit der klinischen Urteile die Ratewahrscheinlichkeit nicht überschreitet (Aamondt &

Custer, 2006, Ekman & O'Sullivan, 1991). Wie Bond und DePaulo (2008) in ihrer Metaanalyse zeigen konnten, ist die auch Treffsicherheit einzelner Experten nicht höher als diejenige ganzer Gruppen aus diversen Professionen.

Doch auch die isolierte Analyse von Testprofilen reicht zur Validierung erhobener Befunde nicht aus. Trueblood und Binder (1997) untersuchten, wie gut Neuropsychologen Testprotokolle von tatsächlichen Simulanten von solchen von Patienten nach schwerer Hirnschädigung voneinander differenzieren konnten. Bis zu 25 % der Psychologen hielten die Testdaten eines Simulanten für diejenigen eines authentischen Hirnverletzten. 8 % der Psychologen nahmen eine falsch positive Zuordnung vor, hielten also die Daten von bona-fide Patienten für die eines Simulanten.

1.2 Lösungsansätze

1.2.1 Konsistenzprüfung

Eine systematische Prüfung der Konsistenz stellt eine notwendige, jedoch keinesfalls hinreichende Methode zur Beschwerdenvalidierung dar. Durch sie kann zwar eine höhere Zuverlässigkeit erreicht werden, doch auch sie bleibt hinter dem, angesichts der mit dem gutachterlichen Urteil verbundenen Konsequenzen für das betroffene Individuum und die Versicherungsgemeinschaft, an die Beurteilungsgenauigkeit zu forderndem Grad an diagnostischer Sicherheit zurück. Eine *Konsistenzanalyse* prüft die Übereinstimmung von Aussagen eines Exploranden über sich selbst. Merten und Dohrenbusch (2016) unterscheiden dabei zwischen einer logischen und einer statistischen Konsistenzprüfung. Die *logische Konsistenzprüfung* hat zum Ziel eventuelle widersprüchliche Aussagen über geltend gemachte Beschwerden innerhalb einer bestimmten Dimension (z.B. depressive Stimmung) aufzudecken. Scheinen Widersprüche evident, so hat der Gutachter sorgfältig zu prüfen, inwieweit sich widersprüchliche Aussagen wirklich auf dieselbe Eigenschaft bzw. Dimension beziehen und ob sich divergente Schilderungen gegebenenfalls aus der Untersuchungssituation oder dem Fragekontext heraus erklären lassen (Hennigsen & Schickel, 2016, S.322; Merten, 2014, S. 24). Da die spezifische Interaktion zwischen Explorand und

Gutachter vielen, auch zufälligen Einflüssen, unterworfen ist, besteht ein erhöhtes Risiko vermeintlicher Inkonsistenzen. Zudem ist diese Form der Konsistenzprüfung in erheblichem Masse von den Beurteilungskriterien des Gutachters abhängig und unterliegt somit einem subjektiven Bias. Die *statistisch abgesicherte Konsistenzprüfung* hingegen basiert auf dem Einsatz psychologischer, auf klassischer Testtheorie basierender Testverfahren, die nachweislich gleiche oder ähnliche Merkmale erfassen. Diese sichern im Falle hoher Testgüte (Objektivität, Reliabilität, Validität) gegen Zufallseinflüsse bei der Beurteilung von Beschwerden, Persönlichkeitseigenschaften, Einstellungen oder Fähigkeiten ab. Besteht in zwei nachweislich konvergent validen (überzufällig übereinstimmenden) Verfahren eine überzufällige (statistisch signifikante) Messwertdifferenz, so kann die Inkonsistenz der Aussagen als statistisch belegt gelten. Das beweist noch nicht, dass der Explorand seine Beschwerden schlimmer darstellt, als es der Realität entspricht, legt jedoch eine bewusstseinsnahe Verzerrung der Antworten nahe. Umgekehrt kann eine konsistente Merkmalsausprägung in zwei konvergent validen Verfahren nicht als Beleg für eine authentische Beschwerdendarstellung gelten. Im Falle einer bewusstseinsnahen Täuschungsabsicht ist eine konsistente Zurücknahme des eigenen Leistungsvermögens über mehrere Leistungstests hinweg einfach zu bewerkstelligen und ohne spezifische Validitätsparameter häufig nicht nachweisbar. Analoges gilt angesichts der hohen Augenscheinvalidität vieler Fragebögen auch für Selbstbeurteilungsinstrumente.

1.2.2 Plausibilitätsprüfung

Zur Validierung beklagter oder demonstrierter Beschwerden gehört auch eine gründliche Plausibilitätsprüfung und kann erfahrungsbasiert und/oder modellbasiert erfolgen (Merten & Dohrenbusch, 2016). Der Rückgriff auf individuelles *Erfahrungswissen* und persönliche Überzeugungen oder Wertungen des Gutachters dürften in der Praxis zu den häufigsten Validierungsstrategien gehören. Obwohl diese Strategien weder prinzipiell falsch noch vollkommen ungeeignet sind, liefern sie doch den schwächsten Beitrag zur Beschwerdenvalidierung. Ihnen wohnt ein hohes Risiko perspektivischer Verzerrung und fehlender Aussenkontrolle sowie die Gefahr von Zirkelschlüssen im Sinne einer „Validierung“ der eigenen Bewertungskriterien inne.

Auf die Problematik der gerade bei erfahrenen Gutachtern, aber auch Rechtsanwendern, häufig anzutreffende Überzeugung, sie/diese seien aufgrund ihrer Routine in der Lage, aggravierende Probanden im Rahmen einer klinischen Untersuchung zu identifizieren, wurde bereits hingewiesen. Die *modellbasierte Prüfung der Plausibilität* bezieht sich auf die Übereinstimmung von Angaben oder Verhaltensweisen eines Exploranden mit Aussenkriterien und setzt somit einen theoretischen Bezugsrahmen voraus. Die Plausibilität geklagter Beschwerden kann durch empirisch gestützte Verteilungsannahmen (z.B. Häufigkeit bestimmter Beschwerden in einer bestimmten Patientenpopulation), Wahrscheinlichkeitsaussagen (z.B. über zusammen auftretende Symptome, Störungsgenese), Funktionszusammenhänge (z.B. von Krankheitsfaktoren und Leistungsmerkmalen) und Kausalzusammenhänge (z.B. lokale körperliche Verletzung u. lokales Schmerzempfinden beim akuten Schmerz, nicht aber beim chronischen Schmerz) bestimmt werden. Selbst hier fließen immer (auch implizite) Annahmen des Gutachters mit ein, die letztlich zu Fehlbeurteilungen führen können. Mehr Objektivität kann durch den Einsatz von psychometrischen Validierungsverfahren erreicht werden, bei deren Konstruktion solche Erkenntnisse mit einfließen. Durch eine sorgfältige Prüfung der Validität sowohl vom Exploranden geschilderter Beschwerden als auch dessen kognitiver und je nach Fragestellung auch körperlicher Leistungsfähigkeit mittels Beschwerdenvalidierungsverfahren, kann die diagnostische Sicherheit im Hinblick auf die anzunehmende Authentizität der beklagten Gesundheitsstörungen wesentlich erhöht werden. Ein entscheidender Vorteil besteht darin, dass der Einsatz dieser Verfahren eine zufalls-kritisch abgesicherte Aussage mit bekanntem Grad an Irrtumswahrscheinlichkeit erlaubt, eine Grösse die bei rein klinischer Beurteilung und Konsistenz- wie Plausibilitätsprüfung niemals bekannt ist.

1.2.3 Beschwerdenvalidierungsverfahren

Aufgrund der starken Verflechtung von Neuropsychologie und Testdiagnostik sahen sich insbesondere nordamerikanische Neuropsychologen seit Ende der 1980er Jahre motiviert, Methoden zur Erkennung von Täuschungsversuchen und ungültiger Testprofile zu entwickeln. Hierzu gehörten auch konzeptionelle Fragen in Zusammenhang mit Simulation, Aggravation und deren

Abgrenzung zu artifiziellen Störungen sowie somatoformen und dissoziativen Störungen. Neben der Entwicklung von Performanzvalidierungsverfahren (PVT) zur Überprüfung einer ausreichenden Anstrengungsbereitschaft und damit der Authentizität von Leistungsprofilen, wurden auch Verfahren zur Validierung der Beschwerdenschilderung, sog. Symptomvalidierungsverfahren (SVT), entwickelt. Gemäss einer Analyse von Sweet und Guidotti Breting (2013) gehört die Beschwerdenvalidierung zu den am intensivsten beforschten Bereichen angewandter Neuropsychologie. 90% der Erstautoren von über 2300 erschienenen Artikel waren Psychologen, die meisten davon Neuropsychologen.

Die verschiedenen methodischen Zugänge zur Beschwerdenvalidierung, welche im deutschsprachigen Raum erhältlich sind, werden von der Autorin in der gemeinsam mit Keppler verfassten und inzwischen publizierten Übersichtsarbeit (Keppler et al., 2017) vorgestellt. Diese Ansätze und Verfahren zur Beschwerdenvalidierung liefern Ergebnisse, auf deren Grundlage sich die oben beschriebenen Konsistenz- und Plausibilitätsprüfungen auf empirischer Grundlage, d.h. mit bekannter Irrtumswahrscheinlichkeit realisieren lassen.

1.2.4 Kriterienkataloge für vorgetäuschte Gesundheitsstörungen

Für einige häufig vorgetragenen, aber auch häufig vorgetäuschten Gesundheitsstörungen wurden inzwischen Klassifikationssysteme entwickelt, die eine Unterscheidung zwischen authentischen und vorgetäuschten Beeinträchtigungen von Körperfunktionen und sich daraus ergebender funktioneller Leistungsfähigkeit erlauben sollen. Vorreiter waren Slick, Sherman und Iverson, die 1999 einen ersten Algorithmus zur Diagnose vorgetäuschter neurokognitiver Störungen (Malingered Neurocognitive Disorder, MND) vorstellten. Dieser erfuhr mit seinen von Larrabee, Greiffenstein, Greve und Bianchini (2007) vorgeschlagenen Modifikationen international breite Akzeptanz und rege Anwendung bei Neuropsychologen, die sich mit dem Problem negativer Antwortverzerrung auseinandersetzen.

Tabelle 1: Kriterien für die Diagnose einer vorgetäuschten neurokognitiven Störung (MND) von Slick, Sherman und Iverson (1999)

Kriterium	Spezifikation
A. Existenz eines äusseren Anreizes für Täuschungsverhalten	Eingangskriterium, das für die Feststellung eines Täuschungsverhaltens (Simulation oder Aggravation) erfüllt sein muss.
B. Belege für Antwortverzerrung aus der Leistungsdiagnostik	B1. Ergebnis in Alternativwahlverfahren unter Zufallsniveau B2. Weitere auffällige Ergebnisse in PVT B3. Diskrepanzen zwischen Testdaten und bekannten Mustern von normalen und pathologischen Hirnfunktionen B4. Diskrepanzen zwischen Testdaten und beobachtbarem Verhalten B5. Diskrepanzen zwischen Testdaten und zuverlässigen Information von Dritten B6. Diskrepanzen zwischen Testdaten und Informationen aus den Akten / der Krankengeschichte
C. Belege für Antwortverzerrung aus den eigenanamnestischen Angaben und/oder aus Selbstbeurteilungsinstrumenten	C1. Diskrepanzen zwischen eigenanamnestischen Angaben und Informationen aus den Akten /der Krankengeschichte C2. Diskrepanzen zwischen der Beschwerdenschilderung und bekannten Mustern von normalen und pathologischen Hirnfunktionen C3. Diskrepanzen zwischen der Beschwerdenschilderung und dem beobachtbaren Verhalten C4. Diskrepanzen zwischen der Beschwerdenschilderung und zuverlässigen Informationen Dritter C5. Hinweise auf übertriebene oder erfundene Darstellung von psychischen Beschwerden aus gut validierten SVT - Skalen
D. Ausschluss anderer Ursachen, welche das Täuschungsverhalten erklären könnten	Die unter B und C aufgeführten Hinweise können nicht vollständig durch psychiatrische, neurologische oder Entwicklungsfaktoren erklärt werden.

Tabelle 2: Diagnostische Sicherheit gemäss Slick-Kriterien

Sicherheitsgrad einer vorliegenden MND	Zu erfüllende Kriterien
Sicher	Kriterien A, B1 und D
Wahrscheinlich	Kriterien A und D sowie <ul style="list-style-type: none"> • mindestens zwei der Kriterien B2-B6 oder • eines der Kriterien B2-B6 und mindestens eines der Kriterien C1-C5
Möglich	<ol style="list-style-type: none"> a. Kriterien A und D sowie mindestens eines der Kriterien C1-C5 oder b. Kriterien für eine sichere oder wahrscheinliche MND sind erfüllt, aber Kriterium D ist nicht erfüllt

Larrabee et al. (2007) schlugen u.a. vor, dass bei mehrfachen auffälligen PVT-Ergebnissen (Kriterium B2) kein weiteres Kriterium für die Feststellung einer wahrscheinlichen MND notwendig sein sollte. Seither verwendeten die meisten Studien mit Known-groups- oder Bootstrapping-Designs diese Modifikation bei der Klassifikation ihrer Stichproben, so auch die Verfasserin dieser Dissertation (Plohmann, eingereicht).

Nach Rogers (2008) werden bei *Known-groups Vergleichen* die Probanden im Rahmen einer umfassenden Untersuchung anhand möglichst trennscharfer Aussenkriterien durch ausgewiesene und unabhängige Experten in die Gruppen authentischer Patienten und solcher mit verzerrter Beschwerdenpräsentation eingeteilt. In einem zweiten Schritt werden diese Probanden mit den eigentlich interessierenden Verfahren untersucht oder andere erhobene Daten im Hinblick auf Gruppenunterschiede analysiert, ohne dass die involvierten Versuchsleiter die Gruppenzugehörigkeit der jeweiligen Probanden kennen. Bei diesem Untersuchungsansatz besteht die Gefahr von Extremgruppenvergleiche und damit eingeschränkter Generalisierbarkeit der Ergebnisse. Beim *Bootstrapping-Design* erfolgt die Gruppeneinteilung unter Rückgriff auf festgelegte Kombinationen und Trennwerte von Beschwerdenvalidierungsverfahren mit dem Ziel maximaler Spezifität und ausreichender Sensitivität. Auch die Verblindung der Untersucher entfällt häufig. In der Praxis bezeichnen allerdings viele Autoren ihre Untersuchungsanordnung als Known-

group Design, obwohl es nach Rogers Definition eher einem Bootstrapping Ansatz entspricht. Beide Ansätze bieten den Vorteil hoher externer Validität, da Studienteilnehmer, Settings, Fragestellung und Anreize realen Gegebenheiten entsprechen. Hingegen ist die interne Validität eher schwach, da die experimentellen Bedingungen nur bedingt zu kontrollieren sind.

Für die grosse Gruppe von chronischen Schmerzpatienten unter zu begutachtenden Patienten wurden von Bianchini, Greve und Glynn (2005) die MND-Kriterien um inkonsistente Befunde aus der somatischen Untersuchung erweitert (Malingered Pain Related Disability, MPRD). Beide Klassifikationssysteme definieren Kriterien, bei deren Erfüllung die Diagnose simulierter/aggravierter neurokognitiver Störungen bzw. schmerzassoziierter Behinderungen mit abgestufter diagnostischer Sicherheit, gestellt werden kann (s. Anhang). Beiden Kriterienkatalogen ist gemein, dass die Frage der Täuschungsabsicht, also der Intention bzw. Motivation indirekt aus dem Untersuchungskontext erschlossen wird. In einem medico-legalen (engl. forensic) Setting beispielsweise nehmen beide Modelle indirekt an, dass die dominierende Motivation für ein gezeigtes Täuschungsverhalten im externen finanziellen Anreiz besteht, womit gleichzeitig ein bewusstes Handeln postuliert wird. In der Realität dürfte sich der Grad der Bewusstheit von Täuschungsverhalten viel eher auf einem Kontinuum von bewusster Täuschung Anderer auf der einen Seite und „Self-deception“ (Boone, 2011) also unbewusster Beschwerdenpräsentation und Motivation auf der anderen Seite bewegen. Unterstützung erfährt dieses dimensionale Konzept von Täuschungsverhalten sowohl für kognitive als auch psychische Störungen durch taxometrische Analysen (Walters, Berry, Rogers, Payne & Granacher, 2009; Walters, et al., 2008).

Um diesem Rechnung zu tragen und damit das Risiko falsch positiver Urteile zu minimieren, aber auch um Forschungsergebnisse besser vergleichen zu können, schlug Young (2014) eine Reihe von Verbesserungen und Ergänzungen vor.

Anders als die Autoren der Kriterienkataloge für MND und MPRD versuchte Young (2014, 2015b) eine möglichst vollständige Liste denkbarer Inkonsistenzen und vorstellbarer Kombinationen von Diskrepanzen zu erstellen und diese möglichst präzise zu definieren, um den Einfluss subjektiver Annahmen von Gutachtern zu minimieren. Darüber hinaus plädiert Young dafür,

genau zu unterscheiden aus welchen Bereichen die Inkonsistenzen entstammen, und auch das Ausmass der Diskrepanzen in der diagnostischen Beurteilung mitzubersichtigen. Der hieraus resultierende Katalog an sehr spezifischen Regeln und Prozeduren für die testpsychologische Untersuchung (unter Aufzählung einsetzbarer Testarten und spezifischer Verfahren) als auch die Konsistenzprüfung ist etwa 10mal so umfangreich wie die ursprünglichen Kriterienkataloge zur Aufdeckung von MND und MPRD. Basierend auf diesen beiden Klassifikationssystemen und einem von Rubenzer (2009) vorgestellten System gewichteter Testverfahren entwickelte Young u.a. einen Kriterienkatalog zur Entdeckung vorgetäuschter Posttraumatischer Belastungsstörung (Feigned Posttraumatic Stress Disorder Disability/Dysfunction, F-PTSDR-D). Zusammen mit modifizierten Diagnosesystemen zu MND und MPRD entstand so ein allgemeines System zur Aufdeckung von vorgetäuschten, simulierten oder verzerrten Darstellungen von psychischen Störungen (Psychological Injury Disability/Dysfunction — Feigning/Malingering/Response Bias System (PID-FMR-S) im forensischen oder medico-legalen Kontext (Young, 2014). Mit seinen sehr präzisen Operationalisierungen einer Vielzahl möglicher Inkonsistenzen hat dieses System das Potential zu einer verbesserten Objektivität der Beschwerdenvalidierung beizutragen. Die empirische Überprüfung der vorgeschlagenen Diagnosesysteme im Hinblick auf Reliabilität und Validität als auch des klinischen Nutzens steht allerdings noch aus. Ein weiterer, wesentlicher Unterschied zwischen Youngs Kriterienkatalogen und den ursprünglichen MND-Kriterien von Slick, Sherman und Iverson (1999) als auch den MPRD-Kriterien Bianchini, Greve und Glynn (2005) besteht in der Empfehlung, den Terminus „Malingering“ nur auf Fälle von nachgewiesener absichtlicher Täuschung (Simulation/Aggravation) zu beschränken und bei wahrscheinlichem Response Bias, bei welchem neben dem äusseren Anreiz auch andere Faktoren bzw. Motive bestehen können, den Terminus Vortäuschung („Feigning“) zu verwenden oder gar nur von Antwortverzerrung zu sprechen. Rogers (2008) definierte „Feigning“ als absichtliche Erzeugung oder starke Übertreibung psychischer oder somatischer Beschwerden ohne irgendwelche Annahmen über Ziel und Zweck dieses Verhaltens („deliberate fabrication or gross exaggeration of psychological or physical symptoms without any assumptions about its goals“, p. 6).

Wie schon zahlreiche Autoren vor ihnen (z.B. Iverson, 2006, Rogers 2008) wiesen Young (2015b) und kürzlich auch Merten und Rogers (2017) zu Recht darauf hin, dass die Frage der Intention hinter einer Antwortverzerrung oftmals nur schwer zu ermitteln ist und in vielen Fällen eine Mischung aus unbewussten und intentionalen Motiven vorliegen dürfte.

Auch Slick und Sherman selbst haben in Reaktion auf kritische Stimmen ihr ursprüngliches Modell grundlegend revidiert (2013). Zu den Neuerungen zählt die Aufgabe des früheren D-Kriteriums zugunsten einer Unterscheidung in primäre und sekundäre Aggravation/Simulation. Sind also die Kriterien A-C für eine wahrscheinliche Aggravation/Simulation neurokognitiver Störungen erfüllt, gilt es zu prüfen, inwieweit der Explorand über die kognitiven Fähigkeiten verfügte, um die moralischen, ethischen und rechtlichen Implikationen seines Verhaltens zu verstehen, und/oder fähig war, sein Verhalten zu kontrollieren. Im positiven Fall würde eine absichtliche (primäre) Simulation/Aggravation angenommen, im anderen Fall eine sekundäre Simulation/Aggravation diagnostiziert.

Änderungen wurden auch für die Definition definitiver und wahrscheinlicher Aggravation/Simulation neurokognitiver Störungen vorgeschlagen. Neben Testleistungen unter Zufallsniveau in mindestens einem Forced-choice-Verfahren gilt neu auch eine hohe A-posteriori-Wahrscheinlichkeit ($\geq 0,95$) in einem oder mehreren gut validierten psychometrischen Indizes aus Selbstbeurteilungsinstrumenten (z. B. Validitätsskalen des MMPI-2) als Beleg für intentionale Aggravation/Simulation. Ziel war es, den Kriterienkatalog auch für Fälle anwendbar zu machen, in denen primär psychologische Symptome übertrieben oder vorgetäuscht werden, während für Befunde aus Leistungstests keine Aggravation/Simulation nachzuweisen ist. Der Nachweis einer hohen A-posteriori-Wahrscheinlichkeit bedingt jedoch das Vorliegen zuverlässiger Grundraten zur Auftretenswahrscheinlichkeit von Aggravation/Simulation in einer dem Exploranden vergleichbaren Population (A-priori-Wahrscheinlichkeiten) und einen gut validierten Prozess zur Bestimmung des positiven prädiktiven Wertes. Auch beklagte Beschwerden, die mit dem beobachteten Verhalten und/oder den Testleistungen eindeutig nicht vereinbar sind oder in direktem Wider-

spruch dazu stehen („compelling inconsistencies“), sollen nun als eindeutige Indikatoren für Aggravation/Simulation gelten können. Schliesslich wurde auch die Kategorie „mögliche“ Aggravation/Simulation bei Vorliegen lediglich eines auffälligen Parameters aufgegeben. Das grundsätzliche Problem, auffällige Validierungsparameter bei gegebenem äusseren Anreiz grundsätzlich und ausschliesslich als Beleg für absichtliche Antwortverzerrung zu werten, bleibt jedoch trotz der Modifikationen bestehen.

1.3 Statistische Kennwerte zur Absicherung von negativer Antwortverzerrung und Aggravation

Bei der Konstruktion psychometrischer Testverfahren und besonders bei Performanz- und Beschwerdenvalidierungsverfahren wird versucht, ein Maximum an *Sensitivität und Spezifität* zu erzielen. Die *Sensitivität* eines Verfahrens bringt zum Ausdruck, wie gut ein Test Personen, die tatsächlich Beschwerden vortäuschen (true positives) erfassen kann. Die *Spezifität* hingegen bezeichnet die Genauigkeit, mit der ein Verfahren authentische Patienten (true negatives) identifiziert. Je höher die Sensitivität, die für einen bestimmten Trennwert eines Verfahrens ermittelt wurde, desto wahrscheinlicher ist es, dass ein negatives Testresultat ein authentisches Antwortverhalten widerspiegelt. Im Gegensatz dazu geht eine höhere Spezifität eines Trennwerts mit einer höheren Wahrscheinlichkeit einher, dass ein positives bzw. auffälliges Testergebnis eine Aggravation/Simulation anzeigt.

Die *Trefferquote* (hit rate) gibt an, wie gross der Anteil korrekt klassifizierter Personen an einer untersuchten Stichprobe ist ($\text{Richtig Positive} + \text{Richtig Negative} / N$).

Ein weiterer Quotient, die sog. *Likelihood ratio* (LR) errechnet sich aus der Sensitivität / (1-Spezifität) und beschreibt die Wahrscheinlichkeit eines Individuums eine bestimmte Diagnose zu haben (z.B. MND). Die Likelihood ratio stellt eine Parallele zur Odds ratio (OR) bei Gruppenvergleichen dar.

Die *positive Vorhersagekraft* bzw. „Positive Predictive Power“ (PPP) bezieht sich auf den Anteil korrekt identifizierter, täuschender Personen (Richtig Positive) an allen Personen mit positivem

bzw. auffälligem Testergebnis (Richtig Positive / (Richtig Positive und Falsch Positive)). Umgekehrt beschreibt die *negative Vorhersagekraft* bzw. „Negative Predictive Power“ (NPP) das Verhältnis von korrekterweise als authentisch klassifizierten Testergebnissen an der Gesamtheit aller unauffälligen Testergebnisse (Richtig Negative / (Richtig Negative + Falsch Negative)).

1.4 Die Bedeutung der Grundrate / Prävalenz

Sowohl die positive als auch die negative Vorhersagekraft hängen von der Auftretenswahrscheinlichkeit der Diagnose (hier Simulation/Aggravation) in der untersuchten Stichprobe ab. Im Falle einer niedrigen Prävalenz bzw. Grundrate ist die Wahrscheinlichkeit, dass ein unauffälliges (negatives) Testergebnis eine korrekte Klassifikation widerspiegelt höher als dies für ein auffälliges (positives) Testergebnis der Fall ist. Bei einer hohen Prävalenz oder Grundrate wird sich eher ein auffälliges (positives) Testergebnis als ein negatives Testergebnis als korrekt erweisen. Dies soll an einem Beispiel verdeutlicht werden:

Wenn die Grundrate für wahrscheinliche und sichere Aggravation/Simulation im versicherungsmedizinischen Bereich wie von einigen Autoren (z.B. Larrabee, Millis & Meyers, 2009) angenommen wird, bei rund 50% liegt, dann wäre bei 500 aus 1000 Personen ein Täuschungsverhalten anzunehmen. Ein Validierungsverfahren, welches eine Spezifität von 90 % und eine Sensitivität von 80 % aufweist, wird 80 % dieser 500 Personen (= 400) korrekt als Simulanten klassifizieren, aber auch 10 % der 500 ehrlichen Personen und damit 50 Personen zu Unrecht als Simulanten einordnen.

Tabelle 3. Einfluss der Grundrate auf die Vorhersagekraft eines Beschwerdenvalidierungsverfahrens (BVT: PVT oder SVT)

Beispiel 1: N = 1000; Grundrate = 50 %

Testergebnis	Unzureichende Anstrengungsbereitschaft	
	Ja	Nein
Auffällig	400	50
Unauffällig	100	450
Positive Vorhersagekraft	88.9 %	
Negative Vorhersagekraft	81.8 %	

Bei einer niedrigen Grundrate von beispielsweise 10 %, wäre bei 100 von 1000 Personen mit wahrscheinlicher oder sicherer Aggravation/Simulation zu rechnen. Dasselbe Verfahren würde wiederum 80 % dieser Personen (80 von 100) korrekt als Simulanten identifizieren, aber eben auch 10 % der 900 ehrlichen Personen fälschlicherweise der Gruppe der täuschenden Personen zuordnen. Damit wären 90 Personen falsch positiv klassifiziert. Die Zahl korrekt und falsch klassifizierter Personen würde die Waage halten und die Quote fehlerhafter Urteile läge bei 50 %. Bei einer noch geringeren Grundrate wäre der Anteil falsch positiver Urteile noch höher.

Tabelle 4. Einfluss der Grundrate auf die Vorhersagekraft eines BVT

Beispiel 2: N = 1000; Grundrate = 10 %

Testergebnis	Unzureichende Anstrengungsbereitschaft	
	Ja	Nein
Auffällig	80	90
Unauffällig	20	810

Positive Vorhersagekraft	47.1 %	
Negative Vorhersagekraft	97.6 %	

Zeigen also in einer bestimmten Population nur 10% der zugehörigen Personen eine unzureichende Anstrengungsbereitschaft, so ist die Wahrscheinlichkeit, dass eine Person mit positivem Testergebnis im selben Verfahren tatsächlich zu dieser Gruppe gehört, nur etwa halb so hoch wie im ersten Beispiel. Weist eine Person hingegen ein negatives Testergebnis auf, so gehört sie mit sehr hoher Wahrscheinlichkeit zum anstrengungsbereiten Teil dieser Population.

Daraus lässt sich ableiten, dass eine bekannte Grundrate bzw. Prävalenz nicht-authentischer Beschwerdenschilderung und unzureichender Leistungsbereitschaft in der interessierenden Population eine notwendige Grundvoraussetzung für eine empirisch abgesicherte Diagnose von negativer Antwortverzerrung und insbesondere Aggravation bzw. Simulation darstellt.

1.5 Erhöhung der diagnostischen Sicherheit

Eine Möglichkeit, insbesondere bei geringer Grundrate die Gefahr falsch positiver Klassifizierungen zu erhöhen, besteht in der Kombination mehrerer, gut validierter und voneinander inhaltlich unabhängiger Performanzvalidierungs- und/oder Beschwerdenvalidierungsverfahren. Bei einer Spezifität der Einzelverfahren von mindestens 90% sinkt das Restrisiko einer Fehlklassifikation beim Einsatz von drei Validierungsverfahren auf $0.10 \times 0.10 \times 0.10$ (= 0.1 %), also auf max. 1 pro 1000 Personen (Prinzip der kombinierten Unwahrscheinlichkeit). Da die Anstrengungsbereitschaft während einer testpsychologischen Untersuchung keine Alles-oder-Nichts-Phänomen darstellt, sondern durchaus Schwankungen unterworfen ist, und Personen mit Täuschungsabsicht durchaus unterschiedliche Strategien verfolgen, also nicht in allen untersuchten Bereichen invalide Leistungen zeigen, empfehlen zahlreiche Autoren multiple PVT aus unterschiedlichen Funktionsbereichen in Kombination mit SVT einzusetzen. Damit steigt aber die Zahl der Verfahren, in welcher potentiell auffällige Ergebnisse erzielt werden können. Befürchtungen, wie von Berthelson, Mulchan, Odland, Miller und Mittenberg (2013) geäußert, dies

könnte zu einer vermehrten Zahl falsch positiver Klassifikationen führen, haben sich nicht bestätigt. So konnte Larrabee (2014) basierend auf Daten aus früheren Untersuchungen (2003, 2009) zeigen, dass beim Einsatz von 7 PVT und gemäss Slick-Kriterien geforderten 2 auffälligen Ergebnissen um wahrscheinliche Aggravation anzunehmen, 88.9 % der bona-fide Patienten und 97.6 % der Personen mit invaliden Befunden korrekt identifiziert wurden. Die Trefferquote lag bei 92.6 %. Erhöhte man die geforderte Zahl an auffälligen Ergebnissen auf 4 von 7 PVT, sank zwar die Zahl falsch positiver Klassifikationen auf Null, die Sensitivität nahm jedoch überproportional ab und die Trefferquote lag insgesamt nur bei 84.2 %.

2 UNTERSUCHUNGSANSATZ UND ZIELE DER EMPIRISCHEN ARBEIT

Während für die USA und Canada multiple Schätzungen der Grundraten vorliegen, gibt es im europäischen Raum kaum Erhebungen. Zwei Arbeiten aus Deutschland kam zu Schätzungen von rund 44% nicht valider Testbefunde in einer Gutachtenspopulation. Diese beiden Studien kombinierten ein oder mehrere PVT mit jeweils einem Screeninginstrument zur Validierung der Beschwerdenschilderung (Merten, Thies, Schneider und Stevens, 2009; Stevens, Friedel, Mehren und Merten, 2008). Für die Schweiz existieren (mit Ausnahme einer Präsentation von Zwischenergebnissen der hier vorgestellten Studie (Plohm, 2013b)) bislang keine empirischen Daten, sondern lediglich Schätzungen im Rahmen von Expertenbefragungen und auf der Basis von aus anderen Daten abgeleiteten Parametern. Erfragt wurde zum einen die von Gutachtern geschätzte Häufigkeit von Aggravation und Simulation, die von IV-Fachleuten und Gutachtern angenommene Häufigkeit von Versicherungsbetrug und nicht zielkonform ausgerichteter Leistungen der Invalidenversicherung. Beide dieser vom Bundesamt für Sozialversicherung in Auftrag gegebenen Studien schätzten die Häufigkeit von Täuschungsverhalten übereinstimmend auf etwa 5-7 % der Fälle.

Mit der hier präsentierten empirischen Arbeit sollte erstmalig die Auftretenshäufigkeit vorgetäuschter kognitiver Störungen bei polydisziplinären Begutachtungen von Gesundheitsstörungen bzw. Unfallfolgen, welche mit kognitiven Leistungseinbussen einhergehen können, für die Schweiz ermittelt werden. Gleichzeitig sollten mögliche Zusammenhänge zwischen nicht-authentischen Testbefunden und verschiedenen soziodemographischen Variablen bzw. neurologischen wie psychiatrischen Diagnosen erhoben werden.

Der Autorin lagen mehr als 500 Datensätze von erwachsenen Exploranden vor, welche in ihrer Praxis zwischen 2005 und Anfang 2016 im polydisziplinären Kontext neuropsychologisch untersucht wurden und dabei auch Performanzvalidierungsverfahren bearbeiteten. Von diesen Datensätzen wurden diejenigen in die Studie aufgenommen, bei welchen die Exploranden mindestens zwei aus einem Pool von sieben Performanzvalidierungsverfahren bearbeitet hatten. Es

wurde angenommen, dass bei dieser Stichprobe „*sicher* vorgetäuschte neurokognitive Störungen“ vorlagen, wenn die Zahl korrekter Antworten in mindestens einem Antwortwahlverfahren (forced-choice) unterhalb der reinen Ratewahrscheinlichkeit und somit dem Zufallsniveau lag (Kriterium B1 nach Slick et al., 1999). Es wurde weiter angenommen, dass bei dieser Stichprobe „*wahrscheinlich* vorgetäuschte bzw. aggravierte neurokognitive Störungen“ bestanden, wenn sich in mindestens zwei Performanzvalidierungsverfahren bzw. testinternen Validierungsparametern Resultate jenseits empfohlener Trennwerte für noch als authentisch zu bewertende Leistungen (aber über dem Zufallsniveau) fanden (Kriterium B2 der nach Larrabee (2007) modifizierten Slick-Kriterien), jedoch keine neurologischen oder psychiatrischen Erkrankungen oder Entwicklungsstörungen, welche das auffällige Testverhalten alternativ vollumfänglich erklären können (Kriterium D). Da es sich bei der untersuchten Stichprobe ausschliesslich um Exploranden handelte, welche Entschädigungs- oder Rentenansprüche geltend machten, war jeweils ein äusserer Anreiz gegeben und das A-Kriterium nach Slick et al. (1999) erfüllt.

Die nach ICD-10 Kriterien erstellten neurologischen und psychiatrischen Diagnosen nach ICD-10 zur Überprüfung des D-Kriteriums nach Slick et al. wurden von den jeweiligen medizinischen Fachgutachten übernommen, die in der Regel zeitgleich zur neuropsychologischen Begutachtung erstellt wurden. Die Ergebnisse der Performanz- und - der teilweise vorgenommenen - Beschwerdenvalidierung mittels MMPI-2 durch die Referentin hatten auf die Diagnosestellung der beteiligten Fachärzte in der Regel keinen Einfluss.

Da die Studie die retrospektive Analyse von patientenbezogenen Daten aus einem Zeitraum von über zehn Jahren beinhaltete, konnte das Einverständnis der Patienten nicht mit vertretbarem Aufwand nachträglich eingeholt werden. Die Ethikkommission Nordwestschweiz hat das vorliegende Forschungsprojekt dennoch bewilligt (EKNZ BASEC 2016-00659).

3 METHODIK

3.1 Stichprobe

Die insgesamt 455 Fälle umfassende Stichprobe bestand zu 48.8% aus Männern und zu 51% aus Frauen, eine Person wurde als Transgender identifiziert. Das Alter der Exploranden variierte zwischen 19 und 75 Jahren, das mittlere Alter lag bei 44.1 (\pm 11.8) Jahren. Der Bildungshintergrund reichte von knapper Grundschulbildung (4 Schuljahre) bis zu einem abgeschlossenen Hochschulstudium (>13 Jahre). Alle Exploranden konnten zumindest rudimentär lesen und schreiben. Die Mehrheit der Exploranden waren gebürtige Schweizer (61.5%), 8.1% stammten aus Deutschland oder Österreich, weitere 20.7% aus dem übrigen Europa. 5.9 % stammten aus Vorderasien (mehrheitlich der Türkei), 3.6% waren aus anderen Teilen der Welt immigriert (s.a. Plohmann, eingereicht).

408 (89,2%) dieser Exploranden erhielten mindestens eine neurologische Diagnose: cervico-spinale Probleme und/oder HWS-Distorsion ($n=208$), leichtes Schädelhirntrauma (SHT) ($n=158$), mittelgradiges SHT ($n=39$), schweres SHT ($n=29$), cerebrovaskulären Ereignisse ($n=33$), Raumforderungen ($n=8$), Epilepsie ($n=24$), Multiple Sklerose ($n=5$), mild cognitive impairment ($n=7$), Andere (Encephalopathie, cerebrale Entzündungen, Hypoxie, toxische Hirnschädigung, etc.) mit möglichem Einfluss auf die geistige Leistungsfähigkeit ($n=68$). In 27.9% aller Fälle konnte mittels Bildgebung (CT/MRT) eine substantielle Hirnschädigung nachgewiesen werden. In der gesamten Stichprobe wurden 421 psychiatrische Diagnosen vergeben, wobei Mehrfachdiagnosen möglich waren: Organische psychische Störungen (F0, $n=245$), Psychische und Verhaltensstörungen durch psychotrope Substanzen (F1, $n=54$), Schizophrenie, schizotype und wahnhaftige Störungen (F2, $n=4$), Affektive Störungen (F3, $n=119$), Neurotische, Belastungs- und somatoforme Störungen (F4, $n=171$), Verhaltensauffälligkeiten mit körperlichen Störungen und Faktoren (F5, $n=23$), Persönlichkeits- und Verhaltensstörungen (F6, $n=27$), Intelligenzstörungen (F7, $n=16$), Entwicklungsstörungen (F8, $n=38$), Verhaltens- und emotionale Störungen mit Beginn in der Kindheit und Jugend (F9), überwiegend Hyperkinetische Störungen (F90, $n=24$), Faktoren, die

den Gesundheitszustand beeinflussen und zur Inanspruchnahme des Gesundheitswesens führen (Z-Kodierungen, $n=75$). 344 (75.6%) aller Exploranden litten unter chronischen Schmerzen.

3.2 Untersuchungsinstrumente

Es wurden nur die Datensätze der Exploranden in die Studie aufgenommen, welche mindestens zwei von sieben Performanzvalidierungstests bearbeitet hatten. Die Begrenzung auf sieben Verfahren, die bereits von Hartmann (2002) qualitativ bewertet wurden, diente dazu, die angesichts der verschiedenen Sensitivitäten und Spezifitäten der Verfahren zu erwartende, erhöhte Variabilität der diagnostischen Sicherheit zu begrenzen (Tabelle 5). Nur eines dieser Verfahren, der *Rey Memory Test (RMT)*, oder *Rey 15-Items Test* (Rey, 1964), erhielt schlechte Bewertungen und wird in seiner ursprünglichen Form in den USA und Canada nicht mehr angewandt. Da der RMT aber von den meisten Psychiatern in der Schweiz, sofern sie denn überhaupt Beschwerdenvalidierung betreiben, routinemässig eingesetzt wird, und daher eine Reihe von Daten vorlagen, und überdies auch die Klassifikationsgenauigkeit der einzelnen Verfahren in einer späteren Arbeit untersucht werden soll, wurde dieses Verfahren dennoch in den Pool der PVT aufgenommen. Die übrigen eingesetzten Verfahren sind sowohl in den USA und Kanada, als auch in Deutschland bereits an verschiedenen Populationen und für verschiedenen Erkrankungsbilder neurologischer wie psychiatrischer Natur validiert worden. Zum Pool der eingesetzten PVT gehörten neben dem RMT fünf Antwortwahlverfahren zu denen an verbalen Verfahren der *Word Memory Test (WMT)*, Green, 2003), der *Medical Symptom Validity Test (MSVT)*, Green, 2004), und der *Amsterdamer Kurzzeitgedächtnistest (AKGT)*, Schmand & Lindeboom, 2005) zählten. Der *Non-Verbal Medical Symptom Validity Test (NV-MSVT)*, Green, 2008) und die *Bremer Symptomvalidierung (BSV - Kurzzeitgedächtnistest A)* aus der *Testbatterie zur Forensischen Neuropsychologie (TFBN)*, Heubrock & Petermann, 2000) gaben vor, nonverbale Gedächtnisleistungen zu erfassen. Ergänzt wurde der Pool an PVT um die *Reliable Digit Span (RDS)*, ein von Greiffenstein, Baker und Gola (1994) aus dem Untertest Zahlennachsprechen der Wechsler Intelligenz- bzw. Gedächtnistests entwickelter Validitätsparameter. Da in der Literatur in Abhängigkeit von der

untersuchten Population unterschiedliche Trennwerte für noch als authentisch zu wertende Leistungen diskutiert werden, wurden die im nächsten Absatz beschriebenen Analysen jeweils für einen RDS-Trennwert von 6 und für einen Trennwert von 7 berechnet.

Tabelle 5. Bewertung der eingesetzten PVT anhand der Hartmann-Kriterien (2002) modifiziert nach Henry (2009)

Verfahren	Sensitivität	Spezifität	Resistenz gg. Coaching	Augenscheinvalidität	Vergleichsdaten für Gesunde, Patienten und Simulanten	Einfachheit der Testanwendung und -auswertung	Einfachheit der Testinterpretation	Gestützt durch aktuelle Forschungsergebnisse	Erfassung simulationsrelevanter Zielfunktion aus Sicht des Exploranden
WMT	++	++	++	++	++	+	+	++	+
MSVT	+	++	+	+	++	++	++	++	++
NV-MSVT	+	++	+	+	+/-	+	+/-	+/-	+
BSV	+	++	-	+	+/-	++	+	-	+
AKGT	++	+	++	++	+	+	++	+/-	++
RMT	-	-	-	-	-	++	-	-	+/-

3.3 Statistische Analysen

Die Prävalenz der beschriebenen Gruppen wurde jeweils als Proportion zusammen mit einem 95 %-Vertrauensintervall nach Blaker (2000) geschätzt. Dies galt auch für die in einem weiteren Schritt ermittelten Prävalenzen für Explorandengruppen, welche nach verschiedenen soziodemographische Variablen, Diagnosen bzw. Diagnosegruppen und den Ergebnissen bildgebender Befunde zusammengestellt wurden. Der Einfluss verschiedener soziodemographischer Faktoren bzw. spezifischer Diagnosen auf das Risiko in mindestens zwei Performanzvalidierungstests ein

auffälliges oder mindestens 1 unterzufälliges Resultat zu erzielen und somit als „wahrscheinlich“ oder „sicher“ aggravierend klassifiziert zu werden wurde mittels einfacher logistischer Regressionsmodelle bestimmt. Die Schätzer der Modellkoeffizienten (als Odds ratios, *OR* zu interpretieren) wurden zusammen mit den zugehörigen 95 %-Vertrauensintervallen und p-Werten für die Nullhypothese, dass das entsprechende *OR* 1 entspricht, ausgewiesen. In einem zweiten Schritt wurden ausgewählte Variablen für die Faktoren Herkunft (aus den Staaten des früheren Jugoslawien und aus der Türkei) und Ausbildung (weniger als 9 bzw. 9-10 Ausbildungsjahre) korrigiert und so adjustierte Odds ratios (*AOR*) mit zugehörigem Vertrauensintervall ermittelt. Korrekturen für multiple Vergleiche wurden nicht angewendet.

4 ERGEBNISSE

Bei einem RDS von ≤ 6 konnte in 27.5 % (95% CI [23.5, 31.7], $n=125$), bei einem Trennwert von ≤ 7 in 34.3 % (95% CI [30.0, 38.8], $n=156$), der Fälle wahrscheinliche oder sichere MND diagnostiziert werden. Unterzufällige Ergebnisse in einem Antwortwahlverfahren und daher sichere Aggravation wurde in 16 Fällen beobachtet. Bei Anwendung eines RDS-Trennwerts von ≤ 6 entsprach dies 12.8 % der Exploranden, welche mehr als zwei auffällige PVT-Ergebnisse aufwiesen, bei Anwendung des strengeren Trennwerts von ≤ 7 waren dies 10.3% dieser Explorandengruppe. 24.8% (95% CI [21.0, 29.0], RDS ≤ 6) und 28.6% (95% CI [24.6, 32.9], RDS ≤ 7) der Exploranden erzielten ein auffälliges Ergebnis in einem PVT und erfüllten so die modifizierten Slick-Kriterien für mögliche MND. Je nach RDS-Trennwert wiesen 47.7% (95% CI [43.1, 52,3]) oder 37.1% (95% CI [32.7, 41.7]), keine psychometrischen Hinweise auf negative Antwortverzerrung auf (Tabelle 6).

4.1 Bezug zu soziodemographischen Faktoren

Es zeigten sich deutliche Zusammenhänge zwischen dem Risiko in mindestens zwei PVTs ein auffälliges Ergebnis aufzuweisen und geringem Bildungsgrad, Zugehörigkeit zur 1. Generation an Migranten sowie Herkunftsland. Im Vergleich zu Exploranden deutscher Muttersprache fand sich das höchste Risiko bzw. die höchsten Odd ratios bei Migranten aus den Ländern des früheren Jugoslawien sowie der Türkei (Abb. 1). Nach Korrektur für Herkunft aus diesen Ländern nahm auch das Risiko für die Migranten der ersten Generation auffällige Ergebnisse zu erzielen signifikant ab, blieb aber immer noch über jenem der 2. Generation von Migranten (Abb. 2). Nach Korrektur für den Faktor Ausbildung (von weniger als 9 Jahren) nahm das Risiko mindestens

Tabelle 6. Häufigkeits- und Prävalenzschätzungen (bei RDS Trennwert ≤ 6)

	Total	Keine oder nur mögliche wahrscheinliche oder sichere negative Antwortverzerrung								
		< 2 auffällige PVTs und kein PVT signifikant unter Zufallsniveau			≥ 2 auffällige PVTs oder 1 PVT signifikant unter Zufallsniveau			≥ 1 PVT signifikant unter Zufallsniveau		
		Häufigkeit n (%)	Prävalenz % CI (%)		Häufigkeit n (%)	Prävalenz % CI (%)		Häufigkeit n (%)	Prävalenz % CI (%)	
Gesamtstichprobe	455	330 (72.5)	72.5	68.3-76.5	125 (27.5)	23.5-31.7		16 (12.8)	12.8	7.7-19.8
Geschlecht										
Männlich	222	155 (34.1)	69.8	63.3-75.7	67 (14.7)	30.2	24.3-36.7	11 (8.8)	16.4	9.0-27.3
Weiblich	232	174 (38.2)	75.0	69.0-80.0	58 (12.7)	25.0	19.7-31.0	5 (4.0)	8.6	3.5-18.5
Transgender	1	1(0.2)	100.0	5.0-100.0	0 (0.0)	0.0	0.0-95.0	0 (0.0)	0.0	
Ausbildung										
< 9 Jahre	40	11 (2.4)	27.5	15.6-43.6	29 (6.4)	72.5	56.4-84.4	5 (4.0)	17.2	7.1-35.7
Regulärer Schulabschluss (9–10 Jahre)	52	36 (7.9)	69.2	54.9-80.8	16 (3.5)	30.8	19.2-45.1	2 (1.6)	12.5	2.3-36.8
Abgeschlossene Ausbildung / Matura (11–13 Jahre)	298	230 (50.9)	77.2	72.1-81.8	68 (14.9)	22.8	18.2-27.9	8 (6.4)	11.8	5.2-21.7
Fachhochschule, Universität, etc. (>13 Jahre)	65	53 (11.6)	81.5	70.3-89.7	12 (2.6)	18.5	10.3-29.7	1 (0.8)	8.3	0.4-36.6
Herkunftsregionen										
Schweiz	280	230 (50.5)	82.1	77.6-86.4	50 (11.0)	17.9	13.6-22.8	3 (2.4)	6.0	1.7-16.4
Deutschsprachige Länder	37	32 (7.0)	86.5	72.1-94.5	5 (1.1)	13.5	5.5-27.9	1 (0.8)	20.0	1.0-65.7
Andere europäische Staaten (ohne Balkan)	39	26 (5.7)	66.7	50.0-80.1	13 (2.9)	33.3	19.9-50.0	1 (0.8)	7.7	0.4-33.6
Balkanstaaten	56	20 (4.4)	35.7	23.7-49.1	36 (7.9)	64.3	50.9-76.3	7 (5.6)	19.4	8.8-35.7
Vorderasien	27	11 (2.4)	40.7	23.3-59.8	16 (3.5)	59.3	40.2-76.7	4 (3.2)	25.0	9.0-50.0
Asien	5	4 (0.9)	80.0	34.3-99.0	1 (0.2)	20.0	1.0-65.7	0 (0.0)	0.0	0.0-95.0
Nordamerika	3	3 (0.7)	100.0	36.9-100.0	0 (0.0)	0.0	0.0-63.1	0 (0.0)	0.0	
Südamerika	4	2 (0.4)	50.0	9.8-90.2	2 (0.4)	50.0	9.8-90.2	0 (0.0)	0.0	0.0-77.6
Afrika	4	2 (0.4)	50.0	9.8-90.2	2 (0.4)	50.0	9.8-90.2	0 (0.0)	0.0	0.0-77.6
Balkanstaaten und Vorderasien	83	31 (6.8)	37.3	27.5-48.2	52 (11.4)	62.7	51.8-72.5			
Balkanstaaten, Vorderasien und übriges, nicht deutschsprachiges Europa	122	57 (12.5)	46.7	37.9-55.8	65 (14.3)	53.3	44.2-62.1			
Migrationsstatus										
NA	279	229 (50.3)	82.1	77.2-86.3	50 (11.0)	17.9	13.7-22.8	3 (2.4)	6.0	1.7-16.4
Erste Generation	145	77 (16.9)	53.1	44.8-61.2	68 (14.9)	46.9	38.8-55.2	13 (10.4)	19.1	11.1-29.9
Zweite Generation	31	24 (5.3)	77.4	59.9-89.7	7 (1.5)	22.6	10.3-40.1	0 (0.0)	0.0	0.0-37.7
Altersklassen										
<20 Jahre	6	6 (1.3)	100.0	59.4-100.0	0 (0.0)	0.0	0-40.6			
20–30 Jahre	66	59 (13.0)	89.4	79.9-95.2	7 (1.5)	10.6	4.8-20.1	0 (0.0)	0.0	0.0- 37.7
30–40 Jahre	91	57 (12.5)	62.6	52.2-72.2	34 (7.5)	37.4	27.8-47.8	3 (2.4)	8.8	2.5-22.9
40–50 Jahre	146	98 (21.5)	67.1	59.0-74.5	48 (10.5)	32.9	25.5-41.0	6 (4.8)	12.5	5.6-24.6
50–60 Jahre	111	87 (19.1)	78.4	69.9-85.4	24 (5.3)	21.6	14.6-30.1	5 (4.0)	20.8	8.6-41.4
60–70 Jahre	34	23 (5.1)	67.6	50.0-81.6	11 (2.4)	32.4	18.4-50.0	2 (1.6)	18.2	3.3-50.0
>70 Jahre	1	0 (0.0)	0.0	0.0-95.0	1 (0.2)	100.0	5.0-100.0	0 (0.0)	0.0	0.0-95.0
Time since injury or onset of disease										
[0–3] Jahre	142	110 (24.2)	77.5	69.9-84.0	32 (7.0)	22.5	16.0-30.1	3 (2.4)	9.4	2.6-24.3
(3–5,6] Jahre	85	61 (13.4)	71.8	61.3-80.9	24 (5.3)	28.2	19.1-38.7	4 (3.2)	16.7	5.9-37.0
(5,6–11] Jahre	117	79 (17.4)	67.5	58.6-75.8	38 (8.4)	32.5	24.2-41.4	4 (3.2)	10.5	3.7-24.4
(11–54] Jahre	109	78 (17.1)	71.6	62.5-79.6	31 (6.8)	28.4	20.4-37.5	5 (4.0)	16.1	6.6-33.4
NA	2	2 (0.2)	100.0	22.4-100.0	0 (0.0)	0.0	0.0-77.6			

Beschwerden / Diagnosen°	Total	Keine oder nur mögliche			wahrscheinliche oder sichere negative Antwortverzerrung					
	N	< 2 auffällige PVTs und kein PVT signifikant unter Zufallsniveau			≥ 2 auffällige PVTs oder 1 PVT signifikant unter Zufallsniveau			≥ 1 PVT signifikant unter Zufallsniveau		
		Häufigkeit n (%)	Prävalenz %	CI (%)	Häufigkeit n (%)	Prävalenz %	CI (%)	Häufigkeit n (%)	Prävalenz %	CI (%)
Neurologische Diagnosen° (n = 408)										
HWS-Problematik, Schleudertrauma Schädelhirntrauma	208	137 (30.1)	65.9	59.2-72.2	71 (15.6)	34.1	27.8-40.8	6 (4.8)	8.5	3.7-17.2
leicht	158	114 (25.1)	72.2	64.6-79.0	44 (9.7)	27.8	21.0-35.4	7 (5.6)	15.9	7.2-29.2
mittelgradig (positive CT/MRI Befunde)*	39	27 (5.9)	69.2	52.6-82.8	12 (2.6)	30.8	17.2-47.4	1 (0.8)	8.3	0.4-36.6
schwer (positive CT/MRI Befunde)*	29	22 (4.8)	75.9	57.1-89.0	7 (1.5)	24.1	11.0-42.9	0 (0.0)	0.0	0.0-37.7
Cerebrovaskuläres Ereignis	33	25 (5.5)	75.8	57.8-88.3	8 (1.8)	24.2	11.7-42.2	3 (2.4)	37.5	11.1-71.1
Cerebrale Raumforderung	8	6 (1.3)	75.0	35.9-95.4	2 (0.4)	25.0	4.6-64.1	1 (0.8)	50.0	2.5-97.5
Epilepsie	24	18 (4.0)	75.0	54.3-88.5	6 (1.3)	25.0	11.5-45.7	2 (1.6)	33.3	6.3-72.9
Multiple Sklerose	5	4 (0.9)	80.0	34.9-99.0	1 (0.2)	20.0	1.0-65.7	0 (0.0)	0.0	0.0-95.0
Mild cognitive impairment, beginnende Demenz	7	5 (1.1)	71.4	34.1-94.7	2 (0.4)	28.6	5.3-65.9	0 (0.0)	0.0	0.0-77.60
Andere mit möglichem Einfluss auf die Kognition	68	53 (11.6)	77.9	66.4-86.4	15 (3.3)	22.1	13.6-33.6	1 (0.8)	6.7	0.3-30.2
Substantielle Hirnschädigung gemäss Bildgebung										
Ja*	127	96 (21.1)	75.6	67.4-82.5	31 (6.8)	24.4	17.5-32.6	6 (4.8)	19.4	8.8-36.7
Nein*	147	99 (21.8)	67.3	59.2-74.6	48 (10.5)	32.7	25.4-40.8	7 (5.6)	14.6	6.6-27.7
Irrelevante Befunde*	64	49 (10.8)	76.6	64.8-85.6	15 (3.3)	23.4	14.4-35.2	1 (0.8)	6.7	0.3-30.2
Kein CT/MRI durchgeführt oder zugänglich	117	86 (18.9)	73.5	64.6-81.0	31 (6.8)	26.5	19.0-35.4	2 (1.6)	6.5	1.2-20.2
Psychiatrische Diagnosen° (n = 421)										
Psychische Störungen nicht hirnganischer Natur (F10–F99)	175	108 (23.7)	61.7	54.3-68.8	67 (14.7)	38.3	32.7-47.6	9 (7.2)	13.4	7.0-23.5
F0 Organische, einschliesslich symptomatischer, psychischer Störungen	245	200 (44.0)	81.6	76.2-86.3	45 (9.9)	18.4	13.3-23.0	5 (4.0)	11.1	4.5-23.9
ohne weitere psychiatrische Komorbiditäten	106	92 (20.2)	86.8	79.0-92.3	14 (3.1)	13.2	7.7-21.0	1 (0.8)	7.1	0.4-31.2
mit weiteren psychiatrischen Komorbiditäten	139	108 (23.7)	77.7	69.9-84.0	31 (6.8)	22.3	16.0-30.1	4 (3.2)	12.9	4.5-28.6
F1 Psychische und Verhaltensstörungen durch psychotrope Substanzen	54	41 (0.9)	75.9	63.2-85.7	13 (2.9)	24.1	14.3-36.8	1 (0.8)	7.7	0.41-33.6
F2 Schizophrenie, schizotypische und wahnhaftige Störungen	4	3 (0.7)	75.0	24.9-98.7	1 (0.2)	25.0	24.9-98.7	0 (0.0)	0.0	0.0-95.0
F3 Affektive Störungen	119	76 (16.7)	63.9	54.7-72.4	43 (9.5)	36.1	27.6-45.3	7 (5.6)	16.3	7.4-29.9
F32-34 Depressive Episode, rezidivierende und anhaltende depressive Störungen	118	75 (16.5)	63.6	54.3-72.2	43 (9.5)	36.4	27.8-45.7	7 (5.6)	16.3	7.4-29.9
F4 Neurotische, Belastungs- und somatoforme Störungen	146	91 (20.0)	62.3	54.1-70.0	55 (12.1)	37.7	30.0-45.9	9 (7.2)	16.4	8.6-28.7
F40–41 Phobische und andere Angststörungen	28	17 (3.7)	60.7	40.8-77.6	11 (2.4)	39.3	22.4-59.2	1 (0.8)	9.1	0.5-40.1
F42 Zwangsstörung	1	0 (0.0)	0.0	0.0-95.0	1 (0.2)	100.0	0.5-100.0	0 (0.0)	0.0	0.0-95.0
F43.0–43.1 Akute Belastungsreaktion und posttraumatische Belastungsstörung	23	13 (13.9)	56.5	36.0-75.3	10 (2.2)	43.5	24.7-64.0	1 (0.8)	10.0	0.5-44.4
F43.2 Anpassungsstörungen	27	21 (4.6)	77.8	58.5-89.8	6 (1.3)	22.2	10.2-41.5	1 (0.8)	16.7	0.9-59.4
F44 Dissoziative Störungen (Konversionsstörungen)	10	7 (1.5)	70.0	38.1-91.3	3 (0.7)	30.0	8.7-61.9	1 (0.8)	33.3	1.7-86.5
F45 Somatoforme Störungen	71	36 (7.9)	50.7	38.6-62.3	35 (7.7)	49.3	37.7-61.4	5 (4.0)	14.3	5.8-29.5
F48 Andere neurotische Störungen	12	9 (2.0)	75.0	45.6-92.8	3 (0.7)	25.0	7.2-54.4	0 (0.0)	0.0	0.0-63.1
F5 Verhaltensauffälligkeiten mit körperlichen Störungen und Faktoren	23	16 (3.5)	69.6	47.7-86.1	7 (1.5)	30.4	13.9-52.3	0 (0.0)	0.0	0.0-37.7
F6 Persönlichkeits- und Verhaltensstörungen	27	18 (4.0)	66.7	46.2-82.4	9 (2.0)	33.3	17.6-53.8	1 (0.8)	11.1	0.6-44.3
F7 Intelligenzstörungen	16	11 (2.4)	68.8	43.4-86.8	5 (1.1)	31.3	13.2-56.6	1 (0.8)	20.0	1.0-65.7
F8 Entwicklungsstörungen	38	35 (7.7)	92.1	79.6-97.8	3 (0.7)	7.9	2.3-20.4	0 (0.0)	0.0	0.0-63.1
F9 Verhaltens- und emotional Störungen mit Beginn in der Kindheit und Jugend	24	22 (4.8)	91.7	73.8-98.5	2 (0.4)	8.3	1.5-26.2	0 (0.0)	0.0	0.0-77.6
F90 Hyperkinetische Störungen	20	19 (4.2)	95.0	76.1-99.7	1 (0.2)	5.0	0.3-23.9	0 (0.0)	0.0	0.0-95.0
F91–98 Andere Verhaltens- und emotionale Störungen	4	3 (0.7)	75.0	24.9-98.7	1 (0.2)	25.0	1.3-75.1	0 (0.0)	0.0	0.0-95.0
Z0-Z9 Faktoren, die den Gesundheitszustand beeinflussen und zur Inanspruchnahme des Gesundheitswesens führen	75	46 (10.1)	61.3	49.9-72.3	29 (6.4)	38.7	27.7-50.1	6 (4.8)	20.7	9.4-39.3
Chronische Schmerzen	344	236 (51.9)	68.6	63.5-73.5	108 (23.7)	31.4	26.5-36.5	12 (9.6)	11.1	6.2-18.3
Chronische Schmerzen unter Ausschluss somatoformer Störungen	273	200 (44.0)	73.3	67.7-78.3	73 (16.0)	26.7	21.7-32.3	7 (5.6)	9.6	4.3-18.4

Note. PVT = Performanzvalidierungstest; NA = not applicable; * = falls CT/MRI durchgeführt; ° = Mehrfachdiagnosen möglich

zwei auffällige PVT-Ergebnisse zu erzielen, bei Exploranden vom Balkan und Vorderasien markant ab, betrug aber immer noch das 5-6fache des Risikos von Schweizer Exploranden (Abb. 3). In Abhängigkeit vom verwendeten RDS-Trennwert war das Risiko von Migranten der ersten Generation, insbesondere wenn sie aus dem Balkan oder der Türkei stammten, in einem Antwortwahlverfahren ein Ergebnis unter dem Zufallsniveau zu erzielen, 3.7- bis 4.3-mal höher als das der Schweizer Exploranden. 68.8% der Exploranden mit unterzufälligem Resultat und damit sicher nachgewiesener Aggravation stammten aus dem Balkan und Vorderasien. Dieses Resultat war unabhängig vom Bildungsniveau (Abb. 1).

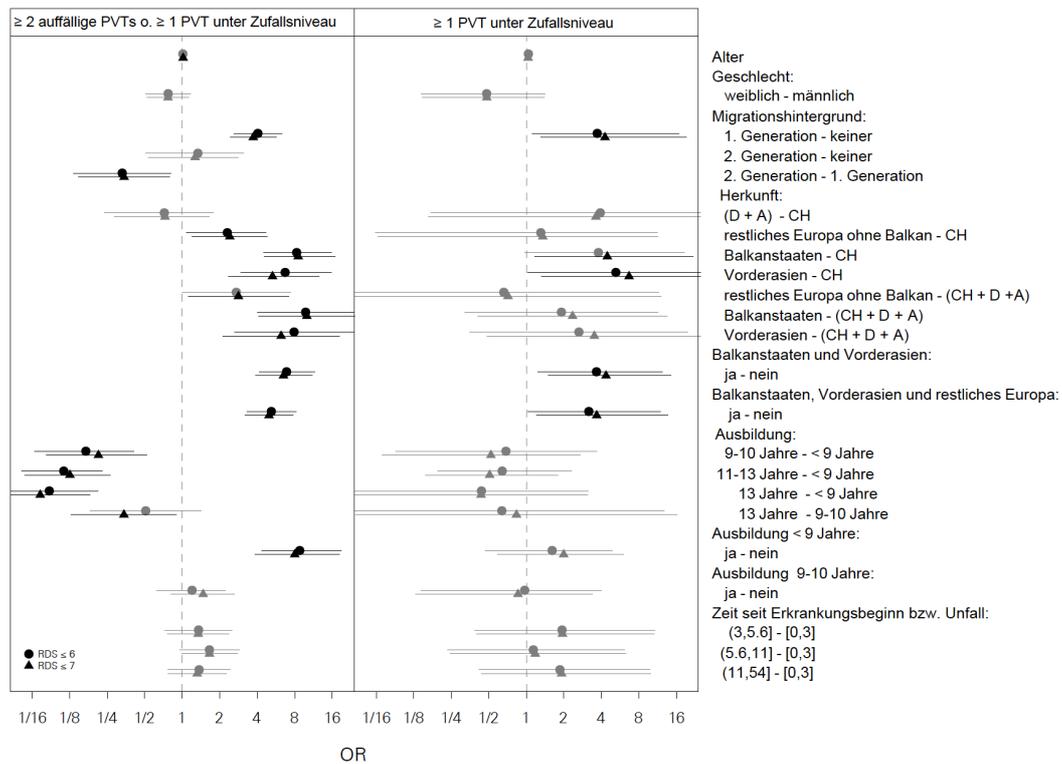


Abbildung 1. Einfluss soziodemographischer Faktoren auf die Wahrscheinlichkeit negativer Antwortverzerrung und sicherer Aggravation in der Gesamtstichprobe.

Anmerkungen: D = Deutschland, A = Österreich, CH = Schweiz; Angabe der Zeit seit Erkrankungsbeginn bzw. Unfall in Quartilen (Jahre); OR = Odds ratio; Linien = 95%-Vertrauensintervalle; schwarze Symbole und Linien = signifikante Unterschiede (d.h. das 95%-Vertrauensintervall schliesst OR = 1 nicht ein).

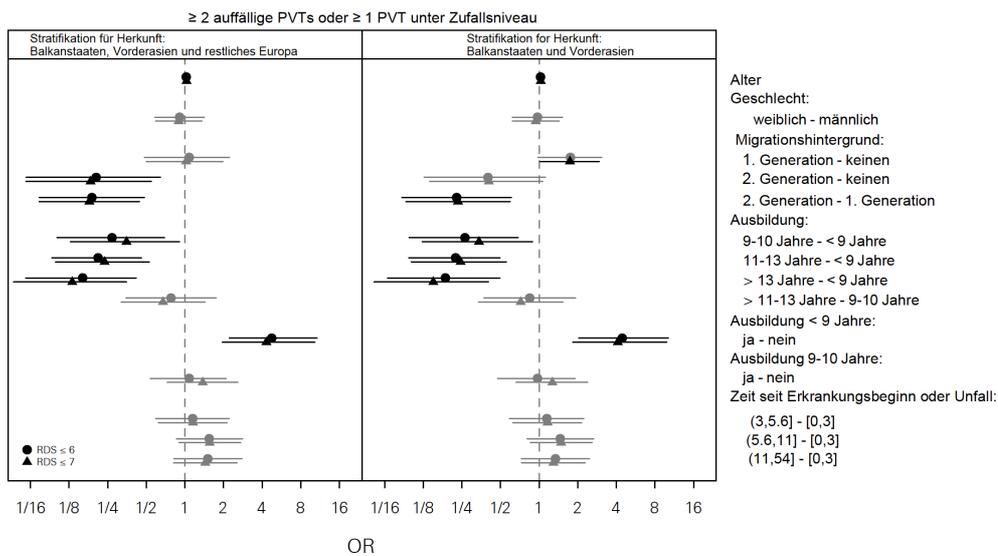


Abbildung 2. Einfluss soziodemographischer Faktoren auf die Wahrscheinlichkeit negativer Antwortverzerrung stratifiziert nach Herkunft.

Anmerkungen: Angabe der Zeit seit Erkrankungsbeginn bzw. Unfall in Quartilen (Jahre); OR = Odds ratio; Linien = 95%. Vertrauensintervalle; schwarze Symbole und Linien = signifikante Unterschiede (d.h. das 95%-Vertrauensintervall schliesst OR = 1 nicht ein).

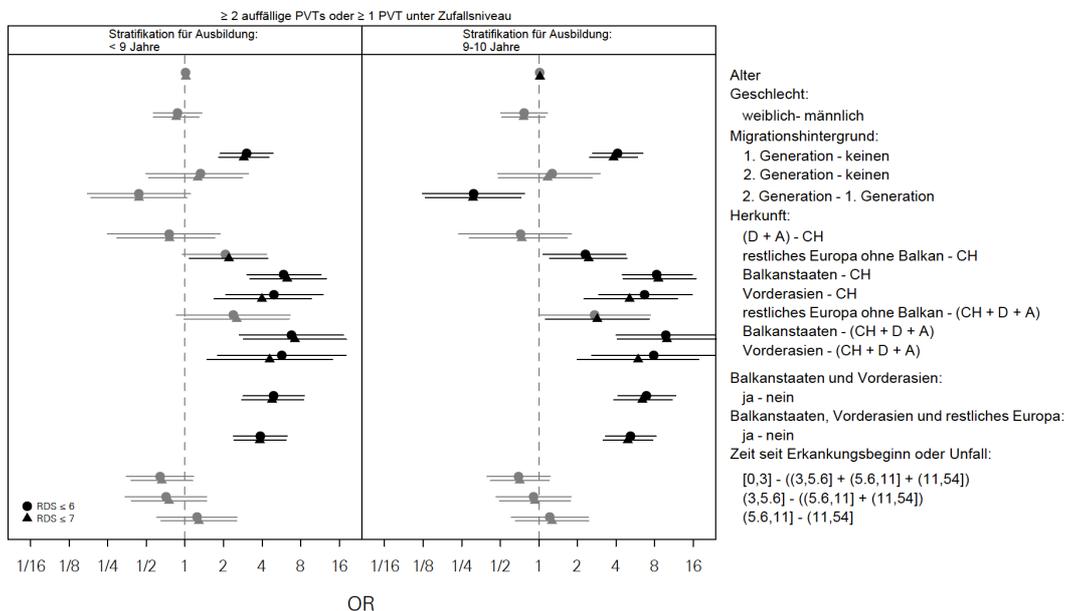


Abbildung 3. Einfluss soziodemographischer Faktoren auf die Wahrscheinlichkeit negativer Antwortverzerrung korrigiert für geringen Bildungsstand.

Anmerkungen: D = Deutschland, A = Österreich, CH = Schweiz; Angabe der Zeit seit Erkrankungsbeginn bzw. Unfall in Quartilen (Jahre); OR = Odds ratio; Linien = 95%-Vertrauensintervalle; schwarze Symbole und Linien = signifikante Unterschiede (d.h. das 95%-Vertrauensintervall schliesst OR = 1 nicht ein).

4.2 Bezug zu neurologischen Störungsbildern

Das Risiko mehr als ein auffälliges Ergebnis in einem PVT zu zeigen war bei Exploranden mit cerviko-spinalen Problemen (HWS-Distorsion und Schleudertrauma), chronischen Schmerzen und fehlendem Nachweis organisch-struktureller Hirnläsionen in der Bildgebung im Vergleich zu Exploranden mit anderen neurologischen Diagnosen etwa um den Faktor 2 erhöht. Bereinigte man die Gruppe chronischer Schmerzpatienten um jene mit einer diagnostizierten somatoformen Störung, bestand kein erhöhtes Risiko für auffällige PVT-Ergebnisse mehr. Einflüsse vom Herkunftsland oder dem Bildungsgrad liessen sich nicht demonstrieren (Abb. 4).

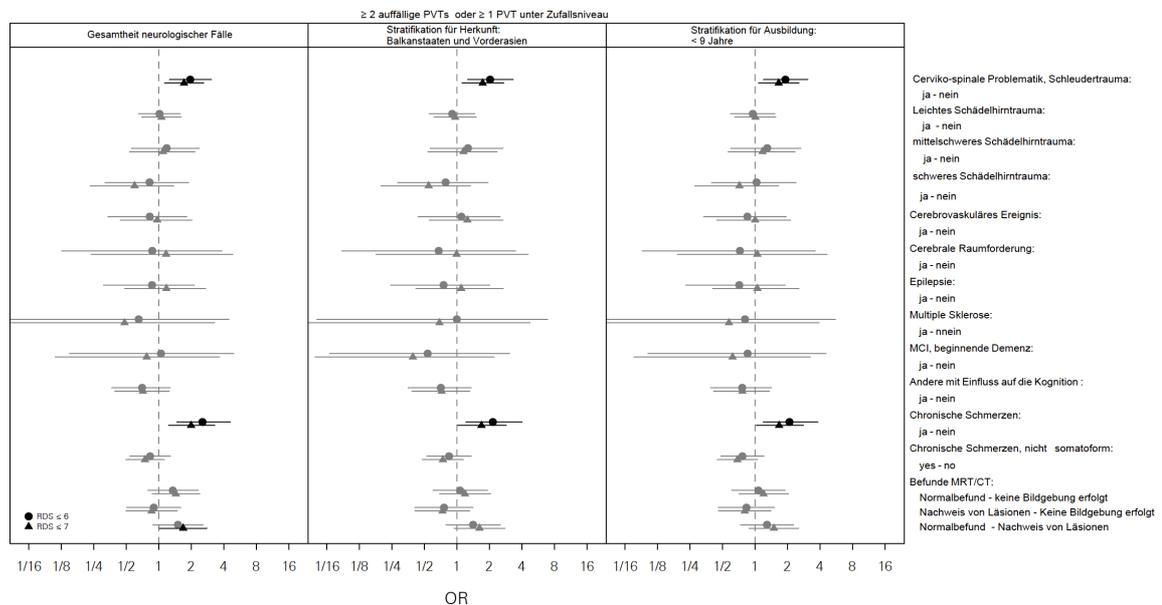


Abbildung 4. Einfluss neurologischer Diagnosen, Schmerzen und bildgebender Befunde auf die Wahrscheinlichkeit negativer Antwortverzerrung für die Gesamtheit neurologischer Patienten sowie korrigiert für Herkunft und Ausbildung.

Anmerkungen: MCI = Mild cognitive impairment; OR = Odds ratio; Linien = 95%-Vertrauensintervalle; schwarze Symbole und Linien = signifikante Unterschiede (d.h. das 95%-Vertrauensintervall schliesst OR = 1 nicht ein).

4.4 Bezug zu psychiatrischen Diagnosen

Eine Beziehung zwischen Diagnosengruppen nach ICD-10 und dem Risiko eines unterzufälligen Ergebnisses in einem PVT liess sich weder für neurologischen Diagnosen noch für psychiatrische Diagnosen nachweisen (aus Platzgründen nicht dargestellt).

Allerdings liess sich bei Exploranden mit psychiatrischer Diagnose im Vergleich zur Gesamtstichprobe ein dreifach höheres Risiko, mindestens zwei auffällige Ergebnisse in PVTs zu erzielen, beobachten. Interessanterweise war das Risiko von Exploranden mit ausschliesslicher F0-Diagnose (Organische, einschließlich symptomatischer psychischer Störungen) dreimal geringer als jenes von Exploranden mit anderen psychiatrischen Diagnosen und nur halb so hoch wie jenes von Exploranden mit einer Hirnschädigung und zusätzlicher psychiatrischer Störung. Dieses Verhältnis blieb auch nach Korrektur für Herkunftsregion und Ausbildung unverändert. Eine im Vergleich zur übrigen Gruppe mit einer psychiatrischen Diagnose versehenen Exploranden geringere Wahrscheinlichkeit mindestens 2 PVT nicht zu bestehen, wiesen auch Exploranden mit Entwicklungsstörungen sowie Verhaltens- und emotionalen Störungen mit Beginn in der Kindheit und Jugend (in dieser Studie überwiegend Hyperkinetische Störungen) auf. Die Wahrscheinlichkeit auffälliger PVT-Ergebnisse erhöhte sich nach Korrektur für Herkunft, nicht aber für den Bildungsgrad (Abb. 5).

Das höchste Risiko, mindestens zwei auffällige Ergebnisse in PVTs zu erzielen, fand sich bei Exploranden mit einer diagnostizierten depressiven, somatoformen Störung oder einer posttraumatischen Belastungsstörung, wobei das höchste Risiko bei den Exploranden mit somatoformer Störung lag. Wiederum minimierte sich die Wahrscheinlichkeit auffälliger PVT-Ergebnisse für Exploranden mit diesen 3 Diagnosen nach Korrektur für die Herkunftsregion, während der Bildungsgrad nur für bei den als depressiv diagnostizierten Exploranden einen Einfluss auf die PVT-Ergebnisse zu haben schien. Ein erhöhtes Risiko, bei mindestens zwei PVTs jenseits kritischer Trennwerte abzuschneiden fand sich auch bei Personen mit Z-Kodierungen (vorwiegend Z73: Probleme mit Bezug auf Schwierigkeiten bei der Lebensbewältigung).

Für eine detaillierte Darstellung der Studienergebnisse sei auf den Zeitschriftenartikel (Plohm, eingereicht) verwiesen.

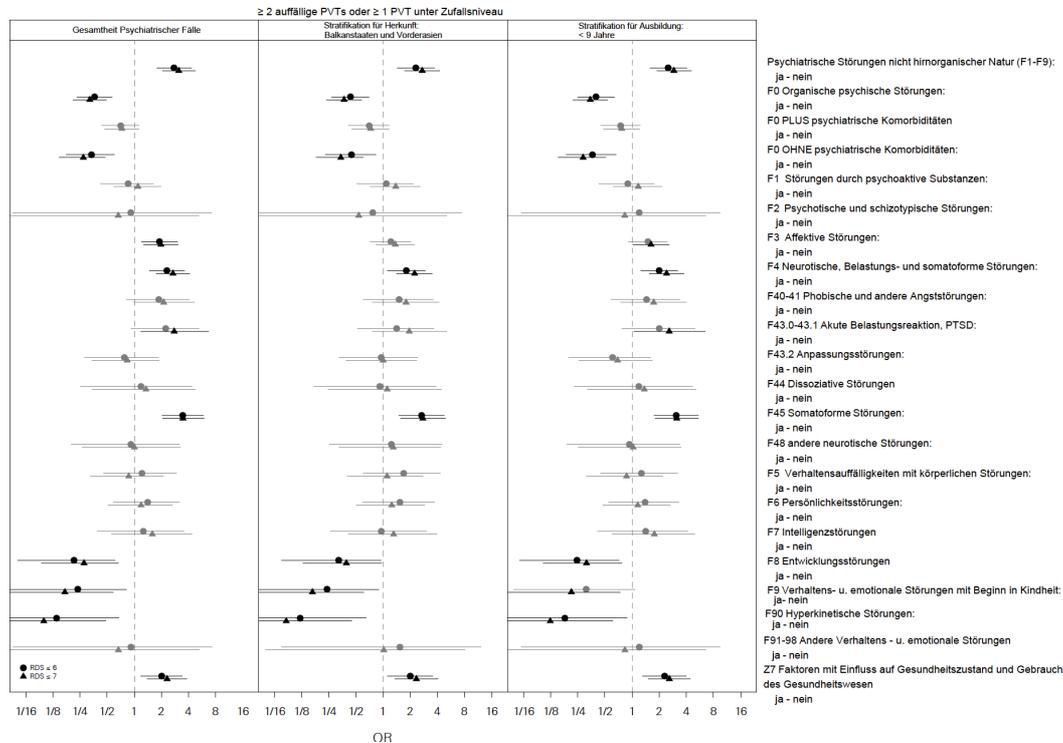


Abbildung 5. Einfluss psychiatrischer Störungen auf die Wahrscheinlichkeit negativer Antwortverzerrung für die Gesamtheit psychiatrischer Patienten sowie korrigiert für Herkunft und Ausbildung.

Anmerkungen: OR = Odds ratio; Linien = 95%-Vertrauensintervalle; schwarze Symbole und Linien = signifikante Unterschiede (d.h. das 95%-Vertrauensintervall schliesst $OR = 1$ nicht ein).

5 DISKUSSION

5.1 Prävalenz unzureichender Anstrengungsbereitschaft und Aggravation

In dieser Stichprobe von 455 Explorand, die sich im Rahmen einer interdisziplinären Begutachtung auch einer neuropsychologischen Untersuchung unterziehen mussten, erfüllten (je nach herangezogenem RDS-Trennwert) 27.5% bis 34.2% Kriterien für eine wahrscheinliche oder sichere Aggravation entsprechend der von Larrabee modifizierten Slick-Kriterien. Die in verschiedenen Subgruppen gefundenen Grundraten erwiesen sich in starkem Mass von bestimmten soziodemographischen Faktoren (Herkunft, 1. oder 2. Einwanderer-Generation und Bildungsgrad) und gewissen neurologischen und psychiatrischen Diagnosegruppen abhängig. Anders als von vielen Experten angenommen erwiesen sich Bildungsgrad und Ausbildung als voneinander unabhängige Faktoren. Die Prävalenz von wahrscheinlicher (inklusive der Fälle sicherer) Aggravation entspricht den in einer Vielzahl internationaler Studien gefundenen Grundraten, welche dieselben Klassifikationskriterien anwandten (z.B. Larrabee, 2003; Greve, Ord, Bianchini, & Curtis, 2009). Gleichzeitig liegt die gefundene Prävalenz in der Schweiz deutlich über den Schätzungen der Schweizer Experten aus den Kreisen der Versicherungsfachleute und medizinischer Gutachter.

3.5% der Gesamtstichprobe und damit 10.3% bis 12.8% der Exploranden mit mindestens 2 auffälligen PVT-Ergebnissen wiesen in einem oder mehreren Antwortwahlverfahren Resultate unterhalb der Ratewahrscheinlichkeit auf und erfüllten somit die Kriterien für sichere Aggravation neurokognitiver Dysfunktionen. Damit lag die Prävalenz sicherer Aggravation von MND in der Gesamtstichprobe unter der von Young (2015) über eine Vielzahl von, mehrheitlich amerikanischen, Studien hinweg ermittelten Grundrate von rund 7%, entsprach aber der Grössenordnung nachgewiesenen Versicherungsbetrugs in der Schweiz (Thommen, 2006). Die höchste Wahrscheinlichkeit unterzufälliger Ergebnisse in PVTs fand sich bei Migranten aus den Nachfolgestaaten des früheren Jugoslawien und der Türkei, jener Gruppe also, die auch das höchste Risiko für wahrscheinliche Aggravation aufwies. Nach der Erhebung von Bolliger, Stadelmann-Steffen,

Thomann und Rüefli (2009) ist dies auch die Bevölkerungsgruppe der Schweiz, welche die höchste Rate an Rentenanträgen aufwies.

5.2 Störungsbilder mit erhöhter Wahrscheinlichkeit nicht-authentischer Befunde

Die nach Diagnosegruppen aufgeschlüsselte Analyse ergab in folgenden Gruppen eine erhöhte Wahrscheinlichkeit nicht-authentische Befunde in mindestens 2 PVT zu finden:

- a. Neurologische Krankheitsbilder
 - Cerviko-spinale Probleme (inkl. HWS-Distorsion und Schleudertrauma)
 - Fehlender Nachweis einer strukturellen Hirnläsion in der Bildgebung
 - Chronische Schmerzen (bei Einbezug der Exploranden mit somatoformen Störungen)
- b. Psychiatrische Störungsgruppen
 - Depressive Störungen
 - Posttraumatische Belastungsstörungen
 - Somatoforme Störungen

Damit ist die Wahrscheinlichkeit von Aggravation besonders bei schwer zu objektivierenden Krankheitsbildern erhöht. Geringe Bildung (< 9 Jahre) und Herkunft aus dem Balkan oder der Türkei erhöhten das Risiko zusätzlich, jedoch unabhängig voneinander. Dass dies für die neurologischen Krankheitsbilder nicht so schien, ist mit hoher Wahrscheinlichkeit auf den mit 16.5% bis 21.8% nur halb so hohe, Anteil an Migranten in den risikobehafteten neurologischen Gruppen zurückzuführen. In den genannten psychiatrischen Störungsgruppen betrug der Anteil der Migranten 35.6% - 43.5%. Für die Explorandengruppen mit depressiven Störungen und posttraumatischen Belastungsstörungen war das erhöhte Risiko nach Korrektur für die Herkunft nicht mehr nachweisbar, wohl aber noch nach Korrektur für niedrigen Bildungsstand. Möglicherweise ist das erhöhte Risiko für auffällige PVT-Befunde in diesen Gruppen dem hohen Anteil an Migranten einer bestimmten Herkunft geschuldet. Dies müsste in weiteren Studien mit grösseren Stichproben geklärt werden. Hingegen blieb das erhöhte Wahrscheinlichkeit negativer Antwortverzerrung für die Exploranden mit diagnostizierter somatoformer Störung auch nach Korrektur für

Herkunft und Bildung bestehen. Damit liegt gerade für eine solche Diagnose eine erhöhte Wahrscheinlichkeit für Täuschungsverhalten vor, bei welcher die Frage, inwieweit negative Antwortverzerrung als ein Symptom der Erkrankung gelten kann, kontrovers diskutiert wird.

5.3 Mögliche falsch positive Klassifikationen

Zur detaillierten Diskussion von Faktoren, die zu möglichen falsch positiven Urteilen bei Migranten beitragen können und der möglichen Beweggründe von Migranten während Begutachtungen keine ausreichende Anstrengungsbereitschaft aufzubringen, sei auf die beiliegende Publikation verwiesen (Plohmann, eingereicht). Falsch positive Klassifikationen aufgrund geringer Bildung sind nach Sichtung der Literatur bei den hier eingesetzten Testverfahren selten. Auch der Umstand, dass die hier untersuchten Exploranden mit Aggravationsverdacht in der Regel mehr als nur zwei auffällige Validierungsparameter aufwiesen, spricht gegen bildungsabhängige Fehlklassifikationen. Jene Fälle, in welchen Exploranden nur zwei Verfahren, nämlich die RDS und ein verbales Antwortwahlverfahren bearbeiten mussten, wurden retrospektiv näher beleuchtet, da es in der Literatur Hinweise gibt, dass Zweisprachigkeit (Englisch und Spanisch bzw. Englisch und indianische Sprachen; zu anderen Sprachen liegen keine Daten vor) in beiden Verfahrenstypen zu schlechteren Leistungen führen kann. Der Rückgriff auf die Akten dieser Fälle ergab, dass angesichts einer vollständigen Durchführung der verbalen Antwortwahlverfahren mit allen dazugehörigen Subtests und der dann möglichen Profilinterpretation und die Berücksichtigung aller zusätzlichen Informationen aus Eigen- und Fremdanamnese, der Verhaltensbeobachtung und den Akten eine Fehlklassifikation dieser konkreten Fälle mit an Sicherheit grenzender Wahrscheinlichkeit ausgeschlossen werden konnte. Grundsätzlich ist es jedoch möglich, dass die geschätzte Prävalenz an unzureichender Anstrengungsbereitschaft, wahrscheinlicher und sicherer Aggravation in der Gruppe der Migranten überschätzt wurde. Dies ist darauf zurückzuführen, dass mangels erhobener und/oder veröffentlichter Daten gerade für Menschen aus dem Kulturraum des Balkans und Vorderasiens, in der vorgelegten Studie auf Trennwerte zurückgegriffen

werden musste, die nur in den USA, Kanada und west- bzw. mitteleuropäischen Staaten validiert wurden.

Ferner gaben alle eingesetzten PVT vor, Gedächtnisleistungen zur erfassen. Es ist zudem denkbar, dass gerade wenig gebildete Migranten bevorzugt Gedächtnisstörungen aggravierten, und somit eher auffällig wurden, während andere Exploranden Täuschungsstrategien einsetzen, welche mit dem gewählten Untersuchungsansatz weniger gut aufgedeckt werden konnten. Andererseits besteht gerade für die Migranten aus dem Balkan und Vorderasien aus einer Reihe von Gründen ein erhöhter Anreiz (internal wie external) zur Aggravation, der auch zur ungewöhnlich hohen Zahl nachgewiesener absichtlicher Antwortverzerrungen beigetragen haben dürfte. Zur ausführlicheren Diskussion siehe Plohmann (eingereicht) und Hausotter und Schouler-Ocak (2007).

5.4 Schwächen der Studie

Da in der vorliegenden Arbeit nur die Exploranden untersucht wurden, die direkt von den Auftraggebern oder von den mitbegutachtenden Ärzten einer neuropsychologischen Begutachtung zugewiesen wurden, ist die Stichprobe nicht repräsentativ für das gesamte Kollektiv an Personen, die sich einer versicherungsmedizinischen Begutachtung unterziehen müssen. Auch aus diesem Grund waren - trotz ausreichend grosser Gesamtstichprobe - einzelne Subgruppen recht klein und die Vertrauensintervalle der Prävalenzschätzungen entsprechend gross. Für Subgruppen mit kleiner Explorandenzahl war somit auch rein statistisch kaum die Möglichkeit gegeben, einen möglichen Einfluss potentiell modulierender Faktoren wie soziodemographische Parameter oder Diagnosegruppen zu ermitteln.

Ferner ist zu betonen, dass der Nachweis von unzureichender Anstrengungsbereitschaft gerade in den psychiatrischen Gruppen das Vorliegen einer depressiven, somatoformen oder posttraumatischen Belastungsstörung zwar erheblich in Zweifel ziehen, jedoch noch nicht vollkommen ausschliessen kann. Hier wären Ergebnisse aus Verfahren zur Validierung der Beschwerdenschilderung wie z.B. aus dem MMPI-2, der derzeit als das beste Instrument auf diesem Gebiet gilt und für viele Kulturen validiert ist (Butcher, 1996), wünschenswert gewesen.

Der Umstand, dass die in die Studie eingeflossenen psychiatrische Diagnosen von einem grossen Pool an Psychiatern mit unterschiedlicher gutachterlicher Erfahrung entstammen und die Diagnosen überwiegend klinisch gestellt wurden, limitiert die Abgrenzung zwischen den diagnostischen Gruppen. Die geringe Interrater-Reliabilität bei der Diagnose psychiatrischer Störungen und der Einschätzung der hieraus resultierenden Einschränkung der Funktionsfähigkeit ist hinlänglich bekannt. Die Vielzahl an Mehrfachdiagnosen konstituiert ein weiteres Problem.

6 SCHLUSSFOLGERUNGEN

In evidenzbasierten, versicherungsmedizinischen Begutachtungen ist eine sorgfältige und genaue Datenerhebung unerlässlich. Psychometrische Testverfahren leisten hier einen wesentlichen Beitrag, wenn es um die zufallskritisch abgesicherte Erfassung der geistigen Leistungsfähigkeit und psychischer Beschwerden und damit der funktionellen Leistungsfähigkeit bzw. Arbeitsfähigkeit geht. Hierzu ist es notwendig, dass die Exploranden ehrlich Auskunft erteilen und bei der Bearbeitung von Leistungstests eine hohe Anstrengungsbereitschaft zeigen. Angesichts der erheblichen finanziellen Anreize für die Exploranden, Gesundheitsstörungen zu betonen, wegen derer sie eine Entschädigungs- oder Rentenzahlung beantragen, ist eine Validierung der beklagten Beschwerden unerlässlich. Der optimale Ansatz hierzu besteht in einem auf mehreren Methoden und Datenquellen beruhenden, evidenzbasierten Beurteilungsprozess, der psychometrische Validitätsindikatoren in Form von PVT und SVT miteinschliesst. Eine enge Zusammenarbeit zwischen Psychiatrie und Neuropsychologie auf Augenhöhe bietet hierzu die besten Voraussetzungen.

Die Validierung geltend gemachter Gesundheitsstörungen ist auch deshalb unerlässlich, da eine fundierte Einschätzung der funktionellen Leistungs- und Arbeitsfähigkeit valide Untersuchungsbefunde voraussetzt (siehe auch Frei et al., 2016). Zumindest was die kognitive Leistungsfähigkeit eines Exploranden anbelangt, ist eine hypothetische Verrechnung des Anteils an übertriebener Beschwerdenpräsentation und unzureichender Anstrengungsbereitschaft nicht möglich. Dazu müssten Daten vorliegen, die das Ausmass an Verdeutlichung und/oder Aggravation in bestimmten Validierungsverfahren mit der zugrundeliegenden realen Leistungsfähigkeit in Beziehung setzen. Im Einzelfall kann eine grobe Abschätzung der funktionellen Leistungs- bzw. Arbeitsfähigkeit über die Dauer der innerhalb der neuropsychologischen Untersuchung gewonnenen Belastbarkeit, einzelne im Normbereich liegende Testergebnisse und berichtete Alltagsaktivitäten möglich sein. Dies bildet aber meist nur die untere Grenze der Leistungsfähigkeit ab. Anders als von vielen Schweizer Experten bislang angenommen, erreicht die Grundrate unzureichender Anstrengungsbereitschaft und Aggravation in Gutachtenspopulationen der Schweiz

durchaus eine Grössenordnung, wie sie aus den USA oder auch Deutschland berichtet werden. Dabei variiert die Prävalenz negativer Antwortverzerrung deutlich bis markant in Abhängigkeit von soziodemographischen Faktoren und geltend gemachten Gesundheitsstörungen.

Mit unterschiedlichen Grundraten verändert sich aber auch die Vorhersagekraft eines beliebigen PVT für die verschiedenen Subgruppen an Exploranden. So ist nach den in der aktuellen Studie erhobenen Daten die positive Vorhersagekraft in den selben Verfahren für Exploranden deutscher Muttersprache deutlich geringer als für Migranten aus den Balkanstaaten und der Türkei. Damit muss das diagnostische Prozedere bei der Beschwerdenvalidierung zwingend an den Einzelfall angepasst werden, um falsch positive wie auch falsch negative Urteile zu minimieren. Der Rückgriff auf die Grundrate auffälliger Testergebnisse ist jedoch nicht nur im Bereich der Beschwerdenvalidierung, sondern grundsätzlich bei der Interpretation psychometrischer Testbefunde angezeigt (Frei et al., 2016; Leitlinien zur Klassifikation und Interpretation neuropsychologischer Testergebnisse, Schweizerische Vereinigung der Neuropsychologen, 2014). Wann immer möglich, sollte neben der Performanz, auch die Beschwerdenschilderung validiert werden. Der hierzu in besonderer Weise geeignete MMPI-2 ist inzwischen in verschiedenen Sprachen und Kulturräumen validiert worden, so dass er auch bei diversen Migrantengruppen eingesetzt werden kann.

Während unterzufällige Testergebnisse (d.h. Ergebnisse, die signifikant unter 50% korrekter Antworten liegen) auch bei Migranten Aggravation, in Einzelfällen vielleicht sogar Simulation, belegen, gestaltet sich die Interpretation von Testleistungen, welche sich zwischen Zufallsniveau und allgemein empfohlenen Trennwerten bewegen, bei Migranten noch schwieriger als es ohnehin der Fall ist. Dies liegt unterem daran, dass für diese Personengruppen bislang weder Analogstudien noch Studien mit Patienten mit und ohne erkennbare äussere Anreize für Täuschungsverhalten und somit auch keine spezifischen Trennwerte für bekannte Performanzvalidierungsverfahren publiziert wurden. Hier bedarf es weiterer Forschungsbemühungen, um zu klären, wie und warum Validitätsindikatoren über Kulturen und Sprachen hinweg variieren und wie deren Klassifikationsgenauigkeit in neuen Populationen verbessert werden kann.

Die hier vorgelegte Prävalenzstudie belegt u.a., dass die unter vielen Psychologen und Psychiatern vertretene Ansicht, dass Performanzvalidierungsverfahren nicht geeignet seien, negative Antwortverzerrungen bei psychiatrischen Patientengruppen aufzudecken, sofern diese keine hirnorganische Schädigung aufweisen oder kognitiven Einschränkungen beklagen (Dressing, Forster, Widder, Schneider und Falkai, 2011), hinterfragt werden muss. Auch in der Literatur wird immer wieder betont, dass Exploranden durchaus unterschiedliche Verfälschungsstrategien verfolgen. Nicht zuletzt deshalb wird in internationalen Leitlinien sowohl der Einsatz von Verfahren, welche die Beschwerdenschilderung prüfen, als auch von Performanzvalidierungsverfahren empfohlen. Hier besteht aus Sicht der Autorin im deutschsprachigen Raum der Bedarf an weiteren empirischen Daten, um diese Verfahren an unterschiedlichen psychiatrischen Diagnosegruppen in verschiedenen Settings zu validieren.

Ein weiteres Problem besteht darin, dass kein Test die Intention erfassen kann, die den invaliden Ergebnissen eines Exploranden zugrunde liegen. Diese können bewusst oder unbewusst entstanden sein (siehe hierzu auch die Diskussion in der Publikation 2). Der beste Beweis, um absichtliche negative Antwortverzerrung belegen zu können, besteht derzeit in unterzufälligen Ergebnissen in Antwortwahlverfahren. Dies kommt aber, wie auch die hier vorgestellte Studie zeigt, in der Gesamtpopulation zu begutachtender Exploranden vergleichsweise selten vor. In den anderen Fällen wird die Validität von Testbefunden am ehesten über den Vergleich mit passenden Vergleichsgruppen eingeschätzt, wie sie sich z.B. im WMT oder dem Dot Counting Test für verschiedene neurologische oder psychiatrische Störungsbilder finden. Doch selbst wenn eine absichtliche Antwortverzerrung nachgewiesen werden kann, bleibt die Frage der Motivation hierfür ungeklärt.

In der Praxis erscheint es deshalb angezeigt, eine probabilistische Sprache zu benutzen, d.h. den Verdacht auf Aggravation oder Simulation in Wahrscheinlichkeitsgraden anzugeben, wie es die Kriterien für MND und MPRD vorsehen. In vielen Situationen ist es angemessen, Begrifflichkeiten zu verwenden, die keine Annahmen über die Motivation des Exploranden, invalide Be-

funde zu generieren, beinhalten. Im englischen Sprachraum wird hier der Begriff „feigning“ vorgeschlagen (Bush, Heilbronner & Ruff, 2014; Rogers, 2008), der am ehesten mit „vortäuschen“ zu übersetzen wäre. Da die sich Motivation für negative wie positive Antwortverzerrungen letztlich zwischen den Dimensionen der „Täuschung anderer“ (Aggravation und Simulation) und der „Selbsttäuschung“ (z.B. im Rahmen ausgeprägter somatoformer Störungen) bewegt, wäre er auch im deutschen Sprachraum passend. Leider weckt der Begriff im deutschen Sprachraum beim Laien rasch Assoziationen mit bewusstem und sogar kriminellen Handeln. Der Autorin erscheinen deshalb Begriffe wie „nicht-authentisch Befunde“, „suboptimale Leistungsmotivation“ oder „negative bzw. positive Antwortverzerrung“ angebrachter zu sein. Umgekehrt sollte jedoch Aggravation klar als solche benannt werden, wenn sie denn z.B. durch unterzufällige Ergebnisse in Antwortwahlverfahren belegt wurde. Neben den zu wahrenen Rechten und der zu beachtenden Würde der Exploranden besteht seitens der Gutachter auch eine Verantwortung gegenüber der Gemeinschaft aller Versicherten. Gutachter können und sollten dazu beitragen, dass finanzielle, diagnostische und therapeutische Ressourcen v.a. jenen zu Gute kommen, die diese auch tatsächlich benötigen.

Wie alle einschlägigen Leitlinien und Positionspapiere betonen, sollten Exploranden aus ethischer Sicht grundsätzlich zu Beginn jeder Untersuchung darüber aufgeklärt werden, dass Beschwerdenvalidierungsverfahren zum Einsatz kommen. Hierbei darf jedoch kein Bedrohungsszenario aufgebaut, sondern im Gegenteil eine möglichst vertrauensvolle Atmosphäre geschaffen und der Explorand zu bestmöglicher Kooperation motiviert werden (Iverson, 2006; Bush, Heilbronner & Ruff, 2014; Heilbronner et al., 2009). Die Auswahl eingesetzter Performanz- und Beschwerdenvalidierungsverfahren hat sich dabei nach der empirischen Evidenz und dem konkreten Fall zu richten. Es sollte also ein grösserer Pool an BVT vorrätig sein, um die jeweils passendsten Verfahren auswählen zu können. Auch ist es unerlässlich, dass der Gutachter die Entwicklungen auf dem Gebiet der Beschwerdenvalidierung kontinuierlich mitverfolgt, da solide Befunde zu verschiedenen Patientengruppen häufig erst nach Erscheinen der jeweiligen Testma-

nuale erhoben werden. Keinesfalls sollten Gutachter die Auswahl ihrer Verfahren vom Auftraggeber abhängig machen (Versicherungswirtschaft vs. Patientenanwälte), was leider in der Praxis nicht ausreichend beachtet wird. Gutachter sind also gut beraten, eigene Voreingenommenheit und Anreize in ihrer beruflichen Tätigkeit wiederholt zu reflektieren, um eine ausbalancierte Perspektive wahren zu können.

7 LITERATUR

- Aamodt, M.G. & Custer, H. (2006). Who can best catch a liar? A meta-analysis of individual differences in detecting deception. *The Forensic Examiner, 15 (1)*, 6-11.
- Ardolf, B.R., Denney, R.L. & Houston, C.M. (2007). Base Rates of Negative Response Bias and Malingered Neurocognitive Dysfunction among Criminal Defendants Referred for Neuropsychological Evaluation. *The Clinical Neuropsychologist, 21 (6)*, 899-916.
- Bartlett, M.S., Littlewort, G.C., Frank, M.G. & Lee, K. (2014). Automatic decoding of facial movements reveals deceptive pain expressions. *Current Biology, 24 (7)*, 738-43.
- Berthelson, L., Mulchan, S.S., Odland, A.P., Miller, L.J. & Mittenberg, W. (2013). False positive diagnosis of malingering due to the use of multiple effort tests. *Brain Injury, 27*, 909–916.
- Bianchini, K.J., Greve, K.W. & Glynn, G. (2005). On the diagnosis of malingered pain-related disability: lessons from cognitive malingering research. *The Spine Journal, 5*, 404-417.
- Bolliger, C., Stadelmann-Steffen, I., Thomann, E. & Rüefli, C. (2009). *Migrantinnen und Migranten in der Invalidenversicherung. Verfahrensverläufe und vorgelagerte Faktoren*. Bern: Bundesamt für Sozialversicherungen
- Bond, C.F. & DePaulo, B.M. (2008). Individual differences in judging deception: accuracy and bias. *Psychological Bulletin, 134 (4)*, 477–492.
- Boone, K B. (2011). Clarification or confusion? A review of Rogers, Bender, and Johnson’s a critical analysis of the MND criteria for feigned cognitive impairment: Implications for forensic practice and research. *Psychological Injury and Law, 4*, 157–162.
- Bush, S. S., Heilbronner, R. L. & Ruff, R. M. (2014). Psychological assessment of symptom and performance validity, response bias, and malingering: Official position of the Association for Scientific Advancement in Psychological Injury and Law. *Psychological Injury and Law, 7*, 197–205.
- Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., ... Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology, 20 (4)*, 419–426.

- Butcher, J.N. (Ed.) (1996). *International adaptations of the MMPI-2*. Minneapolis: University of Minnesota Press
- Dressing, H., Foerster, K., Widder, B., Schneider, F. & Falkai, P. (2011) Zur Anwendung von Beschwerdenvalidierungstests in der psychiatrischen Begutachtung. Stellungnahme der deutschen Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN). Stellungnahme Nr. 3. www.dgppn.de
- Dressing, H., Widder, B. & Foerster, K. (2010). Kritische Bestandsaufnahme zum Einsatz von Beschwerdenvalidierungstests in der psychiatrischen Begutachtung. *Versicherungsmedizin*, 62 (4), 163-167.
- Ekman, P. & O'Sullivan, M. (1991). Who can catch a liar? *American Psychologist*, 46 (9), 913-20.
- European Court of Human Rights. In the case of Vukota-Bojić v. Switzerland, Application no. 61838/10. Judgment, Strassbourg, 18. Oktober 2016. Final, 18.02.2017, Zugriff am 13. April 2017 unter <http://hudoc.echr.coe.int/eng?i=001-167490>
- Frei, A., Balzer, C., Gysi, F., Leros, J., Plohm, A. M., & Steiger-Bächler, G. (2016). Kriterien zur Bestimmung des Schweregrades einer neuropsychologischen Störung sowie Zuordnungen zur Funktions- und Arbeitsfähigkeit. *Zeitschrift für Neuropsychologie*, 27 (2), 107-119.
- Green, P. (2003). *Green's Word Memory Test (WMT)*. Edmonton: Green's Publishing.
- Green, P. (2004). *Green's Medical Symptom Validity Test (MSVT)*. Edmonton: Green's Publishing.
- Green, P. (2008). *Green's Non-Verbal Medical Symptom Validity Test (NV-MSVT)*. Edmonton: Green's Publishing.
- Greiffenstein, M.F., Baker, W.J. & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment*, 6, 218-224.
- Greve, K.W., Ord, J., Bianchini, K.J. & Curtis, K.L. (2009). Prevalence of malingering in patients with chronic pain referred for psychological evaluation in medico-legal context. *Archives of Physical Medicine and Rehabilitation*, 90 (7), 1117-1126.

- Hartman, D.E. (2002). The unexamined lie is a lie worth fibbing: neuropsychological malingering and the Word Memory Test. *Archives of Clinical Neuropsychology*, 17 (7), 709-714.
- Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R. & Participants of the Consensus Conference (2009). American Academy of Clinical Neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, 23, 1093–1129.
- Henningsen, P. & Schickel, S. (2016). Begutachtung bei Schmerz und somatoformen Erkrankungen. In W. Schneider, R. Dohrenbusch, H. Freyberger, P. Henningsen et al. (Hrsg.), *Begutachtung bei psychischen und psychosomatischen Erkrankungen. Autorisierte Leitlinien und Kommentare*, 2. überarbeitete und erweiterte Aufl. (S. 307-326). Bern: Hogrefe.
- Henry, M. (2009). Beschwerdenvalidierungstests in der zivil- und sozialrechtlichen Begutachtung: Verfahrensüberblick. In T. Merten & H. Dettenborn (Hrsg.), *Diagnostik der Beschwerdenvalidität. (Praxis der Rechtspsychologie, S. 118-161)*. Berlin: Deutscher Psychologen Verlag.
- Heubrock, D. & Petermann, F. (2000). *Testbatterie zur Forensischen Neuropsychologie (TBFN)*. Frankfurt a. M.: Swets Test Services.
- Iverson, G.L. (2006). Ethical issues associated with the assessment of exaggeration, poor effort, and malingering. *Applied Neuropsychology*, 13, 77–90.
- Keppler, C., Plohmann, A.M., Pflüger, M., Rabovsky, K., Langewitz, W. & Mager, R. (2017). Beschwerdenvalidierung in der versicherungsmedizinischen Begutachtung. *Fortschritte der Neurologie und Psychiatrie*, 85, 17-33.
- Kool, J., Meichtry, A., Schaffert, R. & Rüesch, P. (2008). *Der Einsatz von Beschwerdenvalidierungstests in der IV-Abklärung*. Bern: Bundesamt für Sozialversicherungen
- Larrabee, G.J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist*, 17, 410-425.

- Larrabee, G.J., Greiffenstein, M.F., Greve, K.W. & Bianchini, K.J. (2007). Refining diagnostic criteria for malingering. In G.J. Larrabee (Ed.), *Assessment of Malingered Neuropsychological Deficits*. New York: Oxford University Press.
- Larrabee, G.J. (2009). Malingering scales for the Continuous Recognition Memory Test and Continuous Visual Memory Test. *The Clinical Neuropsychologist*, 23, 167–180.
- Larrabee, G.J. (2014). False-Positive Rates Associated with the Use of Multiple Performance and Symptom Validity Tests. *Archives of Clinical Neuropsychology* 29, 364–373.
- Larrabee, G.J., Millis, S.R., & Meyers, J.E. (2009). 40 plus or minus 10, a new magical number: Reply to Russell. *The Clinical Neuropsychologist*, 23, 841–849.
- Lilienfeld, S.O., Thames, A.D., & Watts, A.L. (2013). Symptom validity testing: Unresolved questions, future directions. *Journal of Experimental Psychopathology*, 4, 78–87.
- Merten, T. (2014). Beschwerdenvalidierung. In A. Thöne-Otto, H. Flor, S. Gauggel, S. Lautenbacher und H. Niemann (Hrsg.), *Fortschritte der Neuropsychologie*. Göttingen: Hogrefe.
- Merten, T. & Dohrenbusch, R. (2016). Psychologische Methoden der Beschwerdenvalidierung. In W. Schneider, R. Dohrenbusch, H. Freyberger, P. Henningsen et al. (Hrsg.), *Begutachtung bei psychischen und psychosomatischen Erkrankungen. Autorisierte Leitlinien und Kommentare*, 2., überarbeitete und erweiterte Aufl. (S. 152-188). Bern: Hogrefe.
- Merten, T. & Rogers, R. (2017). An international perspective on feigned mental disabilities: Conceptual issues and continuing controversies. *Behavioral Sciences and the Law*. 35 (2), 97–112.
- Merten, T., Thies, E., Schneider, K., & Stevens, A. (2009). Symptom validity testing in claimants with alleged posttraumatic stress disorder: comparing the Morel Emotional Numbing Test, the Structured Inventory of Malingering Symptomatology, and the Word Memory Test. *Psychological Injury and Law*, 2, 284–293.

- Miller, H.A. (2005). The Miller-Forensic Assessment of Symptoms Test (M-FAST): Test Generalizability and Utility Across Race, Literacy, and Clinical Opinion, *Journal of Criminal Justice and Behavior*, 32 (6), 591-611,
- Mittenberg, W., Patton, C., Canyock, E.M., & Condit, D.C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, 24, 1094–1102.
- Ott, W., Bade, S., & Wapf, B. (2007). *Nicht zielkonforme Leistungen in der Invalidenversicherung: Bedeutung und Grössenordnung*. Bern: Bundesamt für Sozialversicherungen.
- Plohmann, A.M. (2008). Relevanz neuropsychologischer Gutachten zur Beurteilung von Arbeitsfähigkeit und Integrität. *Epileptologie*, 25, 182–190.
- Plohmann, A.M. (2013a). Common arguments against symptom validity assessment: The psychiatry debate in Germany and Switzerland (Abstract). *Clinica y Salud / Clinical and Health, Journal of Empirical Research in Psychology*, 24 (3), 198.
- Plohmann, A.M. (2013b). First base rate estimates of performance validity in Switzerland (Abstract). *Clinica y Salud / Clinical and Health, Journal of Empirical Research in Psychology*, 24 (3), 200.
- Plohmann, A.M. (eingereicht). Prevalence of poor effort and malingered neurocognitive dysfunction in a litigating sample in Switzerland. *Zeitschrift für Neuropsychologie*
- Rey, A. (1964). *L'examen clinique en psychologie*. Paris: University Press of France.
- Rogers, R. (2008). An introduction to response styles. In Rogers, R. (Ed.), *Clinical assessment of malingering and deception*, (3rd ed.). New York: Guilford.
- Rosenhan, D.L. (1973). On Being Sane in Insane Places. *Science*, 179 (407), 250–258.
- Rubenzler, S. (2009). Posttraumatic stress disorder: Assessing response style and malingering. *Psychological Injury and Law*, 2, 114–142
- Schmand, B., Lindeboom, J. (2005). *Amsterdam Short-Term Memory Test*. Manual Leiden, NL: PITS.

- Schweizerische Vereinigung der Neuropsychologen. (2014). Leitlinien zur Klassifikation und Interpretation neuropsychologischer Testergebnisse. Zugriff am 31. März 2017 www.neuropsych.ch/w/pages/de/leitlinien.php
- Slick, D.J., & Sherman, E.M.S. (2012). Differential diagnosis of malingering and related clinical presentations. In E. M. S. Sherman & B. L. Brooks (Eds.), *Pediatric forensic neuropsychology* (S. 113–135). New York: Oxford University Press.
- Slick, D.J., & Sherman, E.M.S. (2013). Differential diagnosis of malingering. In D. A. Carone & S.S. Bush (Eds.), *Mild traumatic brain injury: System validity assessment and malingering* (S. 57–72). New York: Springer.
- Slick, D.J., Sherman, E.M.S., & Iverson, G.L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, *13*, 545–561.
- Stevens, A., Friedel, E., Mehren, G., & Merten, T. (2008). Malingering and uncooperativeness in psychiatric and psychological assessment: Prevalence and effects in a German sample of claimants. *Psychiatry Research*, *157*, 191-200.
- Sweet, J.J. & Guidotti Breting, L.M. (2013). Symptom validity test research: Status and clinical implications. *Journal of Experimental Psychopathology*, *4* (1), 6-19.
- Thommen A. (2006): Nicht oder schwer objektivierbare Gesundheitsbeeinträchtigungen im medizinischen Alltag: Erfahrungen einer unabhängigen Beraterin im Gesundheitswesen. In Murer E. (Hrsg.), *Nichtobjektivierbare Gesundheitsbeeinträchtigungen: Ein Grundproblem des öffentlichen und privaten Versicherungsrechtes sowie des Haftpflichtrechtes*, Stämpfli Verlag, Bern, 2006.
- Trueblood, W. & Binder L.M. (1997). Psychologists' accuracy in identifying neuropsychological test protocols of clinical malingerers. *Archives of Clinical Neuropsychology*, *12*(1), 13-27.
- Walters, G.D., Berry, D.R., Rogers, R., Payne, J.W., & Granacher, R.J. (2009). Feigned neurocognitive deficit: Taxon or dimension? *Journal of Clinical and Experimental Neuropsychology*, *31*, 584–593.

- Walters, G.D., Rogers, R., Berry, D.R., Miller, H.A., Duncan, S.A., et al.. (2008). Malingering as a categorical or dimensional construct: The latent structure of feigned psychopathology as measured by the SIRS and MMPI-2. *Psychological Assessment*, 20, 238–247.
- Young, G. (2014). *Malingering, feigning, and response bias in psychiatric/psychological injury: Implications for practice and court*. Dordrecht, Netherlands: Springer Science + Business Media.
- Young, G. (2015a). Malingering in forensic disability-related assessments: Prevalence 15 +/- 15%. *Psychological Injury and Law*, 8, 188–199.
- Young, G. (2015b). Detection system for malingered PTSD and related response biases. *Psychological Injury and Law*, 8, 169–183.
- Young, G. (2017). PTSD in Court III: Malingering, assessment, and the law. *International Journal of Law and Psychiatry*, Zugriff am 31. März 2017 unter <http://dx.doi.org/10.1016/j.ijlp.2017.03.001>

ANHÄNGE

A-1 : Prevalence of poor effort and malingered neurocognitive dysfunction in a litigating sample in Switzerland.

A-2 : Beschwerdenvalidierung in der versicherungsmedizinischen Begutachtung.

A-3 : Kriterien zur Bestimmung des Schweregrades einer neuropsychologischen Störung sowie Zuordnungen zur Funktions- und Arbeitsfähigkeit.

ANHANG A-1

Publikation 1:

Plohmann, A.M., Hurter M. (eingereicht). Prevalence of poor effort and malingered neurocognitive dysfunction in litigating patients in Switzerland. *Zeitschrift für Neuropsychologie*

Prevalence of poor effort and malingered neurocognitive dysfunction in litigating patients in Switzerland

Prävalenz unzureichender Anstrengungsbereitschaft und vorgetäuschter neuropsychologischer Störungen in einer Gutachtenspopulation der Schweiz

Andrea M. Plohmann¹, Max Hurter²

¹Practice for Neuropsychology and Psychotherapy, Basel, Switzerland

²Spital Linth, Memory Clinic, Uznach, Switzerland

Prevalence of poor effort and MND

Korrespondenzadresse:

Dipl. Psych. Andrea M. Plohmann, Praxis für Neuropsychologie und Psychotherapie,

Freie Strasse 59, 4001 Basel

E-Mail: andrea.plohmann@neuropsych.ch

Acknowledgments:

This study was supported by a grant from the SUVA research fund. I thank Thomas Zumbunn (Clinical Trial Unit, University Hospital of Basel) for his help with statistical analysis. There are no conflicts of interest.

Abstract:

To determine the prevalence of inauthentic cognitive test results, the data of 455 examinees who had finished at least two performance validity tests (PVT) were analyzed retrospectively. The PVTs administered were the WMT, MSVT, NV-MSVT, ASTM, BSV, RMT, and RDS. Classification as “definite” or “probable” malingering was done according to the Slick criteria. Sociodemographic variables and diagnoses were described using binary logistic regression. Poor effort in at least two PVTs correlated significantly with education levels, immigration, and origin. Irrespective of education level, the highest risk of definite malingering was found in first-generation migrants. Cervical spine dysfunction, normal cerebral imaging, PTSD, somatoform, and/or depressive disorders also correlated with negative response bias. The probability that psychiatric patients fulfill criteria of probable malingering was higher than in patients with isolated organic mental disorders.

Keywords: Malingered cognitive disorders, mental disorders, sociodemographic parameters, independent medical examinations, insurance medicine

Zusammenfassung:

455 Exploranden, welche mindestens zwei Performanzvalidierungsverfahren (PVT) bearbeitet hatten, wurde retrospektiv hinsichtlich der Prävalenz nicht-authentischer Testbefunde analysiert. Zum Spektrum eingesetzter PVT zählten der WMT, MSVT, NV-MSVT, AKGT, BSV, RMT und die Reliable Digit Span. Anhand der Slick-Kriterien erfolgte eine Klassifikation in „sichere“ oder „wahrscheinliche“ Aggravation/Simulation. Assoziationen dieser Gruppen mit demographischen Faktoren und Diagnosen wurden durch einfache logistische Regressionsmodelle beschrieben. Danach korrelierten unzureichende Anstrengungsbereitschaft in mindestens 2 PVT signifikant mit Bildungsniveau, Immigration und Herkunftsregion. Bildungsunabhängig fand sich bei Migranten der 1. Generation ein signifikant höheres Risiko für sichere Aggravation. HWS-Beschwerden, unauffällige zerebrale Bildgebung, Posttraumatische Belastungsstörungen, somatoforme Störung und/oder depressive Störungen korrelierten ebenfalls mit negativer Antwortverzerrung. Psychiatrische Patienten erfüllten mit höherer Wahrscheinlichkeit die Kriterien einer wahrscheinlichen Aggravation als jene mit einer ausschliesslich hirnganischen Störung.

Schlüsselwörter: vorgetäuschte kognitive Störungen, psychiatrische Störungen, soziodemografische Faktoren, Begutachtung, Versicherungsmedizin

Introduction

Malingering is a central issue in medico-legal contexts associated with financial gains. The Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013) defines malingering as “the intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives” (p. 726). The most frequent malingered symptoms in independent medical examinations (IMEs) are cognitive, emotional and/or physical problems related to injuries and illnesses that are difficult to objectify and potentially alter the capacity to work or to care for oneself and others. In Switzerland, the majority of individuals referred to neuropsychological experts in the past had neurological diseases, brain injuries, and developmental disorders. It is only recently that the remit of neuropsychologists has been extended to include the evaluation of cognitive performance in individuals claiming mental disorders like depression, posttraumatic stress disorder, and attention deficit disorders or chronic pain. In the context of litigation, neuropsychological IMEs typically focus on the presence or degree of neuropsychological impairment and functional disabilities, diagnosis, causality, prognosis, and the need for future neuropsychological services.

In this context, there is an ongoing debate in Switzerland about whether neuropsychologists have the competence and expertise to judge an individual’s capacity to work (Frei, et al., 2016; Kieser, 2012, 2015; Plohmman, 2008; BGE 9C_715/2016) and the authenticity of the symptoms (Hoffman-Richter, Jeger, & Schmidt, 2012; BGE 8C_817/2014). Neuropsychologists have to necessarily rely on neuropsychological test data when examining the presence and degree of cognitive deficits in an individual. For the interpretation of the data to be valid, neuropsychologists must make sure that the examinee applied full effort; otherwise, clinicians risk concluding that someone has neuropsychological dysfunctions when that is not the case. It is not possible to detect poor effort or symptom exaggeration unless applying performance and symptom validity

tests during the examination. Looking for consistency amongst different data sources (i.e., self-report, collateral report from family members, and examination performance) or analysing test profiles is essential, but they are insufficient for validating the information gathered, particularly where personal gain is involved. Until recently, European professionals have considered malingering to be a minor problem despite numerous studies that estimate the base rate of poor effort, symptom exaggeration, and malingering to range between 8% and 60% across different settings (i.e., personal injury, disability, criminal, and medical cases) and populations (Greve, Ord, Bianchini, & Curtis, 2009; Greve, et al., 2006; Hout, Schmand, Weking, & Deelman, 2006; Larrabee, 2003; Mittenberg, Patton, Canyock, & Condit, 2002). A first Swiss study, mandated by the Federal Social Insurance Office (FSIO), estimated that the benefit payments inconsistent with invalidity insurance objectives were 56–%. According to this study, the proportion of invalidity pensions that were awarded unjustly due to malingering or the exaggeration of health problems was quite small (Ott, Bade, & Wapf, B., 2007). In an earlier study, Bachmann and D'Angelo (2006) surveyed the 21 cantonal Swiss Invalidity Insurance Offices as a part of a law thesis, and they estimated that insurance fraud ranged from less than 1% (mainly in the French part of Switzerland) to 12–15%, averaging 5.3% across all Invalidity Insurance Offices. At the same time, Thommen (2006) reported data from a small sample claiming non-objectifiable musculoskeletal problems or psychiatric disorders ($N = 67$); this sample was being observed on behalf of insurance companies. In 28% of these cases, definite malingering was proven (they were illegally working more than 25 hours a week). In another 52% of the claimants, “active leisure activities” were observed, and 27% of those claimants returned to work after case management was implemented. Taken together, 55% of the observed claimants – or 7% of the cases in which observation was initially recommended - were malingering or at least exaggerating symptoms.

But, in a second study initiated by the FSIO, which also relied on interviews with medical

experts, including neuropsychologists, still did not identify definite malingering as a frequently encountered problem (Kool, Meichtry, Schaffert, & Ruesch, 2008). The experts' estimates of the a-priori-probability of malingering ranged from less than 1% to a maximum of 10%. The high prevalence rates reported in the literature were ascribed to: the special characteristics of the samples studied; different thresholds between rather unconscious symptom exaggeration and malingering; neglected psychosocial factors; and different social security and legal systems. Only a very small minority of the experts surveyed declared the use of symptom and performance validity tests (SVTs and PVTs). Many of the physicians doubted that using a single PVT or a combination of tests would enhance the diagnostic accuracy in detecting symptom exaggeration or malingering. On the contrary, they feared that using PVTs would raise the rate of false positive diagnoses, incorrectly classifying authentic patients as malingerers, causing a wide-ranging impact on the individuals and their lives. The study strongly recommended investigating the prevalence of exaggeration and malingering in Switzerland as well as the validation and development of PVTs and SVTs for use in the IV assessment procedure. To date, no attempts have been made to collect this data for base rate estimations of malingering.

Research design and aims of the study

The present study is the first Swiss study that intends to provide empirically based data on effort test performance among examinees who are being evaluated in a medico-legal context. Since the study also tried to gather information on potential variables that influence the probability of poor effort and malingering, it was explorative in nature.

The data source was a pool of patients from civil forensic cases who had undergone a neuropsychological examination as part of an interdisciplinary IME between 2005 and 2016 in the German-speaking part of Switzerland. The retrospective analysis included those who had passed at least both a neurological and a psychiatric examination and had also completed at least two out of a set of seven PVTs in the neuropsychological

evaluation. Psychiatric diagnoses were based on clinical judgement of the psychiatrists involved in the respective IMEs, mostly using the AMDP system. Neurological diagnoses were based on a neurological examination, and, if indicated, additional assessments such as EEG, neurophysiological assessment, and CT or MRI scans. Traumatic brain injuries (TBIs) were classified according to Vos et al. (2002). For study purposes, one retrospective modification was made: individuals who had experienced a mild TBI according to the criteria of the EFNS task force mentioned above, but showed signs of intracerebral affections like haematoma, haemorrhages, or diffuse injuries in brain imaging, were classified as moderate TBI. The physicians' diagnoses all relied on the WHO's ICD-10.

Methods

Participants

The total sample comprised 455 cases. Nearly half of the sample (48.8%) were male, 51% were female, and one person identified as transgender. The mean age was 44.1 years (SD 11.8) and the age span was 19 to 75 years. Educational levels ranged from primary school (≥ 4 years) up to a university degree (≥ 13 years). All participants were literate. Most patients were Swiss natives (61.5%), and 8.1% came from the German-speaking neighbouring countries (Germany and Austria). The remaining patients were mostly migrants from Europe (20.7%) and the Middle East (5.9%); 3.6% immigrated from other parts of the world: (Asia (1.1%), North America (0.7%), South America and the Caribbean (0.9%), and Africa (0.9%).

The IME took place up to 54 years after the injury or the onset of the disease, respectively, leading to the compensation claim. Of all the patients, 408 (89.2%) received one or more neurological diagnoses: cervical spine dysfunctions including whiplash injury ($n=208$); mild traumatic brain injury ($n=158$); moderate traumatic brain injury

(*n*=39); severe traumatic brain injury (*n*= 29); cerebrovascular accident (*n*=33); tumours (*n*=8); epilepsy (*n*=24); multiple sclerosis (*n*=5); mild cognitive impairment (*n*=7), or other injuries and diseases with a possible impact on cognition such as encephalopathy, cerebral inflammation, hypoxia, or toxic exposure (*n*=68). In 27.9% of all examinees, imaging techniques (CT/MRI) documented substantial brain injury. In the whole sample, 421 psychiatric diagnoses were allocated: organic, including symptomatic, mental disorders (*n*=245); mental and behavioural disorders due to psychoactive substance use (*n*=54); schizophrenia, schizotypal, and delusional disorders (*n*=4); mood (affective) disorders (*n*=119); anxiety disorders (*n*=28), dissociative (conversion) disorders (*n*=10), stress-related disorders (*n*=50), somatoform disorders (*n*=71) and other neurotic disorders (*n*=12); behavioural syndromes associated with physiological disturbances and physical factors (*n*=23); disorders of adult personality and behaviour (*n*=27); mental retardation (*n*=16); disorders of psychological development (*n*=38); behavioural and emotional disorders with onset usually occurring in childhood and adolescence, mostly hyperkinetic disorders (*n*=24); factors of external origins that encourage subjects to contact health services (Z-codes, *n*=75). Of all patients, 344 (75.6%) complained of chronic pain.

Assessment procedure

All the examinees underwent an expansive neuropsychological assessment lasting from six to eight hours. Due to the forensic context, the battery was comprehensive and included a broad range of performance validity tests. The selection of test measures and the evaluated domains followed national and international guidelines and recommendations for neuropsychological forensic examinations (Bush et al., 2005; Gesellschaft für Neuropsychologie, 2009, 2015; Heilbronner et al. 2009; Schweizerische Vereinigung der Neuropsychologinnen und Neuropsychologen, 2011). Depending on the referral question and previously applied tests, there was some variation in the tests

administered. Everyone in the sample was confronted with at least two performance validity indicators out of a pool of seven described below. One of these PVTs was administered at the beginning of the assessment. In some cases, additional effort indicators were applied; but to restrict variability these additional indicators were not incorporated in the data analysis. The author of this paper tested all patients individually. She was assisted by a neutral translator when assessing persons not fluent in German. For ethical reasons, part of the routine included an oral explanation of the nature of the neuropsychological evaluation. Examinees were told to respond truthfully and put forth good effort on tests. They were also notified that failure to do so could be detected by validity measures that would be used, as recommended by position statements (Bush, Heilbronner, & Ruff, 2014; Iverson, 2006).

Since the study was retrospective, informed consent about data analysis could not be gathered. This project did, however, receive approval from the Swiss Ethical Committee (EKNZ BASEC 2016-00659).

Performance validity measures

Five out of the seven performance validity tests (PVTs) used were forced-choice stand-alone cognitive effort measures. They present themselves as a memory test using either words - such as the *Word Memory Test (WMT)* (Green, 2003), the *Medical Symptom Validity Test (MSVT)* (Green, 2004), or the *Amsterdam Short-Term Memory Test (ASTM)* (Schmand & Lindeboom, 2005) - or pictures – such as the *Non-Verbal Medical Symptom Validity Test (NV-MSVT)* (Green, 2008), or the *Bremer Symptom Validation (BSV)*, which is part of the Test Battery for Forensic Neuropsychology (TFBN, Heubrock & Petermann, 2000)). Forced-choice PVTs (usually) involve the presentation of a to-be-remembered stimulus (target) that the examinee is to remember. The stimulus is then followed by a two-choice recognition trial. These forced-choice PVTs cannot plausibly be failed unless there is bona-fide dementia, moderate to severe aphasia, visual-spatial or executive

dysfunction, or severe mental disorder such as schizophrenia. But even in the absence of memory ability, persons would be able to select on average 50% of the targets correctly by chance alone. Selecting significantly less than 50% correctly on a PVT requires active avoidance of the correct response (Frederick & Speed, 2007; Reynolds, 1998). This procedure thereby makes it possible to detect deliberate under-performance or definite malingered neurocognitive disorders if a patient scores significantly below random chance within a secondary gain context. Cut-off scores and criteria for passing or failing the PVTs were used according to the respective manuals. Whenever possible, the WMT and MSVT were presented in the patient's first language, the ASTM was only given to German native speakers.

The *Rey Memory Test (RMT)*, also known as Rey 15-Items Test (Rey, 1964), is a widely used stand-alone cognitive effort measure. Despite its shortcomings such as poor sensitivity (Iverson & Binder, 2000; Sweet, 1999), other validity problems, and its spreading on the internet, it was routinely used by the psychiatrists involved in the IMEs. They use the RMT because its administration is quick, and it can be administered on any sort of migrants. The RMT is presented to subjects as a very difficult memory test. In fact, the test is very simple because of the redundancy among items, and even patients with significant impairment can perform the test without difficulty. Cut-off scores have been established for patients with brain injuries, neurological diseases, and a variety of mental disorders (Arnett, Hammeke, & Schwartz, 1995; Bernard & Fowler, 1990; Hays, Emmons, & Lawson, 1993; Lee, Loring, & Martin, 1992; Taylor, Kreutzer, & West, 2003). Based on their data, these investigators proposed a cut-off of less than nine to identify individuals exaggerating memory disorders excluding those with psychiatric disorders and mental retardation ($IQ < 70$), for whom they suggested a cut-off score of seven or less. The latter was also the cut-off used in this study.

The *Reliable Digit Span (RDS)* is a malingering detection measure introduced by Greiffenstein, Baker, and Gola (1994). The RDS used in this study was based on the

standard Digit Span, as found in the German adaptations of the Wechsler Intelligence Scales (WIE, von Aster, Neubauer, & Horn, 2006; WAIS-IV, Petermann, 2012), and Wechsler Memory Scales (WMS-R, Härting, et al., 2000; WMS-IV, Petermann & Lepach, 2012). The digit strings were presented in the patient's first language or, if necessary, by an instructed translator, who also recorded the answers. The RDS is one of the best evaluated effort indicators that has been validated in multiple and diverse samples and diagnostic groups (brain injury, toxic exposure, cerebrovascular events, chronic pain disorders, mental retardation, attention-deficit disorder, psychotic disorders and other psychopathologies), and across different research designs (for a review see Schröder, Twumasi-Ankrah, Baade, & Marshall, 2012). Most authors recommend an RDS cut-off of ≤ 7 in patients with MTBI (Jasinski, Berry, Shandera, & Clark, 2011; Larrabee, 2003, 2008) or ≤ 6 for persons claiming pain-related disability (Greve, et al., 2010), patients with temporal lobectomy, people with a variety of mental diseases (Wechsler, 2009), and low IQ groups (Dean, Victor, Boone & Arnold, 2008). Given a sample of examinees with a broad variety of diagnoses, intellectual functioning, and countries of origin, the present study estimated separate prevalence rates of poor effort using an RDS cut-off score of ≤ 6 and ≤ 7 .

Clinical classification methods

Clinical classification was carried out using the diagnostic criteria for malingered neurocognitive disorder (MND) published by Slick, Sherman, & Iverson (1999). This diagnostic system is a well-established method for operationalizing malingering, not only in neuropsychological but also in psychological and medical conditions, including chronic pain (Bianchini, Greve, & Glynn, 2005). According to Larrabee's modification (2007), probable malingering can be supposed when a person fails two or more well-validated PVTs. This classification method was applied for reasons of comparability, knowing well that significant construct validity issues remain, especially regarding intentionality (e.g.,

whether the examinee exaggerates problems consciously or unconsciously). Even the proposed revision and extension of the Slick et al. criteria did not overcome these difficulties (Slick & Sherman, 2013).

In a first step, the number PVTs each patient failed was determined according to the recommended cut-off scores. This was done twice, using a RDS cut-off of ≤ 6 and ≤ 7 respectively. Depending on these results, examinees were assigned to one of two groups. Group one consisted of patients who had passed all or all but one of the PVTs. Group two comprised patients who had failed at least two PVTs (criterion B2 according to Slick, Sherman & Iverson, 1999) or had scored significantly below chance in a forced-choice PVT. Below-chance performance was considered to be evidence of “definite” MND (criterion B1 according to Slick, Sherman, & Iverson, 1999); failing more than two PVTs was considered suggestive of “probable” MND if this behaviour could not fully be accounted for by psychiatric, neurological or developmental factors (criterion D). Since all the examinees were assessed in a medico-legal context, the presence of an external incentive (criterion A) was assumed. Although the Slick criteria also rely on qualitative findings such as inconsistencies to establish a diagnosis, this study relied solely on psychometric findings.

Statistical Analyses

The prevalence of examinees failing at least two PVTs and/or scoring below random chance was estimated with 95% confidence intervals according to Blaker (2000). The same procedures were applied for the groups of patients graded according to diverse sociodemographic variables and different diagnostic groups. The continuous variables age and time since injury or onset of disease were expressed as mean, standard deviation (SD), median and interquartile range (IQR). Age was also separated by decades. For categorical data and the different sample groups, frequencies and percentages were calculated. To quantify how strongly the presence of a certain

sociodemographic variable or evidence of brain damage from imagery data or diagnosis, respectively, was associated with failure in at least two PVTs and/or below-chance performance in a forced-choice test, binary–logistic-regression analyses were performed. To evaluate the possible effects of confounding variables such as sociodemographic factors on the risk of failing two or more PVTs or scoring below random chance in different subsamples, analyses corrected for education, country of origin and pain were done. Statistical analyses were conducted using R (2017).

Results

Out of the 455 patients, probable or definite MND was observed in 27.5 % (95% CI [23.5, 31.7], $n=125$) when an RDS cut-off score of ≤ 6 was used; it became 34.3% (95% CI [30.0, 38.8], $n=156$), when an RDS of ≤ 7 was applied. Definite malingering was detected from below-chance responding on a forced-choice PVT in 16 cases; which was 12.8% of all the examinees who failed two or more PVTs, when using the more conservative cut-off score, and 10.3% of the subsample when adopting an RDS of ≤ 7 . Of the sample, 24.8% (95% CI [21.0, 29.0], RDS ≤ 6) and 28.6% (95% CI [24.6, 32.9], RDS ≤ 7) failed one PVT consistent with insufficient effort, thus meeting criteria possible MND. Depending on the RDS cut-off score, 47.7% (95% CI [43.1, 52.3]) and 37.1% (95% CI [32.7, 41.7]) did not show any psychometric indicators of negative response bias. For detailed results see Table 1.

- Insert Table 1 here -

Relationship with sociodemographic parameters

Looking at sociodemographic parameters, the risk of probable or definite malingering in terms of failing at least two PVTs was only slightly correlated with older age (only using RDS ≤ 7 ; odds ratio (OR) 1.02, 95% CI [1.00, 1.04], $*p < .05$) but considerably associated with low education, belonging to the first generation of migrants, and country of origin. Compared to the Swiss examinees and the total group of German-speaking patients,

the highest *ORs* were found for migrants from the Balkans and from the Middle East (see Figure 1), while *ORs* for people from non-German-speaking European countries were only twofold.

Adjusted odds ratios (*AORs*) that considered origins in the Balkans or the Middle East as well as from those regions plus non-German-speaking European countries were comparable (Figure 2). But after correction for origin, the risk of failing at least two PVTs for patients belonging to the first generation of migrants was much lower. For the second generation of migrants, the probability of failing at least two PVTs was significantly lower compared to the first generation, even when adjusted for country of origin. Though still significantly enhanced, the *AORs* for lowly educated people seemed to be less prominent in both groups.

Adjusting for the influence of less than nine years of education lowered the risk of failing at least two PVTs by approximately 30% for the patients from the Balkans and by approximately 25% for the patients from the Middle East, but their risk was still five to six times higher than the risk of Swiss examinees (Figure 3). Consequently, for the first generation of migrants, the chance of failing two or more PVTs dropped by 22.5–25%, depending on the RDS cut-off used. For patients from other non-German-speaking European countries, *AORs* were roughly 10% lower, and twice as high as the Swiss' risk. When correcting for 9–10 years of education, *AORs* did not differ from unadjusted *ORs* in any of the sociodemographic variables.

Depending on the RDS cut-off, the risk of definite malingering, as detected by below-chance performance in a forced-choice PVT, was 3.7–4.3 times higher in first-generation migrants compared to Swiss patients, especially for people coming from the Balkans or the Middle East. This did not seem to be influenced by education level, since the *ORs* hardly changed when correcting for less than nine years of education (*OR* 3.85, 95% CI [1.06, 18.34], **p* = 0.05; *OR* 4.14, 95% CI [1.17, 19.39], **p* < .05) as well as 9-10 years of education (*OR* 4.06, 95% CI [1.17, 19.441], *p* < .05; *OR* 4.36, 95% CI [1.32, 19.83],

* $p < .05$). But there was a strong association between coming from the Balkans or the Middle East and the risk of below chance performance in a forced-choice PVT (Figure 1). Of definite malingerers 68.8% immigrated from those countries.

Relationship to neurological diagnoses

A roughly 2 times higher risk of failing at least two PVTs was associated with cervical spine dysfunction or whiplash injuries, claimed chronic pain disorders and lack of relevant findings in CTs or MRIs (Figure 4). Interestingly, chronic pain patients who had not been diagnosed as having a somatoform disorder did not show higher ORs. AORs for country of origin and education of less than nine years appeared similar to the ORs.

Associations to psychiatric diagnoses

As for neurological diseases, there was no association between having received a psychiatric diagnosis and the probability of scoring below chance in a forced-choice PVT. But examinees who had received a psychiatric diagnosis showed a threefold higher risk of failing at least two PVTs than other patients. By contrast, the risk of examinees with none but “mental disorders due to known physiological conditions” (ICD-10 F0) failing two or more PVTs was only a third of the risk of patients who received other psychiatric diagnoses and half of the risk of patients with brain affections and an additional psychopathology (Figure 5), independent of the RDS cut-off score. These relationships showed up in a comparable manner after adjusting for country of origin or education.

Individuals diagnosed as suffering from an affective [depressive] disorder displayed an almost twofold chance of failing at least two PVTs compared to those who had a different psychiatric diagnosis. Correcting for country of origin resulted in 30–35% lower AORs that were no longer significant. Depending on the RDS cut-off score, this loss was less prominent when adjusting for less than nine years of education and completely absent if education lasted from 9–10 years (not shown due to lack of space). The probability of failing at least two PVTs was even higher for those who had received a diagnosis in the

spectrum of “anxiety, dissociative, stress-related, somatoform and other nonpsychotic mental disorders”. Examining the different disorders belonging to this group more closely revealed that individuals who were diagnosed as suffering from a somatoform disorder were 3.5 times more likely to fail at least two PVTs than patients with other mental disorders. This risk decreased about 22% when adjusting for country of origin (and hardly when correcting for education), but it was still significant. Examinees diagnosed with a posttraumatic stress disorder (PTSD) also failed at an almost three times higher rate (*OR* 2.77), when adopting an RDS of ≤ 7 . Adjusting for country of origin lowered this risk by 20–30%, while adjusting for education left the *OR* unaffected. Suffering from “pervasive and specific developmental disorders” (ICD-10 F8) or “behavioural and emotional disorders with onset usually occurring in childhood and adolescence” (ICD-10 F9), especially “hyperkinetic disorders”, was associated with a significantly lower risk of failing two or more PVTs compared to patients with other mental disorders. The lower risk was less prominent in patients suffering from developmental disorders when adjusting for heritage. By contrast, “factors influencing health status and contact with health services” (ICD-10 Z-codes) roughly doubled the risk of failing at least two PVTs, regardless of origin or education.

Discussion

The present study aimed to examine the base rate of PVT failure in an international litigating sample referred to a neuropsychological expert in the context of an IME in Switzerland. It used explicit operationalisations of the malingering of neurocognitive disorders and data from the assessment of patients with financial incentive to estimate the prevalence of malingering and poor effort. It also focused on sociodemographic and clinical parameters that correlated with a higher prevalence of probable or definite malingering. In this litigating sample, the overall rate of probable or definite MND ranged from 27.5% to 34.2% depending on the selected RDS cut-off scores. Base rates varied

significantly as a function of sociodemographic factors, (i.e., education, migration background and country of origin), and neurological, and psychiatric diagnoses. It is important to note that education and origin increased the risk of failure independently from each other. Of the failing patients, 10.3–12.8%, or 3.5% of the total sample, scored below chance in at least one PVT, thus proving definite malingering. Compared to international data, the overall base rate for definite MND found in this study was lower than the average rate of 7% calculated by Young (2015) across a variety of studies, but it was equivalent to the magnitude of confirmed insurance fraud in Switzerland found by Thommen (2006). Regarding the country of origin, the absolute and relative frequency of definite malingering was highest in people from Turkey and the Balkans. As reported by Bolliger, Stadelmann-Steffen, Thomann, and Rüefli (2009), the proportion of applications for disability pensions amongst people who immigrated from those countries is significantly higher as the proportion of applications among migrants from other European countries and twice as high as the proportion of applications among Swiss natives.

The overall base rate of probable and definite MND of roughly 30% conforms to prevalence rates reported in numerous studies with suggested rates of approximately 40% among people with mild traumatic brain injury (TBI) who are receiving disability payments or are in litigation (e.g., Larrabee, 2003) or 20–50% in chronic pain patients (e.g., Greve, Ord, Bianchini, & Curtis, 2009), when using two PVTs. Applying only one PVT, European researchers found cognitive underperformance in 33–61% of litigating patients after whiplash injury (Schmand, et al., 1998), 25.3% of TBI-patients (Bodenburg, 2014), 44.8% of a sample of mixed German patients who were assessed in the context of compensation claims (Stevens, Friedel, Mehren, & Merten (2008), and almost 34% of a sample of mixed psychiatric patients (Dandachi-Fitzgerald, Ponds, Peters & Merckelbach, 2011).

Associations to diagnoses

In contrast to Stevens' data, which does not show a correlation between complaints and the risk of failing a PVT or SVT, in the present study, patients with cervical spine dysfunction or whiplash injuries, claimed chronic pain disorders, or a lack of relevant findings in CT or MRI (when using an RDS cut-off score of ≤ 7) displayed the highest risk of failing two or more PVTs compared to other patients with a neurological diagnosis. This is in line with the literature, which suggests brain damage does not usually account for failing an effort test. Psychiatric diagnoses other than "mental disorders due to physiological conditions" increased the risk of failure threefold. This was especially true for depressive, somatoform and posttraumatic stress disorders.

In the subsample of patients diagnosed as suffering from PTSD, the prevalence of failing two or more PVTs was higher in this study (43.5%) than in a Canadian study that reported failure rates of 14% on two or more PVTs (Demakis, Gervais, & Rohling, 2008), and 29% on at least one PVT. In a sample of litigating patients in Germany (Merten, Thies, Schneider, & Stevens, 2009), 23% failed the RDS and 51% the WMT, at least 13.1%, but no more than 23%, failed both. Unfortunately, the exact rate of examinees failing both PVTs was not reported. The different failure rates between studies reflect differences in the sensitivity and specificity of different PVTs, different strategies used by patients with invalid response styles, and, perhaps, also different compositions of the samples studied.

The risk of failure was not significantly modified by stratification for education in any of these groups. Adjusting for origin lowered the correlation between each of these clinical diagnoses and probable malingering, but patients who had been diagnosed with a somatoform disorder still showed a significantly higher risk of failure than the total group of patients with mental disorders. The lack of an association between the risk of failing PVT and the country of origin in the risky neurological subsamples is most probably tied to the low proportion of migrants from the Balkans and the Middle East (16.5–21.8%) in

these subsamples. The proportion of examinees from those countries in the subsamples diagnosed as having an affective, somatoform or posttraumatic stress disorders had been approximately twice as high (35.6–43.5%).

Influence of sociodemographic features

Immigration, country of origin, and low education significantly increased the risk of failing at least two PVTs. A higher proportion of patients with *low education* (≤ 11 years of education) among those who failed effort tests was also found in other studies (e.g. Stevens, Friedel, Mehren, & Merten, 2008). This raises the question whether low education or being mentally handicapped per se could account for failing the PVTs used in the current study. Regarding MSVT, several researchers have shown (Green, 2004; Richman et al., 2006) that even children with developmental disabilities (mean verbal IQ of 65) exhibited 95% correct responses in MSVT effort subtests. Also, Chafetz, Prentkowsky, and Rao (2011) compared three different groups of low-functioning individuals: (1) people seeking compensation for an inability to work (Social Security Disability group); (2) people seeking to work to obtain compensation (vocational rehabilitation group); and (3) parents with a low IQ in the child-protection system who were seeking to have their children returned to them. There were no PVT failures (including MSVT and RDS) in the child-protection sample. Using profile analysis, the NV-MSVT did not show any false-positive classifications in three independent groups of patients with severe cognitive impairment arising from dementia (Green, 2011). People with low education would thus most probably not be erroneously classified as showing low effort. In summary, in the current sample false-positive (FP) classifications (that is, a patient who is identified as giving suboptimal effort when s/he is actually not) due to low education are highly unlikely.

A slight age effect as seen in the present study, where older individuals have a higher probability of being classified as malingerers, was also reported by Chafetz (2011), Gasquoine, Weimer, and Amador (2017), and others. Since age effects have also been

found in healthy subjects (Giger & Merten, 2013; Zenisek, Millis, Banks, & Miller, 2016), a few PVTs may indeed inherit a risk of FP classification for older subjects.

Regarding *cross-cultural applications of published PVT cut-off scores*, most of which were validated in the United States, a critical attitude seems advisable if no validation studies for the population of interest are available. For a representative, healthy adult sample of the German-speaking Swiss population, Giger and Merten (2013) examined the influence of demographic variables on scores of the effort indicators that were also used in this study. Their sample was comparable to the sample of Swiss natives evaluated here in terms of education, gender distribution, and age. In healthy subjects, sex and education had no effect on PVT scores. Only age was a relevant factor for MSVT IR and CNS scores, but it had no effect on the failure rate. Similarly, Rienstra, Spaan, and Schmand (2009) and Schmand and Lindeboom (2005) demonstrated robustness against such distorting effects for the Dutch versions of WMT and ASTM when adapted into other languages within the Western World. In their healthy Swiss sample, Giger and Merten found possible false-positive rates of only 2% on RDS, ASTM, and none on the MSVT after applying profile analyses as recommended by its author (Green, 2004). Using the RDS, MSVT, and ASTM for German-speaking Swiss and German patients thus seems reasonable according to multiple studies (Blaskewitz & Merten, 2006; Merten, Green, Henry, Blaskewitz, & Brockhaus, 2005). In their recently published review of feigning instruments with linguistically, ethnically, and culturally diverse samples, Nijdam-Jones and Rosenfeld (2017) attributed the strongest classification accuracy to ASTM and MSVT with specificities above 90%. For the WMT, the mean specificity was .84 across multiple studies in different languages: German (Brockhaus & Merten, 2004; Merten, Bossink, & Schmand, 2007; Merten, Henry, & Hilsabeck, 2004; Merten, Lorenz, & Schlatow, 2010), Russian (Tydecks, Merten, & Gubbay, 2006), Portuguese (Martin & Martin, 2010), and Hebrew (Hegedish & Hoofien, 2013). The WMT also showed a greater range of variance than ASTM and MSVT. Due to high failure rates

on the WMT in some studies that had been solely relying on Criterion A (pass/fail easy subtests), there is an ongoing debate on whether WMT cut-off scores should be adjusted, especially for clinical groups with significant cognitive impairment (Greve, Ord, Curtis, Bianchini, & Brennan, 2008). Irrespective of the cut-off used, clinicians should apply profile analyses (Green, 2009) to lower the risk of false-positive classifications. Hampson, Kemp, Coughlan, Moulin, and Bhakta (2014) showed that applying profile analysis indeed almost eliminated the risk of false-positive judgments in a group of presumed bona fide patients with brain injury or epilepsy. But Green, Montijo, and Brockhaus (2011) stated that the profile alone cannot be relied upon to make judgments about poor effort. Only someone with clear clinical signs or a diagnosis that implies severe impairment should be considered as putting forth full effort despite showing a genuine memory-impairment profile. Apart from its borderline acceptable specificity (mean specificity .89), the Rey 15-items test proved to be rather insensitive across three studies of Spanish-speaking people (Burton, Vilar-López, & Puente, 2012; Strutt, Scott, Shrestha, & York, 2011; Vilar-López, Gómez-Río, Caracuel-Romero, Llamas-Elvira, & Pérez-García, 2008) and one German study (Schiemann, 2003). As can be seen, the adequacy of PVT cut-off scores is closely related to the population studied. Unfortunately, there is a big gap of data concerning the classification accuracy of PVTs for people who speak language other than English, German or Spanish. For this reason, the present study's prevalence estimates of poor effort in patients who speak other languages, which also implies another cultural background, might be overestimated.

There is another point to be made here. It is only recently, that *bilingualism* in performance validity assessment has come into the focus of some researchers. While language expression and comprehension in bilinguals seem to activate both languages in parallel (Baum & Titone, 2014), causing interlanguage interference that slows processing time and increases the possibility of errors, performance on nonverbal cognitive tests is supposed to remain unaffected by bilingualism (Gasquoine &

Gonzalez, 2012). Most recently, Gasquoine, Weimer and Amador (2017) have investigated specificity rates for non-clinical, bilingual, Mexican Americans on visual-perceptual tests like the Test of Memory Malingering (TOMM) and Dot Counting as well as for RDS as a verbal test, using cut-scores developed for monolingual English speakers. Failure rates and possible FP classifications were low on the nonverbal tasks. In contrast, for RDSm average failure rates were 38% ($RDS \leq 7$) and 15% ($RDS \leq 6$), which is unacceptably high, even though RDS was administered in the first language. Their results confirmed previous studies of performance validity assessment among people who speak English and another language bilingually, such as Spanish-English speakers or Native Americans (Salazar, Lu, Wen, & Boone, 2007; Schroeder, Twumasi-Ankrah, Baade, & Marshall, 2012). These authors consistently recommended not to use RDS as a performance validity measure with these patients.

I thus seemed necessary to take a closer look on the RDS data on the current subsample of migrants who speak German as a second language. Of 138 patients with a first language other than German, 60 failed RDS using the more conservative cut-off score of ≤ 6 . Twelve of them (20%) only failed the RDS and were thus indeed being potentially misclassified as possible malingerers according to original MND criteria, despite the fact that they possibly might have shown full effort. Nevertheless, these cases did not affect the prevalence estimates, since examinees who only failed one PVT were not included in the analysis. Ten (22.2%) of the remaining 45 migrants had to be classified as true malingerers since they performed significantly below chance in at least one forced-choice PVT. Eight of those who failed a forced-choice test besides the RDS had been examined with either WMT or MSVT in their first language, and were not given a nonverbal PVT. Six of them failed the verbal PVT very clearly, one subject scored below random chance, the remaining five showed results in the guessing range. Applying the recommended profile analysis of the MVST and WMT, and using additional clinical information, none of those cases was inaccurately classified as showing poor effort.

Some researchers have described significant differences for *ethnicity*. Johnson-Greene, Brooks, & Ference (2013) investigated patients who had applied for or were on disability due to a somatoform disorder (fibromyalgia) and found higher proportions of Hispanics failing two PVTs in comparison to Caucasians and Afro-Americans. He suggested a link between lower socioeconomic status, the likelihood of a disability claimm', and failing PVTs. With regard to the current study, one should bear in mind that since most of the migrants are manual workers, they are at a higher risk of losing their jobs, especially with increasing age. Receiving a pension and thus being labelled as ill is more socially acceptable than being unemployed, or receiving money from the social welfare department. In this sense, malingering would be adaptive according to Rogers (1990). As Ott, Bade, and Wapf (2007) stated, receiving a disability pension in Switzerland entails access to further payments from the pension fund and, when one has children, additional annuities for them. As a result, those payments often add up to "an income that exceeds wages and welfare, thus posing an enormous incentive for moral hazard" (pp. 87–88). Alternate life scripts may also trigger malingering: in the Balkans and the Middle East, people tend to start working very early in life, but they often retire starting in the age of 50. Moreover, life expectancy is shorter than in Central or Western Europe. With that said, Ott, Bade, and Wapf (2007) as well as Bolliger, Stadelmann-Steffen, Thomann, and Rüepli (2009) considered poorly trained migrants, aged 50 years and above as a special at-risk group. The present study confirmed this assumption. Additionally, it must be stated that many migrants bear a high psychosocial burden such as living apart from their wives and children, being obliged to take care of family members living in their country of origin, and experiencing the conflicting priorities of different cultures. Perhaps they are exposed to an increased risk of developing stress-related diseases. For a further discussion of possible motivations for exaggeration in migrants, see also Hausotter and Schouler-Ocak (2007).

Limitations

In general, the accuracy of prevalence estimates is heavily dependent on sampling methodology and sample size; larger samples are more likely to be representative of the population of interest than smaller samples. The present sample was large, but some of the subsamples were relatively small (e.g. for certain neurological and mental disorders), so the margin of error will tend to be larger in those subgroups. Furthermore, the variability of sample sizes limits the comparability of prevalence rates between subgroups of patients.

The present findings are supported by the fact that they are in line with the literature. But because the current study only included patients who were referred for neuropsychological evaluations, the prevalence of malingering and poor effort may be lower than in all the patients who are seen for medico-legal examinations in Switzerland. This is especially true for patients who primarily claim mental disorders other than injuries or diseases affecting the brain, since only a small proportion of them were sent for a neuropsychological assessment. On the other hand, in this retrospective analysis, the classification of probable or definite malingerers was mainly based on psychometric data concerning effort, which overestimates the rate of malingering. A measure for psychopathological malingering would have greatly helped to overcome this shortcoming, in particular the Minnesota Multiphasic Personality Inventory-2. It is the best suited instrument for this aspect of controlling, and is well validated across various cultures (Butcher, 1996). This way of establishing criterion-based known groups has its flaws like weak internal validity, while external validity is moderately strong. It must be stressed that for clinical purposes, failing PVTs should not be equated with conscious malingering. Diagnosing malingering in an individual necessarily warrants integrating diverse clinical information.

Another limitation consists in the variety of psychiatrists involved. They established diagnoses by clinical judgment rather than standardized instruments, as for example

SCID or CIDI. Given the fact that, especially in psychiatry, interrater reliability is known to be low, there might have been some overlap between diagnostic groups.

Conclusion

Malingering of cognitive dysfunction in litigating patients is far more frequent than expected by leading experts in Switzerland and seems to be significantly correlated with sociodemographic factors and the kind of claimed health problems. In contrast to the beliefs of many Swiss experts, definite and probable MND is not limited to health problems with potentially inherent brain pathology or cognitive dysfunction. Performance validity assessment (and symptom validation) should therefore become a clinical routine, at least for all kinds of claimed mental disorders as has been strongly recommended in different book chapters, position papers, and guidelines (Bush et al., 2005; Bush, Heilbronner, & Ruff, 2014; Heilbronner et al, 2009; Rogers, 2008). Due to the multiple domains in which symptoms and impairment are presented, these evaluations are complicated. But collaboration between psychiatry and neuropsychology can provide the optimal multi-method approach that is needed for thorough assessment in medico-legal settings. Careful symptom and performance validation is crucial when one considers the costs of healthcare for performing unnecessary treatments, the unequitable distribution of disability and compensation resources, and the reduced access to health resources for patients who really need them. The latter is well reflected in current discussions about the lack of resources for psychiatric and neuropsychological treatment in Switzerland.

The extremely high percentage of definite malingering in first-generation migrants who come from certain countries needs to be verified in further studies. In principle, more research is necessary to determine how and why effort indicators vary across cultures and languages, and what can be done to improve their classification accuracy when used with novel populations. Further stratification of study populations is also needed in order to assess the influence of socioeconomic variables. To quote Merten and Rogers

(2017): “One promising solution would be the development of international research programs that could represent the interests of individual European countries through a systematic collaboration on key issues of language and culture. With the refugee crises and other trends in migration, this international focus would also be capable of addressing critically important transnational issues that will continue to confront Europe in the coming decades.” (“A Snapshot on European Trends”, last para.)

References

- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders (DSM-5) (5th ed.)*. Washington, DC: APA.
- Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie (AMDP) (Hrsg.). (2007). *Das AMDP-System* (8. überarbeitete Auflage). Göttingen: Hogrefe.
- Arnett, P. A., Hammeke, T. A., & Schwartz, L. (1995). Quantitative and qualitative performance on Rey's 15-item test in neurological patients and simulators. *Clinical Neuropsychologist*, 9, 17–26.
- Babikian, T., Boone, K. B., Lu, P., & Arnold, G. (2006). Sensitivity and specificity of various digit span scores in the detection of suspect effort. *The Clinical Neuropsychologist*, 20, 145–159.
- Bachmann, R. & D'Angelo, M. (2006): *Die Bekämpfung des Versicherungsmissbrauches (BVM) in der Invalidenversicherung unter besonderer Berücksichtigung der Beweissicherung vor Ort (BvO)*. Luzern: Diplomarbeit Hochschulzyklus Sozialversicherungsmanagement.
- Baum, S., & Titone, D. (2014). Moving towards a neuroplasticity view of bilingualism, executive control, and aging. *Applied Psycholinguistics*, 35, 857–894.
- Bernard, L. C., & Fowler, W. (1990). Assessing the validity of memory complaints: Performance of brain-damaged and normal individuals on Rey's task to detect malingering. *Journal of Clinical Psychology*, 46, 434–436.
- Bundesgerichtsentscheid, BGE 9C_715/2016 (2017). Retrieved February 27, 2017, from <http://www.bger.ch/index/jurisdiction/jurisdiction-inherit-template/jurisdiction-recht/jurisdiction-recht-urteile2000.htm>
- Bundesgerichtsentscheid, BGE 8C_817/2014 (2015). Retrieved June 3, 2015, from <http://www.bger.ch/index/jurisdiction/jurisdiction-inherit-template/jurisdiction-recht/jurisdiction-recht-urteile2000.htm>
- Bianchini, K. J., Greve, K. W., & Glynn, G. (2005). On the diagnosis of malingered pain-

- related disability: lessons from cognitive malingering research. *The Spine Journal*, 5, 404–417.
- Blaker, H. (2000). Confidence curves and improved exact confidence intervals for discrete distributions. *Canadian Journal of Statistics*, 28, 783–798.
- Blaskewitz, N., & Merten, T. (2006). Validität und Reliabilität von Beschwerdvalidierungstests und -indikatoren: Eine experimentelle Studie. *Zeitschrift für Neuropsychologie*, 17, 35–44.
- Bodenburg, S. (2014). Die Standardabweichung der Reaktionszeiten als ein Maß für negative Antwortverzerrungen. *Zeitschrift für Neuropsychologie*, 25 (2), 89–97.
- Bolliger, C., Stadelmann-Steffen, I., Thomann, E., & Rüefli, C. (2009). *Migrantinnen und Migranten in der Invalidenversicherung: Verfahrensverläufe und vorgelagerte Faktoren*. Bern: Bundesamt für Sozialversicherungen
- Brockhaus, R., & Merten, T. (2004). Neuropsychologische Diagnostik suboptimalen Leistungsverhaltens mit dem Word Memory Test. *Der Nervenarzt*, 75, 882–887.
- Burton, V., Vilar-López, R., & Puente, A. E. (2012). Measuring effort in neuropsychological evaluations of forensic cases of Spanish speakers. *Archives of Clinical Neuropsychology*, 27, 262–267.
- Bush, S. S., Heilbronner, R. L., & Ruff, R. M. (2014). Psychological assessment of symptom and performance validity, response bias, and malingering: Official position of the Association for Scientific Advancement in Psychological Injury and Law. *Psychological Injury and Law*, 7, 197–205.
- Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., ... Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology*, 20 (4), 419–426.
- Butcher, J. N. (Ed.) (1996). *International adaptations of the MMPI-2*. Minneapolis: University of Minnesota Press.

- Chafetz, M. D., Prentkowski, E., & Rao, A. (2011). To work or not to work: Motivation (not low IQ) determines SVT findings. *Archives of Clinical Neuropsychology*, 26, 306–313.
- Dandachi-Fitzgerald, B., Ponds, R. W. H. M., Peters, M. J. V., & Merckelbach, H. (2011). Cognitive underperformance and symptom over-reporting in a mixed psychiatric sample. *The Clinical Neuropsychologist*, 25 (5), 812–828.
- Dean, A. C., Victor, T. L., Boone, K. B., & Arnold G. (2008). The relationship of IQ to effort test performance. *The Clinical Neuropsychologist*, 22, 705–722.
- Demakis, G. J., Gervais, R. O., & Rohling, M. L. (2008). The effect of failure on cognitive and psychological symptom validity tests in litigants with symptoms of post-traumatic stress disorder. *The Clinical Neuropsychologist*, 22, 879–895.
- Frederick, R. I., & Speed, F. M. (2007). On the interpretation of below-chance responding in forced-choice tests. *Assessment*, 14, 31–1.
- Frei, A., Balzer, C., Gysi, F., Leros, J., Plohmann, A. M., & Steiger-Bächler, G. (2016). Kriterien zur Bestimmung des Schweregrades einer neuropsychologischen Störung sowie Zuordnungen zur Funktions- und Arbeitsfähigkeit. *Zeitschrift für Neuropsychologie*, 27 (2), 107–119.
- Gasquoine, P. G., & Gonzalez, C. D. (2012). Using monolingual neuropsychological test norms with bilingual Hispanic Americans: Application of an individual comparison standard. *Archives of Clinical Neuropsychology*, 27, 268–276.
- Gasquoine, P. G., Weimer, A. A., & Amador, A. (2017). Specificity rates for non-clinical, bilingual, Mexican Americans on three popular performance validity measures. *The Clinical Neuropsychologist*, 31(3), 585–597.
- Giger, P., & Merten, T. (2013). Swiss population-based reference data for six symptom validity tests. *Clinica y Salud*, 24, 153–159.
- Green, P. (2003). *Green's Word Memory Test (WMT)*. Edmonton: Green's Publishing.

- Green, P. (2004). *Green's Medical Symptom Validity Test (MSVT)*. Edmonton: Green's Publishing.
- Green, P. (2008). *Green's Non-Verbal Medical Symptom Validity Test (NV-MSVT)*. Edmonton: Green's Publishing.
- Green, P. (2009). *The advanced interpretation program for the WMT, MSVT, NV-MSVT and MCI*. Edmonton: Green's Publishing
- Green, P. (2011). Comparison between the Test of Memory Malingering (TOMM) and the Nonverbal Medical Symptom Validity Test (NV-MSVT) in adults with disability claims. *Applied Neuropsychology, 18* (1), 18–26.
- Green, P., Montijo, J., & Brockhaus, R. (2011). High specificity of the Word Memory Test and Medical Symptom Validity Test in groups with severe verbal memory impairment. *Applied Neuropsychology, 18*, 86–94.
- Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment, 6*, 2182–24.
- Greve, K. W., Bianchini, K. J., Black, F. W., Heinly, M. T., Love, J. M., Swift, D. A., & Ciota, M. (2006). The prevalence of cognitive malingering in persons reporting exposure to occupational and environmental substances. *Neurotoxicology, 27* (6), 940–950.
- Greve, K. W., Bianchini, K. J., Etherton, J. L., Meyers, J. E., & Curtis, K. L. (2010). The Reliable Digit Span Test in chronic pain: Classification accuracy in detecting malingered pain-related disability. *The Clinical Neuropsychologist, 24*, 1371–52.
- Greve, K. W., Ord, J., Curtis, K. L., Bianchini, K. J., & Brennan, A. (2008). Detecting malingering in traumatic brain injury and chronic pain: A comparison of three forced-choice symptom validity tests. *The Clinical Neuropsychologist, 22*, 896–918
- Greve, K. W., Ord, J., Bianchini, K.J., & Curtis, K. L. (2009). Prevalence of malingering in patients with chronic pain referred for psychological evaluation in medico-legal context. *Archives of Physical Medicine and Rehabilitation, 90* (7), 11171–126.

- Härting, C. Markowitsch, H. J., Neufeld, H., Calabrese, P., Deisinger, K. & Kessler, J. (Hrsg.). (2000). *WMS-R. Wechsler Gedächtnistest – Revidierte Fassung*. Bern: Verlag Hans Huber.
- Hampson, N. E., Kemp, S., Coughlan, A. K., Moulin, C. J. A., & Bhakta, B. B. (2014). Effort test performance in clinical acute brain injury, community brain injury, and epilepsy populations. *Applied Neuropsychology: Adult*, 21, 1831–94.
- Hausotter, W., & Schouler-Ocak, M. (2007). *Begutachtung bei Menschen mit Migrationshintergrund unter medizinischen und psychologischen Aspekten*. München: Elsevier.
- Hays, J. R., Emmons, J., & Lawson, K. A. (1993). Psychiatric norms for the Rey 15-item visual memory test. *Perceptual and Motor Skills*, 76, 1331–1334.
- Hegedish, O., & Hoofien, D. (2013). Detection of malingered neurocognitive dysfunction among patients with acquired brain injuries: A Word Memory Test study. *European Journal of Psychological Assessment*, 29, 253–262.
- Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, 23, 1093–1129.
- Heubrock, D., Petermann, F. (2000). *Testbatterie zur Forensischen Neuropsychologie (TBFN)*. Frankfurt am Main.: Swets Test Services.
- Hoffmann-Richter, E., Jeger, J., & Schmidt, H. (2012). *Das Handwerk ärztlicher Begutachtung. Theorie, Methodik, Praxis*. Stuttgart: Kohlhammer.
- Hout, M. S. E., Schmand, B., Weking, E. M., & Deelman, B. G. (2006). Cognitive functioning in patients with suspected chronic toxic encephalopathy: Evidence for neuropsychological disturbances after controlling for insufficient effort. *Journal of Neurology, Neurosurgery and Psychiatry*, 77, 296–303.

- Iverson, G. L. (2006). Ethical issues associated with the assessment of exaggeration, poor effort, and malingering. *Applied Neuropsychology, 13*, 77–90.
- Iverson, G. L., & Binder, L. M. (2000). Detecting exaggeration and malingering in neuropsychological assessment. *Journal of Head Trauma Rehabilitation, 15*, 829–858.
- Jasinski, L. J., Berry, D. T. R., Shandera, A. L., & Clark, J. A. (2011). Use of the Wechsler Adult Intelligence Scale Digit Span subtest for malingering detection: A meta-analytic review. *Journal of Clinical and Experimental Neuropsychology, 33*, 300–314.
- Johnson-Greene, D., Brooks, L., & Ference, T. (2013). Relationship between performance validity testing, disability status, and somatic complaints in patients with fibromyalgia. *The Clinical Neuropsychologist, 27*, 148–158.
- Kieser, U. (2012). Neuropsychologie – Stellenwert und Bedeutung in der sozialversicherungsrechtlichen Rechtsprechung des Bundesgerichts. *Jahrbuch zum Sozialversicherungsrecht, 1671–69*.
- Kieser, U. (2015). *Gutachten zu Fragen des Vorgehens bei der Bestimmung der Arbeitsfähigkeit im Sozialversicherungsrecht und zum allfälligen Beitrag der Neuropsychologie*. Retrieved from <https://www.neuropsychy.ch/de/fachpersonen/qualitaetsicherung>
- Kool, J., Meichtry, A., Schaffert, R., & Rüesch, P. (2008). *Der Einsatz von Beschwerdenuvalidierungstests in der IV-Abklärung*. Bern: Bundesamt für Sozialversicherungen
- Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist, 17*, 410–425
- Larrabee, G. J. (Ed.) (2007). *Assessment of Malingered Neuropsychological Deficits*. New York: Oxford University Press.

- Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist*, 22, 666–679.
- Larrabee, G. J. (Ed.) (2012). *Forensic neuropsychology: A scientific approach*. New York: Oxford University Press.
- Lee, G. P., Loring, D. W., & Martin, R. C. (1992). Rey's 15-item visual memory test for the detection of malingering: Normative observations on patients with neurological disorders. *Psychological Assessment*, 4, 43–46.
- Martins, M., & Martins, I. P. (2010). Memory malingering: Evaluating WMT criteria. *Applied Neuropsychology*, 17, 177–182.
- Merten, T., Bossink, L., & Schmand, B. (2007). On the limits of effort testing: Symptom validity tests and severity of neurocognitive symptoms in nonlitigant patients. *Journal of Clinical and Experimental Neuropsychology*, 29, 308–318.
- Merten, T., Green, P., Henry, M., Blaskewitz, N., & Brockhaus, R. (2005). Analog validation of German-language symptom validity tests and the influence of coaching. *Archives of Clinical Neuropsychology*, 20, 719–726.
- Merten, T., Henry, M., & Hilsabeck, R. (2004). Symptomvalidierungstests in der neuropsychologischen Diagnostik: Eine Analogstudie. *Zeitschrift für Neuropsychologie*, 15, 81–90.
- Merten, T., Lorenz, R., & Schlatow, S. (2010). Posttraumatic stress disorder can easily be faked, but faking can be detected in most cases. *German Journal of Psychiatry*, 13, 140–149.
- Merten, T., & Rogers, R. (2017). An international perspective on feigned mental disabilities: Conceptual issues and continuing controversies. *Behavioral Sciences and the Law*, 35 (2), 97–112.
- Merten, T., Thies, E., Schneider, K., & Stevens, A. (2009). Symptom validity testing in claimants with alleged posttraumatic stress disorder: comparing the Morel Emotional

- Numbing Test, the Structured Inventory of Malingering Symptomatology, and the Word Memory Test. *Psychological Injury and Law*, 2, 284–293.
- Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, 24, 1094–1102.
- Naveh-Benjamin, M. & Ayres, T. J. (1986). Digit span, reading rate, and linguistic relativity. *Quarterly Journal of Experimental Psychology*, 38, 739–751.
- Nijdam-Jones, A. & Rosenfels, B. (2017, January 12). Cross-cultural feigning assessment: A systematic review of feigning instruments used with linguistically, ethnically, and culturally diverse samples. *Psychological Assessment*. Advance online publication. doi: 10.1037/pas0000438
- Ott, W., Bade, S., & Wapf, B. (2007). *Nicht zielkonforme Leistungen in der Invalidenversicherung: Bedeutung und Größenordnung*. Bern: Bundesamt für Sozialversicherungen.
- Petermann, F. (Hrsg.) (2012). *Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV)*. Deutschsprachige Adaptation der WAIS-IV von D. Wechsler. Frankfurt am Main.: Pearson Assessment.
- Petermann, F. & Lepach, A. C. (Hrsg.) (2012). *Wechsler Memory Scale – Fourth Edition (WMS-IV)*. Deutsche Version. Frankfurt am Main: Pearson Assessment.
- Plohmann, A. M. (2008). Relevanz neuropsychologischer Gutachten zur Beurteilung von Arbeitsfähigkeit und Integrität. *Epileptologie*, 25, 182–190.
- R Core Team (2017). R: a language and environment for statistical computing. Retrieved from <https://www.R-project.org/>
- Rienstra, A., Spaan, P. E. J., & Schmand, B. (2009). Reference data for the Word Memory Test. *Archives of Clinical Neuropsychology*, 24, 255–262.
- Rey, A. (1964). *L'examen clinique en psychologie*. Paris: University Press of France.

- Reynolds, C. R. (1998). Common sense, clinicians, and actuarialism. In C. R. Reynolds (Ed.), *Detection of malingering during head-injury litigation*. New York: Plenum.
- Rogers, R. (1990). Development of a new classificatory model of malingering. *Bulletin of the American Academy of Psychiatry and the Law*, 18, 323–333.
- Rogers, R. (2008). (Ed.), *Clinical assessment of malingering and deception*, (3rd ed.). New York: Guilford.
- Salazar, X. F., Lu, P. H., Wen, J., & Boone, K. B. (2007). The use of effort tests in ethnic minorities and in non-English-speaking and English as a second language populations. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 405–427). New York: Guilford Press.
- Schmand, B. & Lindeboom, J. (2005). *Amsterdam Short-Term Memory Test*. Manual. Leiden: PITS.
- Schmand, B., Lindeboom, J., Schagen, S., Heijt, R., Koene, T., & Hamburger, H. L. (1998). Cognitive complaints in patients after whiplash injury: The impact of malingering. *Journal of Neurology, Neurosurgery and Psychiatry*, 64, 339–343.
- Schroeder, R. W., Twumasi-Ankrah, P., Baade, L. E., & Marshall, P. S. (2012). Reliable digit span: A systematic review and cross-validation study. *Assessment*, 19, 21–30.
- Schweizerisches Bundesamt für Statistik (2011). *Statistisches Jahrbuch der Schweiz 2011*. Zürich: Neue Zürcher Zeitung.
- Slick, D. J., & Sherman, E. M. S. (2013). Differential diagnosis of malingering. In D. A. Carone, & S. S. Bush (Eds.), *Mild traumatic brain injury: System validity assessment and malingering* (pp. 57–72). New York: Springer.
- Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, 13, 545–561.

- Stevens, A., Friedel, E., Mehren, G., & Merten, T. (2008). Malingering and uncooperativeness in psychiatric and psychological assessment: Prevalence and effects in a German sample of claimants. *Psychiatry Research*, *157*, 191–200.
- Strutt, A. M., Scott, B. M., Shrestha, S., & York, M. K. (2011). The Rey 15-item memory test and Spanish-speaking older adults. *The Clinical Neuropsychologist*, *25*, 1253–1265.
- Sweet, J. J. (1999). Malingering: Differential diagnosis. In J. J. Sweet (Ed.), *Forensic neuropsychology* (pp. 255–285). Lisse, NL: Swets and Zeitlinger.
- Taylor, L. A., Kreutzer, J. S., & West, D. D. (2003). Evaluation of malingering cut-off scores for the Rey 15-item test: A brain injury case series. *Brain Injury*, *17*, 295–308.
- Thommen A. (2006): Nicht oder schwer objektivierbare Gesundheitsbeeinträchtigungen im medizinischen Alltag: Erfahrungen einer unabhängigen Beraterin im Gesundheitswesen. In E. Murer (Hrsg.), *Nichtobjektivierbare Gesundheitsbeeinträchtigungen: Ein Grundproblem des öffentlichen und privaten Versicherungsrechtes sowie des Haftpflichtrechtes*. Bern: Stämpfli.
- Tombaugh, T. N. (1996). *Test of Memory Malingering*. San Antonio: Psychological Corporation.
- Tydecks, S., Merten, T., & Gubbay, J. (2006). The Word Memory Test and the One-in-Five-Test in an analogue study with Russian-speaking participants. *International Journal of Forensic Psychology*, *1*, 29–37.
- Vilar-López, R., Gómez-Río, M., Caracuel-Romero, A., Llamas-Elvira, J., & Pérez-García, M. (2008). Use of specific malingering measures in a Spanish sample. *Journal of Clinical and Experimental Neuropsychology*, *30*, 710–722.
- von Aster, M., Neubauer, A. & Horn, R. (Hrsg.). (2006). *Wechsler Intelligenztest für Erwachsene WIE*. Deutschsprachige Bearbeitung und Adaptation des WAIS-III von David Wechsler (2., korrigierte Aufl.). Frankfurt am Main: Pearson Assessment.

- Vos, P. E., Battistin, L., Birbamer, G., Gerstenbrand, F., Potapovd, A, Prevec, T. ... European Federation of Neurological Societies (2002). EFNS guideline on mild traumatic brain injury: report of an EFNS task force. *European Journal of Neurology* 9, 207–219.
- Wechsler, D. (2009). *Advanced Clinical Solutions for the WAIS-IV and WMS-IV*. San Antonio, TX: Pearson.
- World Health Organization (1994). *Pocket guide to the ICD-10 classification of mental and behavioral disorders, with glossary and diagnostic criteria for research (ICD-10: DCR-10)*. Compilation and editorial arrangements by J. E. Cooper. Washington, DC: American Psychiatric Press.
- Young, G. (2015). Malingering in forensic disability-related assessments: Prevalence 15 ± 15%. *Psychological Injury and Law*, 8, 188–199.

Table 1. Frequency and prevalence estimates using RDS cut-off score ≤ 6

	Total <i>N</i>	"Pass" < 2 failed PVTs and no PVT below chance			"Fail" ≥ 2 failed PVTs or 1 PVT below chance					
		Frequency <i>n</i> (%)	Prevalence %	CI (%)	Frequency <i>n</i> (%)	Prevalence %	CI (%)	Frequency <i>n</i> (%)	Prevalence %	CI (%)
Total sample	455	330 (72.5)	72.5	68.3-76.5	125 (27.5)	27.5	23.5-31.7	16 (12.8)	12.8	7.7-19.8
Gender										
Male	222	155 (34.1)	69.8	63.3-75.7	67 (14.7)	30.2	24.3-36.7	11 (8.8)	16.4	9.0-27.3
Female	232	174 (38.2)	75.0	69.0-80.0	58 (12.7)	25.0	19.7-31.0	5 (4.0)	8.6	3.5-18.5
Transgender	1	1(0.2)	100.0	5.0-100.0	0 (0.0)	0.0	0.0-95.0	0 (0.0)	0.0	
Education										
< 9 years	40	11 (2.4)	27.5	15.6-43.6	29 (6.4)	72.5	56.4-84.4	5 (4.0)	17.2	7.1-35.7
Regular graduation (9–10 years)	52	36 (7.9)	69.2	54.9-80.8	16 (3.5)	30.8	19.2-45.1	2 (1.6)	12.5	2.3-36.8
Completed vocational training (11–13 years)	298	230 (50.9)	77.2	72.1-81.8	68 (14.9)	22.8	18.2-27.9	8 (6.4)	11.8	5.2-21.7
College, university, etc. (>13 years)	65	53 (11.6)	81.5	70.3-89.7	12 (2.6)	18.5	10.3-29.7	1 (0.8)	8.3	0.4-36.6
Country of origin										
Switzerland	280	230 (50.5)	82.1	77.6-86.4	50 (11.0)	17.9	13.6-22.8	3 (2.4)	6.0	1.7-16.4
German-speaking countries	37	32 (7.0)	86.5	72.1-94.5	5 (1.1)	13.5	5.5-27.9	1 (0.8)	20.0	1.0-65.7
Non-German-speaking European countries excluding the Balkans	39	26 (5.7)	66.7	50.0-80.1	13 (2.9)	33.3	19.9-50.0	1 (0.8)	7.7	0.4-33.6
The Balkans	56	20 (4.4)	35.7	23.7-49.1	36 (7.9)	64.3	50.9-76.3	7 (5.6)	19.4	8.8-35.7
The Middle East	27	11 (2.4)	40.7	23.3-59.8	16 (3.5)	59.3	40.2-76.7	4 (3.2)	25.0	9.0-50.0
Asia	5	4 (0.9)	80.0	34.3-99.0	1 (0.2)	20.0	1.0-65.7	0 (0.0)	0.0	0.0-95.0
North America	3	3 (0.7)	100.0	36.9-100.0	0 (0.0)	0.0	0.0-63.1	0 (0.0)	0.0	
South America	4	2 (0.4)	50.0	9.8-90.2	2 (0.4)	50.0	9.8-90.2	0 (0.0)	0.0	0.0-77.6
Africa	4	2 (0.4)	50.0	9.8-90.2	2 (0.4)	50.0	9.8-90.2	0 (0.0)	0.0	0.0-77.6
The Balkans + the Middle East	83	31 (6.8)	37.3	27.5-48.2	52 (11.4)	62.7	51.8-72.5			
The Balkans, the Middle East and other non-German-speaking European countries	122	57 (12.5)	46.7	37.9-55.8	65 (14.3)	53.3	44.2-62.1			
Migration Status										
NA	279	229 (50.3)	82.1	77.2-86.3	50 (11.0)	17.9	13.7-22.8	3 (2.4)	6.0	1.7-16.4
First generation	145	77 (16.9)	53.1	44.8-61.2	68 (14.9)	46.9	38.8-55.2	13 (10.4)	19.1	11.1-29.9
Second generation	31	24 (5.3)	77.4	59.9-89.7	7 (1.5)	22.6	10.3-40.1	0 (0.0)	0.0	0.0-37.7
Age range										
<20	6	6 (1.3)	100.0	59.4-100.0	0 (0.0)	0.0	0-40.6			
20–30 years	66	59 (13.0)	89.4	79.9-95.2	7 (1.5)	10.6	4.8-20.1	0 (0.0)	0.0	0.0- 37.7
30–40 years	91	57 (12.5)	62.6	52.2-72.2	34 (7.5)	37.4	27.8-47.8	3 (2.4)	8.8	2.5-22.9
40–50 years	146	98 (21.5)	67.1	59.0-74.5	48 (10.5)	32.9	25.5-41.0	6 (4.8)	12.5	5.6-24.6
50–60 years	111	87 (19.1)	78.4	69.9-85.4	24 (5.3)	21.6	14.6-30.1	5 (4.0)	20.8	8.6-41.4
60–70 years	34	23 (5.1)	67.6	50.0-81.6	11 (2.4)	32.4	18.4-50.0	2 (1.6)	18.2	3.3-50.0
>70 years	1	0 (0.0)	0.0	0.0-95.0	1 (0.2)	100.0	5.0-100.0	0 (0.0)	0.0	0.0-95.0
Time since injury or onset of disease										
[0–3] years	142	110 (24.2)	77.5	69.9-84.0	32 (7.0)	22.5	16.0-30.1	3 (2.4)	9.4	2.6-24.3
(3–5,6] years	85	61 (13.4)	71.8	61.3-80.9	24 (5.3)	28.2	19.1-38.7	4 (3.2)	16.7	5.9-37.0
(5,6–11] years	117	79 (17.4)	67.5	58.6-75.8	38 (8.4)	32.5	24.2-41.4	4 (3.2)	10.5	3.7-24.4
(11–54] years	109	78 (17.1)	71.6	62.5-79.6	31 (6.8)	28.4	20.4-37.5	5 (4.0)	16.1	6.6-33.4
NA	2	2 (0.2)	100.0	22.4-100.0	0 (0.0)	0.0	0.0-77.6			

Symptoms/diagnoses°

Neurological diagnoses° (n = 408)

Cervical spine dysfunctions, whiplash injury	208	137 (30.1)	65.9	59.2-72.2	71 (15.6)	34.1	27.8-40.8	6 (4.8)	8.5	3.7-17.2
Traumatic brain injury										
mild	158	114 (25.1)	72.2	64.6-79.0	44 (9.7)	27.8	21.0-35.4	7 (5.6)	15.9	7.2-29.2
moderate (positive CT/MRI findings)*	39	27 (5.9)	69.2	52.6-82.8	12 (2.6)	30.8	17.2-47.4	1 (0.8)	8.3	0.4-36.6
severe (positive CT/MRI findings)*	29	22 (4.8)	75.9	57.1-89.0	7 (1.5)	24.1	11.0-42.9	0 (0.0)	0.0	0.0-37.7
Cerebrovascular accident	33	25 (5.5)	75.8	57.8-88.3	8 (1.8)	24.2	11.7-42.2	3 (2.4)	37.5	11.1-71.1
Cerebral tumour	8	6 (1.3)	75.0	35.9-95.4	2 (0.4)	25.0	4.6-64.1	1 (0.8)	50.0	2.5-97.5
Epilepsy	24	18 (4.0)	75.0	54.3-88.5	6 (1.3)	25.0	11.5-45.7	2 (1.6)	33.3	6.3-72.9
Multiple Sclerosis	5	4 (0.9)	80.0	34.9-99.0	1 (0.2)	20.0	1.0-65.7	0 (0.0)	0.0	0.0-95.0
Mild cognitive impairment, early dementia	7	5 (1.1)	71.4	34.1-94.7	2 (0.4)	28.6	5.3-65.9	0 (0.0)	0.0	0.0-77.60
Others with possible influence on cognition	68	53 (11.6)	77.9	66.4-86.4	15 (3.3)	22.1	13.6-33.6	1 (0.8)	6.7	0.3-30.2

Substantial brain damage according to brain imagery

Yes*	127	96 (21.1)	75.6	67.4-82.5	31 (6.8)	24.4	17.5-32.6	6 (4.8)	19.4	8.8-36.7
No*	147	99 (21.8)	67.3	59.2-74.6	48 (10.5)	32.7	25.4-40.8	7 (5.6)	14.6	6.6-27.7
Irrelevant findings*	64	49 (10.8)	76.6	64.8-85.6	15 (3.3)	23.4	14.4-35.2	1 (0.8)	6.7	0.3-30.2
No CT/MRI done or available	117	86 (18.9)	73.5	64.6-81.0	31 (6.8)	26.5	19.0-35.4	2 (1.6)	6.5	1.2-20.2

Psychiatric diagnoses° (n = 421)

Mental disorders other than organic mental disorders (F10–F99)	175	108 (23.7)	61.7	54.3-68.8	67 (14.7)	38.3	32.7-47.6	9 (7.2)	13.4	7.0-23.5
F00–09 Organic, including symptomatic, mental disorders	245	200 (44.0)	81.6	76.2-86.3	45 (9.9)	18.4	13.3-23.0	5 (4.0)	11.1	4.5-23.9
without other psychiatric comorbidities	106	92 (20.2)	86.8	79.0-92.3	14 (3.1)	13.2	7.7-21.0	1 (0.8)	7.1	0.4-31.2
with other psychiatric comorbidities	139	108 (23.7)	77.7	69.9-84.0	31 (6.8)	22.3	16.0-30.1	4 (3.2)	12.9	4.5-28.6
F10–19 Mental and behavioral disorders due to psychoactive substance use	54	41 (0.9)	75.9	63.2-85.7	13 (2.9)	24.1	14.3-36.8	1 (0.8)	7.7	0.41-33.6
F20–29 Schizophrenia, schizotypal, and delusional disorders	4	3 (0.7)	75.0	24.9-98.7	1 (0.2)	25.0	24.9-98.7	0 (0.0)	0.0	0.0-95.0
F30–39 Mood (affective) disorders	119	76 (16.7)	63.9	54.7-72.4	43 (9.5)	36.1	27.6-45.3	7 (5.6)	16.3	7.4-29.9
F32-34 Depressive episode, recurrent and persistent depressive disorders	118	75 (16.5)	63.6	54.3-72.2	43 (9.5)	36.4	27.8-45.7	7 (5.6)	16.3	7.4-29.9
F40–49 Neurotic, stress-related, and somatoform disorders	146	91 (20.0)	62.3	54.1-70.0	55 (12.1)	37.7	30.0-45.9	9 (7.2)	16.4	8.6-28.7
F40–41 Phobic and other anxiety disorders	28	17 (3.7)	60.7	40.8-77.6	11 (2.4)	39.3	22.4-59.2	1 (0.8)	9.1	0.5-40.1
F42 Obsessive-compulsive disorders	1	0 (0.0)	0.0	0.0-95.0	1 (0.2)	100.0	0.5-100.0	0 (0.0)	0.0	0.0-95.0
F43.0–43.1 Acute stress reactions and post-traumatic stress disorders	23	13 (13.9)	56.5	36.0-75.3	10 (2.2)	43.5	24.7-64.0	1 (0.8)	10.0	0.5-44.4
F43.2 Adjustment disorders	27	21 (4.6)	77.8	58.5-89.8	6 (1.3)	22.2	10.2-41.5	1 (0.8)	16.7	0.9-59.4
F44 Dissociative (conversion) disorders	10	7 (1.5)	70.0	38.1-91.3	3 (0.7)	30.0	8.7-61.9	1 (0.8)	33.3	1.7-86.5
F45 Somatoform disorders	71	36 (7.9)	50.7	38.6-62.3	35 (7.7)	49.3	37.7-61.4	5 (4.0)	14.3	5.8-29.5
F48 Other neurotic disorders	12	9 (2.0)	75.0	45.6-92.8	3 (0.7)	25.0	7.2-54.4	0 (0.0)	0.0	0.0-63.1
F50–59 Behavioral syndromes associated with physiological disturbances and physical factors	23	16 (3.5)	69.6	47.7-86.1	7 (1.5)	30.4	13.9-52.3	0 (0.0)	0.0	0.0-37.7
F60–69 Disorders of adult personality and behaviour	27	18 (4.0)	66.7	46.2-82.4	9 (2.0)	33.3	17.6-53.8	1 (0.8)	11.1	0.6-44.3
F70–79 Mental retardation	16	11 (2.4)	68.8	43.4-86.8	5 (1.1)	31.3	13.2-56.6	1 (0.8)	20.0	1.0-65.7
F80–89 Disorders of psychological development	38	35 (7.7)	92.1	79.6-97.8	3 (0.7)	7.9	2.3-20.4	0 (0.0)	0.0	0.0-63.1
F90–98 Behavioral and emotional disorders with onset usually occurring in childhood and adolescence	24	22 (4.8)	91.7	73.8-98.5	2 (0.4)	8.3	1.5-26.2	0 (0.0)	0.0	0.0-77.6
F90 Hyperkinetic disorders	20	19 (4.2)	95.0	76.1-99.7	1 (0.2)	5.0	0.3-23.9	0 (0.0)	0.0	0.0-95.0
F91–98 Other behavioral and emotional disorders	4	3 (0.7)	75.0	24.9-98.7	1 (0.2)	25.0	1.3-75.1	0 (0.0)	0.0	0.0-95.0
FZ00–Z99 Factors influencing health status and contact with health services	75	46 (10.1)	61.3	49.9-72.3	29 (6.4)	38.7	27.7-50.1	6 (4.8)	20.7	9.4-39.3

Chronic pain

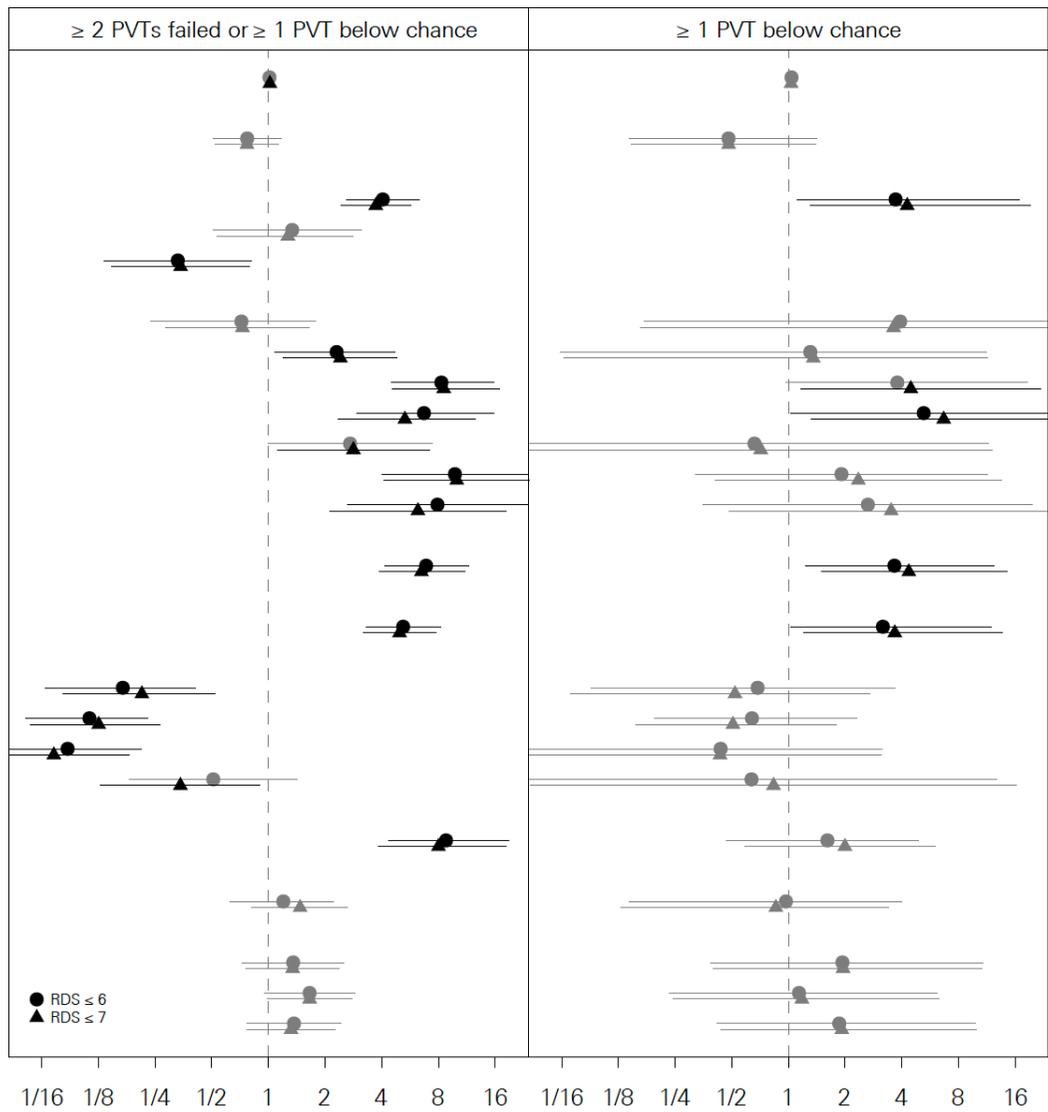
Chronic pain excluding somatoform disorders	273	200 (44.0)	73.3	67.7-78.3	73 (16.0)	26.7	21.7-32.3	7 (5.6)	9.6	4.3-18.4
---	-----	------------	------	-----------	-----------	-------------	-----------	---------	------------	----------

Note. PVT = performance validity test; NA = not applicable; * = if CT/MRI was done; ° = multiple diagnoses possible.

Figure 1.

Influence of sociodemographic factors on the risk of poor effort and definite malingering in the total sample

Note. PVT = performance validity test; D = Germany, A = Austria, CH = Switzerland; OR = Odds ratio; grey and black lines = 95% confidence intervals; black symbols and lines = significant differences (95% confidence intervals don't include OR = 1).



OR

- Age
- Gender:
 - Female - Male
- Migration background:
 - 1. Generation - none
 - 2. Generation - none
 - 2. Generation - 1. Generation
- Country of origin:
 - German-speaking countries (D, A) - Switzerland (CH)
 - non-German-speaking Europe except The Balkans - (CH)
 - The Balkans - CH
 - Middle East - CH
 - non-German-speaking Europe except The Balkans - (CH + D + A)
 - The Balkans - (CH + D + A)
 - Middle East - (CH) + D + A)
- The Balkans / Middle East:
 - yes - no
- The Balkans / Middle East / non-German-speaking Europe:
 - yes - no
- Education:
 - 9-10 years - < 9 years
 - 11-13 years - < 9 years
 - > 13 years (college, university) - < 9 years
 - > 13 years (college, university) - 9-10 years
- Education < 9 years:
 - yes - no
- Education 9-10 years:
 - yes - no
- Time since injury or onset of disease (years):
 - (3,5,6] - [0,3]
 - (5,6,11] - [0,3]
 - (11,54] - [0,3]

Figure 2.

Influence of sociodemographic factors on the risk of poor effort adjusted for country of origin.

Note. PVT = performance validity test; OR = Odds ratio; grey and black lines = 95% confidence intervals; black symbols and lines = significant differences (95% confidence intervals don't include OR = 1).

≥ 2 PVTs failed or ≥ 1 PVT below chance

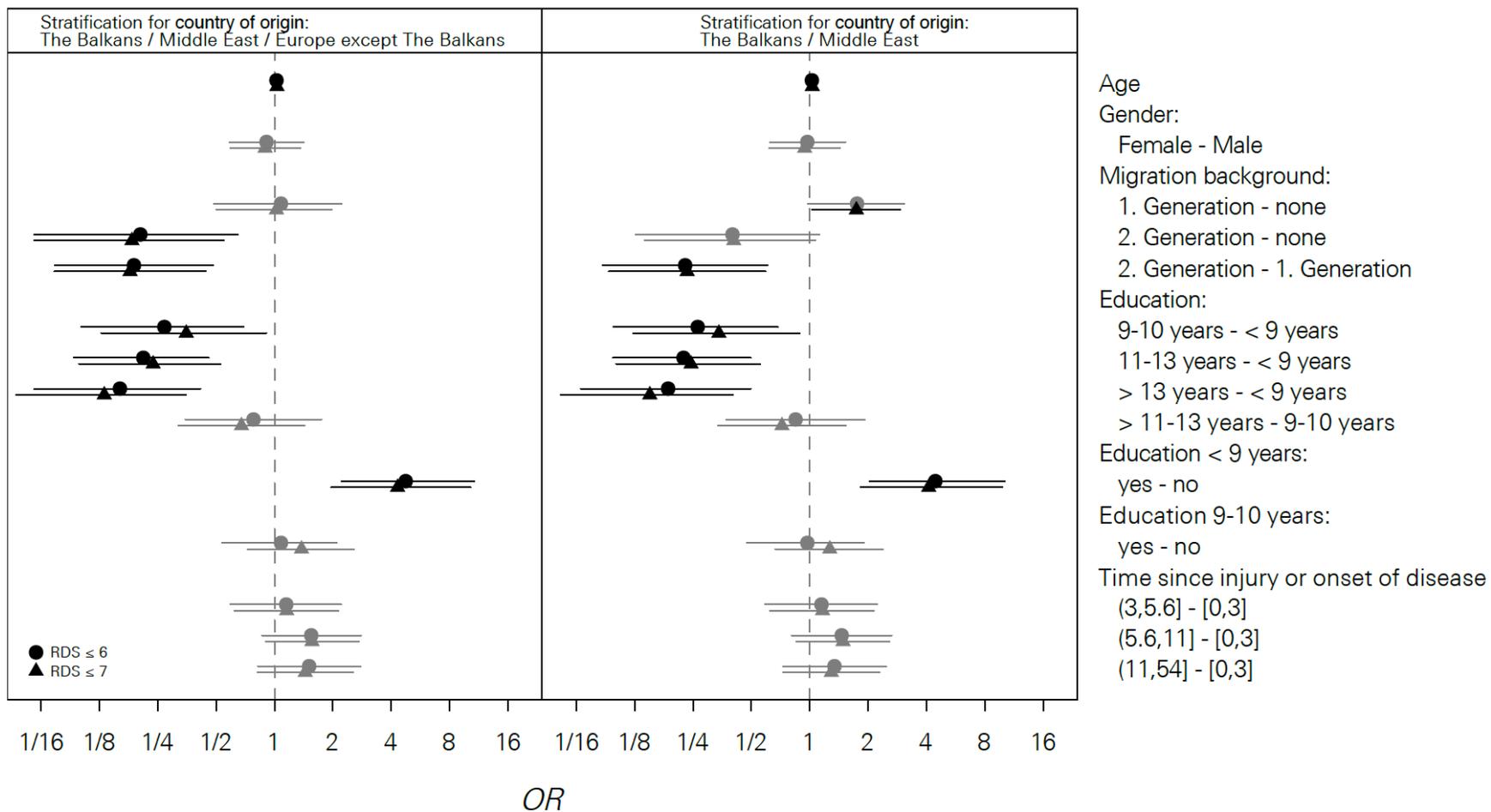


Figure 2

Figure 3.

Influence of sociodemographic factors on the risk of poor effort adjusted for education

Note. PVT = performance validity test; D = Germany, A = Austria, CH = Switzerland; OR = Odds ratio; grey and black lines = 95% confidence intervals; black symbols and lines = significant differences (95% confidence intervals don't include OR = 1).

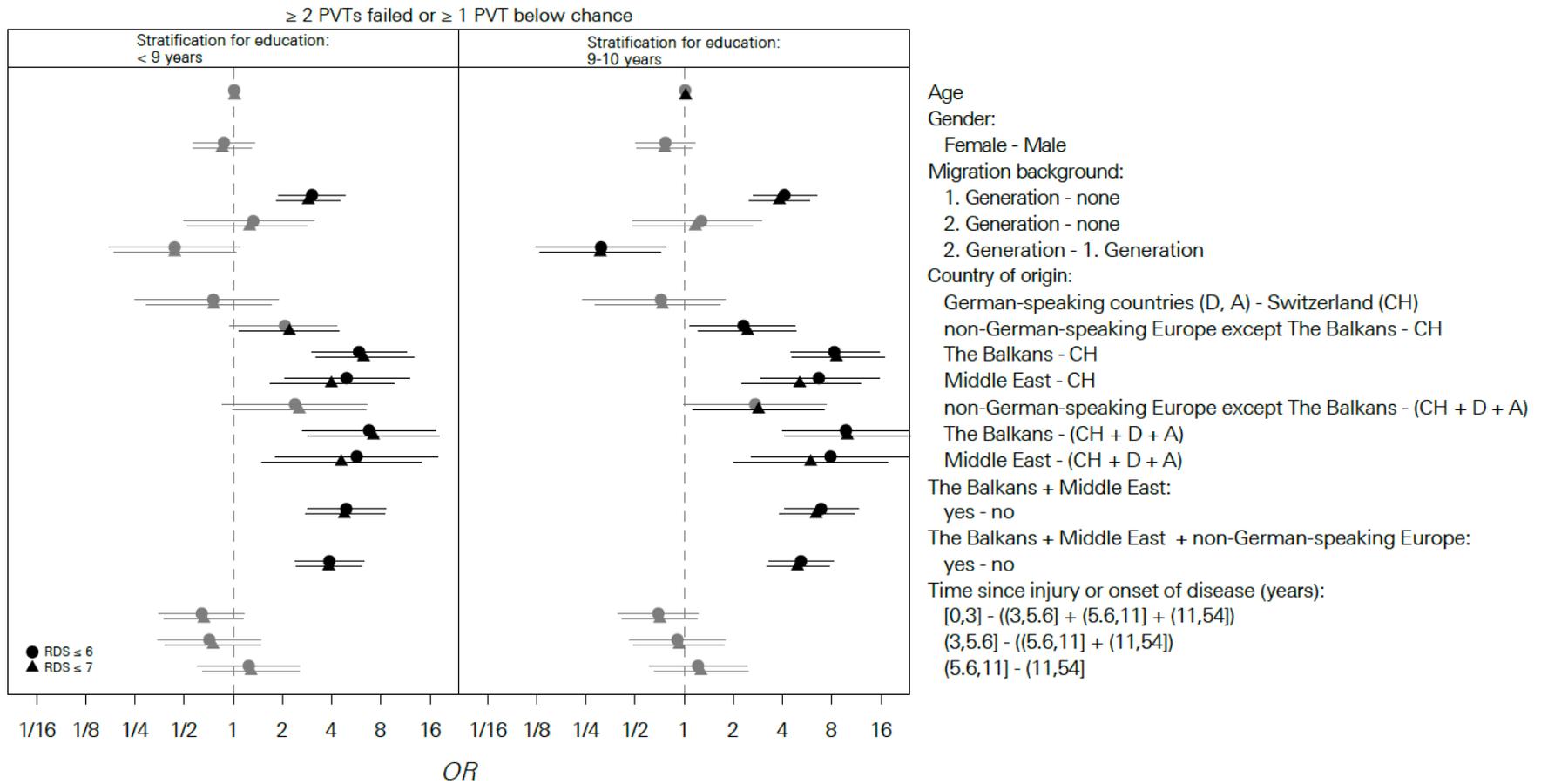


Figure 3

Figure 4.

Influence of neurological diagnosis, chronic pain and findings of MRI/CT scans on the risk of poor effort

Note. PVT = performance validity test; TBI = traumatic brain injury; MCI = mild cognitive impairment; OR = Odds ratio; grey and black lines = 95% confidence intervals; black symbols and lines = significant differences (95% confidence intervals don't include OR = 1).

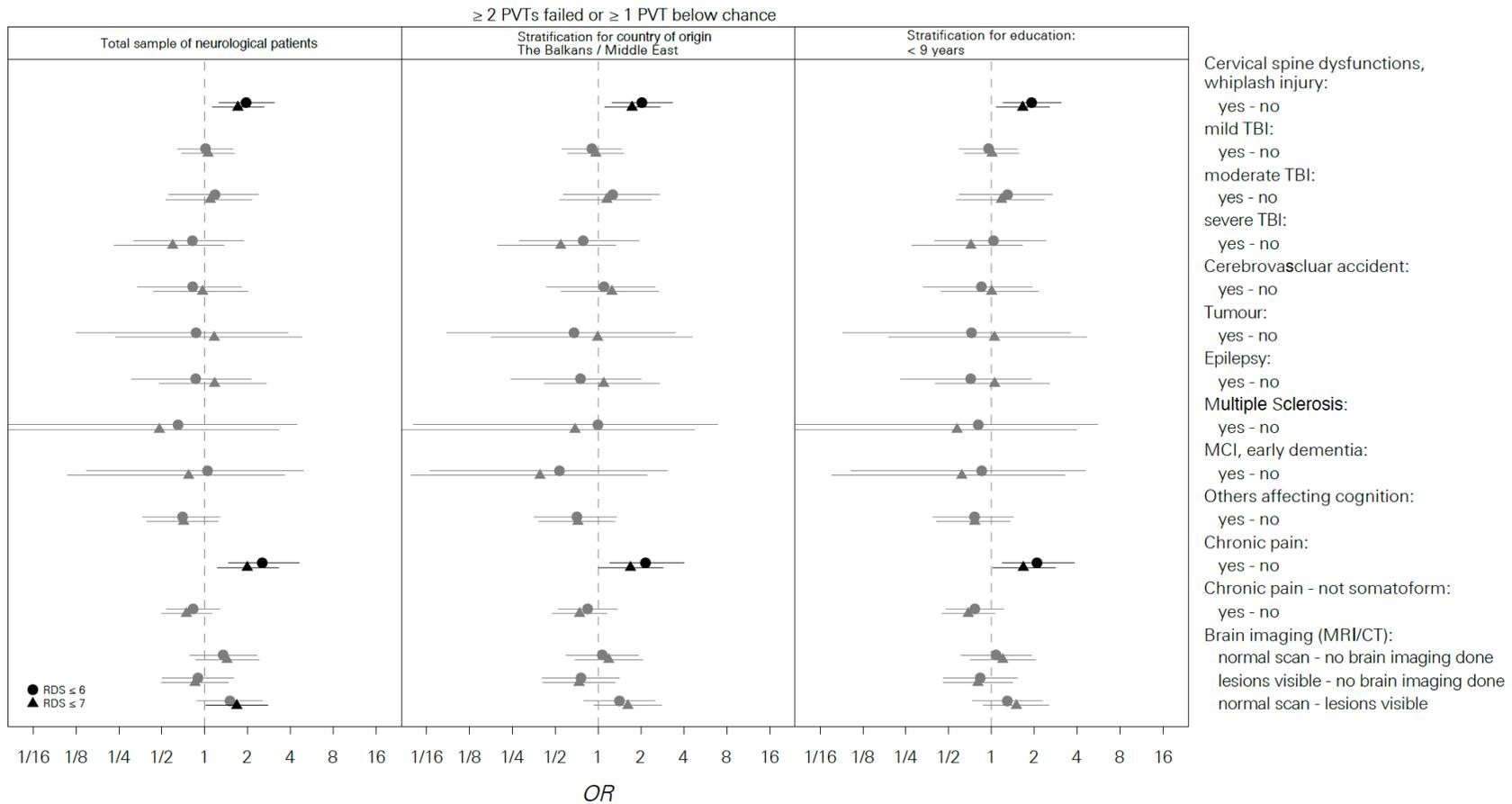
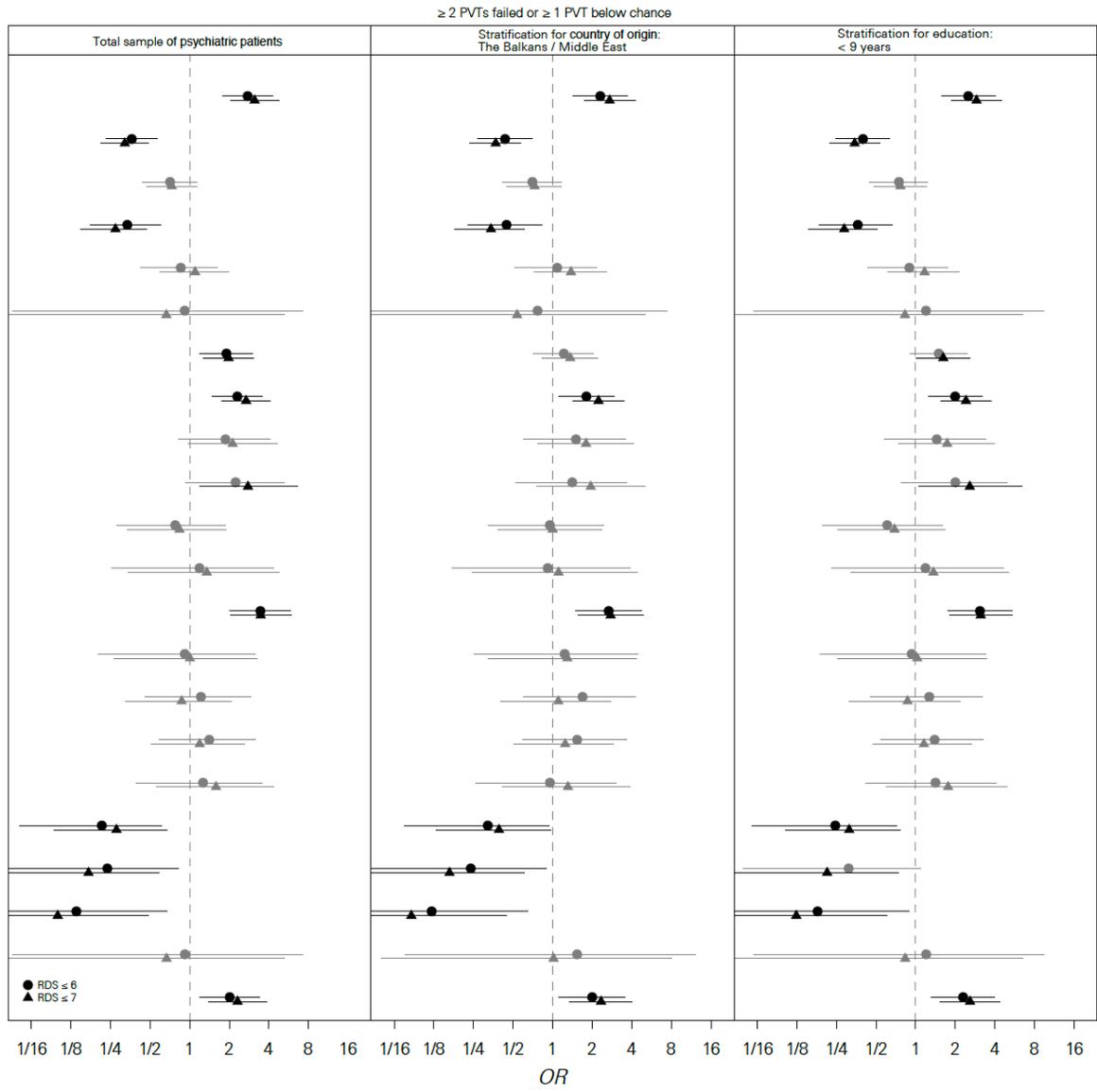


Figure 4

Figure 5.

Influence of psychiatric diagnosis on the risk of poor effort

Note. PVT = performance validity test; PTSD = posttraumatic stress disorder; OR = Odds ratio; grey and black lines = 95% confidence intervals; black symbols and lines = significant differences (95% confidence intervals don't include OR = 1).



- ≥ 1 diagnosis of mental disorder but no F00-09:
 - yes - no
- F00-09 Organic, including symptomatic, mental disorders:
 - yes - no
- F00-09 with additional psychiatric diagnosis:
 - yes - no
- F00-09 without additional psychiatric diagnosis:
 - yes - no
- F10-19 disorders due to psychoactive substance use:
 - yes - no
- F20-29 non-mood psychotic disorders:
 - yes - no
- F30-39 Mood [affective] disorders:
 - yes - no
- F40-49 neurotic, stress-related, somatoform disorders:
 - yes - no
- F40-41 Phobic and anxiety disorders:
 - yes - no
- F43.0-43.1 Acute stress reaction, PTSD:
 - yes - no
- F43.2 Adjustment disorders:
 - yes - no
- F44 Dissociative and conversion disorders:
 - yes - no
- F45 Somatoform disorders:
 - yes - no
- F48 Other neurotic disorders:
 - yes - no
- F50-59 Behavioral syndromes associated with physical factors:
 - yes - no
- F60-69 personality disorders:
 - yes - no
- F70-79 Intellectual disabilities:
 - yes - no
- F80-89 developmental disorders:
 - yes - no
- F90-98 disorders with onset in childhood and adolescence:
 - yes - no
- F90 Hyperkinetic disorder:
 - yes - no
- F91-98 Other disorders with onset in childhood and adolescence:
 - yes - no
- F200-Z99 Factors influencing health status and contact with health services:
 - yes - no

ANHANG A-2

Publikation 2:

Keppler, C., Plohmann, A.M., Pflüger, M., Rabovsky, K., Langewitz, W., & Mager, R. (2017). Beschwerdenvalidierung in der versicherungsmedizinischen Begutachtung. *Fortschritte der Neurologie und Psychiatrie*, 85, 17-33.

Beschwerdenvalidierung in der versicherungsmedizinischen Begutachtung

Symptom Validation in Independent Medical Evaluations

Autoren

Carole Keppler¹, Andrea Maria Plohmann², Marlon Pflueger¹, Kristin Rabovsky³, Wolf Langewitz⁴, Ralph Mager³

Institute

- 1 Forensisch-Psychiatrische Klinik – Forschung Forensik Neurophysiologie, Universitäre Psychiatrische Kliniken Basel
- 2 N/A, Praxis für Neuropsychologie und Psychotherapie, Basel
- 3 Forensisch-Psychiatrische Klinik – Versicherungsmedizin, Universitäre Psychiatrische Kliniken Basel
- 4 Psychosomatik – Innere Medizin, Universitätsspital Basel

Schlüsselwörter

Beschwerdenvalidierung, Performanzvalidierung, Begutachtung, Versicherungsmedizin

Key words

symptom validity testing, performance validity, independent medical evaluation, insurance medicine

Bibliografie

DOI <http://dx.doi.org/10.1055/s-0042-116681>

Fortschr Neurol Psychiatr 2017; 85: 17–33

© Georg Thieme Verlag KG Stuttgart · New York

ISSN 0720-4299

Korrespondenzadresse

Prof. Ralph Mager

Forensisch-Psychiatrische Klinik – Versicherungsmedizin, Universitäre Psychiatrische Kliniken

Wilhelm Klein-Strasse 23

4012 Basel

ralph.mager@upkbs.ch

ZUSAMMENFASSUNG

Hohe Prävalenzraten von nicht authentischen Beschwerden in der versicherungsmedizinischen Begutachtung weisen auf das Risiko hin, dass in einer Gesellschaft zur Verfügung gestellte und limitierte geldwerte Leistungen nicht zielkonform eingesetzt werden. Es liegt dabei wesentlich in der Verantwortung der Gutachter, in geeigneter Form

eine Beschwerdenvalidierung zu erbringen, um dem Rechtsanwender eine zutreffende Entscheidungsgrundlage zu bieten und medizinischen Qualitätskriterien gerecht zu werden. Traditionell geschieht dies im Wesentlichen in einer nicht standardisierten Konsistenzprüfung auf dem Hintergrund der klinischen Untersuchungsergebnisse in Abgleich mit der Aktenlage. Die nicht gesicherte Validität dieses Vorgehens mit hoher Abhängigkeit von individueller Expertise, Datenlage und Haltung hat internationale Fachgesellschaften in jüngerer Zeit dazu veranlasst, Leitlinien mit ihren Empfehlungen zu öffnen für eine standardisiertere Erfassung der Beschwerdenvalidität. Die am besten evaluierten Methoden und Instrumente dazu werden in der vorliegenden Arbeit vorgestellt. Dabei wird versucht, ihre Validität insbesondere in Bezug auf die Anwendung im Begutachtungskontext nach der aktuellen Datenlage darzustellen. Die Autoren kommen zu dem Schluss, dass eine lege artis durchgeführte strukturierte Beschwerdenvalidierung die Qualität klinischer Urteile erheblich verbessern kann. Die qualifizierte Integration dieser zusätzlich erhobenen Befunde in eine psychologische oder ärztliche Expertise ist jedoch anspruchsvoll und bedarf einer weiteren wissenschaftlichen Begleitung. Einer mechanischen Beurteilung nach Punktwerten gilt es vorzuziehen.

ABSTRACT

High prevalence rates of non-authentic complaints identified by experts in the field of insurance medicine draw attention to the risk of services, which are of limited availability and financially compensated, being used in ways that are not goal-oriented. Therefore, symptom validity testing has become a growing issue to prevent non-targeted monetary compensations. This paper outlines the best-evaluated methods and instruments. Based on the data currently available, their validity, in particular in the context of medico-legal assessment, is assessed. It is concluded that symptom validity assessment allows inferences about the degree of certainty of clinical judgements on the authenticity of reported symptoms. Thus, the application of the suggested instruments enhances significantly the quality of medical and psychological expertise. However, the integration of the additional results into the overall assessment is challenging and needs to be further clarified.

Einleitung

In Europa zeigen psychische Störungen ähnlich wie in anderen Industrienationen eine stark ansteigende Bedeutung für die Verursachung von Absenzen am Arbeitsplatz und generieren mittlerweile einen führenden Anteil der entstehenden Versicherungszahlungen im Erwerbsalter. Betroffen sind dabei zunehmend auch jüngere Menschen [1–3]. Die im OECD-Bericht [4] festgehaltene gesamteuropäische Situation bestätigt diese neben dem individuellen Leid auch wirtschaftlich problematische Entwicklung. Aus-

gehend von entsprechenden Untersuchungen in den USA, kommt im deutschsprachigen Raum neben den prioritären medizinischen und betrieblichen Aspekten im Gefolge dieser Entwicklung auch die Frage auf, zu welchem Anteil Arbeitsunfähigkeit und Ausgleichszahlungen wirklich auf tatsächlich bestehenden, also nicht vorgetäuschten Gesundheitsstörungen beruhen. Eine gut belegbare kritische Sicht zeigt dabei, dass trotz der fortschreitenden Erkenntnisse in Medizin, Psychologie und Biologie der Zugang zu großen Anteilen der Beschwerden weiterhin ein sprachlicher ist, der wesentlich bis ausschließlich auf die subjektiven Angaben

des Patienten zurückgeht. Das Risiko einer Fehleinschätzung der Authentizität von Symptomen und ihrem Zusammenhang zu Funktionseinschränkungen ist also in gewissem Sinne medizinimmanent und betrifft insbesondere psychiatrische, aber auch somatische Expertisen. Der Umgang mit nicht authentischen Beschwerden wird dennoch weder in der medizinischen noch in der psychologischen Ausbildung systematisch adressiert.

Einschätzungen mittels Gutachterbefragungen des Bundesamtes für Sozialversicherungen in der Schweiz gehen von etwa 8–10% der IV-Renten aus, bei denen der Verdacht besteht, dass sie nicht zielkonform zugesprochen wurden [5, 6]. Bei diesen nicht zielkonformen Leistungen müsse bei einem kleineren Teil davon ausgegangen werden, dass sie aus falschen Angaben der Exploranden (Aggravation/Simulation) hervorgegangen sind. Diese seien nachvollziehbar hauptsächlich bei Störungsbildern zu erwarten, bei denen ein erheblicher Ermessensspielraum für die Einschätzung der Erwerbsunfähigkeit bestehen kann, etwa Depressionen, posttraumatischen Belastungsstörungen und Schmerzerkrankungen einschließlich Zuständen nach HWS-Distorsionen. Diese Liste ist jedoch um alle Störungsbilder zu erweitern, bei denen der Gutachter bei der Diagnosestellung und der Einschätzung der daraus resultierenden Beeinträchtigungen der Arbeitsfähigkeit im Wesentlichen auf die subjektiven Schilderungen des Betroffenen angewiesen ist. Dazu zählen beispielsweise auch Tinnitus, fragliche Anfallereignisse, nicht prüfbare Störungen des Seh-/Hörvermögens und Sensibilitätsstörungen. Ein Gutachter darf, um seinem Auftrag zu entsprechen, die Authentizität von geschilderten Beschwerden nicht einfach voraussetzen, sondern muss die Validität der geltend gemachten Beschwerden sorgfältig überprüfen. Es wurde mehrfach herausgearbeitet [7–10], dass die Beurteilung der Glaubwürdigkeit von Beschwerdeangaben allein aufgrund klinischer Erfahrung der Gutachter unter Nutzung nicht standardisierter und oft nicht transparenter individueller Analyse von Verhaltensmustern fehleranfällig ist. Zu berücksichtigen ist auch die von dem Gutachter zu fordernde Abgrenzung gegenüber normalpsychologischen Prozessen. Neuere experimentelle Arbeiten im Bereich Neurokognition zeigen, dass Antwortverzerrungen/ Antworttendenzen unbewusst durch im Vorfeld von Testungen erlernte Haltungen und Erfahrungen bedingt sein können [11–13]. Es ist unzweifelhaft, dass der Explorand in einem komplexen Versicherungsverfahren einer Vielzahl von Einflüssen ausgesetzt ist und ein dadurch verändertes Antwortverhalten partiell plausibel ist, ohne dass es als nicht authentisch zu beurteilen wäre. Festgehalten sei aber ebenfalls, dass die Authentizität von subjektiv geklagten Beschwerden auch bei Nachweis von objektiven morphologischen Abweichungen nicht automatisch gegeben ist, da beispielsweise bildgebende Abweichungen von Normalbefunden und daraus resultierende Beschwerden und Funktionseinschränkungen nur sehr unzuverlässig korrelieren [14–17]. Die Praxis zeigt jedoch, dass trotz des häufig niedrigen Beweiswerts eines solchen somatischen Befundes die Rechtsanwender diesem einen unverhältnismäßig hohen Stellenwert einräumen. Auf diesem Hintergrund ergibt sich der Bedarf, Standards für die Beurteilung der Authentizität von Beschwerden zu entwickeln und diese im Gutachtenprozess zu verankern. Insbesondere die Validierung neurokognitiver Beschwerden kann sich dabei bereits auf eine breite Studienlage

und die Entwicklung spezifischer Messinstrumente stützen und hat unzweifelhaft das Potenzial, die Konsistenzprüfung nach klinischen Befunden und Aktenlage sinnvoll zu ergänzen.

In der vorliegenden Arbeit sollen standardisierte Verfahren vorgestellt und methodisch kurz evaluiert werden in ihrer Möglichkeit, angegebene Beschwerden im psychologisch-psychiatrischen Bereich gutachterlich nachzuvollziehen, um verbleibende Funktionen besser einschätzen zu können.

Begriffsbestimmung

Beschwerdenuvalidierung bezeichnet den Prozess, der zur Beurteilung der Beschwerdenuvalidität führt, und somit die Überprüfung der Authentizität von Beschwerdendarstellung und Symptompräsentation [18]. Invalide Beschwerdendarstellung bzw. Antwortverzerrung (response bias) bezeichnet ein Verhalten, das durch unzutreffende Antworten, fälschliche Angaben oder durch ein Testverhalten gekennzeichnet ist, das nicht den tatsächlichen Möglichkeiten einer Person entspricht.

Simulation, Aggravation (engl. unter *malingering* zusammengefasst) und im weitesten Sinne auch Verdeutlichungstendenzen fallen dabei unter die Rubrik negativer Antwortverzerrung. Unter Simulation versteht man die bewusste, auf ein externes Ziel (z. B. finanzielle Entschädigung, Straffreiheit, Dienstunfähigkeit) ausgerichtete Vortäuschung von Krankheit oder Beschwerden bei Abwesenheit von tatsächlichen Beschwerden im beklagten Bereich. In vielen Kontexten dürfte diese eher die Ausnahme sein. Viel häufiger hingegen ist die Aggravation, also die absichtlich übertreibende Darstellung vorhandener Beschwerden zur Erlangung von Vorteilen. Davon wird von einigen Autoren die Verdeutlichungstendenz abgegrenzt, die häufig unbewusst motiviert ist und darauf abzielt, den Gutachter vom Vorhandensein der beklagten Beschwerden zu überzeugen [19]. Auf Leistungsebene tritt negative Antwortverzerrung als unzureichende oder suboptimale Anstrengungsbereitschaft (*poor effort*) auf, während sie bei beklagten psychischen oder somatischen Leiden oft als übertriebene Beschwerdendarstellung (*exaggeration*) in Erscheinung tritt. Der Begriff Anstrengungsbereitschaft ist problematisch, da Formulierungen wie geringe oder inadäquate Anstrengungsbereitschaft implizieren, dass ein Explorand sich bei der Bearbeitung nicht genügend anstrengt. Tatsächlich aber kann die Anstrengung erheblich sein – allerdings dahin gehend, ein schlechtes Testergebnis zu erzielen. Besser wäre es daher, von der Zielrichtung gezeigter Anstrengung zu sprechen.

Positive Antwortverzerrung zeichnet sich demgegenüber durch die Leugnung, Bagatellisierung und Abwehr von Beschwerden (*Dissimulation*), geschönte Selbstdarstellung (*impression management*), aber auch Antwortstile wie die Tendenz zur sozialen Erwünschtheit aus. Diese Form der Antwortverzerrung findet sich vor allem, wenn das Sorge- oder Besuchsrecht zur Disposition steht, das Überwiegen körperlicher gegenüber psychischen Symptomen mit Vorteilen in der Rechtsprechung verbunden ist, bei Substanzmissbrauch oder auch dort, wo psychische Störungen kulturell bedingt transformiert oder stigmatisiert werden. *Dissimulation* kann aber auch psychologisch begründet sein, beispielsweise im Sinne einer Verleugnung von Defiziten bei Angst vor Ausgrenzung.

Selbst bei adäquater Beschwerdendarstellung und Anstrengungsbereitschaft kann es sich formal um Simulation handeln, nämlich dann, wenn authentische Beschwerden anderen Ursprungs bewusst einem nicht ursächlichen, jedoch besser versicherten Ereignis zugeschrieben werden [20]. Auf der anderen Seite ist die einer überzeichneten Symptomdarbietung zugrunde liegende Motivation nicht immer bewusster Natur, sondern kann auch in einem primären Krankheitsgewinn bestehen. So zeigen bei bestimmten psychiatrischen Krankheitsbildern Exploranden ohne Täuschungsabsicht eine verzerrte Beschwerdenschilderung und/oder unzureichende Anstrengungsbereitschaft; möglich ist dies z. B. bei Exploranden mit somatoformen oder dissoziativen Störungen. Eine Auseinandersetzung mit der Problematik, wie auffällige Ergebnisse in der Beschwerdenuvalidierung bei diesen Explorandengruppen zu bewerten sind, haben erst kürzlich Merten und Merckelbach [21] vorgenommen. Die Abgrenzung zwischen Simulation bzw. Aggravation, artifizieller, somatoformer und dissoziativer Störung ist auch deshalb schwierig, weil zwischen Selbst- und Fremdtäuschung fließende Übergänge bestehen, primärer und sekundärer Krankheitsgewinn sich oft vermischen und die Bewusstheit eigener Motive unterschiedlich ausgeprägt ist.

Prävalenzraten negativer Antwortverzerrung

Die generelle Prävalenz von nicht authentischen psychischen Störungen schwankt zwischen 10 und 30 %, je nach Untersuchungskontext und Methodik [22]. Larrabee [23] schätzte anhand einer Analyse von 11 empirischen Studien, dass in etwa 40 % der Begutachtungsfälle relevante Antwortverzerrungen vorkommen. Gemäß einer Erhebung unter amerikanischen Neuropsychologen auf der Basis von 30 000 Begutachtungen wurden bei 32,7 % aller Fälle von Invaliditätsansprüchen „wahrscheinliche nicht authentische kognitive Beschwerden“ (nach Slick, Sherman & Iverson [24]) vermutet, im Bereich Personenschaden bei 30,4 %. Im Rahmen von forensischen Begutachtungen belief sich dieser Wert auf immerhin noch 22,7 %, bei neuropsychologischen Abklärungen im Rahmen medizinischer oder psychiatrischer Untersuchungen außerhalb eines juristischen Kontexts auf 8,1 %. Von den Exploranden, die sich im Rechtsstreit befanden bzw. Schadensersatzansprüche geltend machten, standen nicht nur 41,2 % der Patienten nach leichtem Schädel-Hirn-Trauma unter Täuschungsverdacht, sondern auch ein erheblicher Anteil von Exploranden, die psychische Störungen wie Fibromyalgie oder ein chronisches Erschöpfungssyndrom (38,6 %), Schmerzen oder somatoforme Störungen (33,5 %), depressive Störungen (16,1 %), Angststörungen (13,6 %) oder dissoziative Störungen (10,5 %) beklagten. Unter Patienten nach mittelgradigem oder schwerem Schädel-Hirn-Trauma wurde die Authentizität der beklagten Beschwerden in lediglich 8,8 % der Fälle angezweifelt, bei jenen mit vaskulärer Demenz in 2,4 %. Danach scheint die Auftretenshäufigkeit negativer Antwortverzerrung wesentlich vom Untersuchungskontext und den geltend gemachten Störungen bzw. den gestellten Diagnosen abhängig zu sein [25]. In Deutschland wurden ähnlich hohe Prävalenzraten für Gutachtenpopulationen gefunden. Merten, Friedel und Stevens [26] und Stevens, Friedel, Mehren et al. [27]

konnten jeweils suboptimale Leistungsanstrengungen in Performanzvalidierungsverfahren¹ bei über 45 % aller untersuchten Probanden finden, die wegen psychiatrischer und neurologischer Störungen begutachtet wurden. Nach empirischen Schätzungen von Greve, Ord, Bianchini et al. [28] liegt die Prävalenz von Täuschungsverhalten bei Exploranden mit chronischen Schmerzen im medico-legalen Kontext je nach Klassifikationssystem zwischen 25 % und 36 %. 21 % dieser nicht authentischen Probanden wiesen objektive somatische Befunde auf. Auch bei anderen somatischen Krankheitsbildern finden sich Hinweise auf Aggravation oder Simulation, so z. B. bei 58 % der geltend gemachten Hörstörungen [29].

Methoden zur Beschwerdenuvalidierung

Ergebnisse psychologischer Testverfahren sind leicht zu verfälschen und ohne spezifische Methoden hinsichtlich ihrer Gültigkeit nicht zu beurteilen. Deshalb wurden von Neuropsychologen spezielle Performanzvalidierungsverfahren entwickelt, die inzwischen als am besten untersuchte Methodengruppe gelten. Weitere neuropsychologische Ansätze bestehen in Validitätsindikatoren, die aus Parametern gebräuchlicher Testinstrumente abgeleitet werden können, sowie im Einsatz von Fragebögen und Selbstbeurteilungsinstrumenten, die über spezifisch entwickelte Skalen zur Identifizierung wahrscheinlicher negativer und positiver Antwortverzerrungen verfügen. Letztere dienen vor allem der Validierung der Beschwerdenschilderung. ► **Tab. 1** stellt eine Übersicht über die Methoden der Beschwerdenuvalidierung dar.

A. Verfahren zur Validierung der Anstrengungsbereitschaft

Beschwerdenuvalidierungs- oder nach neuerer Terminologie Performanzvalidierungstests (PVT) sind speziell zur Erfassung von Antwortverzerrungen entwickelt worden. Bei den kognitiven Funktionen operieren sie vorwiegend im Bereich des Gedächtnisses, wobei sie eine hohe Robustheit gegenüber realen kognitiven Defiziten aufweisen. Aufgrund verschiedener Ansätze bei der Testkonstruktion sind die inzwischen verfügbaren Verfahren unterschiedlich gut in der Lage, authentische und negativ verzerrte Testergebnisse zuverlässig zu unterscheiden. Zur Beurteilung der Güte eines PVT entwickelte Hartman [30] einen allgemein akzeptierten Kriterienkatalog. Ein guter Test sollte danach

- Anstrengungsbereitschaft messen, aber nicht für kognitive Störungen anfällig sein (Sensitivität und Spezifität);
- eine gute normative Basis aufweisen. Nur wenn die Ergebnisse von authentischen Patienten verschiedener Diagnosegruppen und mit unterschiedlichem Ausmaß kognitiver Störungen bekannt sind, kann ein zuverlässiger Trennwert festgelegt werden.

1 In der internationalen Literatur etabliert sich zunehmend eine sprachliche Unterscheidung zwischen Performanzvalidierungstests, die die Anstrengungsbereitschaft von Probanden in Leistungstests überprüfen, und Beschwerdenuvalidierungstests, mit denen die Beschwerdenschilderung überprüft werden soll.

► Tab. 1 Methoden zur Beschwerdenvalidierung.

	Grundprinzipien/Detektionsstrategien	Verfahren	Kurze Übersicht	Verfahrensbeispiele	Vorteile	Nachteile	Literatur
Verfahren zur Validierung der Anstrengungsbereitschaft	<ul style="list-style-type: none"> ▪ Alternativwahlverfahren ▪ Testdeckeneffekt oder Prinzip der verdeckten Leichtigkeit ▪ Leistungskurve ▪ Fehlergröße ▪ Identifikation untypischer oder inkonsistenter Leistungsprofile 	Performanzvalidierungstests (PVT)	<ul style="list-style-type: none"> ▪ Höchste Klassifikationsgenauigkeit ▪ Fokus auf vorgetäuschten Gedächtnisdefiziten > hohe Robustheit gegenüber realen kognitiven Defiziten ▪ Große empirische Datenbasis ▪ Ausführliche Validierung ▪ Geringe Sensitivität zur Aufdeckung von negativer Antwortverzerrung in anderen Funktionsbereichen ▪ Erhöhte Interkorrelationen möglich ▪ Soziodemografische Faktoren (Alter, Geschlecht, Bildungsstand) spielen kaum eine Rolle 	Word Memory Test (WMT)	<ul style="list-style-type: none"> ▪ Alle Hartman-Kriterien erfüllt ▪ Hohe Sensitivität 	<ul style="list-style-type: none"> ▪ Bei einzelnen Patientengruppen reduzierte Spezifität ▪ Relativ hoher Zeitaufwand, inzwischen existieren 2 ähnliche, kürzere Verfahren (MSVT, NV-SVT) 	[27, 29, 34–37, 49]
				Test of Malingered Memory (TOMM)	<ul style="list-style-type: none"> ▪ Hohe Augenscheinvalidität ▪ Sprachfreies Material 	<ul style="list-style-type: none"> ▪ Geringe Sensitivität ▪ Anfälligkeit gegenüber Coaching 	[29, 30, 49]
				Amsterdamer Kurzzeitgedächtnistest (AKGT)	<ul style="list-style-type: none"> ▪ Erschwerte Durchschaubarkeit 	<ul style="list-style-type: none"> ▪ Nicht geeignet für Patienten mit schon klinisch feststellbaren Störungen 	[29, 38, 49]
Verfahren zur Validierung der Beschwerdeschilderung	<ul style="list-style-type: none"> ▪ Seltene Beschwerden ▪ Quasi-seltene Beschwerden ▪ Unwahrscheinliche Beschwerden ▪ Beschwerdenkombinationen, die bei echten psychiatrischen Störungen äußerst selten vorkommen ▪ Unübliche psychopathologische Muster 	Beschwerdenvalidierungsparameter in neuropsychologischen Tests	<ul style="list-style-type: none"> ▪ Parameter können direkt aus Verfahren gewonnen werden ▪ Ggf. via zusätzliche Untertests ▪ Geringere Robustheit gegenüber realen kognitiven Defiziten ▪ Überprüfung einer Vielzahl von Funktionsbereichen möglich 	Reliable Digit Span Finger Tapping	<ul style="list-style-type: none"> ▪ Verbesserte Sensitivität ohne wesentliche Erhöhung der Untersuchungszeit ▪ Validierung auch im gutachterlichen Kontext ▪ Erhöhung der Spezifität ▪ Direkte Informationen über Gültigkeit der gezeigten Leistung ▪ Geringere Anfälligkeit gegenüber Coaching 	<ul style="list-style-type: none"> ▪ Gefahr falsch-positiver Zuordnung bei Kombination inhaltlich überlappender Parameter ▪ Klassifikationsgüte variiert in Abhängigkeit vom Störungsbild ▪ Erfordert detaillierte Kenntnis der Primärliteratur 	[25, 29, 39–49]
				Structured Interview of Reported Symptoms (SIRS-2)	<ul style="list-style-type: none"> ▪ Hervorragende Interrater-Reliabilität ▪ Sehr gute Retestreliabilität bei forensischen Patienten ▪ Belege für Konstruktvalidität mit dem MMPI-2 ▪ Deutsche Version in Vorbereitung 	<ul style="list-style-type: none"> ▪ Lange Durchführungszeit ▪ Expertise der Interviewer erforderlich 	[52–59]
				Miller Forensic Assessment of		<ul style="list-style-type: none"> ▪ Reines Screeningverfahren 	[60–63]

► **Tab. 1** (Fortsetzung)

	Grundprinzipien/Detektionsstrategien	Verfahren	Kurze Übersicht	Verfahrensbeispiele	Vorteile	Nachteile	Literatur
	<ul style="list-style-type: none"> ▪ Hohe, undifferenzierte Ausmaße an Beschwerden ▪ Außergewöhnlich hohe Ausprägungen einzelner Symptome ▪ Eindeutige Beschwerden ▪ Diskrepanzen zwischen beobachteten und berichteten Beschwerden ▪ Fehlerhafte Stereotypen 			Symptoms Test (M-Fast)	<ul style="list-style-type: none"> ▪ Sehr gute Reliabilität und Validität in forensischen Populationen ▪ Belege für Konstruktvalidität mit dem SIRS ▪ Reduktion Anteil falsch-negativer Klassifikationen gegenüber klinischem Urteil ▪ Kurze Durchführungs- und Auswertzeit 		
		Selbstbeurteilungsverfahren und andere Screeningtests für psychische Störungen		Strukturierter Fragebogen simulierter Symptome (SFSS > deutsche Version des Structured Inventory of Malingered Symptomatology (SIMS))	<ul style="list-style-type: none"> ▪ Hohe Retestreliabilität ▪ Akzeptable interne Konsistenz ▪ Gute vorhersagende und übereinstimmende Validität ▪ Sensitivität, Spezifität, Positive Predictive Power, Negative Predictive Power über 0,85 in Abhängigkeit vom Trennwert 	<ul style="list-style-type: none"> ▪ Reines Screeningverfahren ▪ Sensitivität abhängig vom Untersuchungskontext bzw. vom angewandten Trennwert 	[58, 64 – 66]
				Minnesota Multiphasic Personality Inventory (MMPI-2)	<ul style="list-style-type: none"> ▪ Aufdeckung von Antwortverzerrung im Bereich psychischer, körperlicher und kognitiver Störungen ▪ Sehr hohe Spezifitäten im forensischen Kontext ▪ Wahrscheinlichkeitsaussage über Unter- oder Übertreibung möglich ▪ Aussagen über psychische Gesundheit und Persönlichkeitsmerkmale möglich 	<ul style="list-style-type: none"> ▪ Lange Durchführungszeit ▪ Expertise erforderlich ▪ Mehrheit der Validitätsskalen nicht Bestandteil der deutschen Version > manuelle Auswertung anhand der Literatur erforderlich 	[67 – 75, 113 – 115]
				<ul style="list-style-type: none"> ▪ Deutlich kürzer als MMPI-2 		[80 – 83]	

► Tab. 1 (Fortsetzung)

	Grundprinzipien/Detektionsstrategien	Verfahren	Kurze Übersicht	Verfahrensbeispiele	Vorteile	Nachteile	Literatur
				Personality Assessment Inventory (PAI) Deutsche Version: Verhaltens- und Erlebensinventar (VEI)	<ul style="list-style-type: none"> ▪ Inhaltliche Ausrichtung an ICD-10 und DSM-IV ▪ Hohe Reliabilität, gute Validität in verschiedenen Settings ▪ Hohe Effektstärken der Validitätsskalen der Originalversion 	<ul style="list-style-type: none"> ▪ Vergleichsweise geringe Sensitivität der deutschen Version 	
				Morel Emotional Numbing Test for Posttraumatic Stress Disorder (MENT)	<ul style="list-style-type: none"> ▪ Forced-choice-Verfahren ▪ Detektionsstrategie fehlerhafte Stereotypen ▪ Hohe Spezifität, gute Sensitivität ▪ In vielen Sprachen vorhanden 	<ul style="list-style-type: none"> ▪ Screeningverfahren 	[84 – 90]
Nachweis bewusster Täuschungsabsicht	<ul style="list-style-type: none"> ▪ Prozess, der sorgfältige Analyse aller verfügbaren Informationen erfordert ▪ Basierend auf objektiven Kriterien ▪ Einschluss von Indikatoren, die über bekannte Klassifikationsgenauigkeit verfügen ▪ Kombination klinisches Urteil mit Ergebnissen wissenschaftlich validierter Verfahren 	Kriterienkataloge	<ul style="list-style-type: none"> ▪ A: Vorhandensein eines substanzialen, externen Anreizes ▪ B: Hinweise aus neuropsychologischer Testdiagnostik ▪ C: Angaben des Exploranden und Validitätsskalen von Selbstbeurteilungsinstrumenten ▪ D: Ausschluss von psychiatrischen, neurologischen oder Entwicklungsfaktoren ▪ Zusätzlich für „sicher“ vorliegende vorgetäuschte, kognitive Störung: Zahl richtiger Antworten in Zwangswahlverfahren unter Zufallsniveau 	Kriterienkatalog für die Diagnose aggravierter/simulierter neurokognitiver Störungen (Malingered Neurocognitive Disorder, MND)	<ul style="list-style-type: none"> ▪ Breit akzeptiertes Konzept ▪ Modifikationen 2013: primäre und sekundäre Aggravation/Simulation, dafür Wegfall Kriterium D/Eindeutige Aggravation/Simulation kann nun auch den Nachweis einer hohen A-posteriori-Wahrscheinlichkeit aus Selbstbeurteilungsinstrumenten bestehen oder aus „compelling inconsistencies“ 	<ul style="list-style-type: none"> ▪ Hoher Zeitaufwand 	[21, 91]
			<ul style="list-style-type: none"> ▪ Zusätzliches B: Hinweise aus somatischen Untersuchungsmethoden 	Kriterienkatalog für die Aufdeckung aggravierter/simulierter schmerzassoziierter Behinderungen (Malingered Pain-Related Disabilities, MPRD)		<ul style="list-style-type: none"> ▪ Hoher Zeitaufwand 	[92, 93]
						<ul style="list-style-type: none"> ▪ Hoher Zeitaufwand 	[94, 95]

► **Tab. 1** (Fortsetzung)

	Grundprinzipien/Detek-tionsstrategien	Verfahren	Kurze Übersicht	Verfahrens-beispiele	Vorteile	Nachteile	Literatur
			<ul style="list-style-type: none"> ▪ A: Evidenz eines externen An-reizes ▪ B: Evidenzen aus Testbefunden ▪ C: Evidenzen aus Inkonsisten-zen/Diskrepanzen ▪ D: Ausschlusskriterien 	Kriterienkatalog für die Diagnose einer unglaubwür-digen, scheinba-ren oder aggra-vierten/simulierten Posttraumati-schen Belastungs-störung (Non-credible, Feigned, or Malingered Posttraumatic Stress Disorder Related Disability/ Dysfunction, F-PTSDR-D)	<ul style="list-style-type: none"> ▪ Eindeutigere Definitio-nen und Abgrenzungen bestimmter Termini und möglicher Inkonsisten-zen ▪ Regeln zur Erhebung und Gewichtung von Testdaten ▪ Separate Graduierung der Wahrscheinlichkeit von Antwortverzerrung für Kriterien B und C ▪ Nutzbar für Diagnose von aggravierten/simu-lierten kognitiven und/ oder schmerzassoziier-ten Störungen ▪ Vereinfachte Gesamt-betrachtung ▪ Klare Regeln mit Bezug zur klinischen Urteils-bildung 		

- beim Exploranden den Eindruck erwecken, dass es sich um eine realistische Messung der untersuchten kognitiven Funktion handelt (Augenscheinvalidität);
- auf Validierungsstudien beruhen, die gesunde Probanden, Patientengruppen und echte sowie instruierte Simulanten einschließen;
- schwierig zu verfälschen oder zu coachen sein, um die Testsi- cherheit zu gewährleisten;
- hinsichtlich Testanwendung, -auswertung und -interpretation möglichst einfach sein;
- Symptombereiche messen, die von Exploranden häufig über- trieben werden;
- von aktuellen Forschungsergebnissen unterstützt werden.

A.1 Detektionsstrategien zur Validierung der Anstrengungsbereitschaft

Grundlage der PVT sind Testprinzipien, die motivationale Einflüsse zufallskritisch und wahrscheinlichkeitsbasiert abschätzen [31]. Nach Rogers, Harrell und Liff [32] lassen sich neuropsychologische Forschungs- und Diagnoseansätze zur Performanzvalidierung wie folgt klassifizieren:

Alternativwahlverfahren

Alternativwahlverfahren stellen die am besten untersuchte Me- thodengruppe mit der höchsten Klassifikationsgenauigkeit bei der Differenzierung von authentischen und nicht authentischen Beschwerden dar. Sie sind so konzipiert, dass ein Zielreiz zusammen mit einem Distraktorreiz angeboten wird, wobei der Explorand den Zielreiz identifizieren soll. Wenn er die richtige Antwort nicht kennt, so muss geraten werden (forced-choice). Statistisch ist bei einer dichotomen Reizverteilung von Ziel- und Distraktorreiz eine 50 %ige Ratewahrscheinlichkeit als minimale Wiedererkennungsrate zu erwarten. Eine wesentlich geringere Trefferquote (signifikant below-chance) ist als Beleg für eine negative Antwortverzerrung zu interpretieren. Damit erlauben Alternativwahlverfahren im Unterschied zu allen anderen Verfahren mit mathematisch begründeter und exakt zu bestimmender Sicherheit willentliche, bewusste Antwortverzerrung nachzuweisen und somit eine sichere Vortäuschung kognitiver Störungen gemäß den weiter unten beschriebenen Slick-Kriterien [24] zu identifizieren. Neben den Antwortmustern unter Zufallsniveau, die eher selten vorkommen, wird mit Trenn- bzw. Grenzwerten (cut-offs) operiert, die Wahrscheinlichkeitsaussagen über das Vorliegen unzureichender Leistungsmotivation erlauben. Ermittelt werden solche Trennwerte bzw. als authentisch zu betrachtende Wertebereiche durch sogenannte „known-groups“- oder „criterion-groups“-Validierungen, in denen als simulierend diagnostizierte Patienten mit ähnlichen authentischen Patienten verglichen werden, die objektiv häufig größere kognitive Einschränkungen, schmerzassoziierte Behinderungen oder psychische Belastungen aufweisen als die Gruppe simulierender Patienten. Eines dieser Alternativwahlverfahren ist beispielsweise der Test of Malingered Memory (TOMM; [33]).

Testdeckeneffekt oder Prinzip der verdeckten Leichtigkeit

Diese Tests täuschen einen Schwierigkeitsgrad vor, der von Exploranden häufig überschätzt wird. Tatsächlich sind die Aufgaben

jedoch so einfach, dass sie von nahezu allen Exploranden bewältigt werden können, auch von solchen mit authentischen kognitiven Defiziten. Aufgrund des niedrigen Schwierigkeitsgrads wird bei anstrengungsbereiten Exploranden eine Punktzahl im Bereich des Maximums (der „Decke“) erwartet. In diese Gruppe fällt z. B. der Rey Fifteen Items Test (FIT; [34]).

Leistungskurve

Dabei wird die Plausibilität des Antwortverhaltens innerhalb eines Tests bzw. zwischen verschiedenen Tests überprüft. Die Leistungskurve eines Exploranden soll dem realen Schwierigkeitsgrad der gestellten Aufgaben oder Anforderungen innerhalb eines Funktionsbereichs entsprechen. Danach sollten einfache Aufgaben eher bewältigt werden als deutlich schwierigere Aufgaben. Beispiele für diesen Ansatz sind der Dot Counting Test [35, 36] oder die Gegenüberstellung von Wiedergabeleistungen zuvor gelernter Inhalte unter verschiedenen Abrufbedingungen. Der typische Leistungsgradient kann bei bestimmten Krankheitsbildern wie z. B. Aphasie oder Akalkulie verändert sein. Wenn keine oder nur wenige empirische Daten zu einem „typischen“ Leistungsprofil vorliegen und die Klassifikationsgüte bestimmter Antwortmuster unbekannt ist, ist ebenfalls Vorsicht geboten.

Fehlergröße

Dieser Ansatz betrifft das gehäufte Auftreten von Knapp-daneben-Antworten. Weicht die Antwort eines Exploranden regelmäßig um eine bestimmte, geringe Konstante von der korrekten Lösung ab, so weist dies darauf hin, dass der Explorand die richtige Antwort kennt, sie aber wissentlich verfälscht. Ein solches Verhalten lässt sich beispielsweise bei Orientierungsfragen oder Rechenaufgaben beobachten.

Identifikation untypischer oder inkonsistenter Leistungsprofile

Darunter fällt die Überprüfung der Konsistenz und Plausibilität eines Testprofils innerhalb einer umfassenden Untersuchung oder über mehrere Testzeitpunkte hinweg. Von Interesse sind dabei v. a. Ergebnisse von Testverfahren, die gleiche oder ähnliche Eigenschaften erfassen. Außerhalb des Einsatzes fixer Testbatterien ist dieser Ansatz jedoch wenig formalisiert, so dass für ihn dieselben Einschränkungen gelten wie für die Analyse der Leistungskurve.

A.2 Spezielle Performanzvalidierungstests

Einige der speziell entwickelten, standardisierten PVT wie der Word Memory Test (WMT; [37]) kombinieren die genannten Grundprinzipien teilweise. Kritisch ist anzumerken, dass die hohe Sensitivität des Verfahrens bei einzelnen Patientengruppen möglicherweise mit einer reduzierten Spezifität einhergeht. So erzielten in einer Untersuchung von Gorissen, Sanz und Schmand [38] 72 % Schizophrener ein Ergebnis unterhalb des empirischen Trennwerts. Negativsymptome (Apathie oder Interessenverlust) korrelierten dabei signifikant mit der Leistungsmotivation. Es wurden von Green noch zwei weitere Kurzverfahren aufgelegt, der Medical Symptom Validity Test (MSVT; [39]) und der Nonverbal Medical Symptom Validity Test (NV-MSVT; [40]), für die, bis auf eine geringere Zahl an Vergleichsdaten, im Wesentlichen die glei-

chen Bewertungen gelten. Der TOMM weist als weltweit verbreitetes Verfahren eine große empirische Datenbasis auf, ist gut validiert, hat eine hohe Augenscheinvalidität und ist als Computer- wie Papier-Bleistift-Version anwendbar. Als Beispiel eines nicht dem Alternativwahlprinzip folgenden Verfahrens sei der Amsterdamer Kurzzeitgedächtnistest (ASTM, dt. AKGT) von Schmand und Lindeboom [41] genannt.

A.3 Eingebettete Beschwerdvalidierungsparameter

Im weiteren Sinne fallen auch Indikatoren in konventionellen neuropsychologischen Tests, sogenannte eingebettete Beschwerdvalidierungsparameter, unter die Kategorie der PVT. Diese Parameter können in verschiedenen Tests direkt aus den Verfahren gewonnen werden, wie beispielsweise die Reliable Digit Span aus dem Untertest „Zahlen nachsprechen“ des Wechsler Intelligenz- oder Gedächtnistests [42]. Andere Variablen erfordern den ergänzenden Einsatz zusätzlicher Untertests, die jeweils speziell für die Beschwerdvalidierung entwickelt wurden. Dazu zählen z. B. zusätzliche Wiedererkennungsdurchgänge im Complex Figure Test [43]. Da die Datenlage zu den meisten dieser Indikatoren fortlaufend ergänzt wird, muss hier auf die jeweilige Originalliteratur verwiesen werden. Englischsprachige Zusammenstellungen finden sich bei Boone [44, 45], Larrabee [46], Morgan und Sweet [47] sowie Carone und Bush [48].

Da die Anstrengungsbereitschaft im Verlauf einer Begutachtung durchaus Schwankungen unterliegt und negative Antwortverzerrung sich selektiv auf einzelne Funktionsbereiche beschränken kann, erhöht sich durch den Einsatz multipler testinterner Validitätsparameter die Wahrscheinlichkeit, solche Verzerrungen im Untersuchungsverlauf zu entdecken [49]. Gleichzeitig lässt sich wegen des Prinzips der kombinierten Unwahrscheinlichkeit auch die Spezifität auf deutlich über 90 % erhöhen [50, 51]. Dabei ist allerdings zu beachten, dass die als authentisch bzw. nicht authentisch herangezogenen Parameter inhaltlich voneinander unabhängig sein müssen, da sich allein aus mathematisch-statistischen Gründen die Wahrscheinlichkeit eines auffälligen Validitätsparameters mit steigender Anzahl erhöht. Ein entscheidender Nachteil besteht darin, dass die Klassifikationsgüte eines einzelnen Indikators nicht an jene eines eigenständigen Performanzvalidierungstests heranreicht und die Anwendung eines einzelnen, eingebetteten Validitätsparameters damit die Gefahr einer falsch-positiven Klassifikation in sich birgt. Ihr Einsatz erfordert daher eine detaillierte Kenntnis der Originalliteratur, da die empfohlenen Grenzwerte für unterschiedliche diagnostische und ethnische Patientengruppen durchaus differieren können.

Ein umfassenderer Überblick über im deutschen Sprachraum verwendete standardisierte PVT und eingebettete Beschwerdvalidierungsparameter findet sich bei Merten [52].

B. Verfahren zur Validierung der Beschwerdentschilderung

Neben der Überprüfung der Anstrengungsbereitschaft (Performanzvalidierung) stellt auch die Validierung der Beschwerdentschilderung einen essenziellen Baustein der Begutachtung dar. Obwohl verfahrensübergreifend das Bild eines mäßigen Zusammenhangs zwischen Performanzvalidierungstests und solchen zur Validierung der Beschwerdentschilderung entsteht, scheinen

beide Ansätze doch verschiedene Konstrukte zu erfassen. Aus diesem Grund empfiehlt es sich, neben der Anstrengungsbereitschaft auch die Beschwerdentschilderung einer Prüfung auf Authentizität zu unterziehen. Grundsätzlich muss in jeder Begutachtungssituation mit einer von verschiedenen Prozessen überlagerten Berichterstattung gerechnet werden. Ein Vorteil von Fragebögen besteht darin, dass einige dieser verzerrenden Einflussfaktoren in einer Begutachtungssituation erfasst und unabhängig von persönlichen Wertungen des Gutachters abgebildet werden können. Auch wenn solche Validitätsskalen keine Aussage darüber erlauben, ob und allenfalls welche Gesundheitsstörungen tatsächlich bestehen, weisen sie den Gutachter doch auf mögliche Verfälschungen der Beschwerdentschilderung hin.

B.1 Detektionsstrategien zur Validierung der Beschwerdentschilderung

Bewährte Strategien zum Nachweis übertriebener Beschwerdentschilderung oder vorgetäuschter psychischer Störungen wurden von Rogers [53] formuliert:

Seltene Beschwerden

Diese werden von authentischen Patienten äußerst selten berichtet. Beispiele für diesen Ansatz stellen u. a. die Infrequency Psychopathology Scale (Fp) und die Infrequent Somatic Complaints Scale (Fs) des Minnesota Multiphasic Personality Inventory (MMPI-2) dar, die sich für die Aufdeckung vorgetäuschter psychischer bzw. somatischer Beschwerden zu bewähren scheinen.

Quasi-seltene Beschwerden (quasi-rare symptoms)

Die Entwicklung solcher Skalen (z. B. MMPI-2 F-Skala und Fb-Skala) beruht auf klinischen Merkmalen, die in der Allgemeinbevölkerung selten vorkommen.

Unwahrscheinliche Beschwerden

Dieser Ansatz (z. B. im Structured Inventory of Reported Symptoms (SIRS)) beruht auf der Präsentation absurder, grotesker Beschwerden. Sein Vorteil liegt darin, dass solche Beschwerden kaum eine andere Interpretation als die einer Täuschungsabsicht zulassen und eine Anwendung auch auf somatische Beschwerden möglich ist.

Beschwerdenkombinationen, die bei echten psychiatrischen Störungen äußerst selten vorkommen

Exploranden mit Täuschungsabsicht geben nicht selten ungewöhnliche Kombinationen von Beschwerden an, wie z. B. Größenideen bei gleichzeitig erhöhtem Schlafbedürfnis. Diese schwer zu durchschauende Detektionsstrategie dürfte nicht leicht zu coachen sein. Sie wird beispielsweise vom SIRS und dem Miller Forensic Assessment of Symptoms Test (M-FAST) benutzt.

Unübliche psychopathologische Muster

Dieser Ansatz, der sich u. a. im Personality Assessment Inventory (PAI) wiederfindet, fokussiert auf Kombinationen erhöhter Skalenwerte, die als charakteristisch für Simulationstendenzen gelten, bei authentischen Patientengruppen jedoch äußerst ungewöhnlich sind.

Hohe, undifferenzierte Ausmaße an Beschwerden

Patienten mit Täuschungsabsichten können oft schon allein durch die Vielzahl beklagter Symptome identifiziert werden. Überprüft wurde diese Detektionsstrategie im Bereich primär psychischer Beschwerden bisher nur in strukturierten klinischen Interviews wie dem SIRS.

Außergewöhnlich hohe Ausprägungen einzelner Symptome

Selbst schwer beeinträchtigte Patienten empfinden nur eine geringe Zahl an Beschwerden als unerträglich oder außergewöhnlich intensiv. Personen mit Täuschungsabsicht hingegen geben häufig für ein weites Spektrum von Symptomen extreme Ausprägungen an. Eine entsprechende Validitätsskala stellt z. B. die Lachar-Wrobel Critical Items Scale des MMPI-2 dar.

Eindeutige Beschwerden

Dieser Ansatz, u. a. umgesetzt in der MMPI-2 Obvious-Subtle Difference Scale, beruht auf dem Gedanken, dass Personen mit Täuschungsabsicht wahrscheinlich Beschwerden beklagen, die als klare Indikatoren einer ernsthaften psychischen Erkrankung bekannt geworden sind.

Diskrepanzen zwischen beobachteten und berichteten Beschwerden

Täuschende Personen können oft durch die Richtung dieser Diskrepanzen identifiziert werden. In diesen Fällen fehlen die zu erwartenden, mit bestimmten beklagten Beschwerden in der Regel einhergehenden, klinischen Beobachtungen in der Untersuchungssituation (SIRS Reported vs. Observed Scale).

Fehlerhafte Stereotypen

Diese Methode nutzt verbreitete, fehlerhafte Annahmen über die mit psychischen Störungen assoziierten klinischen Merkmale. Eine zugehörige Validitätsskala stellt die MMPI-2 Dissimulation Scale dar.

Validiert wurden die erstmals von Rogers [54] vorgestellten Kriterien sowohl in Analogstudien mit instruierten Simulanten als auch in Known-Groups-Designs mit Gutachtenpopulationen und erzielten hohe bis sehr hohe Effektstärken. Für einen detaillierten Überblick sei auf Rogers [22] verwiesen.

B.2 Fremdbeurteilungsverfahren für psychische Störungen

Das Structured Interview of Reported Symptoms (SIRS; [55], SIRS-2, revidierte Version; [56]) setzt die oben beschriebenen Validierungstechniken in der Interviewsituation um [57]. Es umfasst acht Primärskalen und fünf Zusatzskalen, die psychische Störungen erheben sollen. Das Verfahren besticht durch eine hervorragende Interrater-Reliabilität [58] und eine Retestreliabilität zwischen 97,6% und 99,5% bei forensischen Patienten [59]. McCusker, Moran, Serfass et al. [60] fanden auf zwei Skalen hohe Ladungen (Aggravation von Symptomen und unglaubliche Beschwerden), die hoch mit der Fp- und der F-Skala des MMPI-2 korrelierten, was als Beleg für die Konstruktvalidität des SIRS interpretiert wurde. Minuspunkte stellen die relativ lange Durchführungszeit und die nötige Expertise der Interviewer dar [61].

Eine deutsche Version ist als Strukturiertes Interview Berichteter Beschwerden (SIBB; [62]) in Vorbereitung.

Auch der Miller Forensic Assessment of Symptoms Test (M-Fast; [63]) soll simulierte psychische Störungen im Rahmen eines strukturierten Interviews aufdecken. Validiert wurde das Instrument primär in forensischen Kontexten. Das Interview lieferte in einer Studie von Smith [64] bezüglich Reliabilität und Validität sehr gute psychometrische Kennwerte. Guy und Miller [65] konnten mittels hoher Korrelationen mit dem SIRS Belege für die Konstruktvalidität des Interviews erbringen. Bei dem Instrument handelt es sich um ein reines Screeningverfahren, aus dessen isolierten Ergebnissen eine Klassifikation eines Exploranden nicht statthaft ist. Dem klinischen Urteil von Psychologen und Psychiatern gegenüber erwies sich das Instrument jedoch als deutlich überlegen, wobei es v. a. den Anteil falsch-negativer Klassifikationen, also den Anteil übersehener Täuschungsmanöver, reduzierte [66]. Ein weiterer Pluspunkt besteht in der kurzen Durchführungs- und Auswertungszeit von zusammen etwa 15 – 20 Minuten.

B.3 Selbstbeurteilungsverfahren und andere Screeningtests für psychische Störungen

Der Strukturierte Fragebogen simulierter Symptome (SFSS; [61]) als deutsche Version des Structured Inventory of Malingered Symptomatology (SIMS) von Smith und Burger [67] stellt eine leicht handhabbare Screeningmethode zur Erfassung einer Vielzahl simulierter Symptome dar. Für die englische Originalfassung (SIMS) erwies sich der Gesamtwert mit einer Sensitivität von 95,6% und einer Spezifität von 87,9% [67] als ein effizienter Indikator für nicht authentische Beschwerdenschilderung. Die deutsche Version verfügt über eine hohe Retestreliabilität ($r = 0,97$) und eine akzeptable interne Konsistenz (Cronbach's $\alpha = 0,80$). Zudem weist sie eine gute vorhersagende und übereinstimmende Validität auf. Die Werte für Sensitivität, Spezifität, Positive Predictive Power (PPP) und Negative Predictive Power (NPP) liegen in Abhängigkeit vom gewählten Grenzwert über 0,85. Damit ist die diagnostische Genauigkeit recht akzeptabel, solange man den SFSS als Screeningmethode benutzt. Die Falsch-positiv-Rate variiert dabei mit dem Trennwert, der in Abhängigkeit von Untersuchungskontext und Fragestellung unterschiedlich hoch gesetzt werden kann [68]. Ein systematisches Review mit Metaanalyse zum SFSS ist kürzlich erschienen [69].

Das Minnesota Multiphasic Personality Inventory (MMPI-2; [70]) bzw. dessen Nachfolger MMPI-2-RF ist ein umfassendes Persönlichkeitsverfahren mit eingebetteten, gut überprüften Validitätsskalen. Neben den klassischen Validitätsskalen (L, F und K) wurden im Lauf der Zeit weitere entwickelt, die sich die weiter oben beschriebenen, unterschiedlichen Strategien zur Aufdeckung negativer Antworttendenzen zunutze machen. Eine vergleichsweise neu entwickelte Validitätsskala umfasst Items, die eine Unterscheidung zwischen Personen mit hoher und niedriger Anstrengungsbereitschaft in Performanzvalidierungsverfahren ermöglichen (Response Bias Scale; [71, 72]). Eine Stärke des Verfahrens liegt darin, dass die entwickelten Validitätsskalen gemäß verschiedenen Metaanalysen negative Antwortverzerrung im Bereich psychischer Störungen, körperlicher Störungen und kognitiver Störungen aufzudecken vermögen [73, 74]. Im forensischen Kontext fanden Lewis, Simcox und Berry [75] je nach Skala Spezifität

ten zwischen 94 % und 100 %. Auch für die Aufdeckung positiver Antwortverzerrung existiert eine erste Metaanalyse [76], wonach die entsprechenden Skalen tatsächlich zur Unterscheidung von Dissimulation und authentischem Antwortverhalten in der Lage sind. Die jeweils effektivsten Validitätsskalen und Trennwerte variieren in Abhängigkeit von klinischen Gruppen, Untersuchungskontexten und Grundraten von Simulation, Aggravation und Dissimulation. Seit 2000 liegt eine deutsche Version des MMPI-2 vor [77], für die inzwischen eine Schätzung der Gütekriterien der Validitätsskalen an einer realen Gutachtenpopulation erhoben wurde [78]. Unabhängig davon weisen verschiedenste Arbeiten darauf hin, dass der MMPI-2 kulturübergreifend anwendbar ist [79–82]. Eine deutsche Übersetzung des mit nur knapp der Hälfte der Items wesentlich kürzeren MMPI-2-RF ist in Vorbereitung.

Mit dem Personality Assessment Inventory (PAI; [83]) liegt seit einigen Jahren ein weiteres multidimensionales Selbstbeurteilungsinstrument zu klinischen Symptomen und Persönlichkeitsmerkmalen vor, das eine Überprüfung der Beschwerdenschilderung erlaubt. Zur Aufdeckung von vorgetäuschter Psychopathologie bzw. negativer Antwortverzerrung umfasst er drei Indikatoren, die auf der Aufdeckung seltener Beschwerden und unüblichen psychopathologischen Mustern beruhen. Von den acht weiteren Validitätsindikatoren des PAI schätzen drei ein, ob die Beantwortung zufällig erfolgt ist, fünf Maße evaluieren positive Antwortverzerrungen. Vorteile des Verfahrens bestehen in der gegenüber dem MMPI-2 deutlich geringeren Zahl an Fragen sowie der inhaltlichen Ausrichtung an ICD-10 und DSM-IV. Positiv zu werten ist auch die hohe Reliabilität und gute Validität der Skalen in verschiedenen Settings, ebenso wie die hohen Effektstärken der Validitätsskalen [84].

Seit 2013 existiert mit dem Verhaltens- und Erlebensinventar (VEI; [85]) eine deutschsprachige Adaptation des PAI, dessen Normdaten den amerikanischen Normen weitgehend zu entsprechen scheinen. Auch für das VEI wurden an kleinen Gruppen instruierter Simulanten und zu begutachtender Exploranden erste akzeptable Schätzungen der Sensitivität und Spezifität der Skalen für negative Antwortverzerrung vorgelegt [86]. Angesichts ihrer Ergebnisse stellen die Autoren für deutschsprachige Patienten allerdings eine Adaption der von Morey [83] empfohlenen Trennwerte zur Diskussion.

Der Morel Emotional Numbing Test for Posttraumatic Stress Disorder (MENT; [87–89]) ist ein aus 60 Items bestehendes Forced-choice-Verfahren, bei dem anhand von Gesichtsausdrücken Emotionen identifiziert werden sollen. Der Test basiert auf der suggerierten Annahme, dass Personen mit genuiner PTSD dabei Schwierigkeiten aufweisen, und nutzt so die oben erwähnte Detektionsstrategie falscher Stereotype. Ursprünglich an US-Veteranen validiert, umfasst das Verfahren inzwischen Daten zu unterschiedlichen Patientengruppen, auch im medico-legalen Kontext. Es verfügt über eine exzellente Spezifität zwischen 96 und 100 % und eine gute Sensitivität von bis zu 85 % [89, 90] und sollte nach Meinung verschiedener Autoren [91, 92] bei Begutachtungen von PTSD routinemäßigen Einsatz finden. Der MENT ist u. a. in deutscher und türkischer Sprache erhältlich. Normalbefunde von 64 Probanden für den deutschen Sprachraum finden sich bei Wiedmaier [93].

C. Nachweis bewusster Täuschungsabsicht

Die Diagnose von Simulation oder Aggravation umfasst immer eine Aussage über Bewusstheit und Motivation des Täuschungsverhaltens. Um bei nachgewiesener mangelnder Anstrengungsbereitschaft und/oder verzerrter Beschwerdenschilderung eine bewusste Täuschungsabsicht (Aggravation/Simulation) von unbewussten negativen Antwortverzerrungen z. B. im Rahmen einer Konversionsstörung abgrenzen zu können, müssen weitere Informationen mit herangezogen werden. Die Differenzialdiagnose von Simulation bzw. Aggravation umfasst einen Prozess, der eine sorgfältige Analyse aller verfügbaren Informationen erfordert, auf objektiven Kriterien basiert, Indikatoren mit einschließt, die über eine bekannte Klassifikationsgenauigkeit verfügen und das klinische Urteil mit den Ergebnissen wissenschaftlich validierter Verfahren kombiniert. Dabei ist auch den Besonderheiten der Exploranden wie Bildungsstand, kultureller Hintergrund, ethnische Zugehörigkeit und Sprachkompetenz Rechnung zu tragen.

C.1 Kriterienkatalog für die Diagnose aggravierter/ simulierter neurokognitiver Störungen

In diesem Zusammenhang hat sich ein von Slick et al. [24] bereits 1999 vorgestellter Kriterienkatalog für die Diagnose vorgetäuschter neurokognitiver Störungen (Malingered Neurocognitive Disorder, MND) in Forschung wie Einzelfalldiagnostik etabliert. Danach ist eine vorgetäuschte neurokognitive Störung als willentliche Übertreibung oder Vortäuschung kognitiver Dysfunktionen definiert, die zum Ziel hat, einen externen materiellen Gewinn zu erlangen oder aber einer gesetzlichen Verpflichtung (z. B. Militärdienst) oder Verantwortung (z. B. Gerichtsverfahren) zu entkommen. Konzeptuell lässt sich dieser externe Anreiz als sekundärer Gewinn einordnen. Dabei fokussieren die Slick-Kriterien nicht auf psychosoziale Aspekte des Krankheitsgewinns, sondern auf bedeutsame Vorteile wie Entschädigungszahlungen oder Renten. Der Terminus „willentlich“ schließt Bewusstheit und Absicht mit ein. Im Kontext mit der Diagnose von Aggravation/Simulation wird das Ausmaß, in dem ein Verhalten als willentlich deklariert wird, durch Ausschluss plausibler, alternativer Erklärungen bestimmt (D-Kriterium). Mit anderen Worten darf das zu beurteilende Verhalten nicht besser als unbewusstes oder unabsichtliches Handeln im Rahmen von psychiatrischen, neurologischen oder Entwicklungsstörungen erklärt werden können. Falls also ein Explorand in einem Forced-choice-Verfahren eine Leistung zeigt, die unter dem Zufallsniveau liegt, wird angenommen, dass er nicht in der Lage ist, eine korrekte Antwort zu unterdrücken und sie durch eine Falsche zu ersetzen, ohne dass ihm dies bewusst ist. Somit kann dieses Verhalten als vorsätzlich bezeichnet werden.

Slick und Sherman [94] haben kürzlich verschiedene Modifikationen und Erweiterungen des ursprünglichen Kriterienkatalogs vorgeschlagen. Zu den Neuerungen zählt zunächst die Unterscheidung in primäre und sekundäre Aggravation/Simulation. Bei Letzterer sind die Kriterien für eine definitive oder wahrscheinliche Aggravation/Simulation neurokognitiver Störungen bis auf das frühere D-Kriterium erfüllt. Gibt es aber zwingende Gründe anzunehmen, dass der Explorand sekundär zu Unreife (bei Kindern) oder Bona-fide-Entwicklungsstörungen (z. B. geistige Behinderung), psychiatrischen oder neurologischen Erkrankungen oder

Verletzungen mindestens mittelschwerer Ausprägung zum Untersuchungszeitpunkt nicht über die kognitiven Fähigkeiten verfügte, um die moralischen, ethischen und rechtlichen Implikationen seines Verhaltens zu verstehen, und/oder zudem unfähig war, sein Verhalten zu kontrollieren, so wird neu, statt des Ausschlusses der Diagnose MND, die Diagnose sekundäre Aggravation/Simulation vorgeschlagen. Bei Personen mit leichten Erkrankungen/Verletzungen wie z. B. einer MTBI kann diese Diagnose nicht gestellt werden. Der Wegfall des Ausschlusskriteriums (D) soll der Vermeidung von Missverständnissen dienen. So könnte Kriterium D u. a. fälschlicherweise der Annahme Vorschub leisten, Patienten mit psychiatrischen, neurologischen und entwicklungsbedingten Störungen seien unfähig, Beschwerden zu übertreiben oder zu simulieren.

Änderungen wurden auch für die Definition eindeutiger bzw. definitiver und wahrscheinlicher Aggravation/Simulation neurokognitiver Störungen vorgeschlagen. Zu den eindeutigen Indikatoren für eine Aggravation/Simulation neuropsychologischer Defizite zählt neben Testleistungen unter Zufallsniveau in einem oder mehreren Forced-choice-Verfahren nun ebenso eine hohe A-posteriori-Wahrscheinlichkeit ($\geq 0,95$) in einem oder mehreren gut validierten psychometrischen Indizes auch aus Selbstbeurteilungsinstrumenten (z. B. Validitätsskalen des MMPI-2), die belegen, dass die gezeigte Leistung erheblich unter dem aktuellen Fähigkeitsniveau eines Exploranden liegt. Ziel war es, den Kriterienkatalog auch für Fälle anwendbar zu machen, in denen primär psychologische Symptome übertrieben oder vorgetäuscht werden, während für Befunde aus Leistungstests keine Aggravation/Simulation nachzuweisen ist. Der Nachweis einer hohen A-posteriori-Wahrscheinlichkeit bedingt jedoch das Vorliegen zuverlässiger Grundraten zur Auftretenswahrscheinlichkeit von Aggravation/Simulation in einer dem Exploranden vergleichbaren Population (A-priori-Wahrscheinlichkeiten) und einen gut validierten Prozess zur Bestimmung des positiven prädiktiven Wertes. Neu sollen auch beklagte Beschwerden, die mit dem beobachteten Verhalten und/oder den Testleistungen eindeutig nicht vereinbar sind oder in direktem Widerspruch dazu stehen (compelling inconsistencies), als eindeutige Indikatoren für Aggravation/Simulation gelten. Beispielhaft führen die Autoren dazu Erkenntnisse aus Beobachtungen des Gutachters (in Pausen, beim Blick aus dem Fenster usw.), Observationen, Überwachungsvideos und andere Datenquellen an. Die Kriterien für eine wahrscheinliche Aggravation/Simulation neuropsychologischer Defizite bestehen neben der Präsenz eines bedeutsamen externen Anreizes in drei oder mehr Indikatoren möglicher Aggravation/Simulation neuropsychologischer Probleme oder Defizite. Diese können aus Testdaten, die eher für ungläubwürdige Leistungen sprechen sowie markanten und unplausiblen Diskrepanzen zwischen Testleistungen und/oder Beschwerdenschilderung auf der einen Seite und beobachtbarem Verhalten, zu erwartendem Fähigkeitsniveau aufgrund der Krankengeschichte und reliablen Zusatzinformationen auf der anderen Seite bestehen.

C.2 Kriterienkatalog für die Diagnose aggravierter/simulierter schmerzassoziierter Behinderungen

Für die Aufdeckung aggravierter/simulierter schmerzassoziierter Behinderungen (Malingered Pain-Related Disabilities, MPRD) wur-

den die ursprünglichen Slick-Kriterien von Bianchini, Greve und Glynn [95] modifiziert und um ein Kriterium (B) erweitert, das Hinweise auf übertriebene Darstellung oder Vortäuschung körperlicher Behinderung aus der somatischen Untersuchung oder Behandlung (z. B. medizinische Untersuchung, EFL, Physiotherapie) liefert. Dazu zählen Leistungen in einem oder mehreren gut validierten Verfahren zur Erfassung physischer Fähigkeiten (z. B. Jamar Grip Test), die mit der übertriebenen Darstellung oder Vortäuschung körperlicher Einschränkungen vereinbar sind, Diskrepanzen zwischen subjektiven Schmerzschilderungen und physiologischer Reaktivität, eindeutig nicht organische Befunde sowie Diskrepanzen zwischen der Art, wie sich der Patient während der formalen Untersuchung präsentiert, und seinem körperlichen Funktionsvermögen, wenn er sich nicht beobachtet fühlt. Auch für die Diagnose einer definitiven Aggravation/Simulation schmerzassoziierter Behinderungen ist neben einem signifikanten externen Anreiz und dem Ausschluss von Differenzialdiagnosen ein eindeutiger Beweis des Vorsatzes notwendig. Dieser gilt nach Bianchini et al. als erbracht, wenn eine definitive negative Antwortverzerrung, also eine Leistung in einem Alternativwahlverfahren zur Erfassung kognitiver oder perzeptiver Fähigkeiten unterhalb des Zufallsniveaus ($p < 0,05$), oder eindeutige Inkonsistenzen (compelling inconsistencies) vorliegen. Die Autoren betonen, dass Aggravation/Simulation schmerzassoziierter Beeinträchtigungen nicht nur auf der somatischen Ebene, sondern auch auf der emotionalen und kognitiven Ebene zum Ausdruck kommen kann. Das rechtfertigt den Einsatz von Persönlichkeitsinventaren und neuropsychologischen Testverfahren i. e. S. bei Schmerzpatienten.

Zum Nachweis einer bewussten Täuschungsabsicht hielten Larabee, Greiffenstein und Greve [96] fest: „(...) intent is inferred as a result of the combined improbability of events rather than relying on a single definitive indication of intent“ (S. 338). Dieser Ansatz stellte eine wesentliche Erweiterung der ursprünglichen Slick-Kriterien dar, der in deren späterer Überarbeitung Berücksichtigung fand. Durch die Kombination mehrerer, inhaltlich unabhängiger PVT lässt sich nicht nur, wie oben bereits angeführt (siehe A.3), die Sensitivität erhöhen, sondern gleichzeitig auch die falsch-positiv Rate als Ganzes unter diejenige eines einzelnen PVT senken. In der Praxis wird der Grad der Unwahrscheinlichkeit in der falsch-positiv Rate bei Verwendung eines empirisch ermittelten Trennwerts, ab dem ein Testwert nicht mehr als authentisch zu betrachten ist, ausgedrückt. Ermittelt werden solche Trennwerte durch sogenannte „Known-groups“- oder „Criterion-groups“-Validierungen (siehe A.1).

C.3 Kriterienkatalog für die Diagnose einer nicht authentischen posttraumatischen Belastungsstörung

Da PTSD-Patienten häufig über kognitive Beeinträchtigungen klagen, aber auch Exploranden mit nicht authentischer PTSD oft neuropsychologische Leistungseinbußen beklagen, ist es, wie Demakis, Gervais und Rohling [97] aufzeigten, sinnvoll, auch bei dieser Patientengruppe sowohl PVT als auch Verfahren zur Überprüfung der Authentizität der Beschwerdenschilderung einzusetzen.

Young [98] stellte 2014 entsprechende Kriterien eines Diagnosesystems für „Non-credible, Feigned, or Malingered Posttraumatic Stress Disorder Related Disability/Dysfunction (F-PTSDR-D)“

vor. Mit vier Hauptkategorien (A-Evidenz eines externen Anreizes, B-Evidenzen aus Testbefunden, C-Evidenzen aus Inkonsistenzen/Diskrepanzen und D-Ausschlusskriterien) folgte er den Modellen zur Diagnose von aggravierten/simulierten neurokognitiven Störungen (MND) bzw. schmerzassozierten Behinderungen (MPRD), wobei er kritische Äußerungen und Verbesserungsvorschläge für beide Systeme [99 – 101] sowie eine Arbeit von Rubenzer [90] aufgriff und in die Entwicklung seines Diagnosesystems einfließen ließ. Zu den Neuerungen zählen eindeutigeren Definitionen und Abgrenzungen bestimmter Termini und möglicher Inkonsistenzen, Regeln zur Erhebung und Gewichtung von Testdaten sowie eine jeweils separate Graduierung der Wahrscheinlichkeit von Antwortverzerrung für die Kriterien B und C. Das System ist so gestaltet, dass es sich mit gewissen inhaltlichen Anpassungen auch für die Diagnose von aggravierten/simulierten kognitiven Störungen (Feigned Neurocognitive Related Disability/Dysfunction (F-NCR-D)) und/oder schmerzassozierten Störungen (Feigned Pain Related Disability/Dysfunction (F-PR-D)) nutzen lässt. Dadurch lassen sich bei Vorliegen von Komorbiditäten Ratings eines Systems in die beiden anderen integrieren. Eine Gesamtbetrachtung wird so wesentlich vereinfacht. Nach Meinung des Autors stellen die klaren Regeln mit Bezug zur klinischen Urteilsbildung einen bedeutenden Vorteil des Systems dar, da sie auch Berufsgruppen ohne testpsychologische Expertise zu einer reliableren und valideren Beurteilung verhelfen. Die detaillierten Operationalisierungen sollen überdies eine wissenschaftliche Überprüfung der Reliabilität, Validität und Nützlichkeit der vorgestellten Diagnosesysteme erlauben.

Diskussion

Die zentrale Aufgabe in der gutachterlichen Arbeit besteht darin, eine subjektiv vorgebrachte Beschwerdensymptomatik in Abgleich mit medizinischen Befunden in ihrer Relevanz zur Beurteilung von Funktionseinbußen zu gewichten. Dadurch erklärt sich neben der üblichen klinischen Konsistenzprüfung der Bedarf an einer reliablen, gut validierten Methodik, um mehr „Objektivität“ zu erreichen in einem Feld von oft divergierenden Interessen und Haltungen. Es muss allgemein konstatiert werden, dass in der medizinischen Begutachtung die Schlussfolgerungen zwischen subjektiver Beschwerdenangabe, dem erhobenen Befund und der konsekutiv abgeleiteten Funktionseinschränkung bislang wenig kriteriengeleitet erfolgen und sich kaum auf Evidenz stützen können. Dies gilt ausdrücklich für somatische wie auch für psychiatrische Fragestellungen.

Nach dem aktuellen Stand der oben referenzierten Literatur stellen die Überprüfung der Anstrengungsbereitschaft (Performanzvalidierung) und die Validierung der Beschwerdenschilderung in diesem Sinne zwei Bausteine der versicherungsmedizinischen Begutachtung dar, die einen wichtigen Mehrwert generieren und die insbesondere in ihrer Kombination aussagekräftig sind [46, 102]. Der zentrale Vorteil dieser ergänzenden Verfahren besteht in der inhaltlichen und zufallskritischen Absicherung von Informationen. Dabei gelten im Bereich der Performanzvalidierungsverfahren zur Überprüfung kognitiver Beschwerden die Alternativwahlverfahren als die Methode der Wahl. Sie basieren auf großen Datensätzen und zeigen eine hohe

Robustheit gegenüber realen kognitiven Defiziten auf. Sogenannte eingebettete Validitätsparameter, die aus ohnehin eingesetzten Testverfahren entwickelt wurden, bieten den Vorteil der Ökonomie und der spezifischen Beurteilung der Validität der infrage stehenden Funktion. Ihre Anwendung setzt allerdings eine umfassende Sachkenntnis des Gutachters hinsichtlich der Klassifikationsgenauigkeit und der Grenzen in der Anwendbarkeit einzelner Verfahren bei bestimmten Explorandengruppen voraus, ein unkritischer Umgang kann Fehler generieren. Die Klassifikationsgüte eines Performanzvalidierungsverfahrens wird im Wesentlichen durch die Spezifität und die Sensitivität bestimmt. Im Allgemeinen sollten diese Gütekriterien anhand einer Stichprobe von Personen geschätzt werden, die denjenigen Exploranden ähneln, an denen der Test in der klinischen Praxis eingesetzt werden soll [103, 104]. In Erfüllung dieser Vorgabe sind die meisten Performanzvalidierungsverfahren mittlerweile an realen Gutachtenpopulationen überprüft, so dass eine Reihe von Vergleichsdaten zu verschiedenen Explorandengruppen existiert. Demgegenüber sind die Verfahren zur Validierung der Authentizität der Beschwerdenschilderung im deutschsprachigen Raum noch weiterentwicklungsfähig. International wird bei der Validierung diesbezüglicher Instrumente häufig der weiter oben beschriebene SIRS als Referenzstandard eingesetzt. Deshalb ist es bemerkenswert, dass dieses Verfahren derzeit nicht in deutscher Sprache vorliegt. Im deutschsprachigen Raum stehen uns als an hiesigen Gutachtenpopulationen validierte Verfahren derzeit nur der SFSS und der MENT als Screeningverfahren sowie der MMPI-2 und mit noch dürftiger Datenlage das VEI zur Verfügung.

Aus ethischer Sicht ist in der Umsetzung einer strukturierten Beschwerdvalidierung eine minimale Falsch-positiv-Rate zu fordern, was einer submaximalen bis maximalen Spezifität entspricht. Auch wenn die ausgewiesenen Falsch-positiv-Raten der einzelnen Tests statistisch gesehen meist tatsächlich minimal sind, können sie bei singulärer Anwendung zu Fehleinschätzungen führen. Es wird deshalb allgemein empfohlen, sich auf mehrere Indikatoren zur Beschwerdvalidierung zu stützen. Dabei sollten die eingesetzten Verfahren inhaltlich möglichst unabhängig voneinander sein und idealerweise aus unterschiedlichen Funktionsbereichen stammen, damit sie wirklich einen Zugewinn an Information bieten. Mit einer solchen Kombination von Validitätsparametern mit jeweils hoher Spezifität kann die diagnostische Sicherheit beträchtlich erhöht und die Rate falsch-positiver Klassifikationen gegen null gesenkt werden [23]. Ein weiteres Argument für die Kombination mehrerer Validitätsmaße besteht darin, dass Probanden bei der Vortäuschung psychopathologischer, kognitiver und körperlicher Beeinträchtigungen recht unterschiedliche Strategien verfolgen [28, 105, 106]. Zudem lassen sich mit diesem Ansatz auch die Einflüsse von Coaching und erlernten Test-Prozeduren bei wiederholter Begutachtung in Grenzen halten [107]. Eine hohe Klassifikationsgüte wird, wie gesagt, auch mit der Kombination von Performanzvalidierungsverfahren und Instrumenten zur Beschwerdvalidierung erreicht. Beim MMPI-2 und VEI gilt ebenfalls, dass die Klassifikationsgenauigkeit mit der Zahl auffälliger, spezifischer Validitätsskalen steigt, die ebenso inhaltlich möglichst unabhängig voneinander sein sollten.

Neben der methodisch korrekten Durchführung der strukturierten Beschwerdvalidierung ist die fachlich korrekte Überführung

der erhobenen Validierungsdaten in das Gesamtgutachten ein zentraler und kritischer Schritt, der auch immer wieder Irritationen im Umgang mit der beschriebenen Methodik hervorruft. Hinsichtlich der Interpretation auffälliger Validitätsmaße und der daraus abgeleiteten Schlussfolgerungen ist ein hohes Maß an Expertise und Sorgfalt notwendig.

Von Kritikern der Testverfahren zur Authentizitätsprüfung wird häufig angeführt, dass auffällige Ergebnisse zu einem „reflexhaften“ Schluss auf Aggravation oder Simulation [107] verleiten könnten, ohne die vielfältigen potenziellen Ursachen für negative Antwortverzerrung ausreichend zu berücksichtigen. Ein solches Vorgehen hat in der Vergangenheit sicherlich auch dazu beigetragen, die Methodik der Beschwerdvalidierung als reines „Überführungsinstrument“ wahrzunehmen. Für die Annahme eines bewussten Täuschungsverhaltens bedarf es ohnehin weiterer zu erfüllender Kriterien, wie sie für vorgetäuschte kognitive Störungen [24] und vorgetäuschte schmerzassoziierte Störungen [95] mit mittlerweile guter Akzeptanz angewandt werden.

Es gilt in der Tat zu berücksichtigen, dass auch Exploranden mit genuinen psychopathologischen Auffälligkeiten, neuropsychologischen Defiziten, chronischen Schmerzen und/oder organischen Befunden nicht selten aggravieren [91]. Zunächst einmal kann aus den beschriebenen Verfahren nur abgeleitet werden, dass die beklagten Beschwerden und Funktionseinbußen nicht in der angegebenen Ausprägung und/oder Qualität vorliegen. In einer Reihe von Fällen wird der Gutachter auch damit konfrontiert sein, dass eine unzureichende Anstrengungsbereitschaft und eine invalide Selbstbeschreibung in einem Selbstbeurteilungsinventar störungsimmanent sind. So können Exploranden mit schweren kognitiven Störungen je nach Art und Ausmaß der Beeinträchtigung trotz bestmöglicher Kooperation auffällige und in diesem Fall falsch-positive Resultate in Performanzvalidierungsverfahren erzielen. Genannt seien hier beispielhaft schwere Frontallhirnstörungen und akute Psychosen.

Schon Slick berücksichtigte im Rahmen seiner 1999 formulierten Kriterien, dass auffällige Ergebnisse in Beschwerdvalidierungsverfahren durchaus auch Ausdruck einer krankheitswerten Störung sein können. Somit ergibt sich eine ganz grundlegende, bislang aus Sicht der Autoren wenig untersuchte Frage, inwieweit und unter welchen Bedingungen belegbare Beschwerdenverzerrung als krankheitszugehörig gewertet werden kann. Parallel zu einem nachweisbaren externalen Anreiz kann durchaus eine davon unabhängige Motivation zur Einnahme der Krankenrolle bestehen, die weniger der bewussten Kontrolle unterliegt. Diesbezüglich besonders in der Diskussion stehen aber auch im weitesten Sinne „neurotisch“ geprägte, zum Teil ohnehin schlecht validierte, versicherungsmedizinisch aber hochrelevante Diagnosen wie zum Beispiel chronische Erschöpfung, somatoforme Störungen oder auch Schmerzverarbeitungsstörungen. Diese Störungen gehen klinisch fast schon definitorisch zumindest mit einer Verdeutlichungstendenz in der Symptompräsentation einher. Im Hinblick auf psychische Störungen allgemein liegen auf der anderen Seite aber bis dato auch keine empirischen Belege dafür vor, dass psychische Erkrankungen in leichter bis mittelgradiger Ausprägung bei vorhandener Anstrengungsbereitschaft krankheitsbezogen zu auffälligen Resultaten in Performanzvalidierungstests führen. Auch das Vorliegen einer somatoformen

Störung oder einer anderen neurotischen Störung sollte grundsätzlich mit einer unauffälligen Bewältigung üblicher Performanzvalidierungstests vereinbar sein. Unterschreitungen entsprechender Trennwerte stellen in diesen Fällen durchaus einen relevanten, zunächst einmal klärungsbedürftigen Befund dar. Hier braucht es zur reliablen Integration der Beschwerdvalidierungsergebnisse weitere Forschungsbemühungen, um eine Einschätzung ausreichend empirisch untermauern zu können. Wenig beachtet wird aus Sicht der Autoren auch die Abgrenzung gegenüber normalpsychologischen unbewussten Lernprozessen im Rahmen von Versicherungsverfahren, in denen Exploranden unter Umständen repetitiv in ihrem Leid von Ärzten und Institutionen nicht wahrgenommen werden. Bislang unterliegt es häufig dem Expertenurteil des Gutachters, ob ein Explorand auf einem gedachten Kontinuum eher dem Pol der wenig bewusstseinsnahen Selbsttäuschung oder aber dem Pol der Täuschung anderer i. S. einer bewussten Aggravation zuzuordnen ist [18, 108]. Dies gilt neben den häufig in der Diskussion im Vordergrund stehenden neurotischen und psychosomatischen Störungen in gleicher Weise aber auch für Traumafolgestörungen. Auch angesichts der bedeutsamen Zunahme der Zahl potenzieller Trauma-Betroffener in Deutschland (institutioneller Missbrauch, Kampfeinsätze der Bundeswehr, hohe Flüchtlingszahlen aus Kriegsgebieten) ist eine auf der beschriebenen Methodik basierende Plausibilisierung von trauma-bezogenen Symptomkomplexen (z. B. im kognitiven Bereich) wichtig, um Aussagen abzusichern.

Zur Frage, wie sich eine negative Antwortverzerrung in der gutachterlichen Bewertung der Funktions- bzw. Leistungsfähigkeit niederschlagen soll, gibt es bislang unterschiedliche Positionen. Dressing, Widder und Foerster [109] empfehlen, ggf. eine vorhandene „krankheitswertige neurotische Symptomatik“ zu attestieren und festzuhalten, dass „aufgrund der zusätzlich nachweisbaren Übertreibungstendenzen des Probanden das Ausmaß der tatsächlichen Krankheit aber nicht abzuschätzen“ (S. 167) sei. Andere Autoren neigen dazu, zu argumentieren, dass bei negativer Antwortverzerrung bzw. fehlender Kooperativität der Nachweis einer tatsächlich vorliegenden Erkrankung nicht zu führen sei. In jedem Fall haben solche Inkonsistenzen Gewicht, können nicht einfach übergangen werden und machen je nach gutachtlichem Gesamtbild weitergehende Abklärungen unausweichlich. Aus Sicht der Autoren sind hier allgemeine Regeln eher nicht zielführend. Es bleibt unverändert individuell gutachterlich zu entscheiden, ob die Funktions- bzw. Leistungsfähigkeit aufgrund der vorliegenden Informationen trotz Antwortverzerrungen ausreichend sicher beurteilt werden kann. Unsicherheiten bezüglich des Ausmaßes von Einschränkungen sollten jedoch kommuniziert werden, um eine „Pseudogenauigkeit“ zu vermeiden.

Die in der vorliegenden Arbeit im Sinne einer Übersicht vorgestellte derzeit verfügbare Methodik einer Beschwerdvalidierung erbringt gesamthaft bei fachgerechter Anwendung zusätzliche Befunde, die aus Sicht der Autoren die traditionelle klinische Konsistenzprüfung in sinnvoller Weise als annähernd unabhängige Quelle ergänzen. Faktisch stellt die Methodik ein Werkzeug dar, um angegebene Beschwerden und Defizite zu plausibilisieren und gegenüber Aggravation und Simulation abzugrenzen. Dieser Erkenntnis wird zunehmend Rechnung getragen durch fachärztliche und neuropsychologische Fachverbände, die Beschwerdenva-

lidierung als essentiellen Baustein einer Begutachtung in ihre Empfehlungen aufnehmen (Konsensuskonferenz der American Academy of Clinical Neuropsychology [110], Professional Practice Board der British Psychological Society [111], Schweizerische Vereinigung der Neuropsychologinnen und Neuropsychologen [112, 113], Leitlinien für die Begutachtung bei psychischen und psychosomatischen Erkrankungen [114, 115]). Nach Ansicht der Autoren sollten die Fachverbände jedoch dabei auch vermehrt Sorge tragen, dass der klinisch wichtige Nachweis einer Antwortverzerrung/Aggravation nicht als Ausschlusskriterium eines Leistungsdefizits und dadurch mittelbar eines möglicherweise darunter liegenden Leistungsanspruchs missbraucht wird. Jüngere fallbezogene Bundesgerichtsurteile in der Schweiz wie auch die eigene Erfahrung der Autoren zeigen, dass die Relevanz der gutachterlichen Etikettierung „Aggravation“ für das Versicherungsverfahren zuungunsten der Exploranden erheblich sein kann.

In Gesamtsicht kann konstatiert werden, dass die beschriebenen Verfahren zwar in der Lage sind, wichtige und in vielen Fällen auch unverzichtbare, unerwartete Informationen zur Beurteilung der Beschwerdvalidität zu liefern, dabei jedoch zurzeit noch die Integration dieser zusätzlich erhobenen Befunde in das Gesamtgutachten nach der Erfahrung der Autoren wenig strukturiert und uneinheitlich erfolgt und wissenschaftliche Arbeiten diesbezüglich bislang fehlen. Mit anderen Worten besteht die Gefahr, dass in validierter Form erhobene reliable Befunde in invalider Weise interpretiert werden. Aus Sicht der Autoren besteht gerade hier ein erheblicher Entwicklungsbedarf.

Interessenkonflikt

Die Autoren geben an, dass kein Interessenkonflikt besteht.

Literatur

- [1] Huber M. Aspekte der Berufsunfähigkeit bei psychosomatischen Erkrankungen. *Versicherungsmedizin* 2000; 52: (2) 66–75
- [2] Stadland C et al. Kriterien zur Beurteilung der quantitativen Leistungseinschränkung bei der Begutachtung funktioneller körperlicher Störungen. Eine Literaturübersicht. *Versicherungsmedizin* 2003; 55: (3) 111–117
- [3] Schneider W et al. Berufliche Leistungsfähigkeit. Begutachtung bei psychischen und psychosomatischen Erkrankungen. *Psychotherapeut* 2010; 55: (5) 373–379
- [4] OECD. Sick on the Job? Myths and Realities about Mental Health and Work; Mental Health and Work, OECD Publishing; 2012
- [5] Ott W, Bade S, Wapf B. Nicht zielkonforme Leistungen in der Invalidenversicherung: Bedeutung und Größenordnung. Bern: Bundesamt für Sozialversicherungen; 2007
- [6] Kool J et al. Der Einsatz von Beschwerdvalidierungstests in der IV-Abklärung. Bern: Bundesamt für Sozialversicherungen; 2008
- [7] Ekman P, O'Sullivan M. Who can catch a liar? *Am Psychol* 1991; 46: (9) 913–920
- [8] Trueblood W, Binder LM. Psychologists' accuracy in identifying neuro-psychological test protocols of clinical malingerers. *Arch Clin Neuropsychol* 1997; 12: (1) 13–27
- [9] Aamondt MG, Custer H. Who can best catch a liar? A meta-analysis of individual differences in detecting deception. *The Forensic Examiner* 2006; 15: (1) 6–11
- [10] Bartlett MS et al. Automatic decoding of facial movements reveals deceptive pain expressions. *Curr Biol* 2014; 24: (7) 738–743
- [11] Cohn A, Fehr E, Marechal MA. Business culture and dishonesty in the banking industry. *Nature* 2014; 516: 86–U190
- [12] Goldfarb L, Aisenberg D, Henik A. Think the thought, walk the walk – Social priming reduces the Stroop effect. *Cognition* 2011; 118: (2) 193–200
- [13] Zedelius CM et al. A new perspective on human reward research: How consciously and unconsciously perceived reward information influences performance. *Cognitive Affective & Behavioral Neuroscience* 2014; 14: (2) 493–508
- [14] Espeland A, Vetti N, Krakenes J. Are two readers more reliable than one? A study of upper neck ligament scoring on magnetic resonance images. *BMC Med Imaging* 2013; 13: 4
- [15] Khan L et al. Inter-rater reliability between musculoskeletal radiologists and orthopedic surgeons on computed tomography imaging features of spinal metastases. *Curr Oncol* 2011; 18: (6) e282–e287
- [16] Carrino JA et al. Lumbar spine: reliability of MR imaging findings. *Radiology* 2009; 250: (1) 161–170
- [17] Jarvik JG, Deyo RA. Moderate versus mediocre: the reliability of spine MR data interpretations. *Radiology* 2009; 250: (1) 15–17
- [18] Merten T. Beschwerdvalidierung bei der Begutachtung kognitiver und psychischer Störungen. *Fortschr Neurol Psychiatr* 2011; 79: (2) 102–116
- [19] Dohrenbusch R. Symptom- und Beschwerdevalidierung chronifizierter Schmerzen in sozialmedizinischer Begutachtung. Teil I: Terminologische und methodologische Zugänge. *Schmerz* 2009; 23: (3) 231–234, 236–240
- [20] Iverson GL. Ethical issues associated with the assessment of exaggeration, poor effort, and malingering. *Appl Neuropsychol* 2006; 13: (2) 77–90
- [21] Merten T, Merckelbach H. Symptom validity testing in somatoform and dissociative disorders: A critical review. *Psychological Injury and Law* 2013; 6: 122–137
- [22] Rogers R. *Clinical assessment of Malingering and Deception*, 3rd ed; New York: Guilford; 2008
- [23] Larrabee GJ. Detection of malingering using atypical performance patterns on standard neuropsychological tests. *Clin Neuropsychol* 2003; 17: (3) 410–425
- [24] Slick DJ, Sherman EM, Iverson GL. Diagnostic criteria for malingered neurocognitive dysfunction: proposed standards for clinical practice and research. *Clin Neuropsychol* 1999; 13: (4) 545–561
- [25] Mittenberg W et al. Base rates of malingering and symptom exaggeration. *J Clin Exp Neuropsychol* 2002; 24: (8) 1094–1102
- [26] Merten T, Friedel E, Stevens A. Eingeschränkte Kooperativität in der neurologisch-psychiatrischen Begutachtung: Schätzungen zur Auftretenshäufigkeit an einer Begutachtungspopulation. *Versicherungsmedizin* 2006; 58: (1) 19–21
- [27] Stevens A et al. Malingering and uncooperativeness in psychiatric and psychological assessment: prevalence and effects in a German sample of claimants. *Psychiatry Res* 2008; 157: (1) 191–200
- [28] Greve KW et al. Prevalence of malingering in patients with chronic pain referred for psychologic evaluation in a medico-legal context. *Arch Phys Med Rehabil* 2009; 90: (7) 1117–1126
- [29] Strebel M, Brusis T. Zur Problematik der Simulation und Aggravation in der HNO-ärztlichen Begutachtung. *HNO* 2010; 58: 126–131
- [30] Hartman DE. The unexamined lie is a lie worth fibbing: neuropsychological malingering and the Word Memory Test. *Arch Clin Neuropsychol* 2002; 17: (7) 709–714
- [31] Dohrenbusch R, Merten T. Psychologische Mess- und Testverfahren. Aussagekraft in der sozialmedizinischen Begutachtung. *Psychotherapeut* 2010; 55: (5) 389–393
- [32] Rogers R, Harrell EH, Liff CD. Feigning Neuropsychological Impairment – a Critical-Review of Methodological and Clinical Considerations. *Clinical Psychology Review* 1993; 13: (3) 255–274
- [33] Tombaugh TN. *Test of Memory Malingering (TOMM)*. North Tonawanda: Multi-Health Systems; 1996
- [34] Rey A. *L'examen clinique en psychologie*. Paris: Presses Universitaires de Paris; 1958

- [35] Boone KB, Lu P, Herzberg D. The Dot Counting Test™ (DCT™). Torrance (CA): Western Psychological Services (WPS); 2002
- [36] Boone KB. Sensitivity and specificity of the Rey Dot Counting Test in patients with suspect effort and various clinical samples. *Arch Clin Neuropsychol* 2002; 17: (7) 625–642
- [37] Green P. Green's Word Memory Test. User's Manual. Edmonton: Green's Publishing; 2005
- [38] Gorissen M, Sanz JC, Schmand B. Effort and cognition in schizophrenia patients. *Schizophr Res* 2005; 78: (2) 199–208
- [39] Green P. Medical Symptom Validity Test for Windows. User's Manual and Program. 2003, revised 2005, Edmonton: Green's Publishing
- [40] Green P. Non-Verbal Medical Symptom Validity Test for Windows. User's Manual. Edmonton: Green's Publishing; 2008
- [41] Schmand B, Lindeboom J. Amsterdam Short Term Memory Test. Amsterdamer Kurzzeitgedächtnistest. Manual. Handanweisung. Leiden: PITS; 2005
- [42] Schroeder RW. Reliable Digit Span: a systematic review and cross-validation study. *Assessment* 2012; 19: (1) 21–30
- [43] Meyers J, Meyers K. Rey Complex Figure and Recognition Trial: Professional Manual. Odessa, FL: Psychological Assessment Resources; 1995
- [44] Boone KB. Assessment of Feigned Cognitive Impairment: A Neuropsychological Perspective. New York: Guilford Press; 2007
- [45] Boone KB. Clinical Practice of Forensic Neuropsychology. An Evidence-based Approach. New York: Guilford Press; 2013
- [46] Larrabee GJ. Assessment of malingered neuropsychological deficits. Oxford: Oxford University Press; 2007
- [47] Morgan JE, Sweet JJ. Neuropsychology of malingering casebook. New York: Psychology Press; 2009
- [48] Carone D, Bush SS. Mild Traumatic Brain Injury. Symptom Validity Assessment and Malingering. New York: Springer; 2013
- [49] Boone KB. The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *Clin Neuropsychol* 2009; 23: (4) 729–741
- [50] Larrabee GJ. Detection of symptom exaggeration with the MMPI-2 in litigants with malingered neurocognitive dysfunction. *Clinical Neuropsychologist* 2007; 17: 54–68
- [51] Larrabee GJ. Aggregation across multiple indicators improves the detection of malingering: relationship to likelihood ratios. *Clin Neuropsychol* 2008; 22: (4) 666–679
- [52] Merten T. Beschwerdendvalidierung. Fortschritte der Neuropsychologie, Bd. 14; Göttingen: Hogrefe; 2014
- [53] Rogers R. Detection strategies in malingering and defensiveness, in *Clinical assessment of malingering and deception*, 3rd ed; New York: Guilford; 2008: 14–38
- [54] Rogers R. Towards an empirical model of malingering and deception. *Behavioral Sciences and the Law* 1984; 2: 93–112
- [55] Rogers R, Bagby RM, Dickens SE. Structured Interview of Reported Symptoms (SIRS) and professional manual. Odessa, FL: Psychological Assessment Resources; 1992
- [56] Rogers R, Sewell KW, Gillard ND. Structured Interview of Reported Symptoms, 2nd ed. (SIRS-2) and professional manual. Odessa, FL: Psychological Assessment Resources; 2010
- [57] Schmidt T, Lanquillon S, Ullmann U. Kontroverse zu Beschwerdendvalidierungsverfahren bei der Begutachtung psychischer Störungen. *Forensische Psychiatrie, Psychologie, Kriminologie* 2011; 5: (3) 177–183
- [58] Vitacco MJ et al. An evaluation of malingering screens with competency to stand trial patients: a known-groups comparison. *Law Hum Behav* 2007; 31: (3) 249–260
- [59] Rogers R, Vitacco MJ, Kurus SJ. Assessment of malingering with repeat forensic evaluations: patient variability and possible misclassification on the SIRS and other feigning measures. *J Am Acad Psychiatry Law* 2010; 38: (1) 109–114
- [60] McCusker PJ et al. Comparability of the MMPI-2 F(p) and F scales and the SIRS in clinical use with suspected malingerers. *Int J Offender Ther Comp Criminol* 2003; 47: (5) 585–596
- [61] Cima M et al. „Strukturierter Fragebogen Simulierter Symptome“. Die deutsche Version des „Structured Inventory of Malingered Symptomatology: SIMS“. *Nervenarzt* 2003; 74: (11) 977–986
- [62] Lanquillon S, Schmidt T, Stieglitz RD. *Strukturiertes Interview Berichteter Beschwerden SIBB*. Bern: Huber; 2011
- [63] Miller HA. *Miller-Forensic Assessment of Symptoms Test professional manual*. Odessa, FL: Psychological Assessment Resources; 2001
- [64] Smith GP. Brief screening measures for the detection of feigned psychopathology, in *Clinical Assessment of Malingering and Deception*, 3rd ed; R. Rogers, Editor. New York: Guilford; 2008: 323–339
- [65] Guy L, Miller HA. Screening for malingered psychopathology in a correctional setting – Utility of the Miller-Forensic Assessment of Symptoms Test (M-FAST). *Criminal Justice and Behavior* 2004; 31: (6) 695–716
- [66] Miller HA. The Miller-Forensic Assessment of Symptoms Test (M-FAST): Test generalizability and utility across race, literacy, and clinical opinion. *Criminal Justice and Behavior* 2005; 32: (6) 591–611
- [67] Smith GP, Burger GK. Detection of malingering: validation of the Structured Inventory of Malingered Symptomatology (SIMS). *J Am Acad Psychiatry Law* 1997; 25: (2) 183–189
- [68] Wisdom NM, Callahan JL, Shaw TG. Diagnostic utility of the structured inventory of malingered symptomatology to detect malingering in a forensic sample. *Arch Clin Neuropsychol* 2010; 25: (2) 118–125
- [69] van Impelen A et al. The Structured Inventory of Malingered Symptomatology (SIMS): A Systematic Review and Meta-Analysis. *Clin Neuropsychol* 2014; 28: 1336–1365
- [70] Butcher JN et al. *Minnesota Multiphasic Personality Inventory-2 (MMPI-2). Manual for administration, scoring and interpretation. (rev.ed.)*. Minneapolis, MN: University of Minnesota Press; 2001
- [71] Gervais RO et al. Development and validation of a response bias scale (RBS) for the MMPI-2. *Assessment* 2007; 14: (2) 196–208
- [72] Gervais RO et al. Differential sensitivity of the Response Bias Scale (RBS) and MMPI-2 validity scales to memory complaints. *Clin Neuropsychol* 2008; 22: (6) 1061–1079
- [73] Rogers R, Bender SD. Evaluation of malingering and deception, in *Handbook of psychology: forensic psychology* I.B. Weiner and A.M. Goldstein, Editors. Hoboken, NJ: Wiley & Sons; 2003: 109–131
- [74] Nelson NW et al. Updated meta-analysis of the MMPI-2 symptom validity scale (FBS): verified utility in forensic practice. *Clin Neuropsychol* 2010; 24: (4) 701–724
- [75] Lewis JL, Simcox AM, Berry DT. Screening for feigned psychiatric symptoms in a forensic sample by using the MMPI-2 and the structured inventory of malingered symptomatology. *Psychol Assess* 2002; 14: (2) 170–176
- [76] Baer RA, Miller J. Underreporting of psychopathology on the MMPI-2: a meta-analytic review. *Psychol Assess* 2002; 14: (1) 16–26
- [77] Engel RR. *MMPI-2. Minnesota Multiphasic Personality Inventory-2 Manual*. Bern: Hans Huber; 2000
- [78] Thies E. *Der deutsche MMPI-2: Effektivität der Validitätsskalen in der Aufdeckung von Antwortverzerrung*. Marburg: Tectum; 2012
- [79] Butcher JN, Lim J, Nezami E. Objective study of abnormal personality in cross-cultural settings: The Minnesota Multiphasic Personality Inventory (MMPI-2). *Journal of Cross-Cultural Psychology* 1998; 29: 189–211
- [80] Dean AC et al. Examination of the impact of ethnicity on the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) Fake Bad Scale. *Clin Neuropsychol* 2008; 22: (6) 1054–1060
- [81] Hall GCN, Bansal A, Lopez IR. Ethnicity and psychopathology: A meta-analytic review of 31 years of comparative MMPI/MMPI-2 research. *Psychological Assessment* 1999; 11: 186–197
- [82] Knaster CA, Micucci JA. The effect of client ethnicity on clinical interpretation of the MMPI-2. *Assessment* 2013; 20: (1) 43–47
- [83] Morey LC. *The personality assessment inventory professional manual*, 2nd ed; Odessa, FL: Psychological Assessment Resources; 2007
- [84] Hawes SW, Boccaccini MT. Detection of overreporting of psychopathology on the Personality Assessment Inventory: a meta-analytic review. *Psychol Assess* 2009; 21: (1) 112–124
- [85] Engel RR, Groves JA. *Verhaltens- und Erlebensinventar. Deutschsprachige Adaption des Personality Assessment Inventory (PAI) von L.C. Morey*. Bern: Hans Huber; 2013
- [86] Vossler-Thies E et al. Erfassung negativer Antwortverzerrungen mit der deutschen Fassung des „Personality Assessment Inventory“, dem „Verhaltens- und Erlebensinventar“. *Diagnostica* 2013; 59: (2) 73–85

- [87] Morel KR. Development and preliminary validation of a forced-choice test of response bias for posttraumatic stress disorder. *J Pers Assess* 1998; 70: (2) 299–314
- [88] Morel KR. Manual for the Morel Emotional Numbing Test for Posttraumatic Stress Disorder: Psychometric Properties, 3rd ed. rev; Las Vegas: NV: Author; 2012
- [89] Morel KR. Cutoff Scores for the Morel Emotional Numbing Test for PTSD: Considerations for Use in VA Mental Health Examinations. *Psychological Injury and Law* 2013; 6: 138–143
- [90] Rubenzer S. Posttraumatic Stress Disorder: Assessing Response Style and Malingering. *Psychological Injury and Law* 2009; 2: 114–142
- [91] Merten T et al. Symptom Validity Testing in Claimants with Alleged Posttraumatic Stress Disorder: Comparing the Morel Emotional Numbing Test, the Structured Inventory of Malingered Symptomatology, and the Word Memory Test. *Psychological Injury and Law* 2009; 2: 284–293
- [92] Geraerts E et al. Detecting deception of war-related posttraumatic stress disorder. *Journal of Forensic Psychiatry and Psychology* 2009; 20: 278–285
- [93] Wiedmaier P. Schätzung von Normalbefunden für drei Beschwerdenvvalidierungstests: Der Morel Emotional Numbing Test (MENT), der Strukturierte Fragebogen Simulierter Symptome und die List of Improbable Problems (LIMP). Saarbrücken: VDM Verlag Dr. Müller; 2011
- [94] Slick DJ, Sherman EM. Differential Diagnosis of Malingering, in *Mild Traumatic Brain Injury*, in Symptom Validity Assessment and Malingering, D. Carone and S.S. Bush, Editors. New York: Springer Publishing Company; 2013: 57–72
- [95] Bianchini KJ, Greve KW, Glynn G. On the diagnosis of malingered pain-related disability: lessons from cognitive malingering research. *Spine J* 2005; 5: (4) 404–417
- [96] Larrabee GJ et al. Refining diagnostic criteria for malingering, in *Assessment of malingered neuropsychological deficits*, G.J. Larrabee, Editor. Oxford: New York; 2007: 334–371
- [97] Demakis GJ, Gervais RO, Rohling ML. The effect of failure on cognitive and psychological symptom validity tests in litigants with symptoms of post-traumatic stress disorder. *The Clinical Neuropsychologist* 2008; 22: 879–895
- [98] Young G. Malingering, Feigning, and Response Bias in Psychiatric/Psychological Injury. Implications for Practice and Court. *International Library of Ethics, Law, and the New Medicine*, Vol. 56; Dordrecht: Springer SBM; 2014
- [99] Boone KB. Clarification or confusion? A review of Rogers, Bender, and Johnson's a critical analysis of the MND criteria for feigned cognitive impairment: implications for forensic practice and research. *Psychological Injury and Law* 2011; 4: 157–162
- [100] Rogers R, Bender SD, Johnson SF. A critical analysis of the MND criteria for feigned cognitive impairment: Implications for forensic practice and research. *Psychological Injury and Law* 2011; 4: 147–156
- [101] Rogers R, Bender SD, Johnson SF. A commentary on the MND model and the Boone critique: "Saying it doesn't make it so". *Psychological Injury and Law* 2011; 4: 162–167
- [102] Bianchini KJ et al. Classification accuracy of MMPI-2 validity scales in the detection of pain-related malingering: a known-groups study. *Assessment* 2008; 15: (4) 435–449
- [103] Bossuyt PM et al. Toward complete and accurate reporting of studies of diagnostic accuracy. The STARD initiative. *Am J Clin Pathol* 2003; 119: (1) 18–22
- [104] Henry M. Beschwerdenvvalidierungstests in der zivil- und sozialrechtlichen Begutachtung: Verfahrensüberblick, in *Diagnostik der Beschwerdenvvalidität*, T. Merten and H. Dettenborn, Editors. Berlin: Deutscher Psychologen Verlag GmbH; 2009: 118–161
- [105] Greiffenstein MF, Baker WJ, Gola T. Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment* 1994; 6: (3) 218–224
- [106] Nelson NW et al. Response validity in forensic neuropsychology: exploratory factor analytic evidence of distinct cognitive and psychological constructs. *J Int Neuropsychol Soc* 2007; 13: (3) 440–449
- [107] Blaskewitz N, Merten T. Diagnostik der Beschwerdenvvalidität – Diagnostik bei Simulationsverdacht: ein Update 2002 bis 2005. *Fortschr Neurol Psychiatr* 2007; 75: (3) 140–154
- [108] Merten T. Fragen der neuropsychologischen Diagnostik bei Simulationsverdacht. *Fortschr Neurol Psychiatr* 2002; 70: (3) 126–138
- [109] Dressing H, Widder B, Foerster K. Kritische Bestandsaufnahme zum Einsatz von Beschwerdenvvalidierungstests in der psychiatrischen Begutachtung. *Versicherungsmedizin* 2010; 62: 163–167
- [110] Heilbronner RL et al. American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *Clin Neuropsychol* 2009; 23: (7) 1093–1129
- [111] McMillan TM et al. Assessment of effort in clinical testing of cognitive functioning for adults. Professional Practice Board. Leicester: British Psychological Society; 2009: 1–27
- [112] Neuropsychologen, S.V.d.N.u. Leitlinien für neuropsychologische Gutachten. 2011
- [113] Frei A et al. Kriterien zur Bestimmung des Schweregrades einer neuropsychologischen Störung sowie Zuordnungen zur Funktions- und Arbeitsfähigkeit. *Zeitschrift für Neuropsychologie* 2016; 27: 107–119
- [114] Schneider W et al. Begutachtung der Leistungsfähigkeit bei Personen mit psychischen und psychosomatischen Störungen – ein diagnostisches Modell. *Versicherungsmedizin* 2011; 63: (2) 68–75
- [115] Schneider W et al. Begutachtung bei psychischen und psychosomatischen Erkrankungen: Autorisierte Leitlinien und Kommentare, 2., überarbeitete und erweiterte Auflage; Bern: Hogrefe; 2016

ANHANG A-3

Publikation 3:

Frei, A., Balzer, C., Gysi, F., Leros, J., Plohmann, A. M., & Steiger-Bächler, G. (2016). Kriterien zur Bestimmung des Schweregrades einer neuropsychologischen Störung sowie Zuordnungen zur Funktions- und Arbeitsfähigkeit. *Zeitschrift für Neuropsychologie*, 27 (2), 107-119.



Kriterien zur Bestimmung des Schweregrades einer neuropsychologischen Störung sowie Zuordnungen zur Funktions- und Arbeitsfähigkeit

Adrian Frei^{1,2}, Christian Balzer^{1,3}, Françoise Gysi^{1,4}, Julie Leros^{1,5}, Andrea Plohmann^{1,6}, Gregor Steiger^{1,7}

¹Arbeitsgruppe im Auftrag des Vorstandes der Schweizerischen Vereinigung der Neuropsychologinnen und Neuropsychologen (SVNP), Basel

²Clienia Schlössli, Privatklinik für Psychiatrie und Psychotherapie, Oetwil am See

³Reha Rheinfelden, Rheinfelden

⁴Reha Clinic, Bad Zurzach

⁵Consultation de Neuropsychologie, La Tour-de-Peilz

⁶Praxis für Neuropsychologie und Psychotherapie, Basel

⁷asim Begutachtung, Universitätsspital Basel



Zusammenfassung: Im Folgenden wird ein Beurteilungssystem zur Einschätzung des Schweregrades von neuropsychologischen Störungen vorgestellt. Insbesondere umfasst es spezifische Kriterien zur Ermittlung des Schweregrades einer Störung, die wiederum in Beziehung zur Funktions- und Arbeitsfähigkeit eines Patienten oder Versicherten gesetzt werden. Das Beurteilungssystem beruht auf den in der Suva-Tabelle 8 zur Einschätzung der psychischen Folgen einer Hirnverletzung definierten Kriterien. Infolgedessen ist es auch grundsätzlich damit vereinbar, stellt aber aufgrund des Miteinbezugs von neuropsychologischen Testergebnissen in Form von Standardwerten als Beurteilungskriterium sowie verbunden mit den Erläuterungen zu deren Interpretation und Anwendung eine wesentliche Weiterentwicklung dar. Im Gegensatz zur Suva-Tabelle 8 ist das Beurteilungssystem unabhängig vom kausalen Versicherungskontext des UVG. Es kann somit unter anderem auch in Rahmen des IVG und KVG/VVG angewandt werden. In diesem Sinne ist es sowohl unabhängig von der Ätiologie des jeweiligen Beschwerdebildes wie auch unabhängig von der artifiziellen dichotomen Unterscheidung zwischen „organischen“ und „nichtorganischen“ psychischen Störungen. Mit der Erstellung und Publikation dieser neuen Kriterien-Tabelle soll ein Beitrag zur weiteren Verbesserung der Interrater-Reliabilität im Rahmen der Beurteilung des Schweregrades von neuropsychologischen Störungen sowie der Einschätzung der Funktionsfähigkeit geleistet werden. Nicht zuletzt ist das Beurteilungssystem auch kompatibel mit den im DSM-5 festgelegten Kriterien zur Einschätzung des Schweregrades einer neurokognitiven Störung.

Schlüsselwörter: neuropsychologische Störung, Schweregrad, Funktionsfähigkeit, Arbeitsfähigkeit

Criteria for Determining the Severity of a Neuropsychological Disorder as well as Classifications for Functionality and Working Ability

Abstract: An assessment system for assessing the severity of neuropsychological disorders is presented in the following. In particular, it includes specific criteria for determining the severity of a disorder, which are in turn set in relation to the functionality and working ability of a patient or insured person. The assessment system is based on the criteria defined in Suva table 8 for assessing the psychological consequences of a brain injury. As a result, it is fundamentally consistent with it, but constitutes also a significant further development due to the inclusion of neuropsychological test results in the form of standard values as assessment criteria as well as combined with the explanations of their interpretation and application. In contrast to Suva table 8, the assessment system is independent of the insurance context defined in the Swiss Accident Insurance Law (UVG). It can therefore also be applied within the framework of the Swiss Invalidity Insurance Law (IVG) as well as the Swiss Health Insurance Law (KVG) and Swiss Law on Insurance Contracts (IVG). In this sense it is independent of the aetiology of the respective symptoms as well as independent of the artificial dichotomous difference between “organic” and “inorganic” mental disorders. The preparation and publication of this new criteria table shall provide a contribution towards further improvement of interrater reliability in the course of assessing the severity of neuropsychological disorders as well as assessing the

functionality. Last but not least, the assessment system is also compatible with the criteria for assessing the severity of a neurocognitive disorder defined in the Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM-5).

Keywords: neuropsychological disorder, severity, functionality, working ability

Ausgangslage

Nach Art. 24 Abs. 1 UVG hat jeder Versicherte, der durch einen Unfall eine dauernde und erhebliche Schädigung der körperlichen oder geistigen Integrität erleidet, Anspruch auf eine angemessene Integritätsentschädigung. Zur Einschätzung des Schweregrades einer neuropsychologischen Störung sowie zur Bemessung der Integritätsentschädigung bei psychischen Folgen von Hirnverletzungen werden die Kriterien der Suva-Tabelle 8 herangezogen (Suva, 2002). Die Kriterien der Suva-Tabelle 8 haben schweizweit zu einer Verbesserung der Interrater-Reliabilität bei der Einschätzung des Schweregrades von neuropsychologischen Störungen und bei der Beurteilung der Funktionalität der Betroffenen geführt. Nicht zuletzt hat die Suva-Tabelle 8 auch Eingang in deutschsprachige Standard-Werke zum Thema der Begutachtung gefunden (z. B. Stöckli, 2007, S. 473). Seit ihrer Publikation im Jahre 2002 wurden die Kriterien der Suva-Tabelle 8 in Ermangelung anderer Beurteilungssysteme jedoch zusehends auch zur Einschätzung des Schweregrades von neuropsychologischen Funktionseinschränkungen infolge von Krankheiten – also ausserhalb des kausalen Versicherungskontexts des UVG – sowie im Rahmen der Beurteilung der Arbeitsfähigkeit herangezogen, was aber in mancherlei Hinsicht problematisch ist.

Problematik

Insbesondere dient die Suva-Tabelle 8 der Einschätzung des sogenannten Integritätsschadens der neuropsychologischen Folgen einer Hirnverletzung. Während die Höhe des Integritätsschadens somit verbindlich durch die Suva-Tabelle 8 festgelegt ist, gibt es bezüglich der Invaliditätshöhe oder des Grades der Arbeitsunfähigkeit keine Zuordnungen. Die Höhe des Integritätsschadens entspricht auch nicht unbedingt der Invaliditätshöhe beziehungsweise dem Grad der Arbeitsunfähigkeit, was immer wieder zu Verwechslungen und Missverständnissen führt.

Weiter bezieht sich Suva-Tabelle 8 ausschliesslich auf neuropsychologische Störungen, die auf eine medizinisch dokumentierte und *unfallbedingte hirnorganische Schädigung* zurückzuführen sind und die eine dauerhafte Störung

zur Folge haben. Darauf beruhend, wurde von Nichtfachkundigen gelegentlich der falsche Schluss gezogen, dass kognitive Defizite ausschliesslich infolge von nachweislich vorhandenen hirnorganischen Schädigungen zustande kämen. Kognitive Defizite können jedoch auch infolge von körperlichen Erkrankungen entstehen und finden sich auch bei beinahe allen „nichtorganischen“ psychischen Krankheiten – insbesondere bei affektiven Störungen, Schizophrenien, hyperkinetischen Störungen oder Persönlichkeitsstörungen (siehe insbesondere die einschlägige Anthologie von Lautenbacher & Gauggel, 2004, aber auch Falkai & Wittchen, 2015, S. 811). Die Verwendung der Kriterien der Suva-Tabelle 8 im Rahmen der Einschätzung von krankheitsbedingten neuropsychologischen Störungen ist also vor allem deswegen problematisch, weil sich diese auf einen nachweisbaren Organschaden beziehen. Ein organisches Korrelat kann aber bei „nichtorganischen“ psychischen Störungen in der Regel nicht nachgewiesen werden. Die gängige Unterteilung in „organische“ und „nichtorganische“ psychische Störungen ist im Übrigen ohnehin problematisch. Auf dem Hintergrund der aktuellen medizinischen Modelle und Erkenntnisse ist sie sogar als obsolet und artifiziell zu betrachten. In der ICD-10 wird diese Dichotomie hauptsächlich aus historischen und pragmatischen Gründen aufrechterhalten (Dilling, Mombour & Schmidt, 2014, S. 73). Im welchem Ausmass diese Unterteilung paradox erscheinen kann, lässt sich an den Beispielen der depressiven Störungen oder der posttraumatischen Belastungsstörung illustrieren. Bei diesen eigentlichen „nichtorganischen“ Beschwerdebildern sind gemäss neueren Erkenntnissen aus der neuroradiologischen Forschung manchmal durchaus strukturelle Auffälligkeiten – unter anderem in Form einer Volumenminderung der Hippocampi – festzustellen (McKinnon, Yucel, Nazarov & MacQueen, 2009; Bossini et al., 2008).

Auch müssen kognitive Defizite und anderweitige neuropsychologische Störungen nicht unbedingt von Dauer sein, wie dies in der Tabelle 8 im Kontext der Unfallversicherung und der Einschätzung des Integritätsschadens gefordert wird. Beispielweise kann es bei Verhaltensstörungen durch psychotrope Substanzen oder im Rahmen von unerwünschten Medikamentennebenwirkungen durchaus zu vorübergehenden kognitiven Defiziten mit vollständiger Remission kommen. Fluktuierende kognitive Defizite können mitunter sogar eine Kernsymptomatik eines Beschwerdebildes darstellen, wie dies

bei der Lewy-Körper-Demenz der Fall ist. Die Forderung nach einer dauerhaft anhaltenden Störung macht somit ausserhalb des kausalen Versicherungskontextes sowie ausserhalb des Rahmens eines Fallabschlusses häufig keinen Sinn. Insbesondere ist die Forderung inhaltlich nicht vereinbar mit dem Auftrag der Verlaufsdocumentation, wie sie im klinischen Alltag der Neuropsychologen eine wichtige Aufgabe darstellt.

Beruhend auf dem in der Suva-Tabelle 8 geforderten Nachweis einer organischen Ursache einer neuropsychologischen Störung wurde von verschiedenen Seiten noch vor wenigen Jahren implizit und fälschlicherweise der Schluss gezogen, dass die von Patienten mit „nichtorganischen“ Störungen beklagten kognitiven Defizite unweigerlich als nichtauthentische Beschwerden einzuordnen seien. Unter anderem im Zuge der Einführung der sogenannten „Überwindbarkeitspraxis“ wurde diese Sichtweise richtigerweise korrigiert (BGE 130 V 352). Es wurde sodann zwar attestiert, dass organisch nicht begründbare Beschwerden durchaus als authentische Symptome und demzufolge nicht zwangsläufig als Ausdruck einer Aggravation oder Simulation zu werten, in jedem Fall aber überwindbar und somit nicht rentenrelevant seien. Diese Sichtweise, organisch nicht begründbare Beschwerden in jedem Fall als überwindbar anzusehen, ist in Fachkreisen jedoch umstritten (Jeger, 2014). Anfang Juni 2015 hat das Schweizerische Bundesgericht seine seit 2004 geltende Haltung zum Rentenanspruch von Patienten mit organisch nichtbegründbaren Schmerzstörungen revidiert. Ungeachtet der derzeit laufenden Diskussion um die Angemessenheit der „Überwindbarkeitspraxis“, ist es wichtig hervorzuheben, dass die Authentizität eines Beschwerdebildes in vielen Fällen nicht in Abhängigkeit des Nachweises eines organischen Korrelats, sondern – zumindest im Kontext einer neuropsychologischen Untersuchung – anhand der verschiedenen Verfahren der Symptomvalidierung zu beurteilen ist (Symptomvalidierungstests, Fragebögen, „eingebettete“ Faktoren, Slick-Algorithmus von Slick et al., 1999).

Zielsetzungen

Mit der Erstellung und Publikation einer neuen Kriterien-Tabelle wurden somit folgende Ziele verfolgt:

- Die neue Kriterien-Tabelle stellt ein vom jeweiligen *Versicherungskontext unabhängiges Beurteilungssystem* dar. Insbesondere ist das Beurteilungssystem unabhängig vom kausalen Versicherungskontext des UVG. Unter anderem soll es auch in Rahmen des IVG und KVG/VVG angewandt werden können.

- Die neuen Kriterien sind grundsätzlich *vereinbar mit denjenigen der Suva-Tabelle 8*. Das Beurteilungssystem stellt aber – verbunden mit den Erläuterungen zu dessen Interpretation und Anwendung – eine wesentliche Weiterentwicklung dar.
- Die Beurteilungskriterien sind unabhängig von der Ätiologie des jeweiligen Beschwerdebildes und damit *unabhängig von der artifiziellen dichotomen Unterscheidung zwischen „organischen“ und „nichtorganischen“ psychischen Störungen*. Dies bedeutet aber nicht, dass im Rahmen der diagnostischen Einordnung nach ICD oder DSM sowie der neuropsychologischen Berichterstattung keine ätiologischen Zusammenhänge aufgezeigt werden sollten beziehungsweise keine Erläuterungen zur Pathogenese einer Störung erforderlich wären. Entsprechende differenzialdiagnostische Überlegungen sind jedoch Ausdruck einer anderen Reflexions- und Argumentationsebene und haben nichts gemeinsam mit der Einschätzung des Schweregrades einer Störung oder der Beschreibung einer Funktionseinschränkung.
- Aufgrund eines höheren Detaillierungsgrades sowie dank des Miteinbezugs von neuropsychologischen Testergebnissen in Form von Standardwerten als Beurteilungskriterium soll die neue Tabelle einen Beitrag zur weiteren *Verbesserung der Interrater-Reliabilität* im Rahmen der Einschätzung des Schweregrades einer neuropsychologischen Störung leisten.
- Die Art und Weise des Miteinbezugs der ermittelten neuropsychologischen Testergebnisse in die Beurteilung des Schweregrades einer Störung entspricht den im DSM-5 ausformulierten Empfehlungen. Die *Vereinbarkeit mit dem DSM-5* ist somit gewährleistet.
- Das neue Beurteilungssystem dient als Hilfestellung bei der Einschätzung der – aus einer neuropsychologischen Störung resultierenden – Arbeitsunfähigkeit. Es bietet also orientierende Richtwerte bezüglich des Grades der Arbeitsunfähigkeit in Relation zum Ausmass einer jeweiligen neuropsychologischen Störung. Der Grad der Arbeitsunfähigkeit kann jedoch – in Abhängigkeit der Charakteristika einer Störung sowie des jeweiligen beruflichen Anforderungsprofils – erheblich von diesen orientierenden Richtwerten abweichen. Insbesondere ist hierbei zwischen der Einschätzung der Arbeitsunfähigkeit im Rahmen der ursprünglichen Tätigkeit sowie im Rahmen einer leidensangepassten Tätigkeit zu unterscheiden. Die Einschätzung der Arbeitsunfähigkeit erfolgt im Rahmen einer neuropsychologischen Beurteilung unter Berücksichtigung der neuropsychologischen Funktionsstörung. Darüber hinaus kann zusätzlich eine Arbeitsunfähigkeit aufgrund somatischer oder psychiatrischer Faktoren bestehen.

Das Beurteilungssystem

Die Tabelle 1 vermittelt einen Überblick über die Kriterien zur Bestimmung des Schweregrades einer neuropsychologischen Störung. Der ermittelte Schweregrad der Störung wird spezifischen Kriterien der Funktionsfähigkeit zugeordnet sowie – abgeleitet von diesen Kriterien wiederum – die Arbeitsfähigkeit eingeschätzt.

- a) Das jeweilige „Kriterium a“ bezieht sich auf *die kognitiven Funktionen*. Diese sind ungeachtet des Nachweises eines organischen Substrats zu bewerten – also unabhängig davon, ob eine „organische“ oder „nicht-organische“ psychische Störung vorliegt. Die hypothesengeleitete, testpsychologische Abklärung der kognitiven Funktionen stellt die Kernaufgabe der Neuropsychologie dar. Die Interpretation der neuropsychologischen Testergebnisse entspricht dabei der im DSM-5 vorgeschlagenen Vorgehensweise (Falkai & Wittchen, 2015, S. 811 ff.). Testresultate zwischen einer bis zwei Standardabweichungen (SD) unter dem jeweiligen Mittelwert sprechen im Allgemeinen für eine leichte Störung. Liegen die Testergebnissen mehr als zwei Standardabweichungen unterhalb des Mittelwertes, so spricht dies für eine Störung höheren Schweregrads. In Anlehnung zu dieser Vorgehensweise wird auch auf die Leitlinien der Schweizerischen Vereinigung der Neuropsychologen zur Klassifikation und Interpretation neuropsychologischer Testergebnisse verwiesen (Schweizerische Vereinigung der Neuropsychologen, 2014).
- b) Das jeweilige „Kriterium b“ bezieht sich auf *weitere psychische Bereiche* – wie insbesondere die Affektivität, das Sozialverhalten, die Kritikfähigkeit oder die Persönlichkeit. Die Beurteilung dieser zusätzlichen psychischen Bereiche kann im Rahmen einer klinischen Einschätzung, unter Verwendung von Selbst- und Fremdbeurteilungsfragebögen sowie anhand von strukturierten und standardisierten psychopathologischen Instrumenten erfolgen. Beispielsweise können aber Störungen im Bereich des Sozialverhaltens und der sozialen Kognitionen seit einigen Jahren auch mit Hilfe relativ gut normierter Testverfahren erfasst werden (z. B. Faces-Test, Baron-Cohen, Wheelwright & Jolliffe, 1997; Fauxpas-Test, Stone, Baron-Cohen & Knight, 1998; Reading Mind in the Eyes-Test, Bölte, 2005). Im interdisziplinären Kontext ist zu klären, inwieweit diese zusätzlichen psychischen Bereiche entweder durch den Psychiater oder den Neuropsychologen, oder aber in gemeinsamer Zusammenarbeit der psychiatrischen und neuropsychologischen Fachpersonen beurteilt werden sollen. Beim Vorliegen von sogenannten „PÄUSBONOG“, das heisst von pathoge-

netisch-ätiologisch unklaren syndromalen Beschwerdebildern ohne nachweisliche organische Grundlage (BGE 130 V 352; Jeger, 2011) – wie sie beispielsweise die Fibromyalgie, die Neurasthenie, leichtere depressive Störungen, chronifizierte Beschwerden nach Erleiden eines kraniozervikalen Beschleunigungstraumas oder die somatoformen Störungen darstellen – gehört die Beurteilung dieser zusätzlichen psychischen Bereiche in der Regel nicht zu den Hauptaufgaben der Neuropsychologie.

Die folgenden Erläuterungen dienen dem besseren Verständnis der Kriterien und einer einheitlichen Nutzung der Tabelle 1. Insbesondere werden Aspekte in Zusammenhang mit der Testauswahl und den Auswertungsmodalitäten, der Symptomvalidierung, der Einschätzung der Arbeitsfähigkeit sowie der interdisziplinären Zusammenarbeit beleuchtet. Mitunter wird auf die gelegentlich anzutreffende Divergenz zwischen dem Ausprägungsgrad von kognitiven Defiziten und dem Schweregrad von Symptomen in den weiteren psychischen Bereichen eingegangen – ein Umstand, der bei der Gesamteinschätzung des Schweregrades einer Störung manchmal Schwierigkeiten bereiten kann.

Auswahl der Testverfahren, Auswertungsmodalitäten und Darstellung der Befunde

Gemäss Sturm und Wallesch (2007) lassen sich die höheren Hirnleistungen in sogenannte Basisfunktionen (Orientiertheit, Emotionalität, Persönlichkeit, Antrieb, Aufmerksamkeitsfunktionen), deren Störungen auch andere höhere Hirnleistungen beeinflussen können, sowie in spezifische „Werkzeug“-Funktionen (Gedächtnisfunktionen, Sprache, visuokognitive und -konstruktive Funktionen, Praxis, Gnosis, exekutive Funktionen), die weitgehend isoliert gestört sein können, unterteilen. Die Aufgabe des Neuropsychologen besteht unter anderem darin, diese kognitiven und psychischen Funktionen zu erfassen, Störungen in diesen Bereichen zu erkennen und hinsichtlich des Schweregrades ihrer Ausprägung zu bewerten. Von entscheidender Bedeutung ist dabei der Einsatz testpsychologischer Untersuchungsverfahren mit dem Ziel der zuverlässigen Quantifizierung der kognitiven und psychischen Funktionen und der Aufdeckung von regelwidrigen, pathologischen Funktionsbeeinträchtigungen oder Funktionsschwächen. Oft muss auch die Frage des ursächlichen Zusammenhangs zwischen eventuell festgestellten Störungen und einer bestimmten Gesundheitsschädigung, zum Beispiel einer unfallbedingten Hirnschädigung, geprüft und beantwortet werden (Hartje, 2006).

Tabelle 1. Kriterien zur Bestimmung des Schweregrades einer neuropsychologischen Störung, Zuordnungen zur Funktionsfähigkeit und orientierende Richtwerte bezüglich der Arbeitsunfähigkeit

Schweregrad der Störung und diagnostische Kriterien ¹	Funktionsfähigkeit im privaten Alltag und Beruf	Orientierende Richtwerte bezüglich der Arbeitsunfähigkeit ²
<p><u>Minimale neuropsychologische Störung:</u></p> <p>a) Nur unter starker Belastung oder durch neuropsychologische Tests feststellbare leichte Minderleistungen einer oder einzelner kognitiver Teilfunktionen (1 bis 2 SD unter dem Mittelwert) <i>und/oder</i> ...</p> <p>b) Keine fassbaren oder nur unter starker Belastung vorhandene Auffälligkeiten in den Bereichen der Affektivität, des Verhaltens oder der Persönlichkeit</p>	<p>Die Person kann sich subjektiv gestört fühlen. Ihre Funktionsfähigkeit ist aber im privaten Alltag nicht eingeschränkt. Und berufliche Leistungen werden praktisch unvermindert vollbracht. Die Person fällt in ihrem sozialen Umfeld nicht auf. Bei Aufgaben und Tätigkeiten mit sehr hohen Anforderungen kann die Funktionsfähigkeit jedoch leicht eingeschränkt sein.</p>	<p>Grad der Arbeitsunfähigkeit von 0 bis 10 %</p>
<p><u>Leichte neuropsychologische Störung:</u></p> <p>a) Leichte Minderleistungen <u>mehrerer kognitiver Teilfunktionen</u> (1 bis 2 SD unter dem Mittelwert) <i>und/oder</i> ...</p> <p>b) Leichte Auffälligkeiten in den Bereichen der Affektivität, des Verhaltens oder der Persönlichkeit</p>	<p>Die Funktionsfähigkeit ist im Alltag und unter den meisten beruflichen Anforderungen nicht eingeschränkt. Die Person fällt in ihrem sozialen Umfeld auch kaum auf. Bei Aufgaben und Tätigkeiten mit <u>hohen Anforderungen</u> ist die Funktionsfähigkeit aber eingeschränkt.</p>	<p>Grad der Arbeitsunfähigkeit von 10 bis 30 %</p>
<p><u>Leichte bis mittelgradige neuropsychologische Störung:</u></p> <p>a) Eine oder allenfalls zwei kognitive Teilfunktionen sind deutlich (mehr als 2 SD unter dem Mittelwert) sowie weitere leicht vermindert (1 bis 2 SD unter dem Mittelwert), <i>und/oder</i> ...</p> <p>b) Leichte bis mittelschwere Auffälligkeiten in den Bereichen der Affektivität, des Verhaltens oder der Persönlichkeit</p>	<p>Die Funktionsfähigkeit ist im Alltag und unter den <u>meisten</u> beruflichen Anforderungen <u>leicht</u> eingeschränkt. Die Person fällt in ihrem sozialen Umfeld leicht auf. In Berufen oder bei Aufgaben mit <u>hohen Anforderungen</u> ist die Funktionsfähigkeit aber <u>mittelgradig</u> eingeschränkt.</p>	<p>Grad der Arbeitsunfähigkeit von 30 bis 50 %</p>
<p><u>Mittelgradige neuropsychologische Störung:</u></p> <p>a) Mindestens zwei kognitive Teilfunktionen sind deutlich (mehr als 2 SD unter dem Mittelwert) sowie weitere allenfalls leicht vermindert (1 bis 2 SD unter dem Mittelwert), <i>und/oder</i> ...</p> <p>b) Mittelschwere Auffälligkeiten in den Bereichen der Affektivität, des Verhaltens oder der Persönlichkeit</p>	<p>Die Funktionsfähigkeit ist im Alltag und unter den meisten beruflichen Anforderungen <u>deutlich</u> eingeschränkt. Es können nur noch einfachere Arbeiten ausgeführt werden. Die Person fällt in ihrem sozialen Umfeld auch deutlich auf. In Berufen oder bei Aufgaben mit hohen Anforderungen ist die Funktionsfähigkeit sogar stark eingeschränkt.</p>	<p>Grad der Arbeitsunfähigkeit von 50 bis 70 %</p>
<p><u>Mittelgradige bis schwere neuropsychologische Störung:</u></p> <p>a) Die Mehrzahl der kognitiven Teilfunktionen sind deutlich vermindert (mehr als 2 SD unter dem Mittelwert), <i>und/oder</i> ...</p> <p>b) Mittelschwere bis schwere Auffälligkeiten in den Bereichen der Affektivität, des Verhaltens oder der Persönlichkeit</p>	<p>Die Funktionsfähigkeit ist im Alltag und unter sämtlichen beruflichen Anforderungen <u>deutlich</u> eingeschränkt. Es können nur noch <u>sehr einfache</u> Arbeiten unter intensiver <u>Supervision</u> ausgeführt werden. Die Person fällt in ihrem sozialen Umfeld auch deutlich auf. Einfache Tätigkeiten sind unter Umständen in einer geschützten Werkstatt oder einer vergleichbaren Umgebung möglich.</p>	<p>Grad der Arbeitsunfähigkeit von 70 bis 90 %</p>
<p><u>Schwere neuropsychologische Störung:</u></p> <p>a) Beinahe alle kognitiven Teilfunktionen sind deutlich vermindert (mehr als 2 SD unter dem Mittelwert) und können eventuell testpsychologisch gar nicht mehr erfasst werden, <i>und/oder</i> ...</p> <p>b) Schwere Auffälligkeiten in den Bereichen der Affektivität, des Verhaltens oder der Persönlichkeit</p>	<p>Die Funktionsfähigkeit ist im Alltag und unter sämtlichen beruflichen Anforderungen <u>stark</u> eingeschränkt. Weiter fällt die Person in ihrem sozialen Umfeld <u>stark</u> auf. Meist ist der Betroffene <u>voll arbeitsunfähig</u>. Unter Umständen ist eine Tätigkeit in einer <u>geschützten Werkstatt</u> noch möglich.</p>	<p>Grad der Arbeitsunfähigkeit von 100 %</p>
<p><u>Schwerste neuropsychologische Störung:</u></p> <p>Der Patient reagiert kaum oder häufig nicht angepasst auf Umweltreize. Die kognitiven Funktionen und die übrigen psychischen Bereiche sind schwer gestört. Kognitive Leistungen können testpsychologisch nicht erfasst werden.</p>	<p>Die Funktionsfähigkeit ist im Alltag stark eingeschränkt. Der Betroffene ist beinahe rund um die Uhr auf die Hilfe <u>von Drittpersonen angewiesen</u>. Eine Tätigkeit in einer geschützten Werkstatt ist nicht möglich.</p>	<p>Grad der Arbeitsunfähigkeit von 100 %</p>

Anmerkungen. ¹Eine im Einzelfall davon abweichende Einstufung des Schweregrades sollte eingehend begründet werden. ²Bei diesen Richtwerten handelt es sich lediglich um orientierende Angaben. Der Grad der Arbeitsunfähigkeit kann jedoch – in Abhängigkeit der Charakteristika einer Störung sowie des jeweiligen beruflichen Anforderungsprofils – erheblich von diesen Richtwerten abweichen.

Die Vielzahl kognitiver Funktionen können auch in einer ausführlichen Begutachtung nicht vollumfänglich erfasst werden. Für die neuropsychologische Untersuchung

sind daher *Hypothesen* wichtig, die in die Planung der Untersuchung und in die Auswahl der Untersuchungsverfahren einfließen. Eine standardisierte „Testbatterie“

wird diesem Anspruch nicht gerecht und muss im Einzelfall angepasst werden (Sturm & Wallesch, 2007). Beispielweise ist beim Vorliegen einer – insbesondere rechts-hemisphärisch lokalisierten – parietotemporalen Läsion die Durchführung einer spezifischen Neglect-Prüfung erforderlich.

Im Rahmen einer neuropsychologischen Abklärung sollten aber dennoch die wichtigsten kognitiven Funktionsbereiche untersucht werden. Zu einem *Funktionsbereich* gehören meist mehrere *kognitive Teilfunktionen*. So lässt sich zum Beispiel der Gedächtnisbereich weiter aufgliedern in einzelne kognitive Teilfunktionen wie Lernen, Behalten, Wiedererkennen etc. Eine kognitive Teilfunktion kann in einer neuropsychologischen Untersuchung wiederum mit Hilfe mehrerer Testkennwerte abgebildet werden. Aufgrund der Möglichkeit von Testausreissern lässt sich aus einzelnen, unter der Norm liegenden Testkennwerten eines Gesamtprofils oder eines kognitiven Funktionsbereiches nicht unbedingt auf das Vorliegen einer neuropsychologischen Störung schliessen. Zudem sollte jeder Funktionsbereich nach Möglichkeit mit mehr als einem Verfahren erfasst werden. Um das Risiko falsch-positiver Beurteilungen zu minimieren, sind – wenn immer möglich – Informationen zur sogenannten Grundrate auffälliger Testergebnisse beizuziehen (siehe Abschnitt „Interpretation von Testergebnissen“ in den „Leitlinien zur Klassifikation und Interpretation neuropsychologischer Testergebnisse“, Schweizerische Vereinigung der Neuropsychologen, 2014). Zur Vermeidung falsch-negativer Beurteilungen ist der Einsatz einer Profilanalyse hilfreich. Damit lassen sich bei grosser Leistungsvarianz auch Defizite in einem scheinbar unauffälligen Profil einzelfallstatistisch nachweisen (Crawford, Howell & Garthwaite, 1998).

Je nach Anforderungen im Beruf besitzen die einzelnen kognitiven Funktionsbereiche unterschiedliche Bedeutungen im Hinblick auf die Beurteilung der Funktions- und Arbeitsfähigkeit. Deshalb sollte sich die Auswahl der Testverfahren auch an der Analyse der beruflichen oder leistungsbezogenen Anforderungen orientieren (Dohrenbusch, Schneider & Merten, 2012). Grundsätzlich kommt der eingehenden diagnostischen Untersuchung der Aufmerksamkeitsfunktionen aber in der Neuropsychologie eine besondere Bedeutung zu, da diese Basis-Charakter für andere kognitive Funktionen besitzen (Sturm & Wallesch, 2007). Von entscheidender Bedeutung für eine erfolgreiche berufliche Wiedereingliederung haben sich zudem die exekutiven Funktionen, emotionale, soziale und aktivitätsbezogene Regulationsprozesse erwiesen. Eine längere Testsequenz kann zudem als kognitive Belastungserprobung dienen – beispielsweise um das Ausmass einer eventuell vorliegenden „Fatigue“ zu bestimmen (Dohrenbusch, 2007; Lukoschek, Sterr, Claros-Salinas,

Gütler & Dettmers, 2015). Ergänzend kann die „Mini-ICF-APP“ als kurzes Fremdbeurteilungsinstrument dazu verwendet werden, Fähigkeitsstörungen speziell auch in Bezug auf den beruflichen Kontext der untersuchten Person zu operationalisieren und zu quantifizieren (Linden & Baron, 2005). In Bezug auf die Anzahl und die Auswahl der Testverfahren sind die Ressourcen und Einschränkungen des jeweilig zu untersuchenden Probanden aber zu berücksichtigen. Auch sollte der Untersuchende, sich der Problematik bewusst sein, dass mit der Erhöhung der Anzahl durchgeführter Testverfahren, sich auch die Wahrscheinlichkeit der Ermittlung unterdurchschnittlicher Ergebnisse erhöht und sich infolgedessen ab einer gewissen Anzahl verwendeter Tests oft auch bei Gesunden einzelne unterdurchschnittliche Werte nachweisen lassen (Brooks, Sherman, Iverson, Slick & Strauss, 2011). Schliesslich ist bei der Gesamtbewertung zu prüfen, ob die Befunde in sich schlüssig und kongruent sind und deren Bewertung und Interpretation angemessen erfolgt.

Die Einordnung der ermittelten Testergebnisse sollte nach Möglichkeit immer *alters-, geschlechts- und bildungsbezogen* erfolgen (soziale Bezugsnormen). Dies gilt zum Beispiel aber nicht in Bezug auf die Durchführung von Fahreignungstests, da der Vergleich mit Gleichaltrigen in der Regel nicht geeignet ist um zu beurteilen, wie sicher sich eine Person im Strassenverkehr bewegt (kriteriumsorientierter Ansatz).

Da die ermittelten Befunde eindeutig nachvollziehbar sein müssen, sind die verwendeten psychodiagnostischen Verfahren durch Angabe der Testbezeichnung, der Durchführungsversion (Parallelformen) und der Auswertungsmodalitäten (Alters-, Geschlechts- und Bildungskorrektur) sowie die Testergebnisse in Form allgemein bekannter Skalenwerte (Prozentränge, T-Werte, z-Werte, IQ-Punkte etc.) präzise zu dokumentieren, bevor darauf beruhend der Schweregrad einer Störung, wie in Tabelle 1 dargelegt, eingeschätzt wird. Eine Darstellung der Testergebnisse in Form eines Testprofils ist naheliegend und des Überblicks wegen hilfreich. Die Angabe von Rohwerten ist nur dann sinnvoll, falls Tests ohne entsprechende alters-, geschlechts- oder bildungsbezogenen Normen – wie dies beispielsweise bei der Prüfung von Praxien der Fall ist – verwendet werden oder falls die Werte in einem Ausmass von der Norm abweichen, dass diese nicht mehr hinreichend anhand von Skalenwerten dargestellt werden können. Die manchmal vertretene Auffassung, dass die Mitteilung von exakten Testergebnissen zu Problemen oder falschen Interpretationen führen könnte, ist durch nichts gerechtfertigt (Hartje, 2006). Schliesslich sollte im Rahmen der Darstellung und Beschreibung der Befunde klar ersichtlich werden, ob lediglich einzelne kognitive Teilfunktionen (zum Beispiel die tonische Alert-

ness), oder aber ganze Funktionsbereiche (zum Beispiel die Aufmerksamkeitsfunktionen) eingeschränkt sind.

Unauffällige Ergebnisse der Beschwerdvalidierung als Voraussetzung einer aussagekräftigen Einschätzung des Schweregrades einer Störung

Im Berufsalltag des Neuropsychologen stellt die sogenannte Beschwerdvalidierung eine Aufgabe mit zusehends hoher Relevanz dar. Ein Umfrage unter deutschen Neuropsychologen ergab, dass einem Grossteil der Neuropsychologen bewusst ist, dass die Aufrichtigkeit der Beschwerdenschilderung und die Leistungsbereitschaft in neuropsychologischen Untersuchungen nicht stillschweigend vorausgesetzt werden können und sowohl im klinischen als auch – in stärkerem Masse – im gutacherlichen Kontext adäquat untersucht werden müssen (Dandachi-FitzGerald, Merten, Ponds & Niemann, 2015). In der Schweiz gilt eine entsprechende Stellungnahme zur Validität der ermittelten kognitiven Testergebnisse einerseits sowie zu den präsentierten und beklagten Beschwerden andererseits aus juristischer Sicht als unabdingbare Voraussetzung einer zuverlässigen und aussagekräftigen Begutachtung (Germann, 2009). Keine andere Disziplin der Neurowissenschaften hat im Laufe der vergangenen Jahrzehnte einschlägige Instrumente und Verfahren der Validitätsprüfung in solch hoher Zahl und mit solch hoher Aussagekraft hervorgebracht (siehe z.B. Merten & Dettenborn, 2009). Zur Beschwerdvalidierung liegen inzwischen zahlreiche kognitive Testverfahren mit oder ohne dichotomem Antwortformat und Zwangswahlprinzip, aber auch einige gut normierte Fragebögen zur Validierung von psychischen Symptomen vor. Weiter können anhand von sogenannten „eingebetteten“ Faktoren die Ergebnisse herkömmlicher neuropsychologischer Testverfahren zeitökonomisch als Parameter der Beschwerdvalidierung herangezogen werden (Bodenburg, 2014; Fiene et al., 2015). Nicht zuletzt liegen mit den Slick-Kriterien (Slick et al., 1999) für vorgetäuschte neurokognitive Störungen und dem Kriterienkatalog von Bianchini, Greve und Glynn (2005; Young, 2014) für vorgetäuschte schmerzbezogene Funktionseinschränkungen integrative und umfassende neuropsychologische Beurteilungssysteme vor, die nicht nur die Resultate der Performanz- und Beschwerdvalidierungstests, sondern auch eine kriteriengeleitete Konsistenzprüfung bezüglich der Vereinbarkeit der Beschwerden mit dem Grundleiden, der Alltagsfunktionalität und der Lebenssituation der untersuchten

Personen berücksichtigt. Die Beschwerdvalidierung kann unabhängig von der Ätiologie einer Störung – also ungeachtet dessen, ob eine „organische“ oder „nichtorganische“ psychische Störung vorliegt – durchgeführt werden. In bestimmten psychiatrischen Fachkreisen wurde irrtümlicherweise angenommen, dass die Verwendung von Performanzvalidierungstests ausschliesslich bei Vorliegen von „organischen“ psychischen Störungen aussagekräftig sei (Dressing, Foerster, Widder, Schneider & Falkai, 2011). Diese Annahme ist falsch, da die weit überwiegende Anzahl der existierenden neuropsychologischen Performanzvalidierungsverfahren auch mittels Stichproben von Probanden mit „nichtorganischen“ psychischen Störungen normiert wurden. Bei psychiatrischen Fragestellungen kommt zudem der Überprüfung der Beschwerdenschilderung eine zentrale Rolle zu, für welche ebenfalls gut validierte Verfahren wie zum Beispiel der MMPI-2 existieren (Rogers, 2008). Auch für psychiatrische Störungsbilder wie beispielsweise die posttraumatische Belastungsstörung liegen inzwischen spezifische Beurteilungssysteme für die Diagnose vorgetäuschter Psychopathologie vor (Young, 2014). All diese Verfahren sind zwischenzeitlich auch an realen Gutachtenpopulationen validiert worden. Einer der zentralen Punkte bei der Anwendung von Beschwerdvalidierungsverfahren stellt jedoch die Auswahl der für den jeweiligen Exploranden geeigneten Verfahren und die Berücksichtigung störungsspezifischer Trennwerte dar, um falsch positive Klassifikationen zu minimieren. Zum Beispiel ist bekannt, dass bei Personen mit einer nachgewiesenen Intelligenzminderung der Einsatz von Symptomvalidierungsverfahren zu falsch-positiven Ergebnissen führen kann, falls keine entsprechenden Normstichproben oder keine nivellierten Cut-off-Werte zum Vergleich herangezogen werden (Hurley & Deal, 2006).

Beim Vorliegen einer Verdeutlichung, Aggravation oder Simulation auf Performanzebene ist der positive Nachweis einer kognitiven Leistungseinschränkung in der Regel nicht möglich. Ausnahmen hiervon sind zum Beispiel beim Vorliegen von eindeutigen Werkzeugstörungen wie einer Aphasie und validen Vorbefunden gegeben, anhand derer möglicherweise eine grobe Abschätzung der Störung vorgenommen werden kann. Ähnlich problematisch gestalten sich die Diagnosestellung und Beurteilung von Funktionseinschränkungen im psychopathologischen Bereich bei invalider Beschwerdenschilderung in der Exploration. Auf jeden Fall muss aber ein Neuropsychologe, der (bewusstseinsferne) Verdeutlichungstendenzen, eine (bewusstseinsnahe) Aggravation oder eine Simulation begründet vermutet oder nachweist, dies bei der Einschätzung der Störung – wie in der Tabelle 1 dargestellt – entsprechend berücksichtigen (Dohrenbusch, 2007, S. 217f.).

Zur Beurteilung der Arbeitsfähigkeit

Bei der Beurteilung der Arbeitsfähigkeit beziehungsweise Arbeitsunfähigkeit ist die Einschränkung in Bezug auf die bisher ausgeübte Tätigkeit rechtlich relevant. Die Gesundheitsstörung muss Krankheitswert erreichen. Das heisst, sie muss eine Behandlung oder eine teilweise oder volle Arbeitsunfähigkeit zur Folge haben (Swiss Insurance Medicine, 2013). Bei der Einschätzung berücksichtigt man in der Regel zwei Aspekte; erstens die Leistungskomponente – das heisst die Belastbarkeit oder das sogenannte Rendement – sowie zweitens die Zeitkomponente als Arbeitszeit am jeweiligen Arbeitsplatz. Eine Person kann also zum Beispiel während 50 % der Arbeitszeit zu 50 % belastet werden. Daraus ergibt sich eine Arbeitsfähigkeit von 25 % (nämlich 50 % von 50 %) beziehungsweise eine Arbeitsunfähigkeit von 75 % (100 % – 25 %; siehe Swiss Insurance Medicine, 2013, S. 5f.). Die Beurteilung der Arbeitsfähigkeit wird in der Schweiz traditionsgemäss als eine primär medizinische Aufgabe betrachtet. Dies ist für den Bereich der somatischen Medizin nachvollziehbar und sinnvoll. Entsprechend erfolgt die Beurteilung der Arbeitsfähigkeit unter Berücksichtigung der Gesundheitsstörungen in den verschiedenen Teildisziplinen und den daraus resultierenden Einschränkungen aus jeweiliger Fachgutachtersicht im Kontext der polydisziplinären Begutachtung im Rahmen der integrativen Gesamtbeurteilung in der Regel durch den medizinischen Hauptgutachter. Anders verhält es sich jedoch bei der Einschätzung der Arbeitsfähigkeit bei Vorliegen psychischer oder kognitiver Leistungseinschränkungen. Die Vorhersage des beruflichen Erfolgs beziehungsweise der Passung von Anforderungsprofil eines Arbeitsplatzes und dem Leistungsvermögen oder den Fähigkeiten eines Exploranden gehört seit der Entwicklung des „Army Alpha“ und „Army Beta“ für die US-Armee durch Yerkes (1921) während des ersten Weltkrieges – also seit gut 100 Jahren – zum Aufgabengebiet der angewandten Psychologie. Nichts anderes stellt jedoch die Einschätzung der Arbeitsfähigkeit im Hinblick auf die kognitiven Anforderungen, den eine bestimmte Tätigkeit mit sich bringt dar – mit dem Unterschied, dass nicht der berufliche Erfolg eines Gesunden, sondern derjenige eines Menschen mit Beeinträchtigungen der kognitiven Leistungsfähigkeit beurteilt werden muss. 1990 postulierte Lees-Haley elf zentrale Fähigkeiten, die in verschiedener Ausprägung für die Ausübung der 1986 von der US-Regierung veröffentlichten Liste von ca. 13 000 Berufsbildern erforderlich waren. Diese elf Dimensionen (unter anderen die Intelligenz, die verbalen, numerischen und räumlichen Fähigkeiten, die Farbdiskrimination, die Formwahrnehmung und Auge-Hand-Koordination) sind bis heute Bestandteil neuropsy-

chologischer Testbatterien. Mit der Entwicklung spezifischer Testverfahren, die hinsichtlich ihrer Zuverlässigkeit (Reliabilität, Validität, Objektivität) überprüft worden sind, steht der (Neuro-)Psychologie ein hervorragendes Instrumentarium zur Verfügung. Die ökologische Validität dieser Testverfahren – also ihre Vorhersagekraft bezüglich des beruflichen Erfolgs und des Erfolges einer Wiedereingliederung – kann durch die Bestimmung kognitiver Fertigkeiten, welche für eine bestimmte Tätigkeit notwendig sind, den Einsatz berufsspezifischer Fähigkeitstests und somit einer für den Einzelfall massgeschneiderten Testbatterie, der Verhaltensbeobachtung unter arbeitsplatzsimulierenden Bedingungen sowie der Integration von Ergebnissen einer allfälligen beruflichen Abklärung maximiert werden (Guilmette & Pinchot Kastner, 1996).

Ist eine Arbeitsunfähigkeit ausschliesslich oder teilweise neuropsychologisch begründet, ist es somit erforderlich, dass ein Neuropsychologe den Schweregrad der neuropsychologischen Störung bestimmt und Antworten darauf gibt, wie sich diese gegebenenfalls im Detail auf die Funktions- und Leistungsfähigkeit eines Betroffenen auswirkt. Insbesondere im Rahmen von monodisziplinären neuropsychologischen Gutachten, aber auch im klinischen Alltag macht es zudem Sinn, dass der untersuchende Neuropsychologe auch zur Arbeitsfähigkeit Stellung nimmt. Nicht selten werden Neuropsychologen – unter anderem in Form von im Fragekatalog eines Gutachtauftrages ausformulierten Fragen – ohnehin explizit dazu aufgefordert. Neuropsychologische Begutachtungen können wertvolle, fachspezifische Aussagen über das aktuelle Leistungs- und Belastbarkeitsprofil eines Versicherten liefern. Ausserdem können sie notwendige therapeutische Möglichkeiten im Fachgebiet sowie Massnahmen zum Erhalt der Arbeitsfähigkeit und der Unterstützung der Wiedereingliederung aufzeigen. Weiter kann der Neuropsychologe auch in Bezug auf im Praxisalltag ausgestellte Arztzeugnisse wertvolle Hinweise und Empfehlungen zuhanden der verantwortlichen Ärzte geben. Schliesslich vermögen Neuropsychologen auch zur Prognose Stellung zu nehmen (Plohmann, 2008).

Bei den in der Tabelle 1 angegebenen Richtwerten bezüglich der Arbeitsfähigkeit handelt es sich lediglich um orientierende Prozentwerte. Sie dienen der Verbesserung der Interrater-Reliabilität. Der Grad der Arbeitsunfähigkeit kann jedoch im Einzelfall – in Abhängigkeit der Charakteristika einer Störung, des jeweiligen beruflichen Anforderungsprofils oder konkreter Rahmenbedingungen an einem bestimmten Arbeitsplatz – erheblich von diesen Richtwerten abweichen. Zur Illustration werden im Folgenden zwei Beispiele erörtert. So ist zum Beispiel unbestritten, dass das Vorliegen einer Aphasie bei einem Betroffenen, der in einem Beruf mit hohen Anforderungen

an die sprachliche Kommunikation tätig ist, einen höheren Grad der Arbeitsunfähigkeit zur Folge hat, als bei einer im Primärsektor tätigen Person. Eine mutmasslich zwar nur leichte neuropsychologische Störung mit residueller aphasischer Symptomatik hätte im Falle eines untersuchten Radio- oder Fernsehsprechers somit eine 100%ige Arbeitsunfähigkeit im angestammten Beruf zur Folge – und nicht nur eine 10- bis 30%ige Arbeitsunfähigkeit, wie in der Tabelle 1 als Bezugsrahmen angegeben. Oder das Vorliegen eines visuellen Neglects dürfte bei einem Betroffenen, der als Fernkraftfahrer tätig ist, in der Regel einen höheren Grad der Arbeitsunfähigkeit zur Folge haben, als bei einer im Tertiärsektor tätigen Person. Eine mutmassliche mittelgradige neuropsychologische Störung mit persistierendem visuellen Neglect hätte in diesem Fall eine ebenfalls 100%ige Arbeitsunfähigkeit im angestammten Beruf zur Folge – und nicht nur eine 50- bis 70%ige Arbeitsunfähigkeit, wie in der Tabelle 1 richtungsweisend angegeben. Auch wenn die Einschätzung der Arbeitsfähigkeit somit – wie eben beispielhaft aufgezeigt – gelegentlich von den in der Tabelle 1 aufgeführten orientierenden Richtwerten abweichen mag, stellen diese Richtwerte einen wichtigen Bezugsrahmen dar, der zukünftig zur einer Verbesserung der Interrater-Reliabilität im Rahmen der Einschätzung der Arbeitsfähigkeit beitragen dürfte. Schliesslich ist zu bemerken, dass bei der Beurteilung der Arbeitsfähigkeit nicht nur die in der Tabelle 1 aufgeführten Kriterien im Sinne einer pathogenetisch-orientierten Sichtweise, sondern auch die individuellen Ressourcen einer untersuchten Person im Sinne des salutogenetischen Ansatzes zu berücksichtigen sind. Dies gilt insbesondere in Bezug auf die Einschätzung der Arbeitsfähigkeit in einer leidensangepassten Tätigkeit.

Zur Problematik der gelegentlichen Divergenz zwischen dem Schweregrad der kognitiven Defizite und dem Ausprägungsgrad von Symptomen in den weiteren psychischen Bereichen

Bei gewissen psychischen Störungen können der Ausprägungsgrad der vorhandenen kognitiven Defizite einerseits sowie der Schweregrad der übrigen psychischen Symptome andererseits in erheblichem Ausmass voneinander abweichen. Beispielsweise ist bekannt, dass bei Patienten mit einem sogenannten „Dysexekutiv-Syndrom“ oder einer organischen Persönlichkeitsstörung infolge einer Schädigung des präfrontalen Kortexes (ICD-10: F07.0) auf Testebene nicht selten lediglich geringfügige Defizite

nachzuweisen sind, während die Betroffenen aber in Bezug auf ihre Affektregulation, ihr Sozialverhalten und ihre Kritikfähigkeit ausgeprägte Einschränkungen aufweisen können. Hinsichtlich der Einschätzung der Funktionsfähigkeit und somit auch der Arbeitsfähigkeit kann es folglich einen deutlichen Unterschied ausmachen, ob bei einem solchen Beschwerdebild ausschliesslich zu den kognitiven Defiziten, oder aber zu einer neuropsychologischen Störung – als umfassendere psychopathologische Entität – Stellung genommen wird (neurokognitive vs. neuropsychologische Störung).

Die Berücksichtigung und Einschätzung der weiteren psychischen Bereiche der Affektivität, des Verhaltens und der Persönlichkeit ist aber nicht nur in Bezug auf die Diagnostik „organischer“ psychischer Störungen, sondern auch in Bezug auf die Diagnostik „nichtorganischer“ psychischer Störungen bedeutsam. Zur Illustration der ansonsten möglicherweise entstehenden Probleme werden im Folgenden zwei Beispiele von „nichtorganischen“ psychischen Störungen erörtert. Beim Vorliegen eines Aufmerksamkeits- und Hyperaktivitäts-Syndroms würde eine ausschliessliche testpsychologische Abklärung der kognitiven Funktionen – das heisst ohne Berücksichtigung und Einschätzung von etwaigen weiteren Kardinalsymptomen in den Bereichen der Antriebs und des Sozialverhaltens – möglicherweise kein realistisches Bild der Funktionseinschränkungen oder des Grades der Arbeitsunfähigkeit einer betroffenen Person widerspiegeln. Dies wäre der Fall, wenn eine eventuelle Hyperaktivität und Impulsivität – als zwei von drei Kardinalsymptomen des Aufmerksamkeits- und Hyperaktivitäts-Syndroms – nicht berücksichtigt und eingeschätzt würden, obwohl diese im Vergleich mit ebenfalls vorhandenen Aufmerksamkeitsstörungen weit im Vordergrund des Beschwerdebildes lägen. Eine ähnliche Problematik besteht beispielsweise bei der Untersuchung von Störungen aus dem Autismus-Spektrum, wie etwa dem Asperger-Syndrom. Es ergäbe sich kein aussagekräftiges Bild der Funktionseinschränkungen oder des Grades der Arbeitsunfähigkeit, wenn qualitative Beeinträchtigungen der sozialen Interaktionen und stereotype Verhaltensweisen – als charakteristische Symptome des Asperger-Syndroms – nicht mitberücksichtigt und eingeschätzt würden, diese aber im Vergleich mit testpsychologisch fassbaren, jedoch weniger ausgeprägten Defiziten im Vordergrund des Beschwerdebildes stünden. Mit der Unterteilung in Kriterien zur Einschätzung der kognitiven Funktionen einerseits sowie in Kriterien zur Beurteilung der weiteren psychischen Bereiche andererseits wird diesem Umstand aber im Beurteilungssystem gemäss Tabelle 1 Rechnung getragen.

Schlussbemerkungen zur Problematik der Abgrenzung kognitiver Funktionen von weiteren psychischen Bereichen sowie zur interdisziplinären Zusammenarbeit

Die Kriterien der Suva-Tabelle 8 umfassen einerseits den Bereich der kognitiven Funktionen, beziehen sich aber andererseits auch auf weitere psychische Bereiche wie die Affektivität, das Sozialverhalten, die Kritikfähigkeit oder die Persönlichkeit. Im Rahmen der Einschätzung einer neuropsychologischen Störung sind gemäss Erläuterungen zur Suva-Tabelle 8 alle diese erwähnten Aspekte zu berücksichtigen. Dies ist im Kontext der Unfallversicherung – oder genauer ausgedrückt – im Rahmen der Untersuchung der Folgen von traumatisch bedingten Hirnverletzungen grundsätzlich nicht nur sinnvoll, sondern notwendig. Nichtsdestotrotz ist aber zu berücksichtigen, dass sich die in manchen Gutachtensituationen geforderte Abgrenzung der hirnorganisch bedingten Anteile einer Störung oft schwierig gestaltet. Dies gilt speziell in Bezug auf das „organische Psychosyndrom“ nach Schädelhirntrauma, bei dem definitionsgemäss oft erlebnisreaktive Züge und psychogene Symptombildungen mit den organischen Anteilen interagieren (ICD-10: F07.2, Dilling et al., 2014, S. 104). Aus Gründen der Kompatibilität mit der Suva-Tabelle 8 unterscheidet das neue Beurteilungssystem, wie in der Tabelle 1 dargestellt, ebenfalls zwischen kognitiven Funktionen einerseits und den weiteren psychischen Bereichen andererseits. Diese Unterscheidung ist aber – wie oben erwähnt – gelegentlich problematisch und verlangt im interdisziplinären Kontext eine genaue Absprache bezüglich der Aufgabenzuteilung. Im Zuge der in den letzten Jahrzehnten vorstattengegangenen Annäherung der verschiedenen Disziplinen der Neurowissenschaften hat sich auch im praktischen Berufsalltag eine wesentlich verstärkte interdisziplinäre Zusammenarbeit ergeben. Aufgrund dieses verstärkten interdisziplinären Austauschs werden psychiatrische und neuropsychologische Beschwerdebilder zusehends gemeinsam von Psychiatern und Neuropsychologen beurteilt. Immer häufiger umfasst der Auftrag an den untersuchenden Neuropsychologen im Sinne einer Aufgabenteilung sogar ausschliesslich die Abklärung der kognitiven Funktionen. Dies gilt insbesondere im polydisziplinären gutachterlichen Kontext, aber auch im Rahmen von klinischen neuropsychologischen Untersuchungen in der Psychiatrie oder in Memory-Kliniken. Dieser Trend in Richtung einer gesonderten Abklärung der Kognitionen kommt auch in Form des im DSM-5 eigens für den Bereich der „neurokognitiven Störungen“ geschaffenen Kapitels zum Ausdruck (Falkai & Wittchen, 2015, S. 811–860). Im Vorfeld einer interdisziplinären Abklärung mit Beteiligung sowohl

einer psychiatrischen wie auch einer neuropsychologischen Fachperson empfiehlt sich somit sorgfältig abzuklären, ob von Seiten der Neuropsychologie lediglich eine Beurteilung der kognitiven Funktionen, oder aber eine ausführlichere Stellungnahme zu einer neuropsychologischen Störung – als umfassendere psychopathologische Entität – verlangt wird beziehungsweise erforderlich ist.

Autorenhinweise

Dem Antrag um Genehmigung der Inhalte des vorliegenden Artikels wurde an der Generalversammlung der Schweizerischen Vereinigung der Neuropsychologinnen und Neuropsychologen (SVNP) vom 07.11.2015 mit Mehrheitsbeschluss zugestimmt.

Literatur

- Baron-Cohen, S., Wheelwright, S. & Jolliffe, T. (1997). Is there a „language of the eyes“? Evidence from normal adults and adults with autism or Asperger syndrome. *Visual Cognition*, 4, 311–331.
- Bianchini, K., Greve, K. & Glynn, G. (2005). On the diagnosis of malingering pain-related disability: lessons from cognitive malingering research. *The Spine Journal*, 5(4), 404–417.
- Bodenburg, S. (2014). Die Standardabweichung der Reaktionszeiten als ein Maß für negative Antwortverzerrungen. *Zeitschrift für Neuropsychologie*, 25(2), 89–97.
- Bölte, S. (2005). *Reading Mind in the Eyes Test für Erwachsene (dt. Fassung) von S. Baron-Cohen*. Johann Wolfgang Goethe-Universität Frankfurt a.M.
- Bossini, L., Tavanti, M., Calossi, S., Lombardelli, A., Polizzotto, N. R., Galli, R., Vatti, G., Pieraccini, F. & Castrogiovanni, P. (2008). Magnetic resonance imaging volumes of the hippocampus in drug-naïve patients with post-traumatic stress disorder without comorbidity conditions. *Journal of Psychiatric Research*, 42(9), 752–762.
- Brooks, B. L., Sherman, M. S., Iverson, L. G., Slick, D. J. & Strauss, E. (2011). Psychometric foundations for the interpretations of neuropsychological test results. In M. R. Schoenberg and J. G. Scott (Eds.), *The Little Black Book of Neuropsychology. A Syndrome-Based Approach* (pp. 893–922). New York: Springer.
- Bundesgerichtsentscheid, BGE 130 V 352, http://www.polyreg.ch/bgepub/Band_130_2004/BGE_130_V_352.html
- Crawford, J. R., Howell, D. C. & Garthwaite, P. H. (1998). Payne and Jones revisited: estimating the abnormality of test score differences using a modified paired samples t test. *Journal of Clinical and Experimental Neuropsychology*, 20, 898–905.
- Dandachi-FitzGerald, B., Merten, T., Ponds, R. & Niemann, H. (2015). Europäische Umfrage zum Einsatz von Beschwerdengültigkeitstests: Ergebnisse der deutschen Teilnehmer. *Zeitschrift für Neuropsychologie*, 26(2), 99–108.
- Dilling, H., Mombour, W. & Schmidt, M. H. (Hrsg.) (2014). *Internationale Klassifikation psychischer Störungen. ICD-10 Kapitel V (F): Klinisch-diagnostische Leitlinien* (4. überarb. Aufl.). Bern: Huber.

- Dohrenbusch, R. (2007). *Begutachtung somatoformer Störungen und chronifizierter Schmerzen. Konzepte, Methoden, Beispiele*. Stuttgart: Kohlhammer.
- Dohrenbusch, R., Schneider, W. & Merten, T. (2012). Zur Bedeutung der Testpsychologie bei der ICF-orientierten Begutachtung. In W. Schneider et al. (Hrsg.), *Begutachtung bei psychischen und psychosomatischen Erkrankungen. Autorisierte Leitlinien und Kommentare*. Bern: Huber.
- Dressing, H., Foerster, K., Widder, B., Schneider, F. & Falkai, P. (2011). Zur Anwendung von Beschwerdewalidierungstests in der psychiatrischen Begutachtung. Stellungnahme der deutschen Gesellschaft für Psychiatrie, Psychotherapie und Neuroheilkunde (DGPPN). Stellungnahme Nr. 3. www.dgppn.de
- Falkai, P. & Wittchen, H.-U. (Hrsg.) (2015). *Diagnostisches und Statistisches Manual Psychischer Störungen DSM-5*. Göttingen: Hogrefe.
- Fiene, M., Bittner, V., Fischer, J., Schwiecker, K., Heinze, H.-J. & Zaehle, T. (2015). Untersuchung der Simulationssensibilität des Alertness-Tests der Testbatterie zur Aufmerksamkeitsprüfung (TAP). *Zeitschrift für Neuropsychologie*, 26(2), 73–86.
- Germann, T. (2009). Die gerichtlichen Vorgaben an polydisziplinäre Gutachten bei Schleudertraumen und äquivalenten Verletzungen, *Schweizerische Zeitschrift für Sozialversicherung und berufliche Vorsorge*, Nr. 4.
- Guilmette, T. J. & Pinchot Kastner, M. (1996). The prediction of vocational functioning from neuropsychological data. In R. J. Sbordone and C. J. Long (Eds.), *Ecological Validity of Neuropsychological Testing*. (pp. 387–413). Delray Beach, Florida: St. Lucie Press.
- Hartje, W. (2006). Neuropsychologische Begutachtung. In H.-O. Karnath, W. Hartje und W. Ziegler, *Kognitive Neurologie* (S. 235–245). Stuttgart: Thieme.
- Hurley, K. E. & Deal, W. P. (2006). Assessment instruments measuring malingering used with individuals who have mental retardation: potential problems and issues. *Mental Retardation*, 44, 112–119.
- Jeger, J. (2011). Die Entwicklung der „Foerster-Kriterien“ und ihre Übernahme in die bundesgerichtliche Rechtsprechung: Geschichte einer Evidenz. *Jusletter*, Rz. 3.
- Jeger, J. (2014). Die persönlichen Ressourcen und ihre Auswirkungen auf die Arbeits- und Wiedereingliederungsfähigkeit – eine kritische Auseinandersetzung mit der Überwindbarkeitspraxis. In G. Riemer-Kafka (Hrsg.), *Psyche und Sozialversicherung*. Luzerner Beiträge zur Rechtswissenschaft, Bd. 81. Zürich: Schulthess Verlag.
- Lautenbacher, S. & Gauggel, S. (2004). *Neuropsychologie psychischer Störungen*. Berlin: Springer.
- Lees-Haley, P. (1990). Vocational neuropsychological requirements of U.S. occupations. *Perceptual and Motor Skills*, 7, 1383–1386.
- Linden, M. & Baron, S. (2005). Das „Mini-ICF-Rating für psychische Störungen (Mini-ICF-P)“. Ein Kurzinstrument zur Beurteilung von Fähigkeitsstörungen bei psychischen Erkrankungen. *Rehabilitation*, 44, 144–151.
- Lukoschek, C., Sterr, A., Claros-Salinas, D., Güttler, R. & Dettmers, C. (2015). Fatigue in multiple sclerosis compared to stroke. *Frontiers in Neurology*, 6(116). doi: 10.3389/fneur.2015.00116
- McKinnon, M. C., Yucel, K., Nazarov, A. & MacQueen, G. M. (2009). A meta-analysis examining clinical predictors of hippocampal volume in patients with major depressive disorder. *Journal of Psychiatry and Neuroscience*, 34(1), 41–54.
- Merten, T. & Dettenborn, H. (2009). *Diagnostik der Beschwerdewalidität*. Praxis der Rechtspsychologie. Berlin: Deutscher Psychologen-Verlag.
- Plohm, A.M. (2008). Relevanz neuropsychologischer Gutachten zur Beurteilung der Arbeitsfähigkeit und Integrität. *Epileptologie*, 25, 182–190.
- Rogers, R. (Ed.) (2008). *Clinical Assessment of Malingering and Deception* (3rd ed.). New York: The Guilford Press.
- Schweizerische Vereinigung der Neuropsychologen. (2014). Leitlinien zur Klassifikation und Interpretation neuropsychologischer Testergebnisse. www.neuropsych.ch/w/pages/de/leitlinien.php
- Slick, D. J., Sherman, E. & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, 13(4), 545–561.
- Stöckli, H. R. (2007). Besonderheiten der Begutachtung in der Schweiz. In B. Widder und P. W. Gaidzik (Hrsg.), *Begutachtung in der Neurologie* (S. 465–476). Stuttgart: Thieme.
- Stone, V. E., Baron-Cohen, S. & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience*, 10(5), 650–656.
- Sturm, W. & Wallesch, C.-W. (2007). Störungen höherer Hirnleistungen: Aufmerksamkeit, Gedächtnis und exekutive Funktionen. In B. Widder und P. W. Gaidzik (Hrsg.), *Begutachtung in der Neurologie* (S. 203–213). Stuttgart: Thieme.
- SUVA. (2002). Integritätsentschädigung gemäss UVG, Tabelle 8, Integritätsschaden bei psychischen Folgen von Hirnverletzungen. *Medizinische Mitteilungen*, 2870/8.d.
- Swiss Insurance Medicine. (2013). Arbeitsunfähigkeit: Leitlinie zur Beurteilung der Arbeitsunfähigkeit nach Unfall und bei Krankheit (4. Aufl.).
- Yerkes, R. M. (1921). *Psychological Examining in the United States Army*. Memoirs of the National Academy of Sciences, Vol. XV. Washington, DC: U.S. Government Printing Office.
- Young, G. (2014). *Malingering, Feigning, and Response Bias in Psychiatric/Psychological Injury. Implications for Practice and Court*. International Library of Ethics, Law, and the New Medicine, 56. Dordrecht: Springer.

Manuskript eingereicht: 12.11.2015
 Nach Revision angenommen: 29.11.2015
 Interessenskonflikt: Nein

Dr. phil. Adrian Frei
 Clenia Schöllli
 Privatklinik für Psychiatrie und Psychotherapie
 Schöllslistr. 8
 8618 Oetwil am See
 Schweiz
adrian.frei@clenia.ch



CME-Fragen

1. Frage: Was ist im Kontext der Schweizer Rechtsprechung unter dem Begriff der „Überwindbarkeitspraxis“ zu verstehen?

- a. Organisch nicht begründbare Beschwerden werden als Ausdruck einer Aggravation oder Simulation betrachtet.
- b. Der Begriff bezieht sich auf Schmerzen, welche die betroffene Person überwinden sollte.
- c. Der Begriff bezieht sich auf eine Selbstlimitation. Entsprechende Beschwerden können von der betroffenen Person im praktischen Berufsalltag nicht überwunden werden.
- d. Organisch nicht begründbare Beschwerden werden zwar als authentisch, jedoch auf jeden Fall als überwindbar betrachtet und besitzen somit keine Rentenrelevanz.
- e. Patienten lernen, ihre Beschwerden im Klinik- und Praxisalltag zu überwinden.

2. Frage: Was ist mit dem Akronym „PÄUSBONOG“ gemeint?

- a. Pathogenetisch-ätiologisch unbestimmte Syndrome bei orthopädisch nicht objektivierbarer Grundlage
- b. Pathogenetisch-ätiologisch unklare syndromale Beschwerdebilder ohne nachweisliche organische Grundlage
- c. Pathologische, ärztlich unklare Schmerz-Belastungen ohne nachweisbare organische Grundlage
- d. Pathologisch-ätiologisch unbestimmte syndromale Beschwerden ohne nosologische organische Grundlage
- e. Physiologisch-äthyltoxische und schmerzbezogene Beschwerden ohne nachweisbare objektive Gründe

3. Frage: Wie lässt sich gemäss DSM-5 eine leichte kognitive Störung bezogen auf die Ergebnisse kognitiver Tests definieren?

- a. Eine leichte kognitive Störung lässt sich nicht mit Hilfe kognitiver Tests diagnostizieren.
- b. Die Testresultate liegen im Allgemeinen zwischen zwei bis drei Standardabweichungen unter dem jeweiligen Mittelwert.
- c. Im DSM-5 finden sich keine entsprechenden Definitionen oder Empfehlungen.

- d. Die Testresultate liegen im Allgemeinen mehr als zwei Standardabweichungen unterhalb des jeweiligen Mittelwertes.
- e. Die Testresultate liegen im Allgemeinen zwischen einer bis zwei Standardabweichungen unter dem jeweiligen Mittelwert.

4. Frage: Mit der Erstellung und Publikation der Tabelle 1 wurde folgendes hauptsächliches Ziel verfolgt:

- a. Das Ziel ist eine Verbesserung der Interrater-Reliabilität im Rahmen der Einschätzung des Schweregrades einer neuropsychologischen Störung sowie der Beurteilung der Arbeitsfähigkeit.
- b. Das Ziel ist eine Verbesserung der Symptomvalidierung.
- c. Die Tabelle 1 umfasst verpflichtende Vorgaben bezüglich der Einschätzung des Grades der Arbeitsfähigkeit infolge einer neuropsychologischen Störung.
- d. Die Tabelle 1 soll zukünftig die Suva-Tabelle 8 als Referenzgrundlage bei der Beurteilung der Integritätsentschädigung ersetzen.
- e. Das Ziel ist, den Stellenwert der klinischen Expertise im Rahmen der Beurteilung von neuropsychologischen Störungen zu schmälern.

5. Frage: Bestehen zwischen den Begriffen oder psychopathologischen Entitäten einer „kognitiven“ und „neuropsychologischen“ Störung Unterschiede?

- a. Nein, die beiden Begriffe oder Entitäten unterscheiden sich nicht.
- b. In der Tabelle 1 wird zwischen kognitiven Defiziten (Kriterium a) und „weiteren psychischen Bereichen“ (Kriterium b) unterschieden. Entsprechend wird auch zwischen einer kognitiven Störung und einer neuropsychologischen Störung als umfassendere psychopathologische Entität unterschieden.
- c. Es bestehen Unterschiede. Eine kognitive Störung gibt es nur als Teil einer neuropsychologischen Störung.
- d. Es bestehen Unterschiede. Eine neuropsychologische Störung gibt es nur als Teil einer kognitiven Störung.
- e. Es bestehen keine relevanten Unterschiede zwischen den beiden Begriffen oder Entitäten. Kleine Unterschiede sind lediglich historisch begründet.

CURRICULUM VITAE

PERSÖNLICHE DATEN

Name Andrea Maria Plohmann (-Schilling)
Geburtsdatum 25. August 1962

AUSBILDUNG UND WEITERBILDUNGEN

2013 Eidgenössisch anerkannte Psychotherapeutin
Umzertifizierung gemäss PsyG

seit 2012 Doktoratsstudium
Thema „*Bedeutung neuropsychologischer
Beschwerdengültigkeit in der Beurteilung der
funktionellen Leistungsfähigkeit bzw. Arbeitsfähigkeit in
der versicherungsmedizinischen Begutachtung*“
Prof. Stieglitz, Psychologische Fakultät, Universität Basel

2012 Anerkennung der Zusatzbezeichnung Klinische
Neuropsychologie, Landespsychotherapeutenkammer
Baden-Württemberg (LPK-BW)

2010 Zertifizierte Neuropsychologische Gutachterin SIM

2007 Supervisorin GNP (D)

2004 Fachpsychologin für Psychotherapie FSP

2003 Psychologische Schmerztherapie DGPSF
– Theoretischer Teil der Ausbildung (D)

2001 Klinische Neuropsychologin GNP (D)

1999 Fachpsychologin für Neuropsychologie FSP/SVNP

1999 Approbation als Psychologische Psychotherapeutin
Landespsychotherapeutenkammer Baden-Württemberg
(LPK-BW) (D)

1985-1989 Diplom im Fach Psychologie
Studienschwerpunkte: *Klinische Psychologie, Neuropsychologie*
Nachbarfach: *Humanbiologie*
Diplomarbeit: *Ultradiane Lateralitätszyklen der dichotischen
Hörleistung bei Patienten mit Hirnschädigung.*
Abschluss: *Diplom-Psychologin, Note: sehr gut*
Universität Konstanz (D)

1982-1985	Vordiplom im Fach Psychologie Universität Mannheim (D)
1973-1982	Abitur Friedrich-Wöhler-Gymnasium Singen (D)

BERUFLICHER WERDEGANG

02/2015-05/2015	Forensisch-Psychiatrische Universitätsklinik, Abt. Versicherungsmedizin, Universitäre Psychiatrische Kliniken Basel (50%)
WS 2013/14	Dozentin an der Psychologischen Fakultät der Universität Basel, Seminartitel „Neuropsychologische Gutachten“
seit 2012	Dozentin der Akademie König & Müller, Würzburg, Fortbildungskurse für PsychologInnen und Fachinteressierte
2012-2014	Dozentin der Fachgruppe Psychotherapie und Neuropsychologie bei Multipler Sklerose (FGPN-MS) in Kooperation mit Prof. Calabrese
seit 2010	Dozentin der SIM-Gutachterausbildung (Swiss Insurance Medicine)
2008	Erweiterung des Tätigkeitsspektrums in eigener Praxis: <i>Praxis für Neuropsychologie und Psychotherapie</i> in Basel
2008	Definitive Bewilligung zur Ausübung der Psychotherapie im Kanton BS
2007 - 2016	Assessorin und Fachbeirätin bei <i>human life support</i> , einem Unternehmen zur Förderung der medizinischen, beruflichen und sozialen Rehabilitation
2002	Anerkennung der Praxis für Klinische Neuropsychologie als Ausbildungsinstitution Typ B gemäss dem Reglement der postgradualen Weiterbildung ‚Fachpsychologe für Neuropsychologie FSP‘
2001-2009	Psychologische Beratung im Rahmen des „Modells Externer Psychologen (MEP)“ sowie Seminarleitungen für der Schweizerischen Multiple Sklerose Gesellschaft
04/2001	Eröffnung <i>Praxis für Klinische Neuropsychologie</i> in Basel
seit 1999	Freie Mitarbeit als Fachgutachterin bei der Medizinischen Abklärungsstelle des Universitätsspitals Basel (MEDAS), später academy of swiss insurance medicine (asim)
seit 1999	Referentin/Seminarleiterin für die AMSEL, Landesverband Baden-

	Württemberg und Regionalgruppe Südbaden
1995-1998	Referentin der Lehrer- und Lehrerinnenfortbildung Baselland
1992-2000	freiberufliche Tätigkeit in eigener Psychologischer Praxis (D) 10-20%
1992-1999	Wissenschaftliche Assistentin (Drittmittelanstellung), Neurologisch-Neurochirurgische Poliklinik, Universitätsspital Basel (20-50%), Forschungsschwerpunkt „Kognitive Defizite bei MS“ bei Prof. Kappos
1990-1991	Reha B (Zentrum für Querschnittgelähmte und Hirnverletzte), Bürgerspital Basel, Konsiliarische Tätigkeit (50%)
1990-2002	Neurologische Universitätsklinik, Universitätsspital Basel (50%), Profs. Käser, Ettl, Lyrer, Steck
1989-1990	Espan-Klinik, Fachklinik für Erkrankungen der Atmungsorgane, Bad Dürrenheim (100%)

VERÖFFENTLICHUNGEN

Zeitschriftenartikel, Buchbeiträge und Abstracts

Plohm, A.M. (eingereicht). Prevalence of poor effort and malingered neurocognitive dysfunction in a litigating sample in Switzerland. *Zeitschrift für Neuropsychologie*

Keppler, C., Plohm, A.M., Pflüger, M., Rabovsky, K., Langewitz, W., Mager, R. (2017). Beschwerdengültigkeit in der versicherungsmedizinischen Begutachtung. *Fortschritte der Neurologie und Psychiatrie*, 85, 17-33.

Frei A, Balzer C, Gysi F, Leros J, Plohm AM, Steiger G. (2016). Kriterien zur Bestimmung des Schweregrades einer neuropsychologischen Störung sowie Zuordnungen zur Funktions- und Arbeitsfähigkeit. *Zeitschrift für Neuropsychologie*, 27 (2), 107-119.

Plohm, A.M. (2013). First base rate estimates of performance validity in Switzerland (Abstract). *Clinica y Salud / Clinical and Health, Journal of Empirical Research in Psychology*, 24 (3), 200.

Plohm, A.M. (2013). Common arguments against symptom validity assessment: The psychiatry debate in Germany and Switzerland (Abstract). *Clinica y Salud / Clinical and Health, Journal of Empirical Research in Psychology*, 24 (3), 198.

Plohm, A.M., Merten T. (2013). The Third European Symposium on Symptom Validity Assessment – Facts and Controversies. *Clinica y Salud / Clinical and Health, Journal of Empirical Research in Psychology*, 24 (3), 197-203.

Plohm, A.M. (2011). Tagungsbericht: 2. Europäisches Symposium zur Diagnostik der Beschwerdegültigkeit. *Zeitschrift für Neuropsychologie*, 22(4), 303-305.

Plohmann, A.M. (2009). Zahlen-Verbindungs-Test (ZVT). Testrezension. In D. Schellig, R. Drechsler, D. Heinemann, W. Sturm (Hrsg.). *Handbuch neuropsychologischer Testverfahren. Aufmerksamkeit, Gedächtnis und exekutive Funktionen*. Göttingen: Hogrefe

Plohmann, A.M. (2008). Relevanz neuropsychologischer Gutachten zur Beurteilung von Arbeitsfähigkeit und Invalidität. *Epileptologie*, 4 (25), 182-190.

Plohmann, A.M., Kappos, L. (1999). Reply to Goldenberg - A dubious therapy for patients with multiple sclerosis. *Journal of Neurology, Neurosurgery and Psychiatry*, 66, 694-695.

Plohmann, A.M., Kappos, L., Ammann, W., Thordai, A., Wittwer A., et al. (1998). Computer-assisted retraining of attentional impairments in patients with multiple sclerosis. *Journal of Neurology, Neurosurgery and Psychiatry*, 64(4), 455-462

Kappos, L., Plohmann, A.M., Radü, E.W., Roelcke, U. (1997). MRI and PET correlates of cognitive dysfunction. *Electroencephalography and Clinical Neurophysiology*, 103(1), 53.

Roelcke, U., Kappos, L., Lechner-Scott, J., Brunnschweiler, H., Plohmann, A.M., et al. (1997). Reduced glucose metabolism in the caudate nucleus and prefrontal cortex of multiple sclerosis patients with fatigue. *Neurology*, 48(6), 1566-1571.

Plohmann, A.M., Brunnschweiler, H., Steck, A., Kappos, L. (1995). Evaluation eines computergestützten Aufmerksamkeitstrainings für Patienten mit Multipler Sklerose. *Verhandlungen der Deutschen Gesellschaft für Neurologie*, 8, 733-735.

Plohmann, A.M., Kappos, L., Brunnschweiler, H. (1994). Evaluation of a computer-based attention retraining program for patients with multiple sclerosis. *Schweizer Archiv für Neurologie und Psychiatrie*, 145, 55-56

Stahl, H.D., Ettl, T., Plohmann, A. M., Radü, E.W., Müller-Brand, J. et al. (1994). Central nervous system lupus: concomitant occurrence of myelopathy and cognitive function. *Clinical Rheumatology*, 13(2), 273-279.

Kischka, U., Ettl, T., Plohmann, A., Stahl, H. (1994). SPECT findings in patients with whiplash injury of the neck. *Neurology*, 44 (suppl.2), A318.

Ettl, T.M., Heim, S., Kiss, A., Plohmann, A. (1991). Drei Fallbeispiele zur multidimensionalen Ätiologie psychogener Störungen in der Neurologie. *Aktuelle Neurologie*, 18, 9.