

Aus der
Chirurgischen Klinik
Campus Charité Mitte | Campus Virchow-Klinikum
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

Diagnostische Wertigkeit eines elektronischen Tablet-Computer-
basierten Zahlenverbindungstests zur Detektion der minimalen
hepatischen Enzephalopathie bei Patienten mit Leberzirrhose

zur Erlangung des akademischen Grades
Doctor medicinae (Dr. med.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Ferenc Darius Rüther
aus Hamburg

Datum der Promotion: 07.12.2018

Widmung

Meinen Eltern

Vorbemerkung

Die hier gezeigten und ausgewerteten Daten wurden im Rahmen der eNCT-Studie, die im *Deutschen Register Klinischer Studien* registriert ist (DRKS00009432), größtenteils am Campus Virchow-Klinikum der Charité - Universitätsmedizin Berlin erhoben.

Auf Grundlage der gewonnenen Daten ist folgende Publikation entstanden, die von mir zusammen mit meinem Betreuer, meinem Doktorvater und weiteren Ko-Autoren realisiert wurde:

*Wuensch T, **Ruether DF**, Zollner C, Mueller T, Jung T, Kaffarnik M, Kassner U, Schott E, Kiefer S, Pratschke J, Stockmann M, Jara M. Performance characterization of a novel electronic number connection test to detect minimal hepatic encephalopathy in cirrhotic patients. European Journal of Gastroenterology & Hepatology. 2017;29(4):456-63.*

Da in der vorliegenden Dissertation weitestgehend andere und ergänzende Subanalysen durchgeführt wurden, die zu entsprechend neuen Erkenntnissen führten, wurden keine Auswertungen aus der Publikation übernommen. Sämtliche Abbildungen und Textstellen, die Bezug zu den Ergebnissen der Publikation nehmen, sind entsprechend gekennzeichnet.

Inhaltsverzeichnis

1	Abkürzungsverzeichnis	6
2	Kurzzusammenfassung (Abstract)	8
2.1	Deutsch.....	8
2.2	English	9
3	Einleitung.....	11
3.1	Einführung in die Thematik	11
3.2	Geschichte der hepatischen Enzephalopathie	12
3.3	Definition und Einteilung der hepatischen Enzephalopathie.....	13
3.4	Pathophysiologie der hepatischen Enzephalopathie	16
3.5	Diagnostik der hepatischen Enzephalopathie.....	21
3.5.1	Überblick	21
3.5.2	Klinische Testverfahren.....	22
3.5.2.1	West-Haven-Kriterien (WHC)	22
3.5.2.2	Hepatische Enzephalopathie Wertungsalgorithmus (HESA).....	23
3.5.3	Blutparameter.....	23
3.5.4	Psychometrische Testverfahren.....	23
3.5.5	Spezialdiagnostik (CFF, CRT, ICT, EEG)	25
3.5.5.1	Der Stroop-Test	26
3.6	Therapie der hepatischen Enzephalopathie	27
4	Herleitung der Fragestellungen	28
4.1	Problematik der bisherigen Testverfahren	28
4.2	Lösungsansatz und Fragestellungen	29
5	Patienten, Material und Methoden	30
5.1	Studienplanung und Rekrutierung von Probanden	30
5.2	Ein- und Ausschlusskriterien	30
5.3	Zusammensetzung des Probandenkollektivs	31
5.4	Studienaufbau.....	33
5.4.1	Labordiagnostik und Schweregrad der Leberzirrhose	33
5.4.2	Diagnostik der Enzephalopathie mittels HESA und WHC.....	34
5.4.3	Psychometrische Diagnostik.....	35
5.4.3.1	Psychometric Hepatic Encephalopathy Score (PHES)	35
5.4.3.2	Elektronischer Zahlenverbindungstest (eNCT).....	35
5.4.3.3	Stroop-Test.....	39
5.5	Statistische Auswertung	39
5.5.1	Statistische Anforderungen an diagnostische Validierungsstudien.....	41
5.5.1.1	Analyse der eNCT-Testeigenschaften.....	42
5.5.1.2	Analyse der Reliabilität	42
5.5.1.3	Analyse der Validität.....	43
5.5.1.4	Überprüfung der eNCT-Testgenauigkeit	43
5.5.1.5	Überprüfung möglicher Einflussfaktoren auf eNCT-Testzeiten	44

6	Ergebnisse	45
6.1	Demografische Charakterisierung des Probandenkollektivs	45
6.1.1	Spezielle Patientencharakteristika	47
6.2	Eigenschaften psychometrischer Testverfahren.....	49
6.2.1	Psychometric Hepatic Encephalopathy Score (PHES)	49
6.2.1.1	Diskriminierungsfähigkeit der PHES-Einzeltests	53
6.2.2	Elektronischer Zahlenverbindungstest (eNCT)	55
6.2.2.1	Diskriminierungsfähigkeit des eNCTs.....	56
6.2.3	Stroop-Test	57
6.2.3.1	Diskriminierungsfähigkeit des Stroop-Tests	58
6.3	Vergleich von eNCT und PHES-Tests.....	58
6.3.1	Korrelationen der Testergebnisse	58
6.3.2	Testgenauigkeit des eNCTs.....	62
6.4	Wiederholungsgenauigkeit des eNCTs	64
6.5	Veränderungssensitivität des eNCTs	65
6.6	Einflussfaktoren auf die eNCT-Testzeiten von Patienten	67
6.6.1	Abhängigkeit der eNCT-Testzeiten von Laborwerten	68
6.7	Vergleich von Stroop-Test mit PHES und eNCT	70
6.7.1	Korrelationen zwischen Stroop-Test und eNCT bzw. PHES.....	70
6.7.2	Testgenauigkeit des Stroop-Tests	72
7	Diskussion.....	73
7.1	Diskussion der Methodik.....	73
7.2	Diskussion psychometrischer Testergebnisse.....	76
7.2.1	Psychometric Hepatic Encephalopathy Score (PHES)	76
7.2.2	Elektronischer Zahlenverbindungstest (eNCT)	78
7.2.2.1	Diagnostische Wertigkeit des eNCTs	78
7.2.2.2	Vergleich von PHES und eNCT.....	83
7.2.3	Stroop-Test	86
7.2.3.1	Vergleich von Stroop-Test mit PHES und eNCT	86
7.3	Einschränkungen und Limitationen der Testverfahren	87
7.3.1	Methodische Limitationen	87
7.3.2	Limitationen des PHES	90
7.3.3	Limitationen des eNCTs.....	91
7.3.4	Limitationen des Stroop-Tests	93
7.4	Neue Erkenntnisse und möglicher Anwendungsbereich	94
7.5	Ausblick	97
8	Literaturverzeichnis	98
9	Eidesstattliche Versicherung	107
10	Lebenslauf	108
11	Publikationsliste	109
12	Danksagung	110
13	Anhang.....	111

1 Abkürzungsverzeichnis

AIH	Autoimmunhepatitis
ALT	Alanin-Aminotransferase
ANOVA	Varianzanalyse (engl.: analysis of variance)
AP	Alkalische Phosphatase
AST	Aspartat-Aminotransferase
AUC	Fläche unter der Kurve (engl.: Area Under the Curve)
CFF	Critical Flicker Frequency
CRF	Prüfbogen (engl.: Case Report Form)
CRT	Continuous Reaction Time
CT	Computertomografie
DOR	diagnostische Odds-Ratio
DST	Zahlensymboltest (engl.: Digit Symbol Test)
EEG	Elektroenzephalografie
eNCT	elektronischer Zahlenverbindungstest (engl.: electronic Number Connection Test)
FP7	siebtes Rahmenprogramm der Europäischen Union (engl.: 7th Framework Programme)
GABA	Gamma-Amino-Buttersäure
GGT	Gammaglutamyltransferase
Hb	Hämoglobin
HBV	Hepatitis-B-Virus
HCV	Hepatitis-C-Virus
HE	Hepatische Enzephalopathie
HESA	Hepatische Enzephalopathie Wertungsalgorithmus (engl.: Hepatic Encephalopathy Scoring Algorithm)
ICT	Inhibitory Control Test
IKK	Intraklassen-Korrelationskoeffizient
INR	International Normalized Ratio
IQR	Interquartilsabstand (engl.: Interquartile Range)
ISHEN	internationale Gesellschaft für hepatische Enzephalopathie und Stickstoffmetabolismus (engl.: International Society for Hepatic Encephalopathy and Nitrogen Metabolism)
KI	Konfidenzintervall
K-W-Test	Kruskal-Wallis-Test
LNT	Liniennachfahrtstest
LR(+)	positiver Likelihood-Quotient (engl.: positive Likelihood Ratio)
LR(-)	negativer Likelihood-Quotient (engl.: negative Likelihood Ratio)
LZI	Leberzirrhose
MAPK	Mitogen-aktivierte Proteinkinase
MELD	Model for End-Stage Liver Disease
mHE	minimale hepatische Enzephalopathie
MMST	Mini-Mental-Status-Test

MPTP	Mitochondriale Permeabilitäts-Transitions-Pore
MRT	Magnetresonanztomografie
N	Anzahl
NADPH	Nicotinamidadenindinukleotidphosphat
NASH	Nicht-alkoholische Steatohepatitis
NCT	Zahlenverbindungstest (engl.: Number Connection Test)
NMDA	N-Methyl-D-Aspartat
NPW	negativ-prädiktiver Wert
ÖGD	Ösophagogastroduodenoskopie
oHE	offensichtliche hepatische Enzephalopathie
PBC	primär biliäre Cholangitis
PHES	Psychometrischer Hepatische Enzephalopathie-Wert (engl.: Psychometric Hepatic Encephalopathy Score)
PPW	positiv-prädiktiver Wert
PSC	primär sklerosierende Cholangitis
PSE	portosystemische Enzephalopathie
r	Korrelationskoeffizient
r_e	Effektstärkemaß
R^2	Bestimmtheitsmaß
RBANS	Repeatable Battery for the Assessment of Neuropsychological Status
RNA	Ribonukleinsäure
RNOS	reaktive Stickstoffoxidspezies
ROC	Receiver Operating Characteristic
ROS	reaktive Sauerstoffspezies
SD	Standardabweichung (engl.: Standard Deviation)
SDT	Kreise-punktieren Test (engl.: Serial Dot Test)
TIPSS	transjugulärer intrahepatischer portosystemischer Stent-Shunt
U-Test	Mann-Whitney-U-Test
VIF	Varianzinflationsfaktor
WHC	West-Haven-Kriterien
\bar{x}	Mittelwert

Hinweis im Sinne des Gleichbehandlungsgesetzes:

Für eine bessere Lesbarkeit wird auf eine differenzierte Schreibweise männlicher und weiblicher Sprachformen, wie z. B. Patient/Innen, verzichtet. Entsprechende Begriffe gelten gleichwohl für beide Geschlechter.

2 Kurzzusammenfassung (Abstract)

2.1 Deutsch

Titel: Diagnostische Wertigkeit eines elektronischen Tablet-Computerbasierten Zahlenverbindungstests zur Detektion der minimalen hepatischen Enzephalopathie bei Patienten mit Leberzirrhose

Einleitung: Die hepatische Enzephalopathie (HE) ist eine häufige Komplikation bei Patienten mit Leberzirrhose und äußert sich durch ein breites Spektrum neurologischer Symptome, das von leichten Aufmerksamkeitsstörungen bis zum Koma reicht. Die Detektion eines diskreten Anfangsstadiums, der minimalen HE (mHE), kann entscheidend dazu beitragen, durch einen frühzeitigen Therapiebeginn die Ausprägung des Krankheitsbildes zu begrenzen. Aufgrund der variablen Symptomatik ist eine zuverlässige Diagnostik trotz standardisierter Testverfahren jedoch schwierig. Letztere werden daher im klinischen Alltag nicht regelmäßig durchgeführt. Ein Ansatz zur früheren Detektion der mHE wäre ein bereits im häuslichen Umfeld durchführbares Testverfahren zur Selbstkontrolle.

Methodik: Die vorliegende Arbeit beschreibt eine elektronische Version des Zahlenverbindungstests A (NCT-A), bei dem die Zahlen 1-25 zufällig auf einem Blatt Papier angeordnet sind und schnellstmöglich in der richtigen Reihenfolge – aufsteigend von 1-25 – mit einem Stift verbunden werden müssen. Die Studienteilnehmer wurden zunächst mithilfe etablierter Testverfahren auf das Vorliegen einer mHE untersucht. Anschließend wurde der elektronische Zahlenverbindungstest (eNCT) absolviert, bei dem entsprechend der papierbasierten Vorlage die auf einem berührungsempfindlichen Bildschirm eines Tablet-Computers präsentierten Zahlen 1-25 mit dem Finger angetippt werden müssen. Durch eine statistische Analyse zuvor gewonnener Daten sollte die Frage beantwortet werden, ob der eNCT valide Resultate liefert, die einen Einsatz als Screeningtest in der häuslichen Umgebung prinzipiell möglich machen.

Ergebnisse: Es wurden die Daten von 161 Studienteilnehmern ausgewertet – 100 Patienten mit Leberzirrhose und 61 lebergesunde Kontrollprobanden. Aus dem Patientenkollektiv wurden 26 Teilnehmer mit einer mHE identifiziert. Diese erzielten im eNCT sig-

nifikant langsamere Testzeiten (10" Bildschirm: $p < 0,001$ bzw. 18" Bildschirm: $p < 0,01$) als Patienten ohne Zeichen einer mHE (N=57). Die durch ROC-Analyse gewonnenen eNCT-Grenzwerte (38 bzw. 41 Sekunden) zur mHE-Diagnostik erreichten – bezogen auf den Vergleichstest, den Psychometric Hepatic Encephalopathy Score (PHES) – eine Gesamtübereinstimmung von 80,7% bzw. 81,7%, eine Sensitivität von 96,2% bzw. 92,3% und eine Spezifität von 73,7% bzw. 76,8%. Die Wiederholungsgenauigkeit war mit Intraklassen-Korrelationskoeffizienten (IKK) von 0,906 bzw. 0,855 sehr gut bzw. gut.

Schlussfolgerungen: Der eNCT erfüllt als zuverlässiges Instrument zur mHE-Detektion die qualitativen Voraussetzungen für eine Untersuchung der Anwendbarkeit als Selbsttest im häuslichen Umfeld. Der Test könnte zukünftig eine wichtige Rolle für die frühzeitige Detektion einer abnehmenden kognitiven Leistungsfähigkeit bei Patienten mit Leberzirrhose spielen, insbesondere in infrastrukturell schlechter versorgten Regionen. Auf diesem Weg könnten die Angst von Patienten vor akuten bzw. das Risiko für akute Komplikationen und damit verbundene Krankenhauseinweisungen reduziert sowie die Kosten für das Gesundheitssystem gesenkt werden.

2.2 English

Title: Diagnostic value of a tablet-based electronic number connection test for detection of minimal hepatic encephalopathy in patients with liver cirrhosis

Introduction: Hepatic encephalopathy (HE) is a frequent complication of patients with liver cirrhosis and appears with a wide spectrum of neurological symptoms, from a mild disturbance of attention to coma. Detection of a discrete initial stage called minimal HE (mHE) could help crucially to limit the disease severity by an early start of therapy. However, because of the variability of symptoms making diagnosis is difficult despite standardized test methods. Therefore, they are not used regularly in routine clinical work-up. An approach for early mHE-detection would be a self-monitoring system with a test method already feasible at home.

Methodology: This doctoral thesis describes an electronic version of the number connection test A (NCT-A), that comprises the numbers 1-25 placed randomly on a sheet of

paper for connection with a pencil as fast as possible in the correct order. At first, included participants were examined for an undetected mHE by using established test methods. Afterwards, the electronic number connection test (eNCT) was performed equally to the paper-based pattern apart from tapping the numbers 1-25 shown on a touchscreen with a finger. Statistical analysis of before obtained data should answer the question if eNCT produces valid results that could enable the test to be applied as a screening test at home.

Results: Data of 161 participants were analysed – 100 patients with liver cirrhosis and 61 control persons with normal liver function. 26 patients were identified to have an mHE. They showed significant worse eNCT test results (10" screen size: $p < 0.001$ and 18" screen size: $p < 0.01$) than patients without signs of mHE (N=57). Calculation of eNCT cut-off times by ROC-analysis (38 and 41 seconds) demonstrated good agreement (80.7% and 81.7%), sensitivity (96.2% and 92.3%) and specificity (73.7% and 76.8%) in mHE-detection regarding the reference test – the Psychometric Hepatic Encephalopathy Score (PHES). Test-retest-reliability was very good or good with intra-class-correlation coefficients of 0.906 and 0.855.

Conclusion: The eNCT as a reliable method for mHE-detection fulfills qualitative requirements for the investigation of applicability as a self-test at home. The test could become important for an early detection of decreasing cognitive performance in patients with liver cirrhosis, especially in rural areas. This could help to reduce patients' risk and fear of more severe acute decompensation, frequency of related hospital admission and healthcare costs.

3 Einleitung

3.1 Einführung in die Thematik

Die Leber ist als zentrales Stoffwechselorgan maßgeblich für die Aufrechterhaltung der metabolischen Homöostase verantwortlich. Erkrankungen der Leber gehen daher mit vielfältigen Komplikationen einher, die auch Funktionen anderer Organe beeinflussen und zu einer hohen gesundheitlichen Belastung der Patienten führen können.

Im Jahr 2014 starben nach Angaben des Statistischen Bundesamts fast 15.000 Menschen in Deutschland an einer Lebererkrankung. Das entsprach 1,7% aller Verstorbenen (1). Eine Folge vieler Lebererkrankungen, wie infektiöser oder autoimmunogener Hepatitiden, aber auch systemischer Erkrankungen (z. B. Hämochromatose oder Kupferspeicherkrankheiten) und exogener Noxen (z. B. Alkohol oder lebertoxische Medikamente), ist ein chronischer bindegewebiger Umbau der Leber. Das Endstadium dieser Veränderungen wird als Leberzirrhose bezeichnet. Deutschlandweit leiden circa eine Million Menschen an dieser Erkrankung, die neben diversen Komorbiditäten auch mit einer erhöhten Mortalität vergesellschaftet ist (2). Das zeigt sich nicht nur in einer verstärkten Inanspruchnahme des Gesundheitssystems mit vermehrten Krankenhausaufnahmen und einem hohen Pflegebedarf, sondern auch im subjektiv vom Patienten geäußerten Gefühl eines schlechteren Gesundheitszustands (3).

Eine der wesentlichen Komorbiditäten ist die hepatische Enzephalopathie (HE), eine „potenziell reversible, metabolisch bedingte Funktionsstörung des Zentralnervensystems“ auf dem Boden einer unterschiedlich stark eingeschränkten Leberfunktion (4). Bei der HE wird ein manifestes bzw. offensichtliches Stadium (oHE) von einem frühen subklinischen bzw. minimalen Stadium (mHE) unterschieden. Während Patienten mit einer oHE im Rahmen der klinischen Diagnostik durch neurologische bzw. neuropsychiatrische Symptome auffällig werden, können die kognitiven Veränderungen bei einer mHE nur durch spezielle psychometrische und apparative Testverfahren erkannt werden (4-6). Dennoch maßen Koch et al. der mHE bereits 1990 eine große Bedeutung bei, indem sie bei einem Drittel der Patienten mit Leberzirrhose das Vorliegen einer mHE vermuteten (7). Jüngere Studienergebnisse rechtfertigen Prävalenzangaben von 20 bis 80% (8-12).

Die mittleren jährlichen Kosten stationärer Aufenthalte eines Patienten im Jahr der Erstdiagnose einer HE liegen bei 12.220 US-Dollar in Europa bzw. 16.430 US-Dollar in den

USA, in folgenden Jahren bei durchschnittlich 3.810 US-Dollar (13, 14). 2003 gab es in den USA etwa 40.000 Patienten mit der erstmaligen Diagnose einer HE, deren gesundheitliche Versorgung Kosten von 932 Millionen US-Dollar verursachte (15).

Die Erkrankung ist nicht nur für das Gesundheitssystem eine enorme finanzielle Herausforderung, sondern ebenso für den Patienten. Bajaj et al. zeigten, dass nur 37% der Patienten mit Leberzirrhose arbeitstätig sind und längerfristig finanziellen Notsituationen entgehen können. Unbeschäftigte Patienten hatten nicht nur weniger Einkommen, sondern litten auch häufiger unter einer HE (16).

Das Stadium der mHE reicht bereits aus, um die Ergebnisse von Tests zur Erfassung grundlegender Alltagsfunktionen oder Tätigkeiten des prozeduralen Gedächtnisses (wie z. B. Autofahren) negativ zu beeinflussen (9, 17, 18).

Patienten, die darüber hinaus eine oHE entwickeln, können nach Rekonvaleszenz persistierende Defizite des Arbeitsgedächtnisses, der Impulshemmung und der Lernfähigkeit zurückbehalten (10). Die Sterblichkeitsrate ist für Leberzirrhosepatienten mit HE-Episoden erhöht, die Überlebenschancen außerdem geringer als für lebertransplantierte Personen (19).

In Anbetracht dieser weitreichenden Auswirkungen einer dekompenzierten Lebererkrankung ist eine frühzeitige Diagnostik insbesondere der mHE dringend notwendig, um die beschriebenen Langzeitfolgen und Komplikationen zu vermeiden. Mit dieser Dissertation soll ein neues Testverfahren vorgestellt werden, das als Selbsttest zukünftig als Teil eines häuslichen Überwachungssystems der Leberfunktion diese Lücke in der HE-Diagnostik schließen könnte: der *elektronische Zahlenverbindungstest (eNCT)*.

3.2 Geschichte der hepatischen Enzephalopathie

Eine frühe und viel zitierte Beobachtung über den Einfluss der Leberfunktion auf die kognitive Leistungsfähigkeit ist von Hippokrates von Kos (circa 460-375 v. Chr.) überliefert. Er sah, dass Menschen mit einer Erkrankung der Leber einen Zustand wie Tollwut entwickelten: „*[Der Patient] konnte nicht gehalten werden, sagte unverständliche Dinge und bellte wie ein Hund.*“ Weiter heißt es: „*Diese Patienten sind lärmend und böseartig und halten nicht ruhig.*“ Es handle sich um eine fatale Krankheit, in deren Verlauf sich Delir und Ikterus einstellten (20).

Während in Galens (129-201 n. Chr.) Humoralpathologie die Leber noch als ein zentraler Sitz der Seele vermutet wurde und eine Erklärung für Veränderungen des Denkens bot, wurde diese Sichtweise im Spätmittelalter verlassen und die Leber fortan als unbedeutend für die Kognition betrachtet. Eine erste schriftlich festgehaltene Abhandlung über die Verbindung zwischen Lebererkrankungen und kognitiver Leistungsfähigkeit sowie die Differenzierung in ein initiales Stadium gehemmter Sinnesleistungen und ein fortgeschrittenes Stadium mit Delir und Koma gibt es aus dem Jahr 1761 von Morgagni (21). Danach vergingen 200 Jahre bis zu einer wissenschaftlich umfassenden, neuzeitlichen Beschreibung der HE-Symptomatik durch Summerskill und Sherlock (22, 23).

3.3 Definition und Einteilung der hepatischen Enzephalopathie

Die Leitlinien der amerikanischen Vereinigung zur Erforschung von Lebererkrankungen und der europäischen Vereinigung zur Erforschung der Leber von 2014 definieren die HE als *„eine Dysfunktion des Gehirns verursacht durch Leberinsuffizienz und/oder portosystemischen Shunt; sie manifestiert sich mit einem großen Spektrum neurologischer oder psychiatrischer Veränderungen von subklinischen Symptomen bis Koma reichend“* (5). Der Begriff *„Enzephalopathie“* steht für die diffusen Veränderungen des Gehirns, deren kausale Verbindung mit der Leberfunktion durch *„hepatisch“* ausgedrückt wird. Die Bezeichnungen *„portosystemische“* oder *„portokavale“* Enzephalopathie sind nicht mehr aktuell, da sie als Ursache einer HE ausschließlich Umgehungskreisläufe implizieren und die Möglichkeit der eingeschränkten Leberfunktion vernachlässigen (21).

Die Klassifikation der HE hat sich mit steigender Sensitivität der Diagnostik immer wieder verändert. 1970 zeigten sich mit Einführung der Zahlenverbindungstests Hinweise auf eine klinisch nicht offensichtliche, latente bzw. minimale Form der HE (24). Die 2002 von Ferenci et al. veröffentlichten Beschlüsse für eine einheitliche Nomenklatur der HE differenzieren zwischen einer minimalen, einer episodischen und einer persistierenden Form (6). In den Leitlinien zur HE von 2014 werden zur Einteilung der HE von der Arbeitsgruppe um Vilstrup vier Schritte vorgeschlagen (5):

1. Klassifikation der HE entsprechend der Ursache in Typ A (aufgrund reduzierter Entgiftungsleistung der Leber bei akutem Leberversagen), Typ B (bei Umgehung der hepatischen Entgiftungsfunktion durch portosystemischen Bypass oder Shunt) und Typ C (durch Kombination der zuvor genannten Mechanismen bei Leberzirrhose).
2. Unterscheidung der HE entsprechend des Schweregrads: Orientierend an den West-Haven-Kriterien (WHC) von Conn et al. gibt es inklusive der mHE fünf Schweregrade (Tabelle 1):
 - Grad *minimal*: Patienten zeigen leichte Störungen der Aufmerksamkeit und Konzentration, des Antriebs, der Merkfähigkeit, der visuell-räumlichen Kognition und der Feinmotorik. Da diese Symptome durch die klinische Untersuchung nicht erfasst werden können, ist die Anwendung zusätzlicher Methoden, z. B. psychometrischer Testverfahren, notwendig.
 - Grad 1: Patienten klagen über ein verändertes Schlafverhalten und Stimmungsschwankungen, wirken antriebsarm und machen Fehler bei einfachen Rechenaufgaben. Die vollständige Orientierung ist noch erhalten.
 - Grad 2: Eine deutliche Verlangsamung des Denkens mit zeitlicher Desorientierung kann beobachtet werden. Auffällig sind eine zunehmende körperliche Symptomatik mit Asterixis, Hyperreflexie und Dyspraxie.
 - Grad 3: Patienten leiden unter einer hochgradigen Bewusstseinsstörung mit zunehmender Somnolenz und Desorientierung. Eine Reaktion auf Schmerzreize ist erhalten.
 - Grad 4: Es zeigen sich komatöse Patienten mit Bewusstlosigkeit und Aufhebung von Schmerz- und Kornealreflexen (5, 25).
3. Klassifikation der HE entsprechend des Zeitverlaufs in *episodisch* (HE-freies Intervall >6 Monate), *periodisch* (HE-freies Intervall ≤6 Monate) und *persistierend*.
4. Unterscheidung der HE entsprechend des Vorliegens (vor allem bei Typ C) oder Fehlens auslösender Faktoren wie Infektionen, gastrointestinale Blutungen, Elektrolytverschiebungen, Diuretikaüberdosierung oder Einnahme von Benzodiazepinen, Analgetika oder Hypnotika (5).

Tabelle 1: Klassifikation der HE, modifiziert nach Vilstrup et al. (5)

West-Haven-Kriterien	ISHEN	Klinische Symptomatik	Empfohlene diagnostische Kriterien
<i>Nicht pathologisch</i>		Keine HE	
Minimal (=mHE, leichtgradig)	verdeckt	Veränderungen der psychomotorischen Geschwindigkeit oder Exekutivfunktionen ohne klinische Hinweise	Pathologische Ergebnisse in psychometrischen oder neuropsychologischen Tests
Grad 1 (=oHE, leichtgradig)		<ul style="list-style-type: none"> • verkürzte Aufmerksamkeitsspanne • Konzentrationsschwäche • Euphorie oder Angst, Stimmungsschwankungen • Leichtes Bewusstseinsdefizit • Fehler bei einfachen Additions- oder Subtraktionsaufgaben • Veränderter Schlafrhythmus 	Vollständige Orientierung, aber klinische Untersuchung gibt Hinweise auf Veränderungen der Kognition oder des Verhaltens
Grad 2 (=oHE, leichtgradig)		<ul style="list-style-type: none"> • Lethargie oder Apathie • Zeitliche Desorientierung • Persönlichkeitsveränderungen • Unangemessenes Verhalten • Dyspraxie • Asterixis („flapping tremor“) 	Fehler bei drei von: Wochentag, Datum, Jahreszeit, Jahreszahl; ggf. triphasische Wellen im EEG
Grad 3 (=oHE, hochgradig)	sichtbar	<ul style="list-style-type: none"> • Somnolenz bis Stupor • Reaktion auf Schmerzreize • Verwirrt • Ausgeprägte Desorientierung • Bizarres Verhalten 	Fehler bei drei von: Land, Bundesland, Stadt, Ort; ggf. triphasische Wellen im EEG
Grad 4 (=oHE, hochgradig)		Koma (keine Reaktion auf Schmerzreize, Kornealreflexe erloschen)	fehlende Reaktion auf Schmerzreiz; ggf. Deltawellen im EEG
ISHEN: International Society for Hepatic Encephalopathy and Nitrogen Metabolism; EEG: Elektroenzephalografie; mHE/oHE: minimale / offensichtliche hepatische Enzephalopathie			

Da es bei der Einstufung von Patienten mit HE Grad 0-2 in 30-50% der Fälle trotz Anwendung diagnostischer Tests zu unterschiedlichen Ergebnissen kommt, gibt es Ansätze, die Klassifikation für eine Ersteinschätzung der Patienten zu vereinfachen. Die *International Society for Hepatic Encephalopathy and Nitrogen Metabolism* (ISHEN) schlägt für die HE die in Tabelle 1 gezeigte Unterscheidung zwischen Grad mHE sowie Grad 1 (*verdeckt*) und Grad 2 bis 4 (*sichtbar*) vor, wonach eine sichtbare HE ohne spezifische Tests bereits im Patientengespräch zu erkennen ist. An anderer Stelle wird eine hochgradige HE (Grad 3 bis 4) bestehend aus einem geringen Bewusstseinsniveau, einer stark eingeschränkten Handlungsfähigkeit und einer Notwendigkeit zur Hospitalisation von einer leichtgradigen HE (Grad mHE bis 2) unterschieden (26).

3.4 Pathophysiologie der hepatischen Enzephalopathie

Der Pathomechanismus zur Entstehung einer HE ist multifaktoriell. Als zentrale Einflussgröße ist ein Ungleichgewicht zwischen hemmenden und aktivierenden Signalen des zentralen Nervensystems beschrieben; bei akutem Leberversagen mit Ausbildung eines druckrelevanten Hirnödems und bei chronischer Lebererkrankung mit diskreter Zellschwellung und einem Übergewicht der hemmenden Signale (27). Korrelate dieser pathologischen Zustände sind Veränderungen der Blut-Hirn-Schranke, des Energiestoffwechsels und der neuronalen Signalübertragung, welche durch oxidativen Stress und Schwellung der Astrozyten ausgelöst werden. Als zentrale Ursache zur Entstehung der Veränderungen und damit zur Entstehung einer HE werden insbesondere die metabolische Entgleisung des Aminosäurestoffwechsels und die unzureichende Ammoniakentgiftung im Rahmen der Leberzirrhose verantwortlich gemacht (26, 28-30).

Ammoniak ist das zentrale Stoffwechselintermediat, welches beim Abbau von Aminosäuren entsteht und der Bindung des Aminostickstoffs zur weiteren Ausscheidung dient. In der gesunden Leber wird Ammoniak periportal über den Harnstoffzyklus zu Harnstoff und perivenös über den Glutaminzyklus zu Glutamin verarbeitet (31). Ein geringer Anteil des Ammoniakgehalts tritt über die Blut-Hirn-Schranke ins zentrale Nervensystem über. Hier wird Glutamat mit Ammoniak von der astrozytären Glutaminsynthetase unter ATP-Verbrauch in Glutamin umgewandelt, das den Neuronen zur Synthese von Neurotransmittern dient (32-34). Dieser enzymabhängige Schritt ist einer von vielen Funktionen der Astrozyten, den größten und am häufigsten vorkommenden Zellen des glialen

Stützgewebes. Astrozyten sorgen außerdem für die Flüssigkeits- und Kaliumhomöostase, die Abschirmung von Synapsen, die Wiederaufnahme von Transmittern, die Sekretion neurotropher Substanzen, die Ausbildung der Blut-Hirn-Schranke und für Stoffwechselfunktionen wie die Glutaminsynthese (35, 36).

Bei Patienten mit einer verminderten Leberfunktion sind die hepatischen Eliminationsvorgänge beeinträchtigt, was zu erhöhten Ammoniakwerten im Blut führen kann. Dementsprechend gelangt mehr Ammoniak über die Blut-Hirn-Schranke in Astrozyten. Das dort aus Ammoniak und Glutamat entstehende, osmotisch wirksame Glutamin wiederum erhöht die Permeabilität der Blut-Hirn-Schranke, was einen schwer zu durchbrechenden Kreislauf mit weiter steigendem Ammoniakestrom ins zentrale Nervensystem etabliert (37, 38). Durch experimentelle Hemmung der Glutaminsynthetase bei gleichzeitiger Ammoniaküberladung konnte gezeigt werden, dass es ohne die Produktion von Glutamin nicht zu einer gesteigerten Permeabilität kommt (39).

In der Zellschwellung liegt der erste Aspekt der Ammoniaktoxizität: Im akuten Leberversagen und bei rapiden Anstiegen der Ammoniakwerte kann es in kurzer Zeit zu ausgeprägten bis lebensbedrohlichen Hirnödemen kommen. Bei Patienten mit Leberzirrhose sind funktionseinschränkende Schwellungen der Astrozyten mit erhöhter Produktion reaktiver Sauerstoff- (ROS) und Stickstoffoxidspezies (RNOS) sowie diskrete Hirnödeme durch die langfristig verminderte Ammoniakentgiftung möglich (37, 38, 40). Diese Veränderungen sind zwar potenziell reversibel, können aber vorübergehend zur Entstehung einer symptomatischen HE führen (41). In vitro konnte gezeigt werden, dass ein erhöhter Glutaminspiegel im Rahmen der Akkumulation von Ammoniak im Astrozyten zu einer vermehrten Aufnahme von Glutamin ins Mitochondrium führt, wo durch eine Glutaminase erneut Ammoniak gebildet wird. Dieses soll über Calciumsignalwege und oxidativen Stress eine Induktion der mitochondrialen Permeabilitäts-Transitions-Pore (MPTP) bewirken, welche zu weiterer ROS-Bildung und letztlich zur Dysfunktion der mitochondrialen Volumenregulation führt (Abbildung 1) (42). Glutamin wird in dieser Theorie daher als „*trojanisches Pferd*“ für Ammoniak bezeichnet und gilt als potenzieller Auslöser für Astrozytenschwellung und ROS/RNOS-Produktion (30, 32, 43).

Die zweite wesentliche Auswirkung der Ammoniaktoxizität ist die Aktivierung der NADPH-Oxidase bzw. Stickstoffmonoxid-Synthase, die ebenfalls zur erhöhten Bildung von ROS bzw. RNOS beiträgt (28). Auf diesem Weg wird zusätzlich der Energiestoffwechsel des Gehirns und die neuronale Signaltransduktion gehemmt (44). Als Beispiel

sei hier die RNOS-abhängige Mobilisation von Zink genannt. Zink verstärkt inhibierende Gamma-Amino-Buttersäure (GABA)-Signalwege und nimmt auf Ebene der Genexpression Einfluss auf die Produktion von Neurosteroiden (45, 46).

Außerdem kommt es zur Oxidation von RNA-Molekülen, die an Translation und dendritischen Transportprozessen beteiligt sind, wodurch ein beeinträchtigtes Gedächtnis im Rahmen einer HE erklärt werden könnte (47, 48). In Astrozyten von Ratten konnte beispielsweise gezeigt werden, dass die Aktivierung der NADPH-Oxidase die Menge bestimmter micro-RNAs verringert und so zur Seneszenz der Astrozyten führt (49).

Die genannten Folgen der Ammoniaküberladung – Astrozytenschwellung und ROS/RNOS-Bildung – führen letztlich zu der in Abbildung 1 gezeigten Ausbildung eines weiteren Circulus vitiosus (42, 50). Denn auch die Schwellung der Astrozyten erhöht durch konsekutive Aktivierung von N-Methyl-D-Aspartat (NMDA)-Rezeptoren, der NADPH-Oxidase und Calciumsignalkaskaden die Produktion reaktiver Spezies, die wiederum die Zellschwellung verstärken (47, 51, 52).

Allerdings gibt es Patienten mit Leberzirrhose, die trotz hoher Ammoniakspiegel keine HE entwickeln. In diesen Fällen besitzen die Astrozyten vermutlich genügend Osmolyte, um dem durch Glutamin ausgelösten, osmotischen Stress entgegenzuwirken und die Zellschwellung mit Ausbildung HE-typischer Symptome zu verhindern. Hier sind im Wesentlichen die Aminosulfonsäure Taurin und der sechswertige Alkohol Myo-Inositol als protektive, osmotisch wirksame Moleküle zu nennen, die bei vermehrter Glutaminsynthese zum Erhalt des Zellvolumens abgegeben werden können und deren sinkende intrazelluläre Konzentration in Gehirnen von Patienten mit Leberzirrhose mit einem ansteigenden Glutamin/Glutamat-Signal einhergeht (40). Dieses Signal ist insbesondere für Myo-Inositol bildiagnostisch gut darstellbar, sodass es als prognostischer Marker für Hirnödem und kognitive Dysfunktion untersucht wird (53, 54).

Zusätzliche Einflussfaktoren (z. B. Hyponatriämie, durch Infektion oder Trauma erhöhte Produktion proinflammatorischer Zytokine, Einnahme sedativer bzw. diuretischer Medikation) können bei Leberzirrhosepatienten mit verbrauchten Osmolyten hingegen trotz eines normwertigen Ammoniaks im Unterschied zu Lebergesunden zur Schwellung der Astrozyten führen und damit die verschieden rasche Entwicklung der Symptomatik erklären (33, 55). Diese Faktoren wirken synergistisch und tragen wie Ammoniak nicht nur zur Astrozytenschwellung, sondern auch zur ROS/RNOS-Entstehung bei (26). Weiterhin werden aufgrund der erhöhten Permeabilität der Blut-Hirn-Schranke, bei einem teil-

weise gleichzeitig vorliegenden Mangel an verzweigt-kettigen Aminosäuren, vermehrt aromatische Aminosäuren (z. B. Tryptophan, Tyrosin) ins Gehirn aufgenommen. Dort

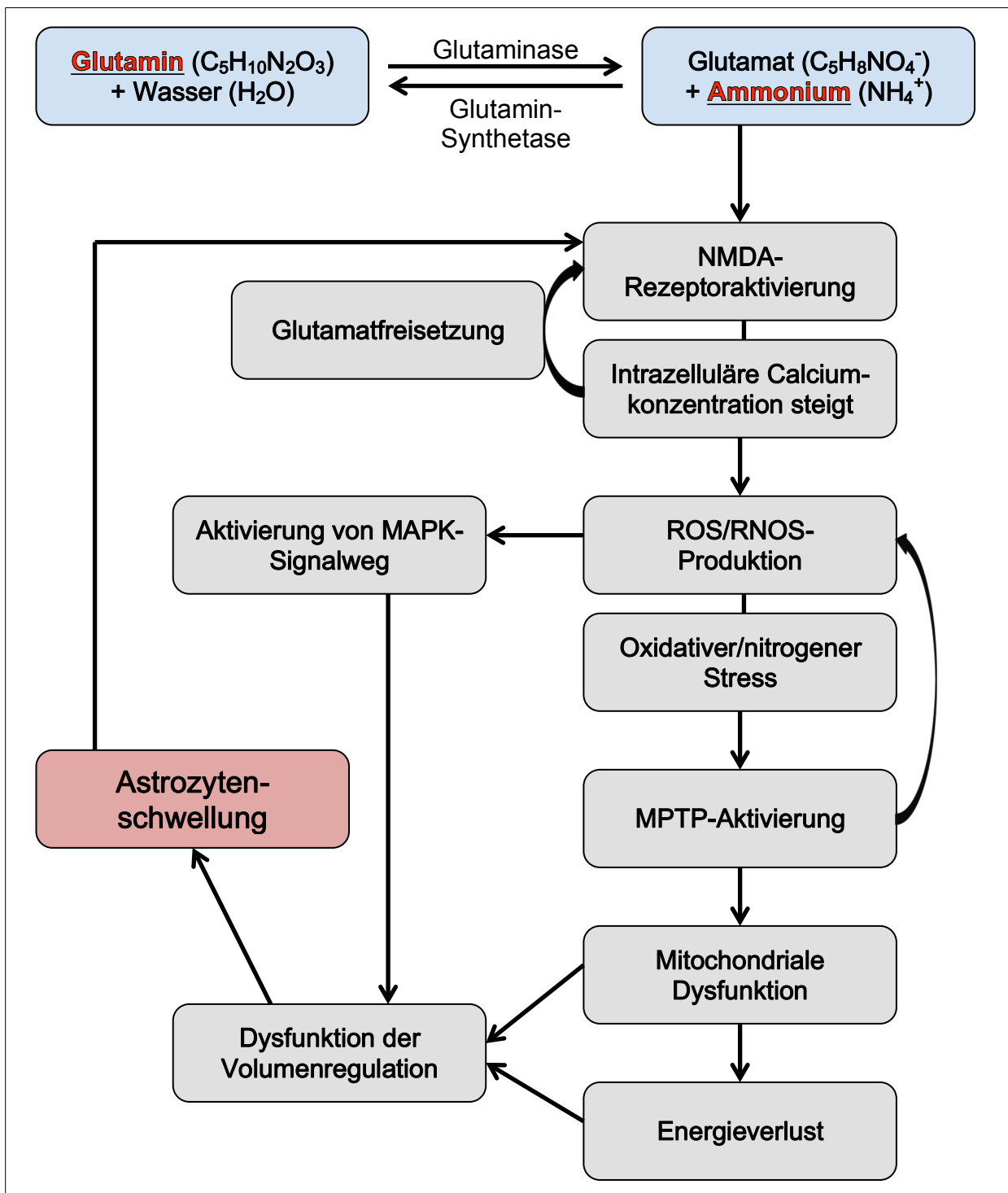


Abbildung 1: Pathogenese der HE auf Mitochondrienebene (modifiziert nach Scott et al. (42)) – Wie ein „trojanisches Pferd“ gelangt Ammoniak per Glutamin ins Mitochondrium und löst dort NMDA-Rezeptorvermittelte Signalkaskaden aus, die sich teilweise gegenseitig verstärken und letztlich zur Schwellung der Astrozyten führen. (MAPK: Mitogen-aktivierte Proteinkinasen; NMDA: N-Methyl-D-Aspartat; ROS/RNOS: reaktive Sauerstoff- bzw. Stickstoffoxidspezies; MPTP: Mitochondriale Permeabilitäts-Transitions-Pore)

bewirken letztere als Vorläufer von inhibierenden Botenstoffen (Serotonin, Dopamin) eine Dysfunktion der neuronalen Transmission und tragen insbesondere zur Manifestation der feinmotorischen Symptomatik bei (Abbildung 2) (27, 56).

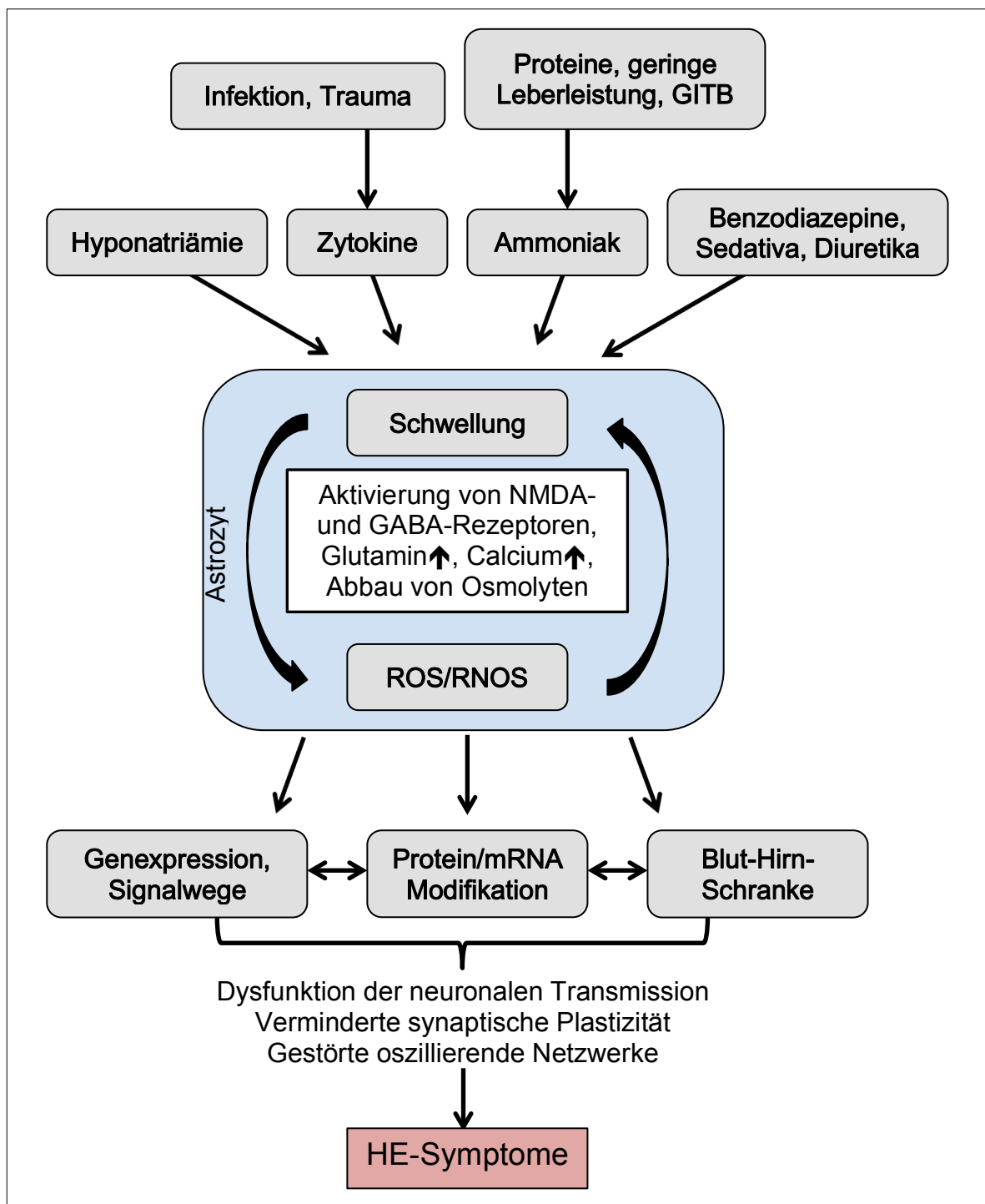


Abbildung 2: Entstehung der HE-Symptomatik (modifiziert nach Häussinger et al. (26)) – Zentral ist die Astrozytendysfunktion, bei Patienten mit Leberzirrhose auslösbar durch die gezeigten Einflussfaktoren. Folge ist auf mehreren Ebenen Beeinträchtigung der neuronalen Funktion und damit die HE-Symptomatik. (GITB: gastrointestinale Blutung; NMDA: N-Methyl-D-Aspartat; GABA: Gamma-Amino-Buttersäure; ROS/RNOS: reaktive Sauerstoff- bzw. Stickstoffoxidspezies; HE: hepatische Enzephalopathie)

An dieser Stelle sollen die verschiedenen Erklärungsansätze zusammengefasst werden: Ammoniak führt zu einer astrozytären Akkumulation von Glutamin, was einen kompensatorischen Abbau osmolytisch wirksamer Substanzen wie Taurin und Myo-Inositol bewirkt. Bei Erschöpfung der volumenregulatorischen Mechanismen kommt es im Gehirn, insbesondere durch Einwirken weiterer HE-auslösender Faktoren, zum vermehrten Einstrom aromatischer Aminosäuren mit konsekutiver Verstärkung hemmender Signale und zur Ausbildung leichtgradiger Ödeme. Diese Ödeme führen zwar noch nicht zu einer Erhöhung des intrakraniellen Drucks, beeinflussen aber die Funktion der Astrozyten und die Genexpression durch – abhängig vom oxidativen Stress – veränderte Proteine und RNA. Eine gegenseitige Verstärkung von astroglialer Schwellung und oxidativem Stress etabliert eine sich selbst erhaltende Signalschleife, die durch die HE-auslösenden Faktoren noch aggraviert wird. Als Konsequenz der primär glialen Dysfunktion durch Astrozytenschwellung werden in Magnet- und Elektroenzephalografie Störungen der zerebralen Oszillationen (niedrigere Frequenzen) messbar, die letztlich ein Äquivalent der HE-Symptome darstellen (Abbildung 2) (26, 27, 57). Diese Funktionseinschränkungen können neben dem Motorkortex weitere Gebiete des Gehirns betreffen und somit die Divergenz der klinischen Symptomatik erklären (58).

3.5 Diagnostik der hepatischen Enzephalopathie

3.5.1 Überblick

Es konnte bisher gezeigt werden, dass sich die HE-Symptomatik aufgrund individuell unterschiedlicher pathophysiologischer Veränderungen sehr variabel präsentiert und entsprechend ihrer Ausprägung in verschiedene Schweregrade eingeteilt wird. Nun sollen diagnostische Testverfahren vorgestellt werden, mithilfe derer Leberzirrhosepatienten einem HE-Schweregrad zugeteilt werden können.

Empfehlungen zur HE-Diagnostik geben die Leitlinien zur HE von Vilstrup et al. (5). Eine offensichtliche HE ist demnach eine Diagnose basierend auf der klinischen Untersuchung, wobei Asterixis und Desorientierung als Leitsymptome gelten. In Studiensituationen bieten die WHC (Tabelle 1) die Möglichkeit einer standardisierten Untersuchung. Die Schwierigkeit besteht darin, vorliegende Symptome eindeutig einer manifesten HE zuzuordnen zu können und nicht als Resultat von Medikamenten, Alkoholmissbrauch,

Hyponatriämie oder psychiatrischer Erkrankung zu werten. Die HE ist demnach eine Ausschlussdiagnose, das gilt für die oHE und die mHE (5).

Für die Diagnose der mHE stehen psychometrische und neuropsychologische Tests zur Verfügung, da die mHE einerseits verschiedene kognitive Domänen und andererseits fast ausschließlich die nicht-verbale Intelligenz betrifft (17, 59). Demnach entzieht sich das minimale Stadium der HE einer Diagnostik im Arzt-Patienten-Gespräch. Blutwerte und computer- oder magnetresonanztomografische Aufnahmen des Gehirns sind bei geringer Aussagekraft ebenso wenig zielführend.

Die ISHEN empfiehlt aufgrund der Vielzahl möglicher Ausprägungen einer mHE bei Mehrzentren-Studien die Anwendung von mindestens zwei Untersuchungsmethoden. Dabei ist eine von zwei gleichwertigen psychometrischen Testbatterien – *Psychometric Hepatic Encephalopathy Score (PHES)* oder *Repeatable Battery for the Assessment of Neuropsychological Status (RBANS)* – und ein weiteres Verfahren (*Critical Flicker Frequency (CFF)*, *Continuous Reaction Time (CRT)*, *Inhibitory Control Test (ICT)*, *Stroop-Test*, *Elektroenzephalografie (EEG)*) auszuwählen (5, 60). Bei Einzelzentrum-Studien ist ein Test ausreichend, mit dem das Studienpersonal vertraut ist und für den Normwerte verfügbar sind (61). Die besagten Tests sollten außerdem nur dann Anwendung finden, wenn alternative Ursachen eingeschränkter kognitiver Leistungsfähigkeit oder Sonderformen der HE (z. B. Reye-Syndrom) ausgeschlossen wurden (5).

3.5.2 Klinische Testverfahren

3.5.2.1 West-Haven-Kriterien (WHC)

Die semiquantitativen WHC beinhalten die Prüfung von Bewusstseinslage, neuropsychiatrischen und neurologischen Symptomen und teilen die oHE in Grad 1 bis 4 ein (Tabelle 1). Zusätzlich existiert Grad 0 für klinisch unauffällige, aber möglicherweise minimal enzephalopathische Patienten. Subjektive Einschätzungen des Schweregrades machen den Test fehleranfällig, weshalb sich in den letzten Jahren unter anderem der *Hepatische Enzephalopathie Wertungsalgorithmus* (engl.: *Hepatic Encephalopathy Scoring Algorithm – HESA*) als Hilfsmittel mit einer Kombination aus subjektiven und objektiven Kriterien etabliert hat. In den Leitlinien zur HE von 2014 wird der HESA noch nicht erwähnt, sodass die WHC als Standard bestehen blieben (5, 60).

3.5.2 Hepatische Enzephalopathie Wertungsalgorithmus (HESA)

Hier werden die neuropsychiatrischen Kategorien Orientierung, Erinnerungsfähigkeit, Aufmerksamkeit, Rechenleistung und Gemütszustand durch standardisierte Fragebögen überprüft (Abbildung 5, Abschnitt 5.4.2). Es erfolgt ebenso eine Einteilung der HE in Grad 0 bis 4. Die Klassifizierung der Patienten mittels HESA ist konservativer als mittels WHC, es werden also durchschnittlich weniger HE-Diagnosen gestellt (62, 63). Prinzipiell kann aber ohne weitere spezielle Tests (vgl. ab Abschnitt 3.5.4) weder mittels HESA noch mittels WHC eine mHE diagnostiziert werden.

3.5.3 Blutparameter

Aufgrund der Beobachtung erhöhter Ammoniakwerte aus den in Abschnitt 3.4 genannten Gründen bei Patienten mit Lebererkrankungen, wurde ab 1938 die diagnostische Abklärung durch Messung der Ammoniakkonzentration im Blut empfohlen.

Obwohl bereits 1952 von Phillips et al. eine geringe Bedeutung des Blutwerts für die HE-Diagnostik beschrieben wurde, galt Ammoniak weiterhin als potenzieller Marker (64). Die Aussagekraft dieses Parameters ist bei niedriger Sensitivität und Spezifität allerdings gering (65). Des Weiteren sind keine Blutwerte bekannt, die für die HE-Diagnostik von pathognomonischer Bedeutung wären. Es gibt zuverlässige Leber-, Nieren- und Gerinnungsfunktionswerte, die ebenso wie die Elektrolyte zur Ausschlussdiagnostik relevant sind. Es bietet sich aber kein Parameter an, von dessen Wert direkt auf eine HE geschlossen werden kann.

3.5.4 Psychometrische Testverfahren

Ein potenzieller Goldstandard in der psychometrischen Diagnostik der mHE sollte neben einer hohen Genauigkeit viele kognitive Domänen erfassen, universell anwendbar sein, alterskorrigierte Testwerte beinhalten, von angemessener Dauer sein und aus Einzeltests mit alternativen Anordnungen der Testelemente bestehen. Diese Voraussetzungen werden laut Randolph et al. gleichermaßen von den genannten Testbatterien PHES und RBANS erfüllt (60).

Der in Deutschland etablierte und aus dem PSE-Syndrom-Test weiterentwickelte PHES besteht aus fünf papierbasierten Tests, die zusammen die kognitiven Domänen *Aufmerksamkeit*, *visuell-räumliche Orientierung*, *visuelles Kurzzeitgedächtnis*, *psychomoto-*

rische Geschwindigkeit und *feinmotorische Präzision* erfassen und deren alterskorrigierbare Einzelergebnisse zu einem Gesamtwert zusammengerechnet werden (66, 67). Die Testbatterie besteht aus dem Zahlensymboltest (DST), dem Zahlenverbindungstest A (NCT-A), dem Zahlenverbindungstest B (NCT-B), dem Kreise-punktieren-Test (SDT) und dem Liniennachfahrttest (LNT). Die Tests werden in der genannten Reihenfolge durchgeführt und jeder Test beginnt zunächst mit einem Übungsteil.

(1) Im DST ist den Zahlen 1 bis 9 jeweils ein Symbol zugeordnet. Diese Zuteilung befindet sich oberhalb des Testbereichs und ist während der Testdurchführung jederzeit sichtbar. Der Testbereich besteht aus einer zufälligen Abfolge der Zahlen 1 bis 9. Unter den Zahlen sind jeweils leere Kästchen, in die innerhalb von 90 Sekunden so viele (den Zahlen korrekt entsprechende) Symbole wie möglich eingezeichnet werden sollen.

(2) Der NCT-A besteht aus 25 zufällig auf einem Blatt Papier angeordneten Zahlen von 1 bis 25, die so schnell wie möglich mithilfe eines Stifts in der normalen Zählreihenfolge, auf kürzestem Weg und ohne abzusetzen miteinander verbunden werden sollen.

(3) Im NCT-B gibt es die Zahlen 1 bis 13 und die Buchstaben A bis L, die ebenso zufällig auf einem Blatt Papier angeordnet sind und so schnell wie möglich, auf kürzestem Weg und ohne abzusetzen mit einem Stift alternierend (Zahl-Buchstabe-Zahl-...) in der richtigen Zähl- bzw. Buchstabierreihenfolge miteinander verbunden werden sollen.

(4) Ziel des SDTs ist es, schnellstmöglich alle auf einem Blatt Papier gezeigten Kreise der Reihe nach in ihrer Mitte mit einem Punkt zu versehen.

(5) Beim LNT muss eine dargestellte Strecke mit dem Stift so schnell und genau wie möglich und ohne abzusetzen nachgefahren werden. Fehler (Berührungen oder Überquerungen der Streckenmarkierungen) werden mit einer Schablone ausgewertet.

Am Ende gehen sechs Einzelergebnisse in den Gesamtwert ein, vier Zeiten und zwei Mengen (Paare im DST und Fehler im LNT). Jedes Einzelergebnis wird anhand von altersentsprechenden Normwerttabellen einem Bereich der Standardabweichung (SD) vom Mittelwert zugeordnet, was den resultierenden Punktwert eines Einzelergebnisses bildet (Einzelergebnis schlechter als: 3 SD = -3 / 2 SD = -2 / 1 SD = -1 Punkte; so gut wie Mittelwert \pm 1 SD = 0 Punkte; besser als 1 SD = 1 Punkt). Dementsprechend kann der Gesamtwert des PHES zwischen -18 und +6 liegen, eine mHE wird bei Werten kleiner als -4 diagnostiziert. Mit diesem Grenzwert erreichte der PHES eine Sensitivität von 96% und eine Spezifität von 100%, wobei der SDT mit einer Sensitivität von 84% und einer Spezifität von 98,3% die besten Einzelwerte erzielte. Unter Anwendung des PHES

wurde bei 22,2% der klinisch unauffälligen Patienten mit Leberzirrhose eine mHE diagnostiziert (67). Die Entwicklung und die Testbögen des PHES sowie deren korrekte Anwendung und Auswertung sind ausführlich in der Dissertation von J. C. Ennen und im Handbuch von Schomerus et al. dargestellt (68, 69).

Der RBANS wurde in den USA primär zur Demenzdiagnostik und zur Überwachung neurokognitiver Leistungen entwickelt. Die Testbatterie ist ebenso papierbasiert, prüft aber zusätzlich zum PHES die bei der mHE weniger betroffenen Domänen *verbale Intelligenz* und *Langzeitgedächtnis*. Allerdings ist der RBANS mit zwölf Einzeltests zeitaufwendiger und teilweise sprachbasiert, also zur internationalen Anwendung auf Übersetzungen und entsprechende Validierung angewiesen (70, 71).

3.5.5 Spezialdiagnostik (CFF, CRT, ICT, EEG)

In den letzten Jahren wurde zunehmend der Versuch unternommen, ein technisches Messgerät für die HE zu entwickeln, das die Papiertests ablösen oder ergänzen könnte.

- Critical Flicker Frequency (CFF): Die CFF ist die Frequenz, ab der ein Lichtstrahl flackernd wahrgenommen wird. So wird dem Probanden beim CFF-Test über eine nach außen geschlossene Brille eine definierte Lichtquelle von 60 Hertz präsentiert, deren Frequenz langsam reduziert wird. Der Proband drückt einen Knopf sobald er ein Flackern erkennt. Patienten mit einer mHE nehmen dieses Flackern verzögert wahr.

Eine eigenständige Anwendung des CFF-Tests ist nur bedingt möglich, da ein spezielles Gerät notwendig ist und die Sehfähigkeit nicht eingeschränkt sein darf (72, 73).

- Continuous Reaction Time (CRT): In ähnlicher Weise soll der Proband beim CRT-Test schnellstmöglich auf ein akustisches Signal reagieren. Das Resultat ist der CRT-Index, der unabhängig von Alter, Geschlecht und Lerneffekten eine Aussage über die Stabilität der Reaktionszeit trifft, welche bei Patienten mit einer mHE wechselhaft ist (74).

- Inhibitory Control Test (ICT): Beim computerbasierten ICT soll der Proband dargestellten Aufforderungen zum Klicken widerstehen, falls diese nicht einer übergeordneten Testaufgabe entsprechen. Je häufiger ein Patient fehlerhafte Antworten gibt, desto größer ist die Wahrscheinlichkeit für eine bestehende mHE (75).

- Elektroenzephalografie (EEG): Die EEG nimmt als einziges Verfahren, das Ergebnisse unabhängig von der Motivation des Patienten liefert, eine Sonderposition ein. Bei Patienten mit einer mHE können mit der EEG Veränderungen der Hirnaktivität gemessen werden, die allerdings nicht spezifisch sind (76).

3.5.5.1 Der Stroop-Test

Der erstmals 1935 im *Journal of Experimental Psychology* beschriebene und nach seinem Erfinder, John Ridley Stroop, benannte Stroop-Test basiert auf dem Effekt, dass Einschränkungen der Aufmerksamkeit zu Problemen bei ungewohnten Handlungen führen (77). Der Test konnte in seiner elektronischen Variante, der *EncephalApp*, wie die vorher genannten Tests für die mHE-Diagnostik validiert werden (78). Dabei werden dem Probanden auf einem berührungsempfindlichen Bildschirm eines Tablet-Computers zunächst Rauten-Symbole in einer von drei möglichen Farben (rot, blau, grün) präsentiert. Die Aufgabe besteht darin, den korrekten Farbbegriff zu der Farbe des jeweiligen Symbols aus den drei angebotenen Farbbegriffen auszuwählen und anzuklicken (Abbildung 3A). Dieser so genannte *Stroop-Off*-Modus wird nach zwei Probe- und fünf Testdurchgängen mit je zehn Aufgaben vom *Stroop-On*-Modus für weitere zwei Probe- und fünf Testdurchgänge abgelöst. Hier sieht der Proband nun neben Symbolen auch Farbbegriffe („rot“, „blau“, „grün“), die aber teilweise nicht in der dem Begriff entsprechenden Farbe abgebildet sind (Abbildung 3B). Dennoch soll der Proband unabhängig vom Begriff und im Sinne einer ungewohnten Handlung die korrekte Farbe auswählen. Werden Fehler gemacht, so ist der jeweilige Durchgang zu wiederholen. Insgesamt sind sechs Testergebnisse messbar: die *Stroop-Off*-Zeit, die *Stroop-On*-Zeit, die Gesamtzeit (*On+Off*), die Differenz (*On-Off*), die Anzahl der Versuche im *Stroop-Off*- und im *Stroop-On*-Modus. Bisher sind *On+Off*-Grenzwerte für zwei verschiedene Altersgruppen (<45 Jahre: $\geq 145s$, ≥ 45 Jahre: $\geq 190s$) validiert (78, 79). Der Test ist kostenlos als Smartphone- und Tablet-Computer-Anwendung erhältlich.



Abbildung 3: Bildschirmpräsentation des Stroop-Tests im *Stroop-Off*- (A) und *Stroop-On*- (B) Modus der *EncephalApp*. In (A) wäre Grün die richtige Wahl, in (B) dagegen Rot.

3.6 Therapie der hepatischen Enzephalopathie

Der folgende Abschnitt soll aktuelle Behandlungskonzepte der HE aufzeigen, die durch ihren positiven Einfluss auf den Krankheitsverlauf die Relevanz einer zuverlässigen HE-Diagnostik begründen. In den Leitlinien zur HE von Vilstrup et al. sind Ziele und Empfehlungen für die Behandlung einer oHE formuliert. Die Therapie von Patienten mit einer mHE wird dagegen nicht exakt festgelegt, da wenig Langzeitdaten existieren und die Diagnostik in den verschiedenen Studien uneinheitlich war (5).

Zunächst sollen alternative Ursachen veränderten Denkens ausgeschlossen bzw. behandelt werden, da hiermit bereits 90% der Fälle aufgeklärt werden können. Daraufhin soll gleichzeitig mit der Identifikation und Korrektur eventuell vorliegender auslösender Faktoren (Abbildung 2) eine aktive Behandlung der oHE begonnen werden. Es sollte keine proteinarme Diät zur Verminderung der Ammoniakaufnahme durchgeführt werden, da der daraus folgende Abbau von Muskelmasse zusätzlich Ammoniak freisetzt. Eine gesteigerte Zufuhr verzweigkettiger Aminosäuren kann im Gegenteil die Ammoniakentgiftung und damit die Symptomatik der HE verbessern (80, 81).

Das Wirkprinzip der Medikamente ist entsprechend der Pathogenese einer HE im Wesentlichen auf eine verminderte Ammoniakaufnahme ausgelegt. Eine astrozytäre Akkumulation von Glutamin, ein Verbrauch von Osmolyten und eine Astrozytenschwellung sollen so verhindert werden. Die Erstlinientherapie geschieht durch nicht-resorbierbare Disaccharide (z. B. Laktulose), die durch Reduktion des pH-Werts im Kolon und laxierende Wirkung die Aufnahme von Ammoniak und Glutamin verringern (82). Als Alternative oder Ergänzung kann ein schwer resorbierbares Antibiotikum wie Rifaximin verwendet werden, das gegen ammoniakbildende Darmbakterien wirksam ist. Bei kurzen Behandlungsintervallen können ebenso Neomycin oder Metronidazol verordnet werden, die für eine Langzeitbehandlung aufgrund ihrer Toxizität nicht in Frage kommen. Bei Patienten mit persistierender HE kann die zusätzliche intravenöse Gabe von Ornithin-Aspartat indiziert sein, das die Metabolisierung von Ammoniak zu Harnstoff beschleunigt. Die orale Gabe von Ornithin-Aspartat erwies sich hier bisher als unterlegen (4, 5). Während eine sekundäre medikamentöse Prophylaxe nach einer erstmaligen oHE-Episode empfohlen wird, soll eine primäre Prophylaxe nur bei Hochrisikopatienten durchgeführt werden. In Kombination mit Leberversagen sind therapierefraktäre, rezidivierende oHE-Episoden eine Indikation zur Lebertransplantation (5, 83).

4 Herleitung der Fragestellungen

4.1 Problematik der bisherigen Testverfahren

Die Anwendung etablierter Methoden zur mHE-Diagnostik benötigt entweder geschultes Personal (PHES, RBANS) und/oder unifunktionales Spezialgerät (CFF, CRT, EEG). Außerdem ist die Durchführung und Auswertung der genannten Tests zeitintensiv (je nach Test 10-30 Minuten). Der ICT und der Stroop-Test sind bereits computerbasiert und damit über das klinische Umfeld hinaus anwendbar und automatisiert auswertbar. Beide Tests sind aber durch ihren erhöhten Schwierigkeitsgrad limitiert (5). Dementsprechend werden die Tests bisher vor allem zu Studienzwecken durchgeführt (84). Eine routinemäßige Kontrolle von Patienten mit Leberzirrhose im ambulanten Umfeld ist kaum etabliert, obwohl sie bereits 1990 empfohlen wurde (7). Ein Selbsttest durch den Patienten ist aufgrund der genannten Limitationen mit den aktuell verfügbaren Tests nicht möglich. Dabei weisen Studienergebnisse ein erhöhtes Risiko für die Entwicklung einer oHE und ihrer Komplikationen aus einer vorbestehenden mHE nach und liefern damit Argumente für eine frühzeitige Diagnostik und Therapie (11, 85). Außerdem führt bereits eine mHE zu relevanten Einschränkungen von Alltagstätigkeiten. Eine resultierende Arbeitsunfähigkeit kann finanzielle Notsituationen mit sozialem Abstieg bedingen (9, 16). Es ist davon auszugehen, dass bei Prävalenzangaben der mHE bei Patienten mit Leberzirrhose von 20-80% mindestens jeder vierte Patient tatsächlich von einer mHE betroffen ist. Das wären in Deutschland ungefähr 250.000 Menschen, die nicht nur unter diskreten kognitiven Dysfunktionen leiden, sondern auch eine hohe Rate an Arbeitsunfähigkeit aufweisen und häufig Leistungen des Gesundheitssystems in Anspruch nehmen müssen (3). Auch aus ökonomischer Sicht sollte damit die Indikation für die Evaluation eines zuverlässigen, niedrighwelligen Screeninginstruments gegeben sein. Es konnte bereits 1993 gezeigt werden, dass die Anwendung einfacher psychometrischer Tests in der ambulanten Praxis möglich und sinnvoll ist (8). Die Leitlinien zur HE von 2014 empfehlen, dass idealerweise jeder Patient getestet werden sollte, der potenziell eine mHE haben könnte. Diese Empfehlung wird an gleicher Stelle aufgrund des derzeit damit verbundenen Kosten- und Zeitaufwands auf Patienten mit Beschwerden reduziert (5). Der Ansatz ist problematisch, da Patienten mit einer mHE subjektiv nicht notwendigerweise unter Beschwerden leiden. Ein in das häusliche Umfeld und die alltägliche Routine des Patienten integrierbarer Test könnte die Vorbehalte gegen eine

generelle Untersuchung von Risikopatienten beseitigen und jedem Betroffenen die Möglichkeit zur Selbstkontrolle bieten. Dieses Ziel wird mit dem neu entwickelten elektronischen Zahlenverbindungstest (eNCT) verfolgt. Der eNCT wird nachfolgend vorgestellt und auf seine Wertigkeit für die mHE-Diagnostik hin analysiert werden.

4.2 Lösungsansatz und Fragestellungen

Der eNCT soll durch ein vereinfachtes und elektronisches Testdesign leicht verständlich, im ambulanten und häuslichen Umfeld einsetzbar und automatisch auswertbar sein. Durch die regelmäßige Anwendung dieses zeitgerechten mHE-Screeningverfahrens sollen frühzeitig minimale kognitive Dysfunktionen festgestellt werden, um der Entwicklung einer oHE oder anderer Komplikationen vorbeugen zu können.

Zur erstmaligen Charakterisierung des eNCTs müssen zunächst Aussagen zu Gütekriterien psychodiagnostischer Tests getroffen werden. In einer weiteren Dissertation zum eNCT werden dafür die Resultate von Kontrollprobanden unter anderem hinsichtlich Objektivität und Reliabilität ausgewertet. Außerdem ist zu prüfen, ob der eNCT in der *Zielgruppe* (Patienten mit Leberzirrhose) zu einer zuverlässigen Detektion der mHE führt und die Voraussetzungen für einen zukünftigen Einsatz als mHE-Screeningtest im häuslichen Umfeld erfüllt. Das Ziel der vorliegenden Arbeit ist daher, durch Rekrutierung von Patienten mit Leberzirrhose eine erstmalige Einschätzung der diagnostischen Wertigkeit anhand verschiedener Gütekriterien in einem Patientenkollektiv und im direkten Vergleich mit etablierten Testverfahren vorzunehmen.

Daraus lassen sich folgende Fragestellungen für die vorliegende Arbeit ableiten:

1. Ist der eNCT ein zuverlässiges Instrument zur Einschätzung der kognitiven Leistungsfähigkeit von Leberzirrhosepatienten?
2. Lassen sich für das neue Testverfahren sinnvolle Grenzwerte definieren? Lassen sich klinische Einflussfaktoren auf die Testzeiten bestimmen?
3. Wie hoch ist die Testgenauigkeit (Sensitivität, Spezifität, Vorhersagewerte) des eNCTs bei der Detektion von Patienten mit einer mHE im Vergleich mit etablierten Testverfahren?
4. Rechtfertigen die Werte der Gütekriterien eine Anwendung des eNCTs als Screeningtest und welche Einschränkungen sind beim Einsatz des Tests zu beachten?

5 Patienten, Material und Methoden

5.1 Studienplanung und Rekrutierung von Probanden

Von November 2014 bis Mai 2015 wurden nach Genehmigung durch die Ethikkommission der Charité (Antragsnummer EA1/263/14) Probanden für die eNCT-Studie rekrutiert. Die Studie wurde unter der ärztlichen Leitung von Prof. Dr. med. Martin Stockmann durchgeführt und ist im *Deutschen Register Klinischer Studien* unter der Nummer DRKS00009432 registriert. Der eNCT wurde im Rahmen des FP7-EU-Projekts *d-LIVER* in Kooperation mit Projektpartnern der Universität Newcastle und dem Fraunhofer-Institut für Biomedizinische Technik (IBMT) entwickelt. Perspektivisch könnte ein solches elektronisches Testverfahren Bestandteil eines häuslichen Überwachungssystems für chronisch leberkranke Patienten sein, dessen Funktionalität und Nutzen durch das *d-LIVER*-Projekt evaluiert wurden. Die zentrale Integration aller Funktionen dieses Systems, inklusive der Ergebnisse des kognitiven Tests, der Verwaltung von aktuellen Medikamentenpläne, Blutwerten und Körpergewichtsdaten sowie einer Kommunikation mit betreuenden Ärzten, soll über einen Tablet-Computer geschehen (Abbildung 6D).

Im Sinne einer Querschnittsstudie wurden eingeschlossene lebergesunde Probanden und Patienten mit Leberzirrhose einmalig untersucht. Lediglich eine kleine Untergruppe wurde zu einem zweiten Zeitpunkt im Verlauf nachuntersucht. Die Patienten wurden über die Leberambulanz und die gastroenterologischen Stationen der Medizinischen Klinik mit Schwerpunkt Hepatologie und Gastroenterologie der Charité - Universitätsmedizin Berlin (Campus Virchow-Klinikum) rekrutiert. Darüber hinaus wurden lebergesunde Kontrollprobanden klinikintern über die Lipidambulanz des Arbeitsbereichs Lipidstoffwechsel, über das Centrum für Muskuloskelettale Chirurgie und aus dem persönlichen Umfeld der Doktoranden eingeschlossen. Für die Durchführung der Studie wurde ein Studienprotokoll mit einheitlichem Dokumentationsbogen (Case Report Form (CRF), siehe Anhang 2) erstellt, das den Vorschriften der Deklaration von Helsinki in seiner (zum Zeitpunkt der Studienplanung aktuellen) Version von 2013 entspricht (86).

5.2 Ein- und Ausschlusskriterien

Einschlusskriterien der lebergesunden Probanden waren ein Mindestalter von 18 Jahren und fehlende Hinweise auf Lebererkrankungen. Probanden des Patientenkollektivs mussten neben dem Mindestalter von 18 Jahren eine histologisch gesicherte Leberzir-

rhose haben oder es musste ein hochgradiger Verdacht auf eine Leberzirrhose bestehen, der bei Vorliegen von mindestens zwei der folgenden Befunde ausgesprochen wurde: Aszites, Splenomegalie (>12 cm), Ösophagusvarizen, HE, Thrombozytenzahl <120/nl, Zirrhoseaspekte in radiologischer Bildgebung oder erhöhter Bindegewebeanteil der Leber (>17,6 kPa im Fibroscan). Dabei wurde jegliche Genese der Lebererkrankung akzeptiert. Alle Studienteilnehmer mussten für einen Studieneinschluss nach Aufklärung über Inhalt, Ablauf und Zweck der Studie, mögliche Risiken und Freiwilligkeit der Studienteilnahme eine Einwilligungserklärung unterzeichnen.

Ausgeschlossen wurden Probanden, bei denen eine beeinträchtigte Testzeit nicht hepato-enzephalopathischer Ursache zu erwarten war. Das betraf Personen mit:

- floridem Alkohol- oder Drogenabusus im letzten halben Jahr vor Studieneinschluss,
- einer bestehenden neurologischen bzw. psychiatrischen Erkrankung, die zu einer eingeschränkten kognitiven Leistung führt,
- einem Resultat im Mini-Mental-Status-Test (MMST) von weniger als 26 Punkten (wenn nicht gleichzeitig ein Verdacht auf eine HE vorlag),
- einer bekannten Rot-Grün-Sehschwäche (für den Stroop-Test) oder
- aktueller sedierender Medikation mit Benzodiazepinen oder verwandter Substanzen.

Von den Ausschlusskriterien ausgenommen waren dabei die im Zentrum der Studie stehende HE sowie die regelmäßige Einnahme von Antidepressiva bei stabilem Krankheitsverlauf und von verschriebener Schlafmedikation (Quetiapine oder Melperon).

Außerdem konnten folgende Kriterien zu einer Verzögerung des Einschlussdatums führen: eine schwere HE mit fehlender Einwilligungsfähigkeit, ein operativer Eingriff unter Vollnarkose, ein endoskopischer (oder vergleichbarer) Eingriff unter Analgosedierung, die Anlage eines transjugulären intrahepatischen portosystemischen Stent-Shunts (TIPSS) oder eine neu aufgetretene Niereninsuffizienz mit unklar veränderter Entgiftungsleistung (87-91). Patienten, die im testfreien Intervall zwischen Einschluss und Nachbeobachtung lebertransplantiert wurden, wurden nachträglich ausgeschlossen.

5.3 Zusammensetzung des Probandenkollektivs

Insgesamt wurden initial 238 Probanden eingeschlossen, von denen 14 Probanden nachträglich von der Studienteilnahme ausgeschlossen werden mussten (Abbildung 4). Die Gründe waren ein Punktwert kleiner als 26 im MMST von Kontrollprobanden (N=5), florider Alkoholkonsum (N=3), Patienten mit fehlendem Nachweis einer Leberzirrhose

(N=2), neurologische Erkrankungen (N=2), Diazepam-Einnahme (N=1) und Lebertransplantation im testfreien Intervall (N=1). Die übrigen 224 Probanden verteilten sich auf vier Studienarme:

- (1) Patienten mit Leberzirrhose und ohne klinische Zeichen einer HE (*LZI*; N=88),
- (2) Patienten mit Leberzirrhose und klinischen Zeichen einer HE (*oHE*; N=12),
- (3) lebergesunde Probanden mit einem Durchschnittsalter vergleichbar mit der Patientengruppe ($\geq 39J$; N=61) und
- (4) lebergesunde Probanden mit einem Alter zwischen 18 und 38 Jahren ($< 39J$; N=63), wobei das Alter des jüngsten LZI-Patienten (39 Jahre) als Grenzwert für die Gruppeneinteilung diente (Abbildung 4).

Die junge Kontrollkohorte wurde hier nur für eine vollständige Übersicht aufgeführt. Die entsprechenden Daten sind für die Fragestellungen dieser Arbeit nicht relevant.

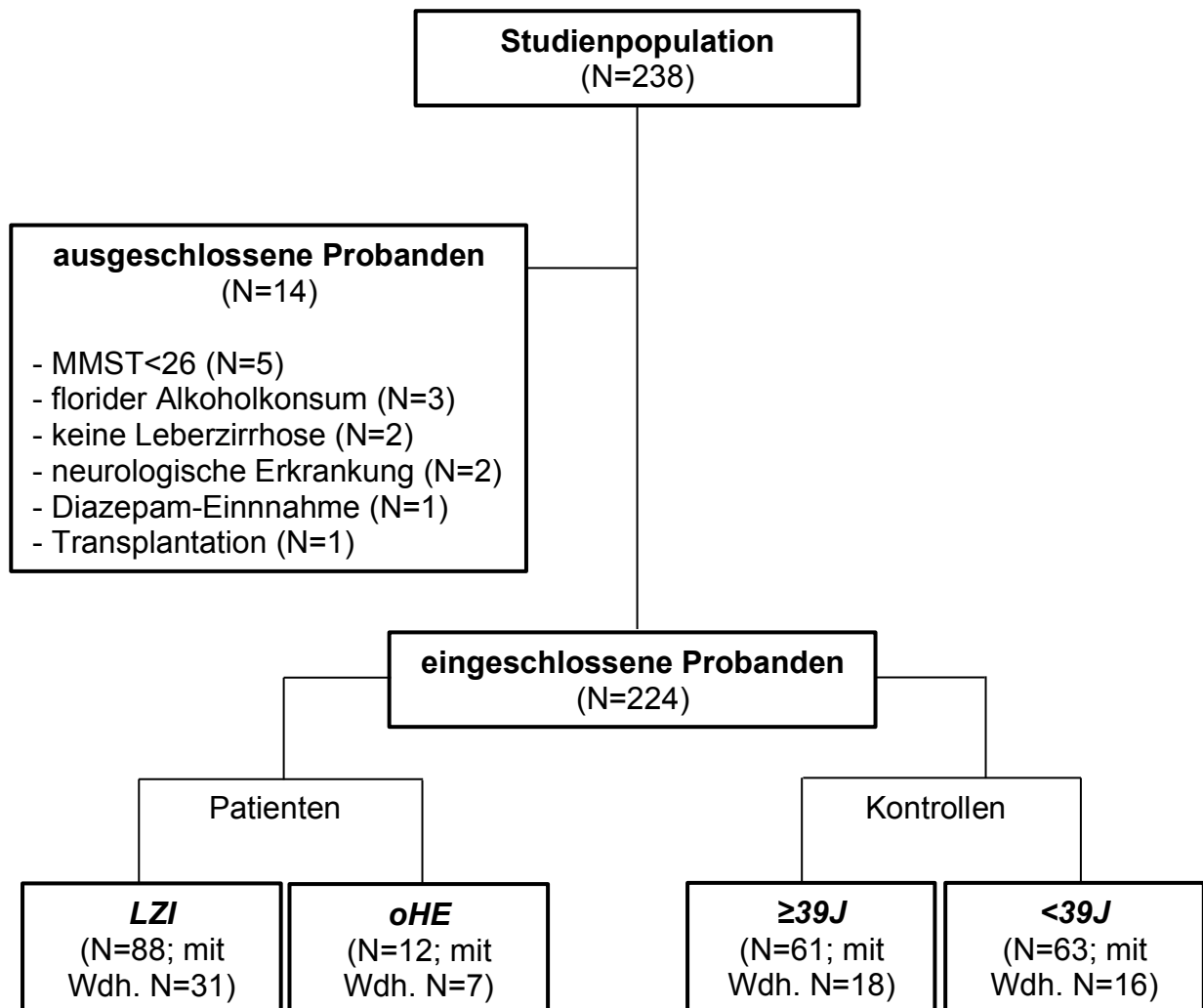


Abbildung 4: Aufteilung der Probanden in die verschiedenen Studienkohorten. (MMST: Mini-Mental-Status-Test; LZI: Patienten mit Leberzirrhose ohne klinische Zeichen einer HE; oHE: Patienten mit Leberzirrhose und klinischen Zeichen einer HE; $\geq 39J$: lebergesunde Probanden ab 39 Jahren; $< 39J$: lebergesunde Probanden zwischen 18 und 38 Jahren; mit Wdh.: mit Wiederholungstest im Verlauf)

5.4 Studienaufbau

Folgende Materialien wurden zur vollständigen Erhebung aller Daten benötigt: Aufklärungsbogen mit Einwilligungserklärung, CRF-Protokoll, MMST-Erhebungsbogen, PHES-Testbatterie, Stift, Stoppuhr, Tablet-Computer für den eNCT und den Stroop-Test, Blutabnahmesystem und Evaluationsbogen. Soweit medizinisch vertretbar wurden die Tests sitzend, am Tisch und in heller und ruhiger Umgebung durchgeführt. Nach Aufklärung, Einwilligung und Erhebung der demografischen und patientenbezogenen Daten wurde zunächst venös Blut entnommen und mit der Durchführung des modifizierten MMST nach Folstein fortgefahren, welcher zur Früherkennung von Demenzerkrankungen entwickelt wurde und durch 30 einzelne Items zeitliche und örtliche Orientierung, Merkfähigkeit und Kurzzeitgedächtnis, Sprach- und Textverständnis sowie exekutive Funktionen prüft (92, 93). Mithilfe von HESA und WHC wurde der Schweregrad der kognitiven Funktionseinschränkung eingeschätzt. Danach absolvierten die Probanden die psychometrischen Tests in der Reihenfolge PHES, eNCT und Stroop-Test. Zum Schluss wurde eine schriftliche Evaluation des neuen Testverfahrens erbeten.

Die beiden an der Studie beteiligten Doktoranden wurden in der Durchführung der Fragebögen und Tests trainiert, sodass ein standardisierter Versuchsablauf erreicht wurde. Für den PHES lieferte das entsprechende Handbuch ausformulierte Testinstruktionen. Jeder Test begann mit einem Übungsteil, durch den Verständnisfehler minimiert werden sollten. Fehler im Versuchsteil wurden vom Versuchsleiter umgehend korrigiert.

5.4.1 Labordiagnostik und Schweregrad der Leberzirrhose

Mithilfe der Blutentnahme und Messung von Serumkreatinin, Gesamtbilirubin, Serumalbumin, der INR und des Quick-Werts wurde bei Patienten der Grad der Lebererkrankung untersucht. Gemeinsam mit den weiteren Leberfunktionswerten Alkalische Phosphatase (AP), Gammaglutamyltransferase (GGT), Aspartat-Aminotransferase (AST), Alanin-Aminotransferase (ALT) und Faktor II wurden bei lebergesunden Probanden bisher unerkannte Lebererkrankungen ausgeschlossen. Allgemeine Ausschlusskriterien wie eine neu aufgetretene Niereninsuffizienz, frische Blutungen oder Elektrolytstörungen konnten durch die Erhebung der Kreatinin-, Natrium-, Kalium-, Hämoglobin (Hb)- und Thrombozytenwerte geprüft werden. Für die Bestimmung fehlerfreier Ammoniakwerte wurde die Blutentnahme an nicht bzw. möglichst kurzzeitig gestauten Venen durchgeführt und die Blutprobe gekühlt umgehend ans Labor geschickt (94). Eine Auf-

listung der gemessenen Laborwerte ist unter Anhang 1 zu finden, eine genaue Analyse relevanter Parameter für die statistische Auswertung des eNCTs folgt in Abschnitt 6.6.1. Zur Einschätzung des Schweregrads der Lebererkrankung wurden anschließend der MELD-Score (Model for End-stage Liver Disease) und der Child-Pugh-Score berechnet. Für die Erhebung des MELD-Scores wurde folgende Formel verwendet:

$$10 \times [0,957 \times \ln(\text{Serumkreatinin [mg/dl]}) + 0,378 \times \ln(\text{Gesamtbilirubin [mg/dl]}) + 1,12 \times \ln(\text{INR}) + 0,643]$$

Blutwerte kleiner als eins sind mit eins gleichzusetzen und für Patienten mit einer Dialyse in den letzten sieben Tagen ist der Kreatininwert auf 4 mg/dl anzupassen. Der MELD-Score kann Werte zwischen 6 und 40 annehmen (95, 96).

Der Child-Pugh-Score berechnet sich wie in Tabelle 2 gezeigt. Dabei gelten die gezeigten alternativen Bilirubin-Grenzwerte für cholestatische Lebererkrankungen wie die primär biliäre Zirrhose (PBC) und die primär sklerosierende Cholangitis (PSC). Alternativ zur geforderten Einteilung des Asziteschweregrads per Ultraschall wurde auch eine klinische Diagnostik zugelassen. Insgesamt kann also ein ganzzahliger Punktwert von 5 bis 15 erreicht werden, wobei Werte von 5-6 (Stadium A) als niedrig gelten und ein längeres mittleres Überleben prognostizieren als Werte von 7-9 (B) bzw. 10-15 (C) (97).

Ähnliches gilt für den MELD-Score, der mit einer höheren Mortalität bei steigenden Punktwerten assoziiert ist. Darüber hinaus findet der MELD-Score insbesondere zur Allokation von Spenderorganen für Lebertransplantationen Anwendung (95).

Tabelle 2: Child-Pugh-Score nach Pugh et al. (97)

Kriterium	1 Punkt	2 Punkte	3 Punkte
Gesamtbilirubin im Serum (mg/dl) (bei PSC/PBC)	<2,0 (<4,0)	2,0-3,0 (4,0-10,0)	>3,0 (>10,0)
Albumin im Serum (g/L)	>35	28-35	<28
Quick-Wert (%)	>70	40-70	<40
Aszites	Keiner	Leicht	Mittelgradig
Hepatische Enzephalopathie	Keine	Stadium 1-2	Stadium 3-4

5.4.2 Diagnostik der Enzephalopathie mittels HESA und WHC

Patienten mit Leberzirrhose wurden durch klinisch-neurologische Untersuchungen unter Verwendung der modifizierten WHC von Conn et al. (Tabelle 1) und des HESA (Abbil-

dung 5) auf Symptome einer oHE untersucht (4, 25, 62). Da die WHC auf subjektiven Einschätzungen basieren, der HESA aber in den Leitlinien zur HE von 2014 nicht erwähnt wird, wurden in der vorliegenden Arbeit beide Verfahren kombiniert und Patienten der oHE-Kohorte zugeteilt, wenn eines der beiden Testergebnisse auffällig war. Insgesamt wurde bei zwölf Patienten eine oHE diagnostiziert, die damit die Gruppe *oHE* bildeten. Von diesen zwölf Patienten zeigten sieben Patienten in beiden Verfahren die Diagnose oHE, zehn Patienten unter Anwendung der WHC (je 5x oHE Grad 1 und 2) und neun Patienten unter Anwendung des HESA (4x oHE Grad 1 und 5x oHE Grad 2).

5.4.3 Psychometrische Diagnostik

Die psychometrischen Tests beinhalteten den PHES, eNCT und Stroop-Test.

5.4.3.1 Psychometric Hepatic Encephalopathy Score (PHES)

Der PHES kam wie in Abschnitt 3.5.4 beschrieben und ohne Festlegung maximaler Testzeiten zur Anwendung. Änderungen betrafen die Reihenfolge der Tests (1. NCT-A, 2. NCT-B, 3. DST, 4. LNT, 5. SDT) und die Durchführungshäufigkeit des NCT-A (3x). Damit sollte die Vergleichbarkeit mit dem eNCT verbessert werden, der ebenso dreimal absolviert werden sollte und für den der NCT-A als Vorlage diente. Für die drei Durchgänge des NCT-A wurden vor Studienbeginn drei Paralleltestversionen mit unterschiedlichen Zahlenanordnungen ausgewählt, die von jedem Probanden in der gleichen Reihenfolge bearbeitet wurden. Abgesehen von leichten Abweichungen der Kopiervorlagen im Vergleich zur Originalversion (z. B. Anordnung der Zahlen im NCT-A, Symbole im DST), die bei Studienbeginn noch nicht unmittelbar verfügbar war, wurden die Tests entsprechend der präzisen Vorgaben des Handbuchs durchgeführt und ausgewertet. Die Berechnung der Ergebnisse erfolgte dem Alter entsprechend nach dem in Abschnitt 3.5.4 vorgestellten Punktesystem des PHES-Handbuchs (68). Aus den drei Testzeiten des NCT-A wurde dafür der Mittelwert gebildet.

5.4.3.2 Elektronischer Zahlenverbindungstest (eNCT)

Der eNCT ist eine Tablet-Computerbasierte Version des papierbasierten NCT-A und wurde in Kooperation mit den Projektpartnern des Fraunhofer-Instituts (IBMT) und der Universität Newcastle zu Forschungszwecken in Eigenherstellung entwickelt.

Auf einem berührungsempfindlichen Bildschirm werden die Zahlen von 1 bis 25 zufällig

HESA (Hepatic Encephalopathy Scoring Algorithm)			
Grade	Klinisch	Neurophysiologisch	Punktevergabe
IV	<input type="checkbox"/> Keine verbale Antwort <input type="checkbox"/> Keine Augenöffnung <input type="checkbox"/> Keine Reaktion auf einfache Aufforderung	<input type="checkbox"/> Nicht durchführbar	<input type="checkbox"/> Alle 3 klinischen Punkte positiv
III	<input type="checkbox"/> Somnolenz <input type="checkbox"/> Konfusion <input type="checkbox"/> Desorientiert zum Ort <input type="checkbox"/> Bizarres Verhalten/ Aggressivität <input type="checkbox"/> Klonus/ Babinsky <input type="checkbox"/> Muskelsteifigkeit/Nystagmus	<input type="checkbox"/> Mentaler Kontroll-Test = 0	<input type="checkbox"/> Zumindest 3 Punkte positiv
Mentaler Kontroll-Test:			
		<i>Kein Fehler</i>	<i>Im Zeitlimit</i>
	Zurückzählen von 20 bis 1 (Zeitlimit 30s)	<input type="checkbox"/> 1 Punkt	<input type="checkbox"/> 1 Punkt
	Vorwärtszählen von 1 bis 100 in Dreier Schritten (Zeitlimit 45s)	<input type="checkbox"/> 1 Punkt	<input type="checkbox"/> 1 Punkt
	Alphabet aufsagen (Zeitlimit: 30s)	<input type="checkbox"/> 1 Punkt	<input type="checkbox"/> 1 Punkt
Bei < 4 Punkten „Konfusion“ ankreuzen, ohne Fehler, aber überm Zeitlimit „Verlangsamte Antwort“ ankreuzen.			
Tag, „Welches Datum haben wir heute?“ Jahr „Welches Jahr haben wir heute?“ „Haben wir 2013?“ und Ort abfragen: Wenn Ort nicht korrekt, dann „Desorientiert zum Ort“ ankreuzen.			
II	<input type="checkbox"/> Lethargisch <input type="checkbox"/> Desorientiert zur Zeit <input type="checkbox"/> Verwaschene Sprache <input type="checkbox"/> Verstärkte Reflexe <input type="checkbox"/> Unpassendes Verhalten	<input type="checkbox"/> Verlangsamte Antwort <input type="checkbox"/> Angstniveau größer 4 <input type="checkbox"/> Probleme mit dem Kurzzeitgedächtnis <input type="checkbox"/> Einfache Berechnungen	<input type="checkbox"/> Mindestens 2 positive klinische und 3 Neurophysiologische Punkte
Probleme mit dem Kurzzeitgedächtnis: 100% (kein Fehler) normal			
Auf einer Liste stehen 12 Wörter. 6 davon werden vorgelesen und der Patient gebeten diese zu wiederholen. Die Reihenfolge ist dabei nicht wichtig! Anschließend soll der Patient die vorgelesenen Wörter auf der Liste wiedererkennen. Dazu liest der Untersucher dem Patienten alle 12 Wörter vor. Bei den 6 richtigen Wörtern soll der Patient „Ja“ sagen, bei den falschen 6 Wörtern mit „Nein“ antworten.			
1. Sonne 2. Krankenhaus 3. Fußball 4. Radio 5. Löffel 6. Nase 7. Schule 8. Katze 9. Bild 10. Auto 11. Regenschirm 12. Blume			
Einfache Berechnungen: Drei mal Ein-Ziffer-Addition oder Subtraktion. Kein Fehler normal.			
Addition: (I) 6 + 9=15 <input type="checkbox"/> (II) 4 + 7=11 <input type="checkbox"/> (III) 5 + 8=13 <input type="checkbox"/>			
Subtraktion: (I) 7 – 4=3 <input type="checkbox"/> (II) 9 – 6=3 <input type="checkbox"/> (III) 8 – 5=3 <input type="checkbox"/>			
I	<input type="checkbox"/> Schlafprobleme <input type="checkbox"/> Tremor	<input type="checkbox"/> Komplexe Berechnungen <input type="checkbox"/> Konstruktion Fähigkeit <input type="checkbox"/> Verkürzte Aufmerksamkeitsspanne <input type="checkbox"/> Depression größer 4	<input type="checkbox"/> Zumindest 4 Punkte positiv
Komplexe Berechnungen: Drei mal Zwei-Ziffer Addition oder simple Multiplikation/Division. Kein Fehler normal.			
Addition: (I) 16 + 18=34 <input type="checkbox"/> (II) 12 + 17=29 <input type="checkbox"/> (III) 14 + 19=33 <input type="checkbox"/>			
Multiplikation: (I) 6 x 8=48 <input type="checkbox"/> (II) 4 x 9=36 <input type="checkbox"/> (III) 7 x 5= 35 <input type="checkbox"/>			
Division: (I) 16/4=4 <input type="checkbox"/> (II) 12/3= 4 <input type="checkbox"/> (III) 15/5=3 <input type="checkbox"/>			
Konstruktion Fähigkeit:			
Der Patient schreibt seinen Vor- und Zunamen auf ein Blatt Papier. Korrekt und lesbar! <input type="checkbox"/>			
Verkürzte Aufmerksamkeitsspanne: Länger werdende Ziffernreihen (2-9) wiederholen lassen. 5 ist normal.			
(I) 6 9 3 7 2 <input type="checkbox"/>			
(II) 3 8 5 7 2 6 <input type="checkbox"/>			
(III) 5 8 3 1 4 9 6 7 <input type="checkbox"/>			

Abbildung 5: Der in der eNCT-Studie verwendete HESA-Erhebungsbogen

angeordnet abgebildet. Die Aufgabe für den Probanden besteht darin, die Zahlen in der richtigen Zählreihenfolge nacheinander anzutippen. Bei Berührung der korrekten Ziffer ändert der die Ziffer umgebende Kreis die Farbe von grau zu gelb und der Proband muss die nächste Ziffer aufsuchen. Es kann immer nur die aktuell korrekte Zahl markiert werden. Mit Berührung der ersten Zahl (=1) wird automatisch die Messung der Zeit begonnen, die mit Berührung der letzten Zahl (=25) gestoppt und auf dem Bildschirm angezeigt wird. Ein Versuchsdurchlauf besteht aus einem Übungsteil mit fünf Ziffern und ohne Zeitmessung (Abbildung 6A) sowie drei verschiedenen Durchgängen von 1 bis 25 mit Zeitmessung (Abbildung 6B und C). Die drei Testzeiten werden addiert und der Mittelwert gebildet. Die sich ergebende Zeit ist das Resultat des eNCTs. Prinzipiell bietet der eNCT nahezu unendlich viele Möglichkeiten, die 25 Zahlen auf dem Bildschirm anzuordnen. Im Rahmen der Studie wurden aber entsprechend der drei verschiedenen NCT-A-Testbögen nur drei verschiedene eNCT-Parallelversionen verwendet, um die Versuchsbedingungen für jeden Probanden möglichst konstant zu halten.

Der eNCT wurde auf Tablet-Computern mit 10“-Bildschirmdiagonale (Aspire Switch 10, Acer, Taipeh, Taiwan) und 18“-Bildschirmdiagonale (XPS One, DELL Inc., Round Rock, TX, USA) durchgeführt, um den Einfluss der Bildschirmgröße auf die Testzeiten zu untersuchen. In Abbildung 7, die der Publikation von Wuensch et al. entnommen ist, zeigt der schwarze Pfeil den eNCT 10“ und der blaue Pfeil den eNCT 18“ (98).

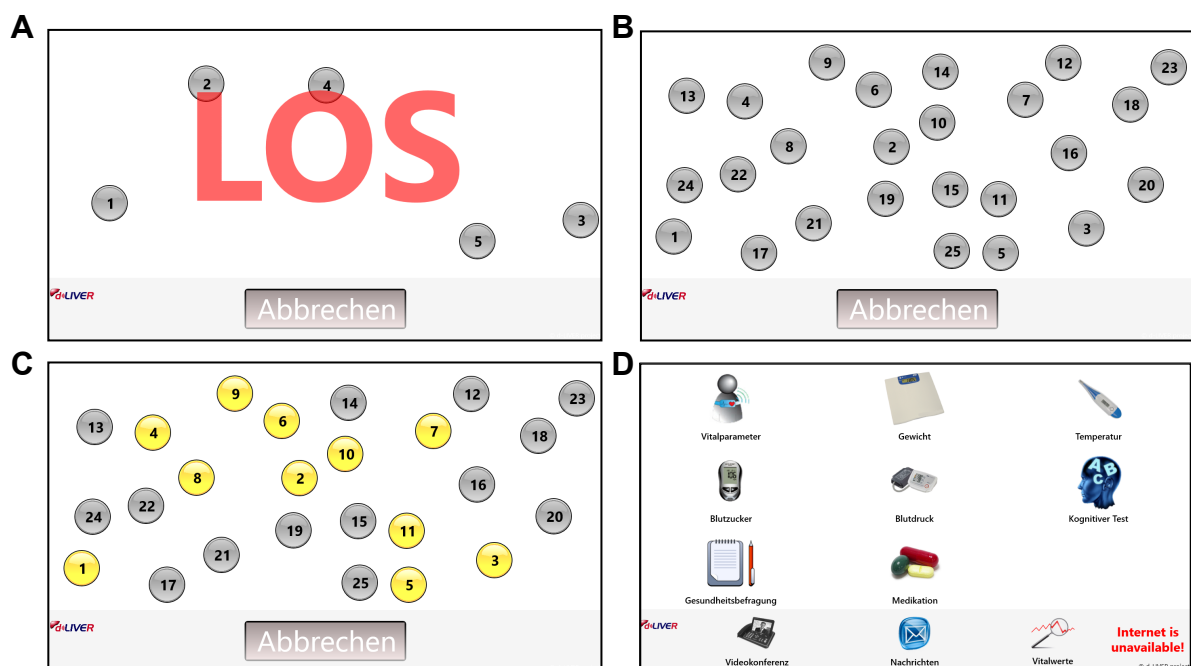


Abbildung 6: Bildschirmpräsentation des eNCTs. Übungsteil (A), vor Versuchsbeginn (B), während eines Versuchs (C), zentrale Verwaltungsoberfläche mit diversen Wahlmöglichkeiten (D).

Bei der Entwicklung des eNCTs diente der NCT-A als Vorlage. Dieser Papiertest half bereits in den Fünfziger Jahren des 20. Jahrhunderts bei der Evaluation von Hirnschäden (99). In der Anwendung zur mHE-Diagnostik kann allerdings erst durch die Kombination mit anderen psychometrischen Tests eine ausreichende Genauigkeit erzielt werden, sodass von Ferenci et al. zunächst eine Durchführung von mindestens zwei Papiertests der PHES-Testbatterie verlangt wurde und mit verbesserter Normierung sich die Durchführung der gesamten PHES-Testbatterie zur mHE-Diagnostik etablierte (5, 6, 67). Dennoch wurde bei der Entwicklung des eNCTs gezielt nur ein einziger, simpler und niedrigschwelliger Papiertest als Grundlage für den elektronischen Test verwendet, um eine regelmäßige Durchführung unabhängig von Sprache und betreuendem Fachpersonal im Gegensatz zum PHES und anderen Testverfahren der mHE-Diagnostik auch außerhalb von Studienzwecken zu ermöglichen (84). Da der eNCT als Screeningtest in erster Instanz Risikopatienten identifizieren soll, ohne etablierte Testverfahren zu ersetzen und bei der mHE zuerst feinmotorische Handlungen beeinträchtigt sind, die in den Studien zur Validierung des PHES am ehesten durch Zahlenverbindungstests und den SDT messbar waren, galten diese Tests als geeignete Modelle bei der Entwicklung des eNCTs, der neben Elementen des NCTs durch das *Tippen* auf Zahlen auch Elemente des SDTs besitzt (67, 69, 100). Als endgültiger Testwert wurde der Mittelwert aus drei Versuchen unterschiedlicher Parallelversionen festgelegt, da hierdurch eine bessere Reliabilität zu erwarten ist als durch einzelne Messwerte (101).

Die Instruktionen für den Probanden lauteten:

„Sie sehen auf diesem Bildschirm gleich die Zahlen 1 bis 5. Es handelt sich dabei um einen Übungsdurchlauf ohne Zeitmessung. Bitte versuchen Sie die Zahlen der Reihe nach so schnell wie möglich anzutippen. Tippen Sie die richtige Zahl an, ändert der umgebende Kreis die Farbe von grau zu gelb. Tippen Sie die falsche Zahl an, passiert nichts und Sie können erst fortfahren, wenn die ausgelassene Zahl berührt wurde. Im Anschluss an den Übungsdurchlauf und nach Klärung etwaiger Fragen folgen drei Durchgänge mit den Zahlen 1 bis 25, die Sie bitte in der gleichen Art und Weise schnellstmöglich berühren. Die Messung der Zeit beginnt jeweils automatisch mit Berührung der 1 und endet mit Berührung der 25. Nach erfolgreicher Beendigung jedes einzelnen Testdurchgangs wird die entsprechende Testzeit auf dem Bildschirm angezeigt. Durch erneutes Tippen auf den Bildschirm erscheint der nächste Testdurchgang. Haben Sie alles verstanden? Dann beginnen Sie bitte jetzt.“

5.4.3.3 Stroop-Test

Der ebenfalls Tablet-Computerbasierte Stroop-Test wurde auf einem iPad Air 2 (Apple Inc., Cupertino, CA, USA) durchgeführt (grüner Stern in Abbildung 7).

Damit sollten die eNCT-Resultate mit einem bereits etablierten elektronischen Diagnostest verglichen werden.

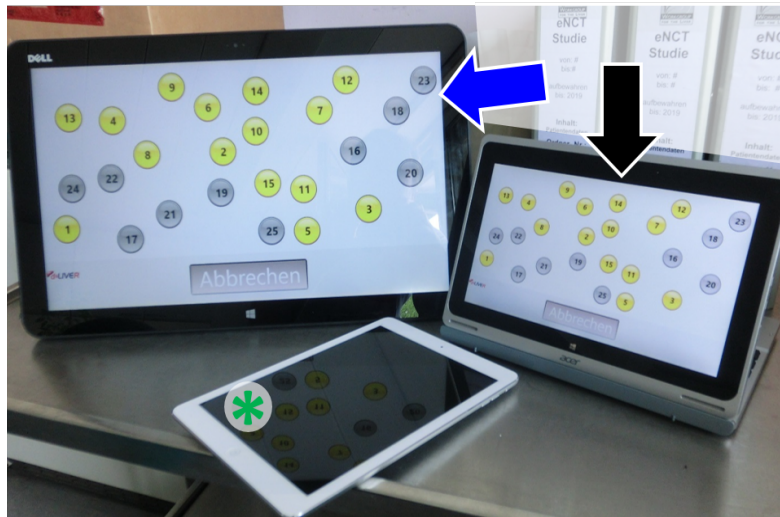


Abbildung 7: Die verwendeten Tablet-Computer für den eNCT (schwarzer Pfeil: 10“; blauer Pfeil: 18“) und den Stroop-Test (grüner Stern).

5.5 Statistische Auswertung

Die statistische Auswertung der Daten erfolgte mithilfe von Version 22 des Computerprogramms *Statistical Package for the Social Sciences* (SPSS, IBM, Armonk, New York, USA). Die in dieser Arbeit angewandten statistischen Tests wurden auf ihre Eignung für die jeweilige Analyse durch das Institut für Biometrie und Klinische Epidemiologie der Charité – Universitätsmedizin Berlin abschließend geprüft. Signifikanztests wurden entsprechend einer vorhandenen oder fehlenden Normalverteilung der Daten ausgewählt, wobei für die Annahme einer Normalverteilung folgende Kriterien zutreffen mussten: Keine Signifikanz im Shapiro-Wilk-Test, symmetrische und eingipflige Histogramme (Abbildung 8) und eine Schiefe zwischen -1 und +1, die nicht signifikant von 0 abweicht (102, 103). Entsprechend des Ergebnisses wurden die Unterschiede metrischer, normalverteilter Daten durch den t-Test oder bei mehr als zwei Stichproben durch eine Varianzanalyse (ANOVA) untersucht, Zusammenhänge wurden mittels Pearson-Korrelation geprüft. Bei metrischen, nicht-normalverteilten oder ordinalen Da-

ten kamen für unabhängige Stichproben der Mann-Whitney-U-Test (bei >2 Stichproben: Kruskal-Wallis-Test) und für abhängige Stichproben der Wilcoxon-Test als Signifikanztests sowie die Spearman-Korrelation zur Analyse von Zusammenhängen zur Anwendung. Der Chi-Quadrat-Test (bei unabhängigen Stichproben) und der Vorzeichentest (bei abhängigen Stichproben) wurden bei kategoriellen Daten verwendet (101).

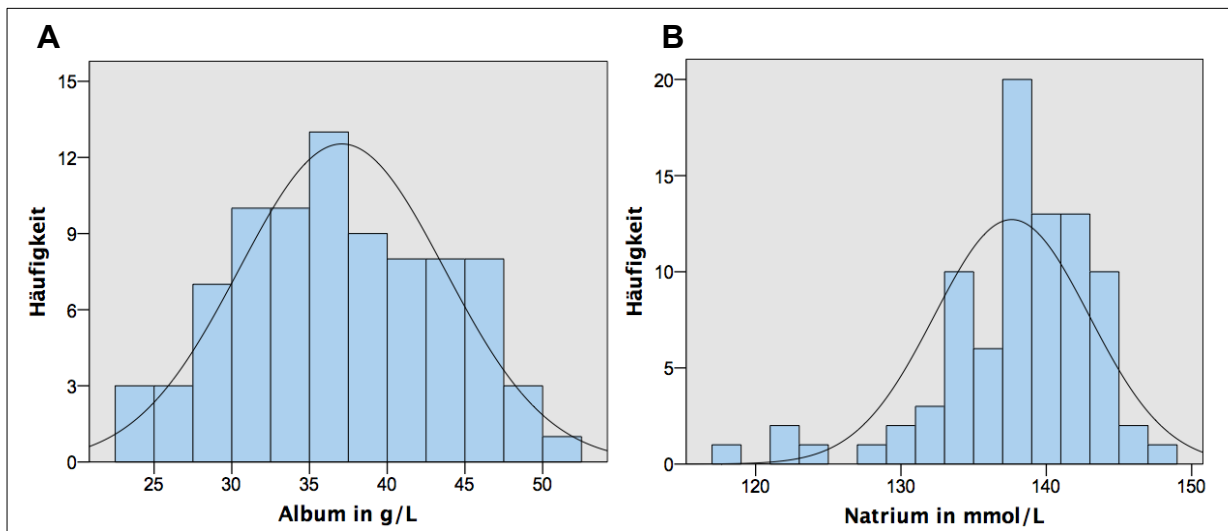


Abbildung 8: Vergleich von Normalverteilung (A) und schiefer Verteilung (B) – Beispielhaft für die LZI-Kohorte konnte für die Verteilung der Albuminwerte (A) eine Normalverteilung gezeigt werden (symmetrisches Histogramm, Schiefe=-0,050, p-Wert im Shapiro-Wilk-Test=0,263), während die Natriumwerte (B) linksschief verteilt sind (Schiefe=-1,331, p-Wert im Shapiro-Wilk-Test<0,001).

Statistisch signifikante Unterschiede wurden für *p*-Werte kleiner als ein *Signifikanzniveau* α von 0,05 angenommen, das einer Wahrscheinlichkeit unter 5% entspricht, die Nullhypothese fälschlicherweise zu verwerfen (=Fehler 1. Art). Werden Lageunterschiede von Testwerten für mehr als zwei Stichproben gleichzeitig untersucht (z. B. ANOVA) und damit eine Hypothese durch die Kombination mehrerer einzelner Signifikanztests geprüft, so wird ein Fehler 1. Art wahrscheinlicher. Dieses Problem des *multiplen Testens* wird durch eine Fehlerkorrektur gelöst, für die bei einer ANOVA neben der Normalverteilung auch die Varianzhomogenität mit dem Levene-Test überprüft wurde. Entsprechend der unbalancierten Stichproben wurden der Scheffé-Test bei Varianzhomogenität und der Games-Howell-Test bei Varianzheterogenität durchgeführt. Vergleichbare post-hoc-Verfahren wurden mithilfe von SPSS auch für den Kruskal-Wallis-Test angewandt (104, 105). In Abbildungen wurde für *p*-Werte <0,05 folgende Schreibweise verwendet: $p < 0,05 = *$, $p < 0,01 = **$, $p < 0,001 = ***$. Dabei wird auf eine Wertung der Signifikanz verzichtet, da von niedrigeren *p*-Werten nur bedingt auf stärkere Effekte geschlossen werden kann (106). *Korrelationskoeffizienten* *r* wurden ab 0,7 als hoch und ab 0,9

als sehr hoch interpretiert. Die in den Streudiagrammen der Korrelationsanalysen dargestellten Regressionsgeraden entsprechen der mathematischen Funktion, deren Graph in der Summe den geringsten Abstand zu allen Punkten hat (101).

Eine häufige Darstellungsweise von Ergebnissen ist in dieser Arbeit, wie erstmals in Abbildung 11 gezeigt, die so genannte Kastengrafik (engl.: *Box-Whisker-Plot*). Sie zeigt den Median (schwarze Querlinie innerhalb der blauen Kästen), den Interquartilsabstand IQR (Höhe der Kästen) zwischen dem 25%-Perzentil (untere Begrenzung der Kästen) und dem 75%-Perzentil (obere Begrenzung der Kästen), über oder unter dem IQR liegende Werte bis maximal zum 1,5-fachen IQR („Antennen“ der Kästen) und die darüber hinausgehenden Ausreißer (Kreise: 1,5 bis 3-facher IQR) und Extremwerte (Kreuze: >3-facher IQR).

5.5.1 Statistische Anforderungen an diagnostische Validierungsstudien

Diagnostische Validierungsstudien, zu denen die eNCT-Studie zu zählen ist, versuchen die Güte eines Testverfahrens mithilfe verschiedener Kriterien zu beschreiben und bestehen aus mehreren Phasen.

Unter den Hauptgütekriterien werden Objektivität, Reliabilität und Validität zusammengefasst. Als Nebengütekriterien werden Normierung, Vergleichbarkeit, Ökonomie, Nützlichkeit, Zumutbarkeit, Fairness und Nicht-Verfälschbarkeit betrachtet (107). Diese Kriterien sollten zur Beurteilung der diagnostischen Wertigkeit des eNCTs untersucht werden. In einer ersten Phase der eNCT-Studie wurden daher an Kontrollprobanden und Patienten mit Leberzirrhose grundlegende Charakteristika des eNCTs, wie Anwendbarkeit, Objektivität und Reproduzierbarkeit von Testergebnissen, mögliche Störfaktoren, Aufwand für das Erlernen des Testablaufs und die Akzeptanz des Tests durch die Studienteilnehmer überprüft (98, 108). Diese Aspekte sind im Detail Gegenstand einer weiteren Dissertation zum eNCT. Im Sinne der zweiten Phase diagnostischer Validierungsstudien wird in dieser Promotionsarbeit der diagnostische Nutzen des eNCTs zur mHE-Detektion im Fokus stehen (vgl. Abschnitt 4.2). Dazu zählt die Festlegung eines eNCT-Grenzwerts zur mHE-Diagnostik, die durch eine Receiver Operating Characteristic (ROC) Analyse mit Bestimmung der Fläche unter der Kurve (AUC) als ein Maß der Testgenauigkeit erfolgen kann. Weitere Kennzeichen der Genauigkeit wie Sensitivität, Spezifität, Vorhersagewerte und Likelihood-Quotienten sind über Kreuztabellen im Ver-

gleich mit einem etablierten Testverfahren (hier: PHES) zu berechnen. Außerdem sollten Aussagen über die Hauptgütekriterien Reliabilität und Validität getroffen werden. Analysen von Testeigenschaften und Einflussfaktoren können Bestandteil erster und zweiter Phasen diagnostischer Validierungsstudien sein (108, 109).

5.5.1.1 Analyse der eNCT-Testeigenschaften

Psychometrische Testverfahren unterscheiden sich hinsichtlich der Aufgabenstellung und damit teilweise auch bezüglich der beanspruchten kognitiven Domänen. Für Aussagen über die Eigenschaften des eNCTs wurde der Test durch Korrelationsanalysen mit den PHES-Einzeltests und dem Stroop-Test verglichen. Zusätzlich wurde die Stärke des Effekts einer mHE-Diagnose auf die Leistungsfähigkeit in den Tests analysiert und im Vergleich bewertet. Dafür wurde das *Effektstärkemaß* r_e für nicht-parametrische Daten aus dem Quotienten der standardisierten *Teststatistik* z und der Wurzel aus der *Fallzahl* N berechnet. Werte ab 0,1 zeigen einen kleinen, Werte ab 0,3 einen mittleren und Werte ab 0,5 einen großen Effekt (105, 110).

5.5.1.2 Analyse der Reliabilität

Als Maß der Interrater-Reliabilität zwischen eNCT und PHES wurde der *Kappa-Wert* (κ) berechnet, der die gemessene Übereinstimmung in ein Verhältnis zur erwarteten zufälligen Übereinstimmung setzt. Werte ab 0,41 definieren eine moderate, Werte ab 0,61 eine sehr gute und Werte ab 0,80 eine ausgezeichnete Zuverlässigkeit des eNCTs bezogen auf den PHES-Vergleichstest (111).

Die Wiederholungsgenauigkeit kann durch die Intraklassen-Korrelation beurteilt werden, bei der im Gegensatz zu den klassischen Zusammenhangsmaßen nach Pearson und Spearman der Einfluss systematischer und zufälliger Fehler durch Transformation der Daten (z. B. Vertauschen der Untersuchungszeitpunkte) reduziert wird. Der *Intraklassen-Korrelationskoeffizient* (IKK) lässt sich interpretieren als der Anteil der wahren Varianz der Messwerte zwischen verschiedenen Messungen an der Gesamtvarianz und kann Werte zwischen null (ausschließlich Messfehler) und eins (keine Messfehler) annehmen. Dabei werden Werte ab 0,9 bei einem unteren Wert des 95%-Konfidenzintervalls oberhalb von 0,75 als sehr gut bezeichnet. Für die Berechnung des IKK wurde in den Einstellungen von SPSS die gemischte Zwei-Wege-Analyse verwendet und auf absolute Übereinstimmung getestet (101, 111, 112).

5.5.1.3 Analyse der Validität

Es lassen sich Inhalts-, Konstrukt- und Kriteriumsvalidität unterscheiden. Zur Beurteilung der *Inhaltsvalidität* muss die Frage auf Grundlage von Expertenmeinungen beantwortet werden, ob die Messwerte inhaltlich das wiedergeben können, was der Test erfassen soll. Eine hohe *Konstruktvalidität* liegt vor, wenn die Messwerte nur mit Messwerten inhaltsähnlicher Testverfahren zusammenhängen. Neben theoretischen Überlegungen können Korrelationsanalysen zur Bewertung beitragen.

Die *Kriteriumsvalidität* gibt Aufschluss darüber, ob die anhand der Testwerte getroffenen Aussagen mit der Realität übereinstimmen. Dafür müssen eNCT und PHES von einer Person zu einem Zeitpunkt hintereinander absolviert werden und die Stärke der Korrelation beider Testverfahren überprüft werden. Als Validitätsmaß wird die Fläche unter der Kurve (AUC), die bei Werten ab 0,9 als hoch definiert wird ($\geq 0,7$ = moderat, $\geq 0,5$ = niedrig), der folgend erläuterten ROC-Analyse empfohlen (108, 113).

Als Teil der Kriteriumsvalidität wurde außerdem noch die Veränderungssensitivität untersucht, mit der die Fähigkeit eines Testverfahrens beschrieben werden kann, Veränderungen des zu messenden Kriteriums im zeitlichen Verlauf zu detektieren (111).

5.5.1.4 Überprüfung der eNCT-Testgenauigkeit

Zur Grenzwertoptimierung des neuen elektronischen Zahlenverbindungstests wurde eine ROC-Analyse durchgeführt. Hierbei werden alle möglichen Testzeiten im Bereich der erzielten Testzeiten als Grenzwert für ein pathologisches Resultat geprüft und Sensitivität und Spezifität dieser Grenzwerte berechnet, hier bezogen auf das Ergebnis im PHES. Die Ergebnisse der ROC-Analyse werden in einer Kurve abgebildet (Abbildung 18), bei der Sensitivität gegen 1-Spezifität aufgetragen ist und die mit einer eingezeichneten diagonalen Linie verglichen werden kann, die einen Test ohne jegliche diskriminatorische Fähigkeiten beschreibt (101). Aus den theoretischen Berechnungen sollte in dieser Arbeit der Grenzwert mit der höchsten Sensitivität bestimmt werden, für den die Spezifität nicht unter 70% fällt.

Für den festgelegten Grenzwert wurden neben Sensitivität und Spezifität noch folgende Kriterien zur Bestimmung der Testgenauigkeit berechnet: Die Vorhersagewerte positiv-prädiktiver Wert (PPW) und negativ-prädiktiver Wert (NPW) sowie die Präzision (richtig-positive Resultate + richtig-negative Resultate). PPW und NPW sind prävalenzabhän-

gig, ein korrektes Vorgehen bezieht die Vortestwahrscheinlichkeit (Prä.) für eine Krankheit mithilfe der Bayes-Formel in die Berechnungen ein (108). Demnach gilt:

$$1. PPW = (Sensitivität \times Prä.) / [Sensitivität \times Prä. + (1 - Spezifität) \times (1 - Prä.)]$$

$$2. NPW = Spezifität \times (1 - Prä.) / [Spezifität \times (1 - Prä.) + (1 - Sensitivität) \times Prä.]$$

Ermittelt wurden außerdem die Wahrscheinlichkeitsverhältnisse (Likelihood-Quotient bzw. Ratio = LR) für ein positives ($LR(+)$; >10 : sehr hohe Evidenz, >5 : hohe Evidenz) und für ein negatives Testergebnis ($LR(-)$; $<0,1$: sehr hohe Evidenz, $<0,2$: hohe Evidenz) (114). Die diagnostische Odds-Ratio (DOR) beschreibt „das Verhältnis der Chance eines positiven Testergebnisses bei erkrankten Personen zur Chance eines positiven Testergebnisses bei nicht erkrankten Personen“ (115). Je höher die DOR über eins liegt, desto besser kann der eNCT zwischen *mHE* und *keine mHE* diskriminieren. Bei einer DOR von eins hingegen hätte der eNCT keine diagnostische Trennfähigkeit.

5.5.1.5 Überprüfung möglicher Einflussfaktoren auf eNCT-Testzeiten

Durch eine lineare, schrittweise Regressionsanalyse wurde untersucht, welche patientenspezifischen Faktoren neben dem HE-Status die Testzeiten zusätzlich beeinflussen. Patienten mit *mHE* oder *oHE* wurden ausgeschlossen, um das Regressionsmodell nicht durch pathologische Werte zu beeinflussen. Voraussetzungen für aussagekräftige Ergebnisse linearer Regressionsanalysen mussten überprüft werden:

- Durch Visualisierung der Daten als Streudiagramme (Abbildungen 21-22) wurden *Nicht-Linearität* und *Heteroskedasizität* (= Varianzheterogenität) ausgeschlossen.
- Zum Auffinden unerwünschter *Autokorrelationen*, also Korrelationen zwischen aufeinanderfolgenden Residualgrößen, wurde der Durbin-Watson-Test durchgeführt (wobei $d=2$: keine Autokorrelation; $d=0$: perfekte positive Autokorrelation; $d=4$: perfekte negative Autokorrelation).
- *Multikollinearität* (= lineare Zusammenhänge zwischen unabhängigen Variablen) wurde mittels Varianzinflationsfaktor (VIF) geprüft (VIF=1: keine Korrelation; VIF<5: mäßige Korrelation; VIF>5: starke Korrelation). Starke Zusammenhänge unabhängiger Variablen vermindern die Aussagekraft von Regressionsanalysen.

Die zu berichtenden Ergebnisse einer Regressionsanalyse bestehen aus:

- dem Regressionsmodell (mit einer Konstanten und Variablenkoeffizienten)
- dem *Bestimmtheitsmaß* R^2 (= durch Modell erklärter Anteil an Gesamtvarianz)
- Beta-Werten (= relative Einflussstärke der einzelnen Variablen) (104).

6 Ergebnisse

6.1 Demografische Charakterisierung des Probandenkollektivs

Alle Probanden wurden vor Beginn der Testreihe mithilfe des CRF-Dokumentationsbogens einer ausführlichen und standardisierten Anamnese unterzogen, die sowohl biografische Fragen als auch krankheitsspezifische Aspekte abdeckte (siehe Anhang 2). Im nächsten Schritt wurde eine grobe Einschätzung der kognitiven Leistungsfähigkeit aller Probanden unter Verwendung des modifizierten MMSTs nach Folstein erhoben. Damit sollten unbekannte neurologische Defizite von Probanden aufgedeckt werden, was bei einem Resultat im MMST von weniger als 26 Punkten zum Ausschluss aus der Studie führte. Bei Patienten ist dieser Grenzwert nur mit Einschränkungen anwendbar, da eine vorliegende HE die Leistung im MMST beeinflusst. In diesen Fällen erfolgte eine differenzialdiagnostische Abklärung durch die Ergebnisse der klinischen Diagnostik und die medizinische Vorgeschichte.

LZI-Kohorte: Es konnten 88 Patienten mit Leberzirrhose und ohne oHE-Symptomatik eingeschlossen werden. Der Anteil männlicher Studienteilnehmer lag bei 61,4% (N=54), das mittlere Alter bei 59 Jahren mit einer Standardabweichung (SD) von 8 Jahren und einer Altersspanne zwischen 39 und 76 Jahren. Es wurden durchschnittlich 10 (Mittelwert = \bar{x}) \pm 2 (SD) Schuljahre und 13 \pm 2 Ausbildungsjahre (inklusive Schulzeit) absolviert. Der Maximalwert im MMST (30 Punkte) wurde von 23 Patienten (26%) erreicht, der Mittelwert lag bei 29 Punkten (Tabelle 3, Abbildung 9).

oHE-Kohorte: 12 Probanden (6 Männer und 6 Frauen) mit Leberzirrhose zeigten zum Zeitpunkt des Einschlusses eine oHE-Symptomatik Grad 1 oder 2. Das mittlere Alter lag hier bei 62 \pm 8 Jahren mit einer Altersspanne zwischen 42 und 75 Jahren. Es wurden durchschnittlich 10 \pm 1 Schuljahre und 13 \pm 3 Ausbildungsjahre absolviert. Im MMST lag der maximal erreichte Wert bei 27 Punkten, der Mittelwert bei 21 Punkten (Tabelle 3, Abbildung 9).

≥39J-Kohorte: Bei den 61 rekrutierten lebergesunden Probanden lag der Anteil männlicher Studienteilnehmer bei 49,2% (N=30). Die Probanden waren zwischen 39 und 74 Jahre alt (\bar{x} : 59, SD: 11) und absolvierten 11 ± 2 Schuljahre und 14 ± 3 Ausbildungsjahre. Der Maximalwert im MMST wurde von 44 Probanden (72%) erreicht und bildete damit auch den Mittelwert (Tabelle 3, Abbildung 9).

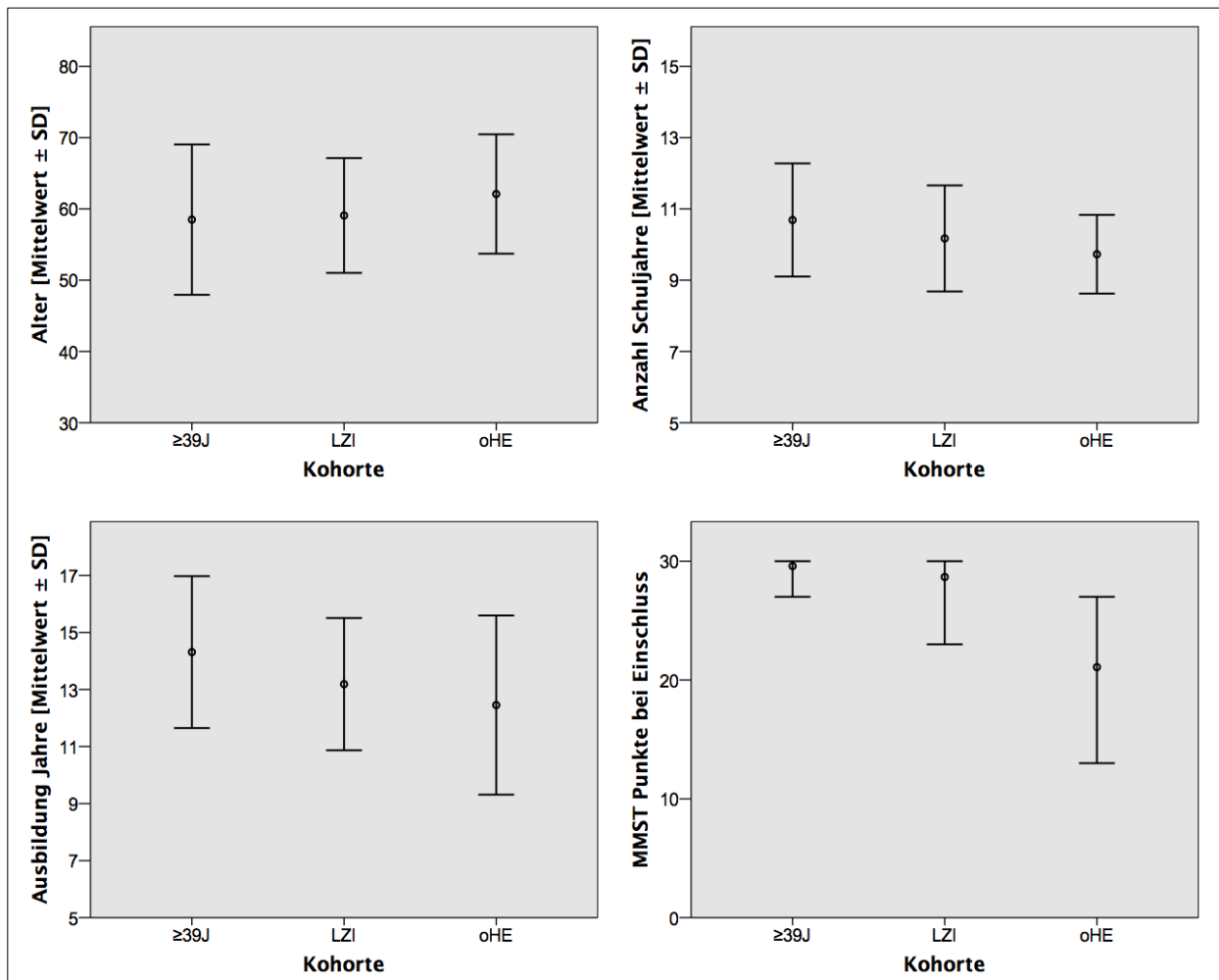


Abbildung 9: Darstellung der Mittelwerte mit dem Bereich der einfachen Standardabweichung für die Variablen Alter, Schuljahre und Ausbildungsjahre sowie mit der maximal und der minimal erreichten Punktzahl im MMST. (SD: Standardabweichung; MMST: Mini-Mental-Status-Test; LZI: Patienten mit Leberzirrhose ohne klinische Zeichen einer HE; oHE: Patienten mit Leberzirrhose und klinischen Zeichen einer HE; ≥39J: lebergesunde Probanden ab 39 Jahren)

Während die Unterschiede in den Variablen „Geschlecht“, „Alter“, „Schuljahre“ und „Ausbildungsjahre“ zwischen den drei Kohorten mit $p > 0,05$ jeweils nicht signifikant waren und damit eine ausgewogene Zusammensetzung der Kohorten angenommen werden konnte, waren die Differenzen im MMST jeweils signifikant ($p < 0,001$).

Tabelle 3: Demografische Daten der verschiedenen Studienkohorten

	N	Geschlecht	
		männlich	weiblich
Gesamt	161	90	71
LZI	88	54	34
oHE	12	6	6
≥39J	61	30	31

N: Anzahl; LZI: Patienten mit Leberzirrhose ohne klinische Zeichen einer HE; oHE: Patienten mit Leberzirrhose und klinischen Zeichen einer HE; ≥39J: lebergesunde Probanden ab 39 Jahren.

6.1.1 Spezielle Patientencharakteristika

Für jeden Patienten wurden die Genese der Leberzirrhose, der Grund des aktuellen Arztkontakts, bisher durchgeführte Untersuchungen (Leberbiopsie, Fibroscan, Ösophagogastroduodenoskopie (ÖGD), Bildgebung) mit Ergebnissen, aktuelle Symptome und Medikation, ein klinischer Untersuchungsbefund (Aszites, Ödeme) und die Vorgeschichte (vorige Dekompensationen, spontan bakterielle Peritonitiden, sonstige Eingriffe) erhoben. Die Ergebnisse sind in Tabelle 4 zusammengefasst.

Tabelle 4: Zusammensetzung der Patientenkohorte (N=100)

	Anzahl		Anzahl
Einschluss ambulant	53	Fibroscan	10
Einschluss stationär	47	HE in Vorgeschichte	21
Stationärer Aufnahmegrund:		TIPSS	13
HE	13	Medikamentöse HE-Prophylaxe (Laktulose, Rifaximin, Neomycin)	41
Dekompensierter Aszites	12	Spontan bakterielle Peritonitis in Vorgeschichte	9
Ösophagusvarizenligatur	6	Diuretikaeyinnahme	71
Evaluation LTx	4	Ödeme	49
Sonstige	12	Aszites	48
LZI durch Biopsie gesichert	22	Therapierefraktärer Aszites	21
Radiologische LZI-Hinweise	97	ÖGD in Vorgeschichte	77
CT	17	Dabei Varizen gesehen	67
MRT	21	Thrombozyten <120/nl	61
Sono	59		
Splenomegalie (>12cm)	70		

LTx: Lebertransplantation; LZI: Leberzirrhose; TIPSS: transjugulärer intrahepatischer portosystemischer Stent-Shunt; ÖGD: Ösophagogastroduodenoskopie.

Bei 21 Patienten wurde ein therapierefraktärer Aszites nach den Definitionskriterien von Arroyo et al. festgestellt. Nach diesen Kriterien ist bei einer Behandlungsdauer mit einer intensivierten Therapie von über einer Woche mit unzureichender Abnahme des Körpergewichts, eines Aszitesrezidivs innerhalb von vier Wochen und diuretikabedingten Nebenwirkungen (Elektrolytverschiebungen, Anstieg des Kreatinins, HE) von einem refraktären Aszites auszugehen (116).

Die häufigsten Ursachen für Leberzirrhose waren Alkoholabusus (N=37/100) und Hepatitis C (N=19/100). In 13 von 100 Fällen musste die Diagnose einer kryptogenen Leberzirrhose gestellt werden, bei den restlichen 31 Patienten lagen seltenere Ursachen vor (Abbildung 10).

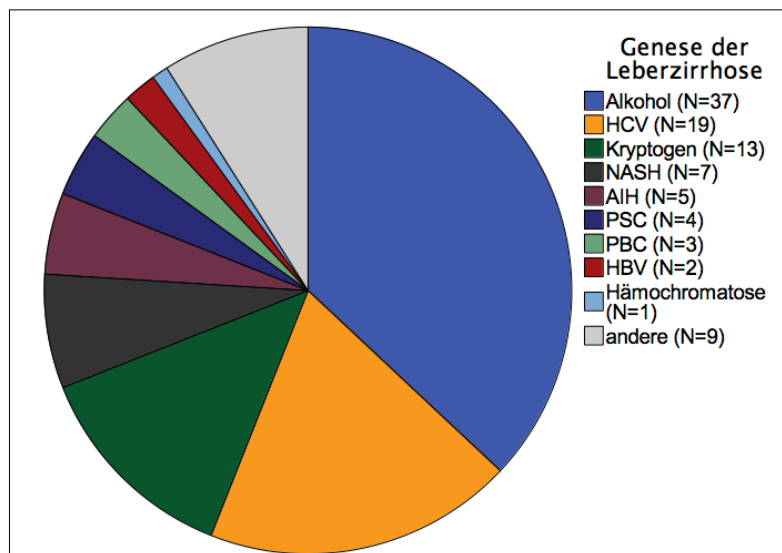


Abbildung 10: Ursachen der Leberzirrhose. (HCV: Hepatitis-C-Virus; NASH: Nicht-alkoholische Steatohepatitis; AIH: Autoimmunhepatitis; PSC: primär sklerosierende Cholangitis; PBC: primär biliäre Cholangitis; HBV: Hepatitis-B-Virus)

6.2 Eigenschaften psychometrischer Testverfahren

Für die statistischen Analysen dieser Dissertation wurden die Testergebnisse der 100 eingeschlossenen Patienten mit Leberzirrhose (88x LZI, 12x oHE) und der 61 im Alter vergleichbaren Kontrollprobanden (≥ 39 J-Kohorte) ausgewertet.

6.2.1 Psychometric Hepatic Encephalopathy Score (PHES)

Insgesamt absolvierten 145 von 161 Probanden die komplette PHES-Testbatterie und zwar 57 von 61 lebergesunde Probanden, 83 von 88 Patienten ohne oHE und 5 von 12 Patienten mit oHE.

Im Vergleich der Patienten mit LZI und *ohne* oHE (LZI-Kohorte) mit der Kontrollkohorte waren die Mittelwerte (\bar{x}) der LZI-Kohorte in den Tests auf Zeit um 26% (SDT, \bar{x} : 54s vs. 43s) bis 63% (NCT-B, \bar{x} : 147s vs. 90s) langsamer. Die Testwerte im DST und im LNT-Fehler fielen für die LZI-Kohorte 27% (\bar{x} : 33 vs. 45) bzw. 16% (\bar{x} : 29 vs. 25) schlechter aus. Der PHES-Gesamtwert der LZI-Kohorte lag durchschnittlich bei -3 Punkten und damit vier Punkte unter den Resultaten der ≥ 39 J-Kohorte (Tabelle 5). Diese Unterschiede waren im Kruskal-Wallis-Rangsummentest (K-W) bzw. der Varianzanalyse (ANOVA, nur für DST) in allen PHES-Einzeltests (Abbildung 11) sowie dem PHES-Gesamtwert ($p < 0,001$) signifikant.

Im Vergleich der Patienten mit LZI *und* oHE (oHE-Kohorte) mit der LZI-Kohorte war die oHE-Kohorte in den Tests auf Zeit durchschnittlich 46% (LNT-Zeit, \bar{x} : 179s vs. 123s) bis 176% (SDT, \bar{x} : 149s vs. 54s) langsamer als die LZI-Kohorte. Im DST und im LNT-Fehler erreichte die oHE-Kohorte um 45% (\bar{x} : 18 vs. 33) bzw. 159% (\bar{x} : 75 vs. 29) schlechtere Resultate. Der PHES-Gesamtwert der oHE-Kohorte lag mit einem Mittelwert von -11 Punkten durchschnittlich acht Punkte unter den Resultaten der LZI-Kohorte (Tabelle 5). Diese Unterschiede waren im K-W-Test bzw. der ANOVA (nur für DST) im DST, SDT (Abbildung 11) und PHES-Gesamtwert ($p = 0,031$) signifikant.

Ergebnisse

Tabelle 5: PHES-Testergebnisse der einzelnen Kohorten

		N	Mittelwert ± SD	[95%-KI]	Median [IQR]	Minimum	Maximum
NCT-A (s)	LZI	88	50,4 ± 22,8	[45,6-55,2]	45,2 [26,3]	20,3	152,3
	oHE	9	137,7 ± 123,7	[42,7-232,8]	90,0 [171,2]	40,0	353,0
	≥39J	61	36,3 ± 12,3	[33,1-39,4]	32,0 [15,3]	21,0	77,0
NCT-B (s)	LZI	85	147 ± 105	[124-169]	117 [80]	29	691
	oHE	5	231 ± 84	[126-335]	261 [152]	117	334
	≥39J	59	90 ± 40	[80-100]	80 [49]	42	214
DST	LZI	83	33 ± 11	[30-35]	33 [17]	9	55
	oHE	7	18 ± 9	[10-26]	22 [16]	4	28
	≥39J	59	45 ± 11	[42-47]	44 [15]	19	70
LNT-Zeit (s)	LZI	85	123 ± 48	[113-133]	110 [72]	52	257
	oHE	8	179 ± 101	[95-263]	141 [170]	62	343
	≥39J	60	92 ± 34	[83-101]	88 [52]	37	170
LNT-Fehler	LZI	88	29 ± 38	[21-37]	17 [29]	0	262
	oHE	11	75 ± 119*	[0-156]	38 [55]	0	382
	≥39J	60	25 ± 32*	[16-33]	13 [28]	0	138
SDT (s)	LZI	85	54 ± 20	[49-58]	50 [25]	25	136
	oHE	9	149 ± 155*	[30-269]	89 [106]	62	536
	≥39J	60	43 ± 14	[39-46]	40 [14]	22	92
PHES-Gesamtwert	LZI	83	-3 ± 5	[-4-(-2)]	-2 [7]	-16	4
	oHE	5	-11 ± 3	[-14-(-7)]	-11 [5]	-13	-8
	≥39J	57	1 ± 2	[0-1]	1 [3]	-5	5

*: bei Einheiten, die keine negativen Werte annehmen können, ist eine Standardabweichung größer als der Mittelwert ein rein mathematisches Konstrukt. Im Weiteren wurde hier der Median verwendet. NCT-A/B: Zahlenverbindungstest A/B; DST: Zahlensymboltest; LNT: Liniennachfahrttest; SDT: Kreispunktieren-Test; PHES: Psychometric Hepatic Encephalopathy Score; LZI: Patienten mit Leberzirrhose ohne klinische Zeichen einer HE; oHE: Patienten mit Leberzirrhose und klinischen Zeichen einer HE; ≥39J: lebergesunde Probanden ab 39 Jahren; N: Anzahl; SD: Standardabweichung; KI: Konfidenzintervall; IQR: Interquartilsabstand; s: Sekunden.

Von 16 Studienteilnehmern konnte kein PHES-Gesamtwert berechnet werden, da einzelne Testergebnisse aufgrund stark eingeschränkter kognitiver Leistungsfähigkeit oder eines Mangels an Motivation nicht erhoben werden konnten. Die Resultate erfolgreich beendeter Einzeltests dieser Probanden sind in Tabelle 5 mit dargestellt und erklären die variierende Anzahl (N) an Probanden zwischen den verschiedenen Tests.

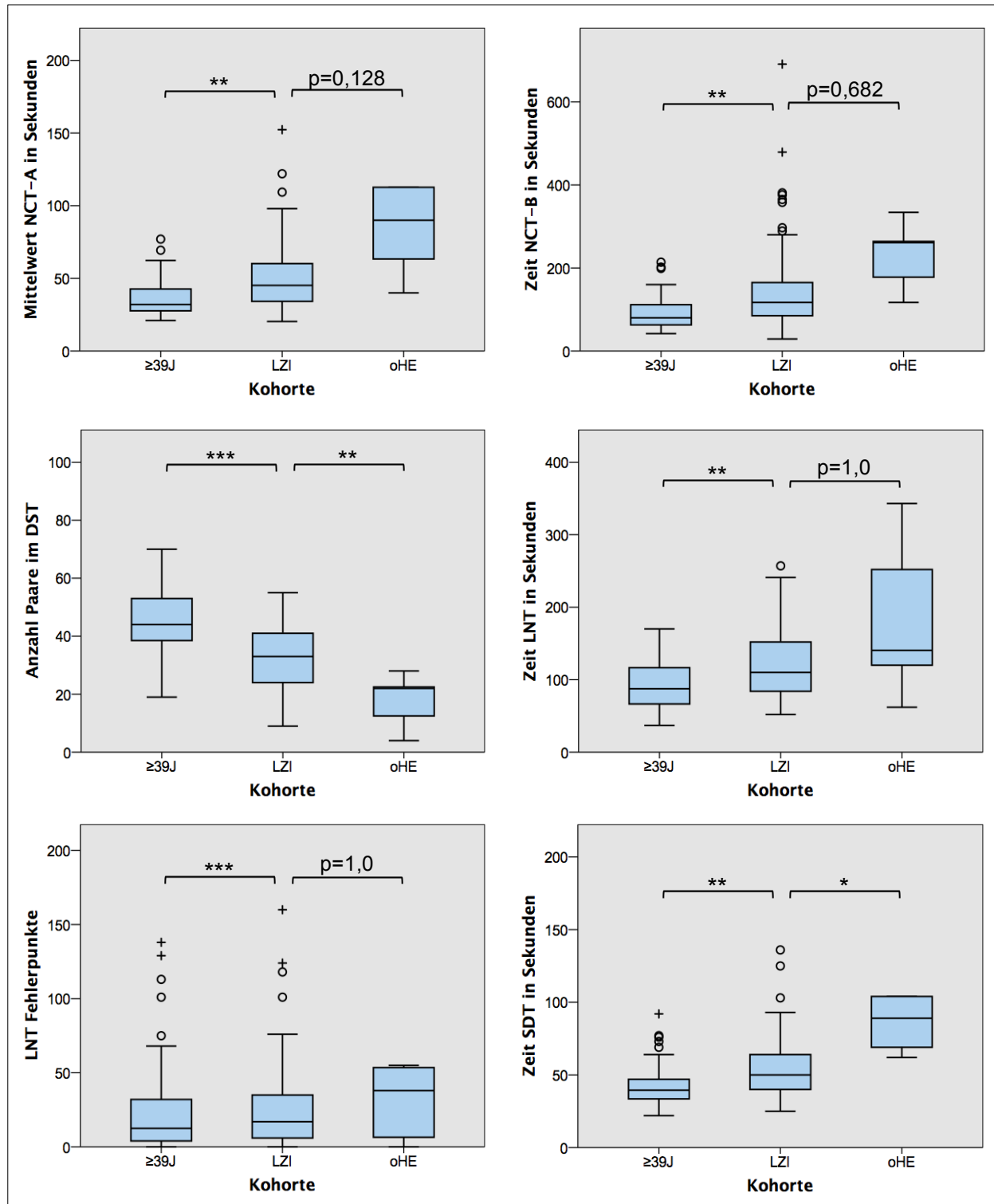


Abbildung 11: PHES-Testergebnisse der einzelnen Kohorten als Kastengrafik. (NCT-A/B: Zahlenverbindungstest A/B; DST: Zahlensymboltest; LNT: Liniennachfahrtstest; SDT: Kreise-punktieren-Test; LZI: Patienten mit Leberzirrhose ohne klinische Zeichen einer HE; oHE: Patienten mit Leberzirrhose und klinischen Zeichen einer HE; $\geq 39J$: lebergesunde Probanden ab 39 Jahren. Aus Gründen der Übersichtlichkeit nicht abgebildet sind folgende Extremwerte der Patientenkohorten: im NCT-A 353s, 350s; im LNT-Fehler 382, 262, 222; im SDT 536s, 240s. $***p<0,001$, $**p<0,01$, $*p<0,05$.)

Dem PHES-Grenzwert von -4 Punkten folgend, der per Definition Patienten mit mHE von Patienten ohne mHE differenziert, wurden die Resultate von 26 Patienten mit Leberzirrhose als mHE eingestuft. Fünf Patienten, bei denen bereits durch die klinische Untersuchung eine oHE diagnostiziert werden konnte, konnten den PHES komplett absolvieren und erzielten alle ein Ergebnis von kleiner -4. Darüber hinaus unterschritt ein lebergesunder Proband den Grenzwert (-5 Punkte), was für die weiteren Analysen aber kein Ausschlusskriterium darstellte (Tabelle 6, Abbildung 12A).

Tabelle 6: Anzahl pathologischer (<-4 Punkte) PHES-Testergebnisse

		≥39J	LZI	oHE	Gesamt
PHES kleiner als Grenzwert (-4)	ja	1	26	5	32
	nein	56	57	0	113
Gesamtsumme		57	83	5	145
Kein vollständiger PHES		4	5	7	16

PHES: Psychometric Hepatic Encephalopathy Score; LZI: Patienten mit Leberzirrhose ohne klinische Zeichen einer HE; oHE: Patienten mit Leberzirrhose und klinischen Zeichen einer HE; ≥39J: lebergesunde Probanden ab 39 Jahren.

Auf Grundlage der PHES-Ergebnisse wurde die LZI-Kohorte für die nachfolgenden Berechnungen unterteilt in die Untergruppen „keine mHE“ (N=57; 68,7%) und „mHE“ (N=26; 31,3%). Außerdem sind nur die Testergebnisse der Studienteilnehmer in die weitere statistische Auswertung des PHES eingegangen, die den PHES vollständig durchgeführt und erfolgreich beendet haben (N=145). Daher besteht die LZI-Kohorte im Folgenden nur noch aus 83 Patienten („mHE+keine mHE“), die oHE-Kohorte aus 5 Patienten und die ≥39J-Kohorte aus 57 Probanden.

Durch die grenzwertbezogene Aufteilung der LZI-Gruppe entstand ein signifikanter Unterschied im PHES zwischen den neuen Gruppen *keine mHE* und *mHE* ($p < 0,001$). Aber auch die Gruppen ≥39J und *keine mHE* unterschieden sich im K-W-Test signifikant voneinander ($p = 0,045$; Abbildung 12B).

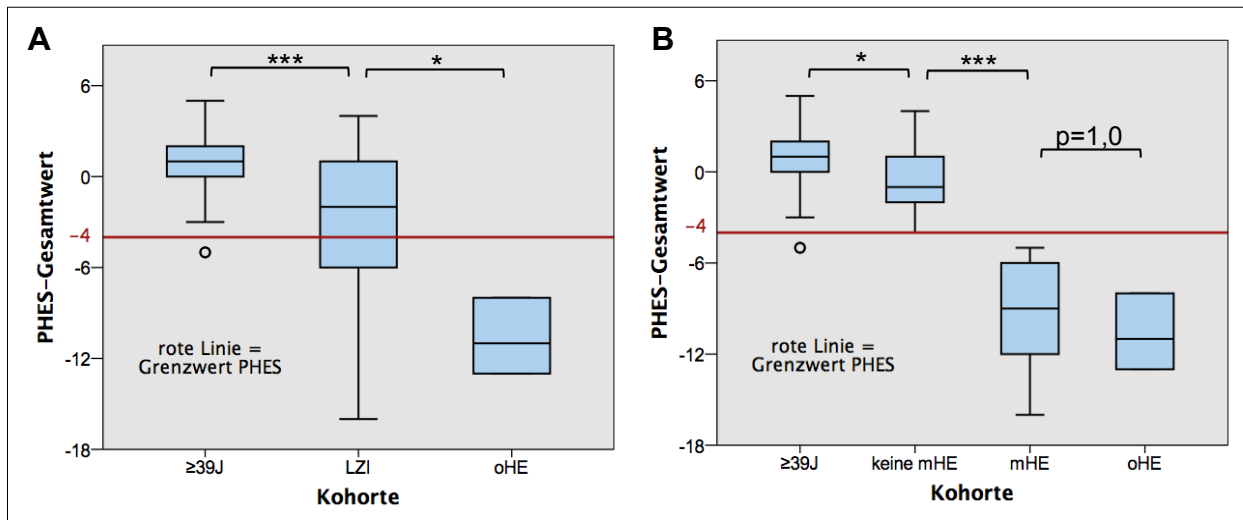


Abbildung 12: Aus den Einzeltests errechneter PHES-Gesamtwert – Die Kohorte LZI aus A wird entsprechend des PHES-Grenzwerts von -4 (veranschaulicht durch die horizontale rote Linie) in B in die Gruppen *keine mHE* und *mHE* aufgeteilt. (LZI: Patienten mit Leberzirrhose ohne klinische Zeichen einer HE; oHE: Patienten mit Leberzirrhose und klinischen Zeichen einer HE; $\geq 39J$: lebergesunde Probanden ab 39 Jahren; PHES: Psychometric Hepatic Encephalopathy Score. *** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$.)

6.2.1.1 Diskriminierungsfähigkeit der PHES-Einzeltests

Bei der Analyse der PHES-Einzeltests zeigten sich mit Ausnahme des Fehlerpunkte-werts im LNT alle Testresultate der mHE-Kohorte signifikant schlechter als die Resultate der Gruppe ohne mHE (Abbildung 13, Tabelle 7). Der größte relative Unterschied der Mittelwerte konnte im NCT-B gemessen werden. Hier waren die Testzeiten der mHE-Kohorte um 153% langsamer als die Testzeiten der Gruppe ohne mHE (\bar{x} : 205,2s vs. 98,9s).

Hohe Werte für die Effektstärke (hier als Maß für die Auswirkungen einer mHE auf die Testresultate interpretierbar) wurden im DST ($r_e=0,718$), NCT-B ($r_e=0,712$), NCT-A ($r_e=0,699$) und SDT ($r_e=0,557$) berechnet (Tabelle 7).

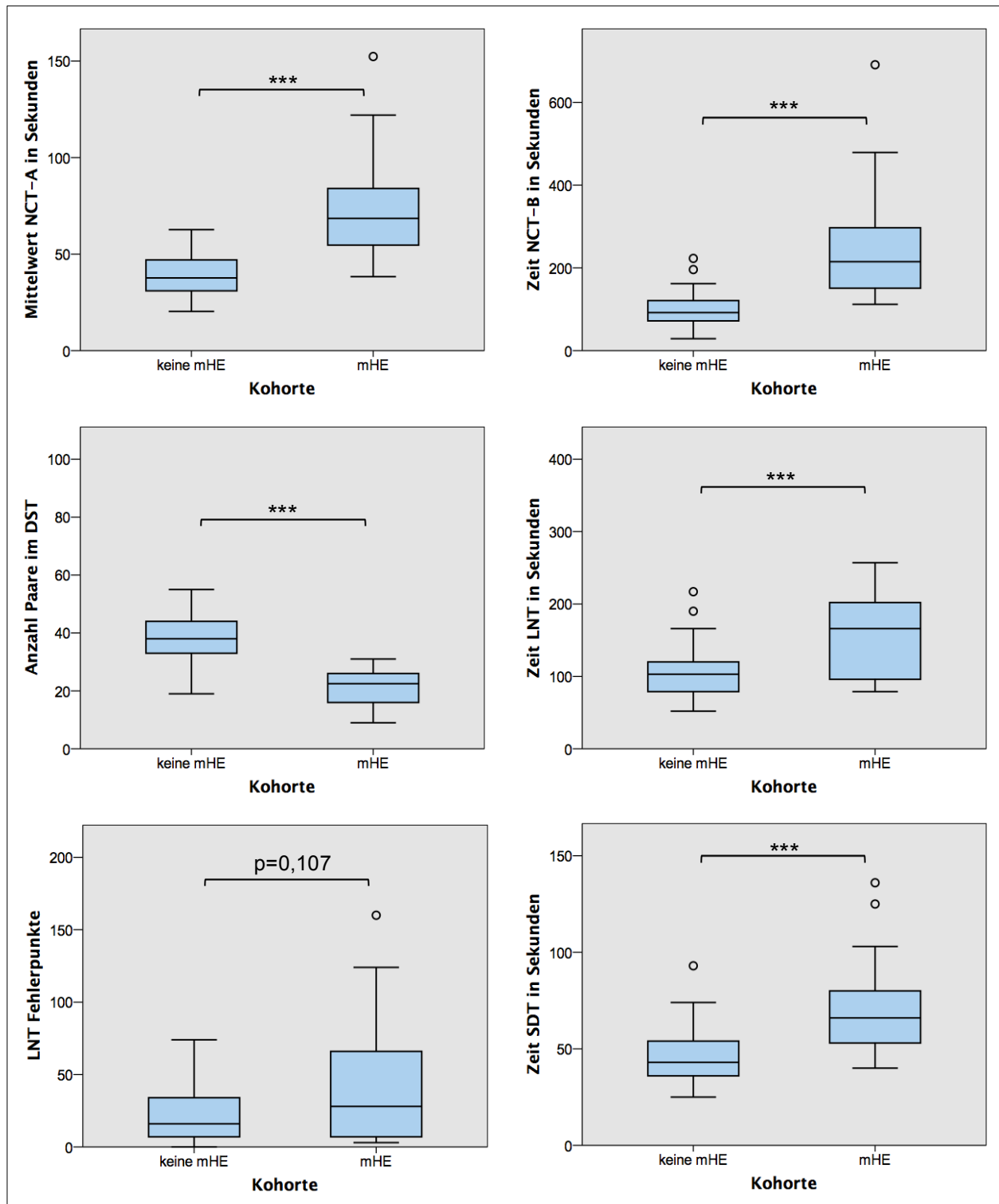


Abbildung 13: Vergleich der PHES-Testergebnisse für *keine mHE* und *mHE* – In jedem Test des PHES (Ausnahme: Fehlerpunktwert des LNTs) ist ein signifikanter Unterschied zwischen den beiden Kohorten *keine mHE* und *mHE* messbar. (NCT-A/B: Zahlenverbindungstest A/B; DST: Zahlensymboltest; LNT: Liniennachfahrttest; SDT: Kreise-punktieren-Test; mHE: minimale hepatische Enzephalopathie. ***p<0,001, **p<0,01, *p<0,05.)

Tabelle 7: Vergleich der PHES-Testergebnisse für *keine mHE* und *mHE*

	Median [IQR] (keine mHE)	Median [IQR] (mHE)	p-Wert	Effekt- stärke r_e
NCT-A (s)	37,7 [17,5]	68,5 [19,7]	<0,001 (U-Test)	0,699
NCT-B (s)	92,0 [51,0]	215,0 [163,5]	<0,001 (U-Test)	0,712
DST	38 [12]	23 [11]	<0,001 (U-Test)	0,718
LNT-Zeit (s)	103,0 [44,5]	166,0 [107,0]	<0,001 (U-Test)	0,454
LNT-Fehler	16 [28]	28 [62]	0,107 (U-Test)	0,177
SDT (s)	43,0 [19,5]	66,0 [27,8]	<0,001 (U-Test)	0,557

NCT-A/B: Zahlenverbindungstest A/B; DST: Zahlensymboltest; LNT: Liniennachfahrttest; SDT: Kreispunktieren-Test; U-Test: Mann-Whitney-U-Test; IQR: Interquartilsabstand; keine mHE: Patienten mit Leberzirrhose ohne mHE; mHE: Patienten mit Leberzirrhose und mHE; s: Sekunden.

6.2.2 Elektronischer Zahlenverbindungstest (eNCT)

Die Studienteilnehmer absolvierten im Anschluss an die papierbasierten Tests des PHES den eNCT auf jeweils einem Tablet-Computer mit 10“ und 18“ Bildschirmdurchmesser. Resultate im eNCT 10“ bzw. eNCT 18“ konnten insgesamt von 156 bzw. 155 Probanden erhoben werden und von 145 bzw. 144 Probanden mit vollständig absolvierter PHES-Testbatterie ausgewertet werden (83 bzw. 82x LZI, 5x oHE, 57x $\geq 39J$).

Die durchschnittlichen Testresultate der LZI-Kohorte waren in beiden Tests schlechter als in der $\geq 39J$ -Kohorte und besser als in der oHE-Kohorte (Tabelle 8).

Tabelle 8: eNCT-Testergebnisse der einzelnen Kohorten

		N	Mittelwert \pm SD	[95%-KI]	Median [IQR]	Mini- mum	Maxi- mum
eNCT 10“ (s)	LZI	88	42,8 \pm 15,7	[39,4-46,1]	38,5 [16,7]	19,7	110,3
	oHE	7	115,1 \pm 113,6	[10,1-220,2]	78,0 [67,3]	37,7	364,3
	$\geq 39J$	61	31,5 \pm 7,5	[29,6-33,4]	30,7 [10,3]	18,7	55,0
eNCT 18“ (s)	LZI	87	43,8 \pm 13,8	[40,9-46,8]	41,0 [18,7]	23,7	90,0
	oHE	9	119,8 \pm 117,8	[29,2-210,3]	74,0 [110,0]	38,7	389,7
	$\geq 39J$	59	32,5 \pm 7,9	[30,5-34,6]	30,7 [12,3]	22,0	55,3

eNCT: elektronischer Zahlenverbindungstest; $\geq 39J$: lebergesunde Probanden ab 39 Jahren; LZI: Patienten mit Leberzirrhose ohne oHE; oHE: Patienten mit Leberzirrhose und oHE; N: Anzahl; SD: Standardabweichung; KI: Konfidenzintervall; IQR: Interquartilsabstand; s: Sekunden.

Im Detail betrachtet und nach Prüfung auf Normalverteilung wurden signifikante Unterschiede für den eNCT zwischen den Gruppen $\geq 39J$ und *LZI* festgestellt (Tabelle 9). Dabei brauchten Patienten mit Leberzirrhose in beiden Tests durchschnittlich 11,3 Sekunden länger als Probanden der Kontrollkohorte. Zwischen den Gruppen *LZI* und *oHE* waren die absoluten Differenzen der Testzeiten mit 72,3 Sekunden (eNCT 10“) bzw. 76,0 Sekunden (eNCT 18“) zwar größer, signifikant waren diese Unterschiede aber nicht (Tabelle 9).

Tabelle 9: Ergebnisse der Signifikanztests (eNCT)

	p-Wert (oHE vs. LZI)	p-Wert (LZI vs. $\geq 39J$)	Test
eNCT 10“	0,265	<0,001	K-W
eNCT 18“	0,168	<0,001	K-W

eNCT: elektronischer Zahlenverbindungstest; LZI: Patienten mit Leberzirrhose ohne klinische Zeichen einer HE; oHE: Patienten mit Leberzirrhose und klinischen Zeichen einer HE; $\geq 39J$: lebergesunde Probanden ab 39 Jahren; K-W: Kruskal-Wallis.

6.2.2.1 Diskriminierungsfähigkeit des eNCTs

Nach der Neuaufteilung der LZI-Kohorte durch die PHES-Auswertung in *keine mHE* (N=57) und *mHE* (N=26) wurde analysiert, ob sich die eNCT-Testzeiten zwischen den beiden zu diskriminierenden Kohorten unterschieden.

Die *mHE*-Kohorte war im eNCT 10“ um 71% (\bar{x} : 59,0s vs. 34,5s) und im eNCT 18“ um 55% (\bar{x} : 57,3s vs. 36,8s) langsamer als die Kohorte ohne *mHE*. Die Ergebnisse des Mann-Whitney-U-Tests (U-Test) ergaben einen signifikanten Unterschied zwischen diesen beiden Patientengruppen (Tabelle 10, Abbildung 14). Das Effektstärkemaß r_e fiel für den eNCT 10“ ($r_e=0,698$) und den eNCT 18“ ($r_e=0,675$) hoch aus (Tabelle 10).

Tabelle 10: Vergleich der eNCT-Testergebnisse für *keine mHE* und *mHE*

	Median [IQR] (keine mHE)	Median [IQR] (mHE)	p-Wert	Effekt- stärke r_e
eNCT 10“ (s)	34,0 [9,3]	57,0 [19,2]	<0,001 (U-Test)	0,698
eNCT 18“ (s)	35,8 [10,3]	56,2 [19,6]	<0,001 (U-Test)	0,675

eNCT: elektronischer Zahlenverbindungstest; mHE: minimale hepatische Enzephalopathie; U-Test: Mann-Whitney-U-Test; IQR: Interquartilsabstand; s: Sekunden.

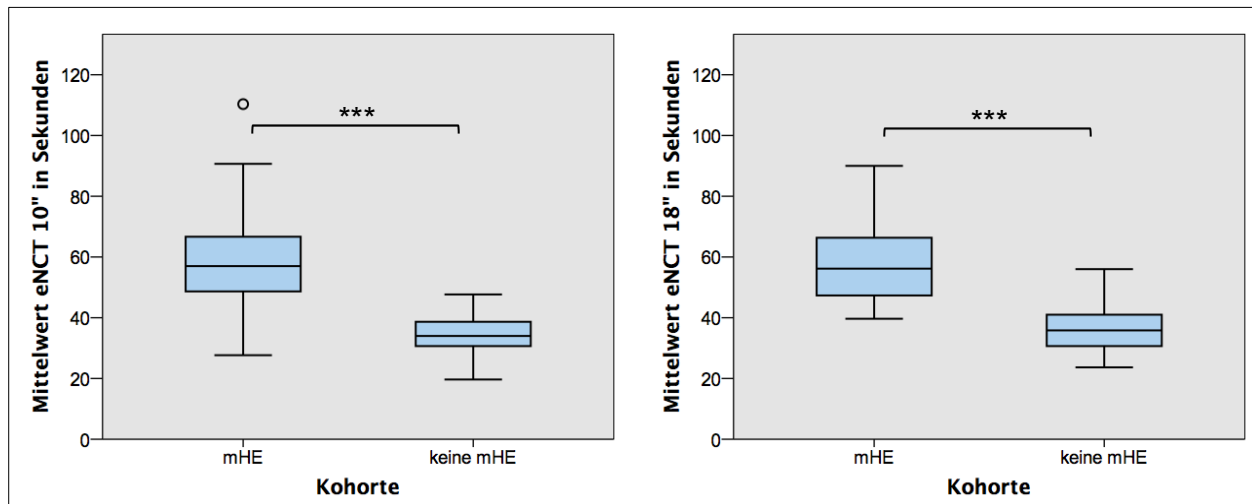


Abbildung 14: Vergleich der eNCT-Testergebnisse für *keine mHE* und *mHE* – In beiden Testversionen des eNCTs ist ein signifikanter Unterschied zwischen den Kohorten *keine mHE* und *mHE* messbar. (eNCT: elektronischer Zahlenverbindungstest; mHE: minimale hepatische Enzephalopathie. *** $p < 0,001$.)

6.2.3 Stroop-Test

Den Stroop-Test beendeten insgesamt 104 Studienteilnehmer, von denen 102 auch die PHES-Testbatterie vollständig absolvierten (51x LZI, 3x oHE, 48x $\geq 39J$). Die Summe der *Stroop-On-* und *Stroop-Off-*Testzeiten (*On+Off*) wurde als wesentliches Maß für die Vergleiche mit PHES und eNCT verwendet.

In der LZI-Kohorte war diese Gesamtzeit signifikant ($p = 0,020$ im K-W-Test) langsamer als in der $\geq 39J$ -Kohorte (\bar{x} : 195,9s vs. 165,0s), wohingegen sich die Testzeiten der oHE-Kohorte und der LZI-Kohorte (\bar{x} : 253,3s vs. 195,9s) trotz durchschnittlich größerer Differenz nicht signifikant ($p = 1,0$ im K-W-Test) voneinander unterschieden (Tabelle 11).

Tabelle 11: Stroop-Testergebnisse der einzelnen Kohorten

		N	Mittelwert \pm SD	[95%-KI]	Median [IQR]	Mini- mum	Maxi- mum
Stroop On+Off (s)	LZI	53	195,9 \pm 49,5	[182,2-209,5]	187,0 [36,0]	123	404
	oHE	3	253,3 \pm 126,2	[0*-566,7]	182,0 [-]	179	399
	$\geq 39J$	48	165,0 \pm 33,7	[155,2-174,8]	158,5 [39,0]	107	268

*: da die Einheit *Sekunden* keine negativen Werte annehmen kann, wurde der als negativ berechnete untere Wert des 95%-Konfidenzintervalls auf 0 gesetzt.

LZI: Patienten mit Leberzirrhose ohne oHE; oHE: Patienten mit Leberzirrhose und oHE; $\geq 39J$: lebergesunde Probanden ab 39 Jahren; N: Anzahl; SD: Standardabweichung; KI: Konfidenzintervall; IQR: Interquartilsabstand; s: Sekunden.

6.2.3.1 Diskriminierungsfähigkeit des Stroop-Tests

Durch Anwendung des PHES-Grenzwerts wurden zwölf Patienten der LZI-Kohorte mit Stroop-Testwerten der Kohorte *mHE* zugeordnet (23,5%) und 39 der Kohorte *keine mHE* (76,5%). Zwischen diesen beiden Gruppen konnte im Stroop-Test ein signifikanter Unterschied ($p < 0,001$ im U-Test) der Testzeiten berechnet werden, die in der *mHE*-Kohorte durchschnittlich 69,2 Sekunden langsamer ausfielen als in der Kohorte ohne *mHE* (\bar{x} : 248,5s vs. 179,3s; Median: 250,0s vs. 179,0s; Abbildung 15). Für die *mHE*-Diagnose konnte eine Effektstärke auf die Testzeiten von $r_e = 0,551$ gemessen werden.

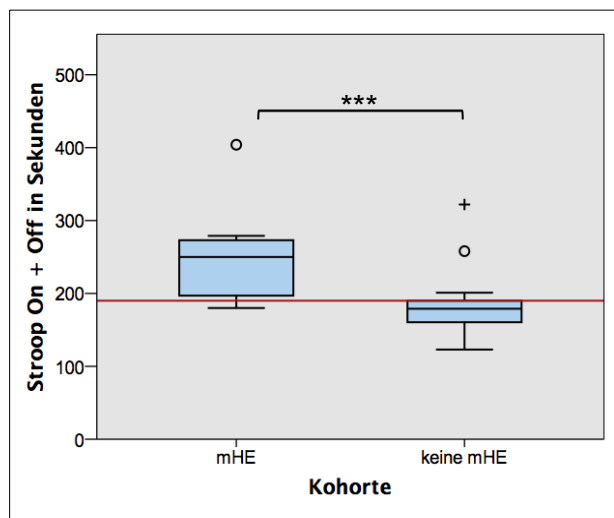


Abbildung 15: Vergleich der Stroop-Testergebnisse für *keine mHE* und *mHE* – Ein signifikanter Unterschied zwischen den Kohorten *keine mHE* und *mHE* ist messbar. Die horizontale rote Linie entspricht dem Stroop-Grenzwert von 190 Sekunden für Probanden ≥ 45 Jahren. (eNCT: elektronischer Zahlenverbindungstest; *mHE*: minimale hepatische Enzephalopathie. *** $p < 0,001$.)

6.3 Vergleich von eNCT und PHES-Tests

6.3.1 Korrelationen der Testergebnisse

Zur Untersuchung der Validität und möglicher Zusammenhänge der eNCT-Resultate mit einzelnen PHES-Testergebnissen wurden Korrelationsanalysen durchgeführt.

Unabhängig vom Bildschirmdurchmesser des Tablet-Computers und von der untersuchten LZI-Untergruppe (*mHE* oder *keine mHE*) oder der Gesamtgruppe *mHE+keine mHE* wurden die stärksten Korrelationen zwischen eNCT-Testzeiten und den Resultaten in den PHES-Einzeltests NCT-A, NCT-B und DST gemessen (Abbildungen 16 und 17).

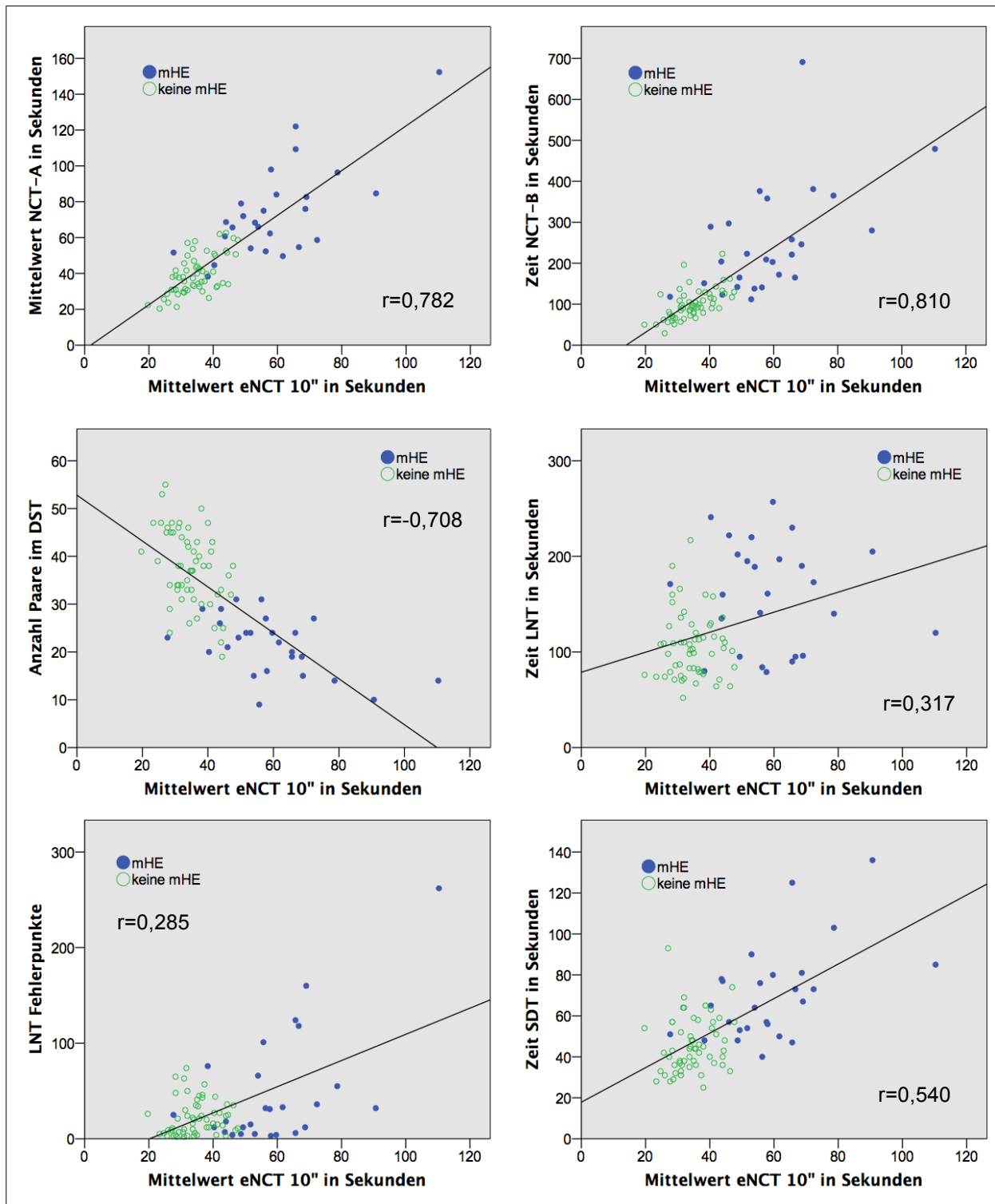


Abbildung 16: Streudiagramme der Korrelationsanalysen vom eNCT 10'' mit den PHES-Einzeltests – Jeder Kreis steht für die Testergebnisse eines Probanden im eNCT 10'' (x-Achse) und im jeweiligen Papiertest (y-Achse), die diagonale Linie entspricht der Regressionsgeraden. (NCT-A/B: Zahlenverbindungstest A/B; DST: Zahlensymboltest; LNT: Liniennachfahrttest; SDT: Kreise-punktieren-Test; r: Spearman-Korrelationskoeffizient; eNCT: elektronischer Zahlenverbindungstest; mHE: minimale hepatische Enzephalopathie)

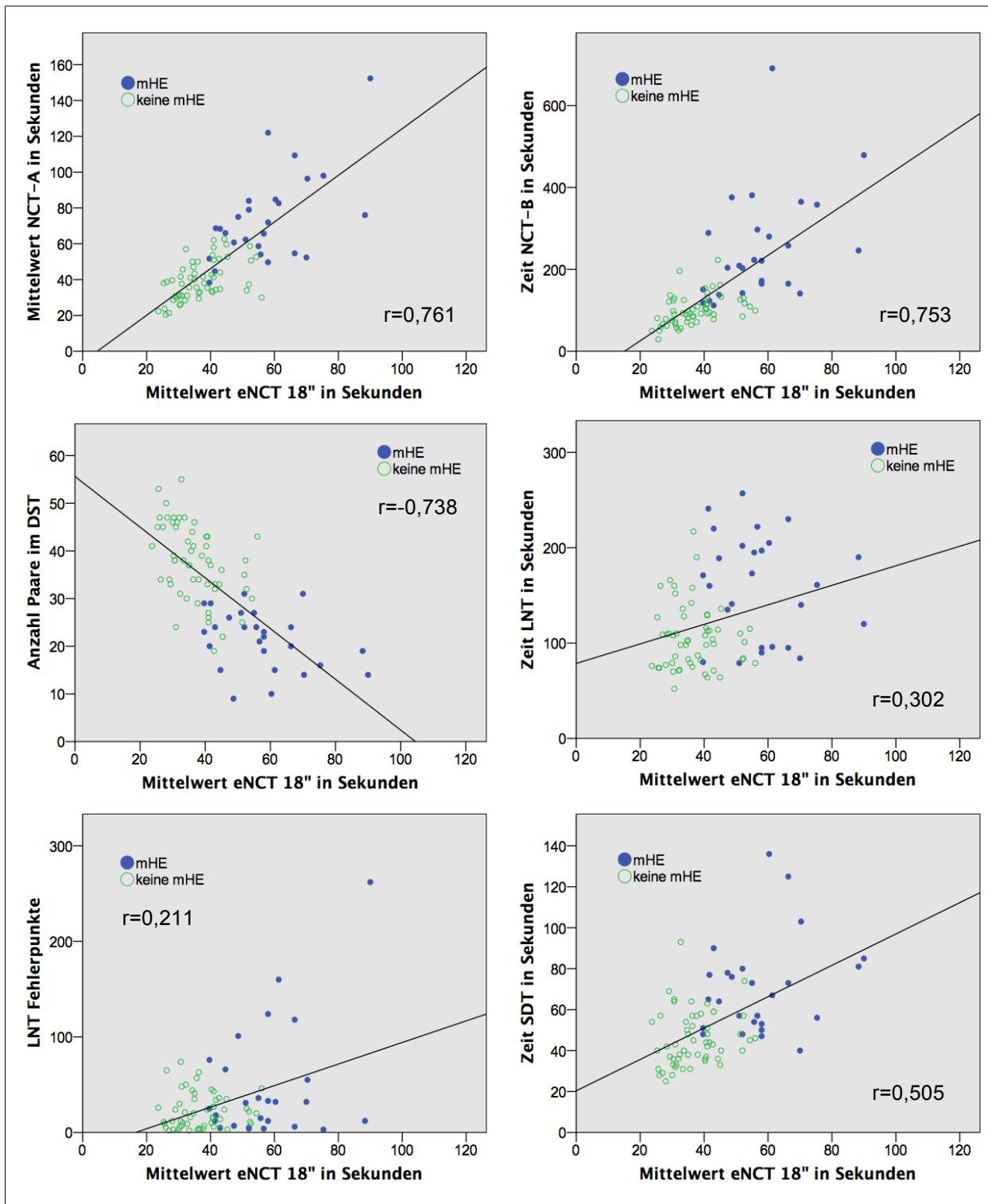


Abbildung 17: Streudiagramme der Korrelationsanalysen vom eNCT 18'' mit den PHEs-Einzeltests – Jeder Kreis steht für die Testergebnisse eines Probanden im eNCT 18'' (x-Achse) und im jeweiligen Papiertest (y-Achse), die diagonale Linie entspricht der Regressionsgeraden. (NCT-A/B: Zahlenverbindungstest A/B; DST: Zahlensymboltest; LNT: Liniennachfahrttest; SDT: Kreise-punktieren-Test; r: Spearman-Korrelationskoeffizient; eNCT: elektronischer Zahlenverbindungstest; mHE: minimale hepatische Enzephalopathie)

Hohe Korrelationskoeffizienten ($r > 0,7$) ergaben sich bezüglich dieser PHES-Einzeltests (NCT-A, NCT-B, DST) für die Ergebnisse der Gruppe *mHE+keine mHE*, wobei die Koeffizienten für den eNCT 10“ im NCT-A ($r=0,782$ vs. $r=0,761$) und NCT-B ($r=0,810$ vs. $r=0,753$) größer und im DST ($r=-0,708$ vs. $r=-0,738$) kleiner waren als für den eNCT 18“ (Tabellen 12 und 13).

Die Korrelationskoeffizienten für eNCT und PHES-Gesamtwert lagen in der relevanten Patientengruppe *mHE+keine mHE* im hohen Bereich (eNCT 10“: $r=-0,747$; eNCT 18“: $r=-0,717$).

Tabelle 12: Spearman-Rangkorrelationskoeffizienten (eNCT 10“)

	mHE (N=26)	Keine mHE (N=57)	mHE+keine mHE (N=83)
NCT-A	$r=0,585$ ($p=0,002$)	$r=0,566$ ($p<0,001$)	$r=0,782$ ($p<0,001$)
NCT-B	$r=0,569$ ($p=0,002$)	$r=0,652$ ($p<0,001$)	$r=0,810$ ($p<0,001$)
DST	$r=-0,492$ ($p=0,011$)	$r=-0,425$ ($p=0,001$)	$r=-0,708$ ($p<0,001$)
LNT-F	$r=0,416$ ($p=0,035$)	$r=0,257$ ($p=0,053$)	$r=0,285$ ($p=0,009$)
LNT-Z	$r=-0,064$ ($p=0,756$)	$r=0,025$ ($p=0,852$)	$r=0,317$ ($p=0,004$)
SDT	$r=0,461$ ($p=0,018$)	$r=0,249$ ($p=0,062$)	$r=0,540$ ($p<0,001$)
PHES	$r=-0,716$ ($p<0,001$)	$r=-0,416$ ($p=0,001$)	$r=-0,747$ ($p<0,001$)

NCT-A/B: Zahlenverbindungstest A/B; DST: Zahlensymboltest; LNT-F: Liniennachfahrtest Fehlerpunkte; LNT-Z: Liniennachfahrtest Zeit; SDT: Kreise-punktieren-Test; PHES: Psychometric Hepatic Encephalopathy Score; mHE: minimale hepatische Enzephalopathie; r: Spearman-Korrelationskoeffizient; N: Anzahl.

Tabelle 13: Spearman-Rangkorrelationskoeffizienten (eNCT 18“)

	mHE (N=26)	Keine mHE (N=56)	mHE+keine mHE (N=82)
NCT-A	$r=0,553$ ($p=0,003$)	$r=0,564$ ($p<0,001$)	$r=0,761$ ($p<0,001$)
NCT-B	$r=0,500$ ($p=0,009$)	$r=0,534$ ($p<0,001$)	$r=0,753$ ($p<0,001$)
DST	$r=-0,402$ ($p=0,042$)	$r=-0,532$ ($p<0,001$)	$r=-0,738$ ($p<0,001$)
LNT-F	$r=0,137$ ($p=0,504$)	$r=0,150$ ($p=0,271$)	$r=0,211$ ($p=0,057$)
LNT-Z	$r=-0,133$ ($p=0,518$)	$r=0,023$ ($p=0,864$)	$r=0,302$ ($p=0,006$)
SDT	$r=0,203$ ($p=0,320$)	$r=0,280$ ($p=0,037$)	$r=0,505$ ($p<0,001$)
PHES	$r=-0,468$ ($p=0,016$)	$r=-0,407$ ($p=0,002$)	$r=-0,717$ ($p<0,001$)

NCT-A/B: Zahlenverbindungstest A/B; DST: Zahlensymboltest; LNT-F: Liniennachfahrtest Fehlerpunkte; LNT-Z: Liniennachfahrtest Zeit; SDT: Kreise-punktieren-Test; PHES: Psychometric Hepatic Encephalopathy Score; mHE: minimale hepatische Enzephalopathie; r: Spearman-Korrelationskoeffizient; N: Anzahl.

6.3.2 Testgenauigkeit des eNCTs

Zur Einschätzung der eNCT-Testgenauigkeit, Patienten mit mHE mithilfe eines Grenzwerts als solche zu identifizieren, wurde eine ROC-Analyse in der LZI-Kohorte durchgeführt, da diese Kohorte der Zielgruppe des eNCTs entspricht. Patienten mit oHE sind aufgrund klinischer Symptome zu diagnostizieren und daher von den Grenzwertberechnungen auszuschließen.

Die gemessenen Flächen unter den Kurven (AUC) des eNCT 10“ bzw. eNCT 18“ waren mit 0,937 (95%-KI: 0,867-1,0) bzw. 0,921 (95%-KI: 0,865-0,978) hoch (Abbildung 18). Diese Werte waren vergleichbar mit der AUC des NCT-A (0,936; 95%-KI: 0,884-0,988).

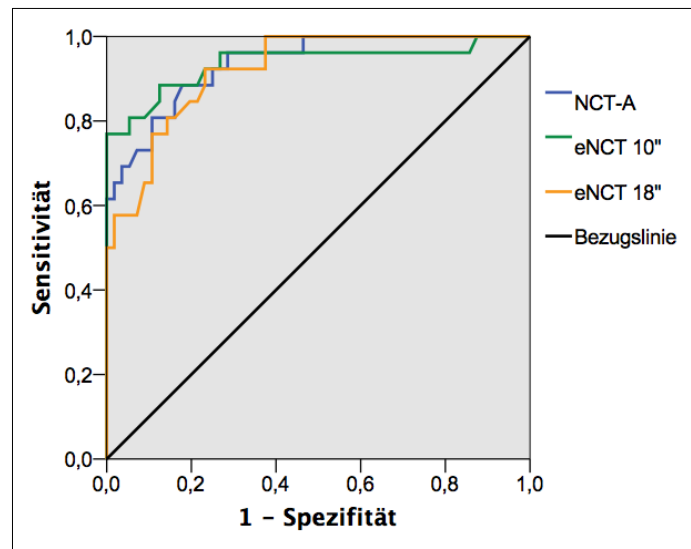


Abbildung 18: Kurven der ROC-Analysen für die Testzeiten im NCT-A (blau), eNCT 10“ (grün) und eNCT 18“ (orange) bezogen auf die Testergebnisse im PHES – Die Fläche unter der Kurve (AUC) ist für den eNCT 10“ (AUC=0,937) etwas größer als für den NCT-A (AUC=0,936) und den eNCT 18“ (AUC=0,921). (NCT-A: Zahlenverbindungstest A; eNCT: elektronischer Zahlenverbindungstest)

Als Grenzwerte für eine mHE wurden in dieser Arbeit Zeiten gewählt, die entsprechend der ROC-Analyse eine sehr hohe Sensitivität im Sinne des beabsichtigten Einsatzgebiets des eNCTs ermöglichten und dennoch eine Spezifität von mindestens 70% gewährleisten. So wurde für den eNCT 10“ festgelegt, dass Testzeiten (Mittelwert der drei Versuche) länger als 38 Sekunden als pathologisch gelten (Sensitivität laut ROC-Analyse 96%, Spezifität 70%). Für den eNCT 18“ wurden 41 Sekunden (Sensitivität 92%, Spezifität 70%) als Grenzwert gewählt.

An diesen Grenzwerten wurden die Testergebnisse der LZI-Kohorte bemessen, um die Testgenauigkeitskriterien in der praktischen Anwendung zu untersuchen. Wie einleitend

erwähnt, ist eine Prävalenz der mHE von 20-80% bei klinischer Symptombfreiheit zu erwarten und konnte vom PHES bei 31,3% der LZI-Kohorte nachgewiesen werden.

Insgesamt klassifizierte der eNCT 10“ mit dem Grenzwert von 38 Sekunden 80,7% (N=67/83) der Fälle richtig. Eine mHE wurde in 40 von 83 Fällen diagnostiziert, was im Vergleich mit dem PHES (N=26/83) einer Zunahme um 17 Prozentpunkte entspricht. Falsch-negative Diagnosen traten in einem Fall (=3,8%) auf. Dementsprechend konnte eine Sensitivität von 96,2% (95%-KI: 84,1-99,8%) berechnet werden. Die Spezifität betrug 73,7% (95%-KI: 61,4-83,9%) und die Vorhersagewerte lagen bei 97,7% (NPW) und 62,5% (PPW) (Tabelle 14).

Der eNCT 18“ klassifizierte mit dem Grenzwert von 41 Sekunden 81,7% (N=67/82) der Fälle richtig. Mit 37 diagnostizierten Fällen nahm die mHE-Diagnose im Vergleich mit dem PHES um 14 Prozentpunkte zu. In zwei Fällen (7,7%) kam es zu falsch-negativen Diagnosen, was einer Sensitivität von 92,3% (95%-KI: 78,1-98,7%) entspricht. Während die Sensitivität damit niedriger war als im eNCT 10“, war die Spezifität mit 76,8% (95%-KI: 64,7-86,5%) höher. Der NPW betrug 95,6% und der PPW 64,9% (Tabelle 14).

Da die mHE-Prävalenz in der LZI-Kohorte der eNCT-Studie laut PHES 31,3% betrug und in einer vergleichbaren deutschen Studie bei 25% lag, wurden die prävalenzabhängigen Vorhersagewerte für den eNCT 10“ bzw. eNCT 18“ ergänzend mithilfe der Bayes-Formel (vgl. Abschnitt 5.5.1.4) für eine Prävalenz von 25% berechnet (117). Daraus resultierte ein PPW von 54,9% bzw. 57,0% und ein NPW von 98,3% bzw. 96,8%.

Die Wahrscheinlichkeitsverhältnisse für ein positives (über dem Grenzwert liegendes) Testergebnis (=LR(+)) bzw. für ein negatives Testergebnis (=LR(-)) lagen im eNCT 10“ bei 3,65 bzw. 0,05 und im eNCT 18“ bei 3,98 bzw. 0,10, womit nur die negativen Likelihood-Quotienten eine hohe bis sehr hohe Evidenz erreichten. Der Quotient von LR(+) und LR(-), die diagnostische Odds-Ratio (DOR), war im eNCT 10“ mit einem Wert von 70,0 größer als im eNCT 18“ mit einem Wert von 39,7 (Tabelle 14).

Für Kappa (κ) als Maß der Zuverlässigkeit der eNCT-Testresultate zur mHE-Diagnostik in Bezug auf den PHES (Interrater-Reliabilität) ergaben sich für den eNCT 10“ mit $\kappa=0,609$ und den eNCT 18“ mit $\kappa=0,621$ sehr hohe Werte (Tabelle 14). Die Interrater-Reliabilität zwischen eNCT 10“ und eNCT 18“ fiel mit einem Kappa-Wert von 0,702 ebenso sehr hoch aus.

Tabelle 14: Kriterien der Testgenauigkeit (eNCT) zur Detektion einer mHE

	mHE+keine mHE eNCT 10“ (N=83)	mHE+keine mHE eNCT 18“ (N=82)
Grenzwert	38 Sekunden	41 Sekunden
SENS [95%-KI]	96,2% [84,1-99,8%]	92,3% [78,1-98,7%]
SPEZ [95%-KI]	73,7% [61,4-83,9%]	76,8% [64,7-86,5%]
PPW [95%-KI]	62,5%* [47,1-76,4%]	64,9%** [48,9-78,9%]
NPW [95%-KI]	97,7%* [90,2-99,9%]	95,6%** [86,9-99,2%]
Präzision	80,7%	81,7%
Kappa (p-Wert)	0,609 (p<0,001)	0,621 (p<0,001)
LR(+) [95%-KI]	3,65 [2,35- 5,68]	3,98 [2,44-6,49]
LR(-) [95%-KI]	0,05 [0,01-0,36]	0,10 [0,03-0,38]
DOR [95%-KI]	70,0 [8,7-562,5]	39,7 [8,3-190,8]

*prävalenzkorrigierter Wert: PPW=54,9%, NPW=98,3%. **prävalenzkorrigierter Wert: PPW=57,0%, NPW=96,8%. SENS: Sensitivität; SPEZ: Spezifität; PPW: positiv-prädiktiver Wert; NPW: negativ-prädiktiver Wert; LR: Likelihood-Quotient; DOR: diagnostische Odds-Ratio; mHE: minimale hepatische Enzephalopathie; eNCT: elektronischer Zahlenverbindungstest; N: Anzahl; KI: Konfidenzintervall.

6.4 Wiederholungsgenauigkeit des eNCTs

Zur Beurteilung der Wiederholungsgenauigkeit des eNCTs in der Zielgruppe wurden die Testergebnisse der Probanden der LZI-Kohorte analysiert, die den PHES an zwei verschiedenen Zeitpunkten jeweils vollständig durchgeführt haben.

Von den 83 Probanden der LZI-Kohorte wiederholten 25 den kompletten Testdurchlauf ein zweites Mal mit einem zeitlichen Abstand von mindestens zwei bis maximal 134 Tagen (Median: 7 Tage, \bar{x} : 34 Tage). Die intraindividuellen Unterschiede der Testzeiten zwischen Zeitpunkt 1 und Zeitpunkt 2 waren in der mHE-Kohorte mit durchschnittlich 6,6 (eNCT 10“) bzw. 3,2 (eNCT 18“) Sekunden größer als in der Gruppe ohne mHE (0,9 bzw. 3,1 Sekunden). Für den IKK als Maß der Wiederholungsgenauigkeit konnte in beiden eNCT-Versionen ein höherer Wert für die Gruppe ohne mHE (eNCT 10“: 0,857; eNCT 18“: 0,508) als für die Gruppe mit mHE (eNCT 10“: 0,659; eNCT 18“: 0,438) berechnet werden. Wurden die beiden Kohorten zusammengefasst (*mHE+keine mHE*), nahm der IKK im eNCT 10“ mit 0,906 (95%-KI: 0,789-0,959) einen sehr guten und im eNCT 18“ mit 0,855 (95%-KI: 0,672-0,936) einen guten Wert an (Abbildung 19).

Die Anwendung des Grenzwerts an den Wiederholungsergebnissen (N=25, 7x mHE) zeigte für beide eNCT-Versionen eine Sensitivität und einen NPW von 100%, eine Spezifität von 64,7% (95%-KI: 41,1-84,2%), einen PPW von 57,1% (95%-KI: 31,7-80,2%) und eine Gesamtübereinstimmung von 76,0%.

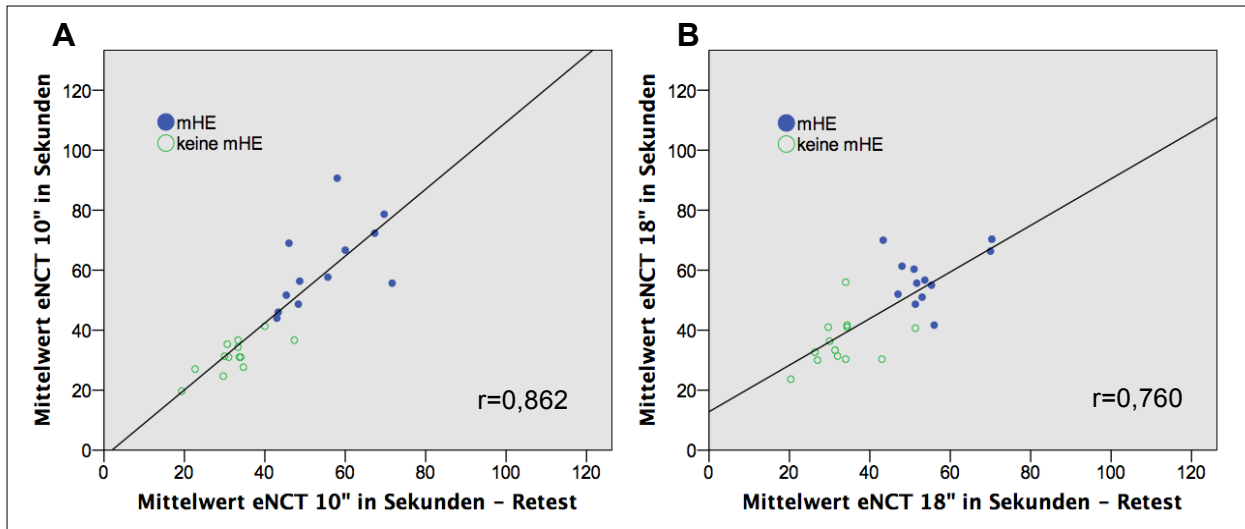


Abbildung 19: Streudiagramme der Test-Retest-Reliabilitätsanalyse vom eNCT 10" (A) und vom eNCT 18" (B) – Jeder Kreis steht für die Testergebnisse eines Probanden bei Einschluss (y-Achse) und bei Nachbeobachtung (x-Achse), die diagonale Linie entspricht der Regressionsgeraden. (eNCT: elektronischer Zahlenverbindingstest; r=Pearson-Korrelationskoeffizient; mHE: minimale hepatische Enzephalopathie)

6.5 Veränderungssensitivität des eNCTs

Zur Untersuchung der Veränderungssensitivität, also der Fähigkeit des eNCTs, Änderungen der kognitiven Leistungsfähigkeit im Verlauf zu erkennen, wurden von allen Patienten mit Leberzirrhose und mHE oder oHE (N=38) diejenigen ausgewählt, deren HE-Status beim Wiederholungstest besser als zum Zeitpunkt des Einschlusses war (N=8). Als Erklärung der Gruppengröße von N=8 sei darauf hingewiesen, dass nur fünf der Patienten mit oHE die komplette PHES-Testbatterie absolvieren konnten, von denen vier im Verlauf klinische Symptombefreiheit und damit eine Verbesserung des HE-Schweregrads zeigten. Zusätzlich erreichten vier der sieben Patienten mit mHE und Nachbeobachtung im Wiederholungstest einen PHES oberhalb des Grenzwerts und damit eine Verbesserung des HE-Schweregrads. An dieser Gruppe aus insgesamt acht Patienten wurden die Unterschiede in den eNCT-Testzeiten zwischen Einschluss und Nachbeobachtung mit den Unterschieden im PHES verglichen (Abbildung 20).

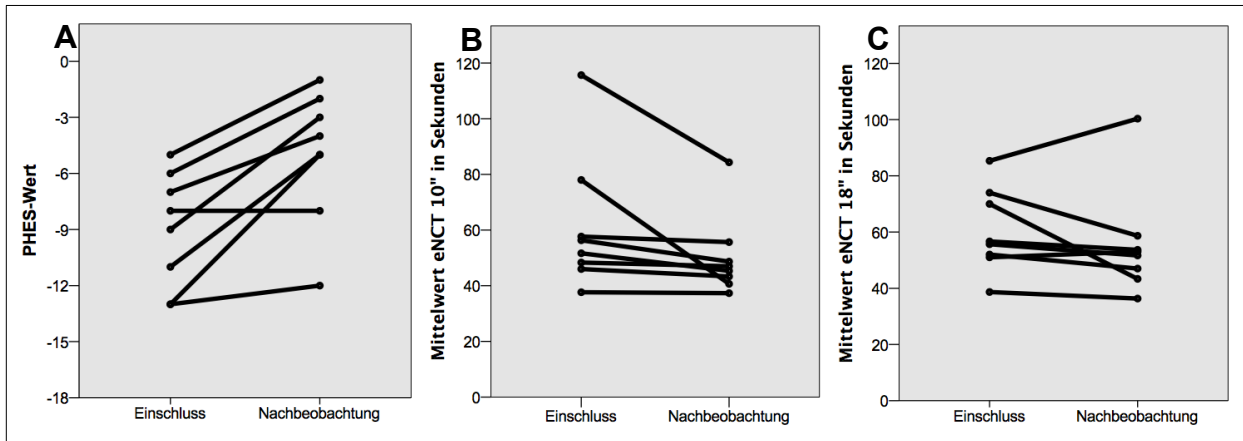


Abbildung 20: Vergleich der Testwertveränderungen von Probanden mit oHE oder mHE und Verbesserung der HE-Diagnose im Verlauf (N=8) für den PHES (A), den eNCT 10'' (B) und den eNCT 18'' (C) – Ausgangspunkt einer Linie entspricht Testwert eines Probanden bei Einschuss, Endpunkt der Linie entspricht Testwert des Probanden im Verlauf. (PHES: Psychometric Hepatic Encephalopathy Score; eNCT: elektronischer Zahlenverbindungstest)

Im Wilcoxon-Rangtest konnte mit $p=0,018$ ein signifikanter Unterschied zwischen PHES-Gesamtwert bei Einschuss und bei Nachbeobachtung festgestellt werden, der mit $p=0,012$ auch für den eNCT 10'' nachweisbar war. Im eNCT 18'' allerdings konnte mit einem p-Wert von 0,123 nur eine Tendenz zur Verbesserung festgestellt werden:

$\bar{x} \pm SD$ (PHES): $-9 \pm 3,1$ vs. $-5 \pm 3,5$ (Einschluss vs. Nachbeobachtung)

$\bar{x} \pm SD$ (eNCT 10''): $61,4s \pm 24,9s$ vs. $50,3s \pm 14,8s$

$\bar{x} \pm SD$ (eNCT 18''): $60,4s \pm 15,0s$ vs. $55,5s \pm 19,4s$.

Da einer der acht Patienten, deren HE-Schweregrad sich bis zur Nachbeobachtung verbessert hatte und deren Testergebnisse in Abbildung 20 gezeigt sind, dennoch bei Nachbeobachtung im eNCT 18'' ein deutlich schlechteres Resultat erzielte als bei Einschuss und dieser Patient ebenso im eNCT 10'' und im PHES mit den schlechtesten Werten auffiel, wurden unter Ausschluss des genannten Patienten erneut Berechnungen zur Veränderungssensitivität durchgeführt. In diesem Fall zeigten auch die Ergebnisse im eNCT 18'' mit $p=0,028$ einen signifikanten Unterschied zwischen Einschuss und Nachbeobachtung. Die Mittelwerte beider eNCTs waren dann um fünf bis zehn Sekunden, die Standardabweichungen um drei bis zwölf Sekunden geringer.

Keiner der vier mHE-Patienten, von denen bei Einschuss jeder mit eNCT-Testzeiten auffiel, die langsamer als der Grenzwert waren und die im Verlaufstest alle einen PHES ≥ -4 erzielten (=keine mHE), verbesserte sich in einem der beiden eNCT-Versionen auf Testzeiten unterhalb des Grenzwerts.

6.6 Einflussfaktoren auf die eNCT-Testzeiten von Patienten

Durch eine lineare, schrittweise Regressionsanalyse in der LZI-Untergruppe *keine mHE* wurde untersucht, welche patientenspezifischen Faktoren zusätzlich zum HE-Status die Testzeiten beeinflussen.

Neben den Variablen „Alter“, „Ausbildung“ und „Geschlecht“ wurden die patientenspezifischen Variablen „Aszites“, „HE-prophylaktische Medikation“, „vorige oHE“ und der „Einschlussstatus“ (ambulant oder stationär) in die Regressionsanalyse miteinbezogen. In das endgültige Regressionsmodell wurde nur die Variable „Alter“ (eNCT 10“: Beta=0,351, p=0,007; eNCT 18“: Beta=0,360, p=0,006) aufgenommen, da die anderen Variablen keinen signifikanten Einfluss auf das Regressionsmodell zeigten. Die angepassten Bestimmtheitsmaße R^2 der Modelle betragen 0,107 im eNCT 10“ (p=0,007, d=1,429, VIF=1,0) und 0,113 im eNCT 18“ (p=0,006, d=1,813, VIF=1,0). Die p-Werte lagen für die ausgeschlossenen Koeffizienten zwischen 0,319 und 0,946 und damit deutlich über dem Signifikanzniveau α . Es resultierten folgende Gleichungen mit einem Standardfehler der Schätzung von 6,05 (eNCT 10“) bzw. 7,51 (eNCT 18“):

$$18,213 + 0,279 \times (\text{Alter in Jahren}) \text{ für den eNCT 10" und} \\ 15,290 + 0,370 \times (\text{Alter in Jahren}) \text{ für den eNCT 18"}$$

Entsprechend errechneten sich für die Variable „Ausbildung“ nicht signifikante (eNCT 10“: p=0,41; eNCT 18“: p=0,41) und im Vergleich mit der Variable „Alter“ geringere Korrelationskoeffizienten (eNCT 10“ bzw. eNCT 18“: r=-0,112 bzw. r=-0,111 vs. r=0,412 bzw. r=0,439; Abbildung 21A-B).

Der fehlende Einfluss der übrigen ausgeschlossenen Variablen auf die Testzeiten zeigte sich auch durch sehr kleine *Effektstärken* r_e und nicht signifikante Unterschiede der Testzeiten zwischen den jeweiligen Gruppen im T-Test, der bei Normalverteilung der Ergebnisse verwendet wurde („Einschlussstatus“, „HE Prophylaxe“ und „Aszites“ im eNCT 10“), bzw. im U-Test (Abbildung 21C-F).

Bei fast identischen Werten im eNCT 10“ („Einschlussstatus“: p=0,807, $r_e=0,031$; „vorige oHE“: p=0,441, $r_e=0,108$; „HE Prophylaxe“: p=0,674, $r_e=0,078$; „Aszites“: p=0,295, $r_e=0,145$) beschränkt sich Abbildung 21 exemplarisch auf die Darstellung der Daten im eNCT 18“.

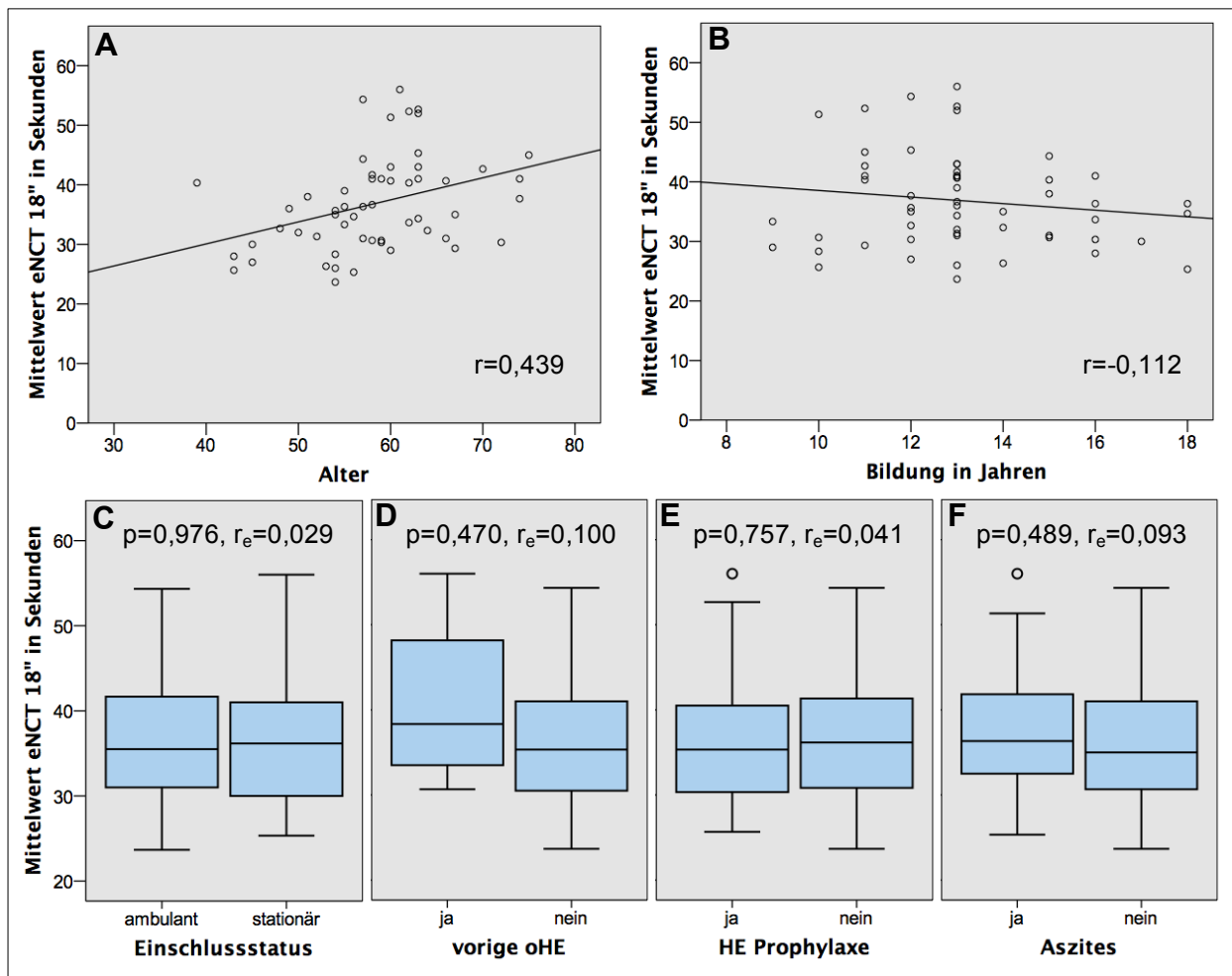


Abbildung 21: Die eNCT 18"-Testzeiten einzelner Probanden (Kreise in A und B) sind in Abhängigkeit von den in der Regressionsanalyse untersuchten Variablen Alter (A) und Bildung (B) als Streudiagramm sowie von den Variablen Einschlussstatus (C), vorige oHE (D), HE Prophylaxe (E) und Aszites (F) kohortenweise als Kastengrafik gezeigt – das Alter zeigte als einzige Variable statistisch signifikanten Einfluss auf die eNCT-Testzeiten. Die diagonale Linie in A und B entspricht der Regressionsgeraden. (eNCT: elektronischer Zahlenverbindungstest; oHE: offensichtliche hepatische Enzephalopathie; r: Spearman-Korrelationskoeffizient; r_e : Effektstärke)

6.6.1 Abhängigkeit der eNCT-Testzeiten von Laborwerten

Die Leberzirrhose spiegelte sich in den gemessenen Blutwerten der Patienten insbesondere im Vergleich mit lebergesunden Probanden wider (siehe Anhang 1). Die Werte des MELD-Scores lagen in der LZI- bzw. der oHE-Kohorte durchschnittlich 4,7 bzw. 11 Punkte über den Werten der ≥ 39 J-Kohorte. Der Child-Pugh-Score konnte für 90 Patienten mit Leberzirrhose ausgerechnet werden, 30 Patienten wurden als Grad A, 41 Patienten als Grad B und 19 Patienten als Grad C eingestuft. Die Ammoniakwerte waren in der LZI- bzw. oHE-Kohorte um das 2-fache bzw. 2,4-fache erhöht.

Um den Einfluss des laborchemisch durch MELD- und Child-Pugh-Score feststellbaren Schweregrads der Lebererkrankung auf eNCT-Testzeiten zu überprüfen, wurde eine Regressionsanalyse für diese Parameter in der LZI-Untergruppe *keine mHE* durchgeführt. Zusätzlich wurden die Laborwerte von Ammoniak und Natrium auf einen möglichen Einfluss hin untersucht, da Ammoniak in direktem Zusammenhang mit der Pathogenese der HE steht und extreme Natriumwerte im Sinne einer Elektrolytstörung als differenzialdiagnostisches Kriterium gelten. Die genannten Laborparameter nahmen Werte von physiologisch bis hochpathologisch an (Tabelle 15). Die Spearman-Korrelationsanalyse ergab für die untersuchten Laborwerte (Natrium, Ammoniak, Child-Pugh-Score und MELD-Score) jeweils nicht-signifikante Korrelationskoeffizienten im eNCT 10“ (r: -0,029, -0,110, 0,119 und 0,160) bzw. eNCT 18“ (Abbildung 22).

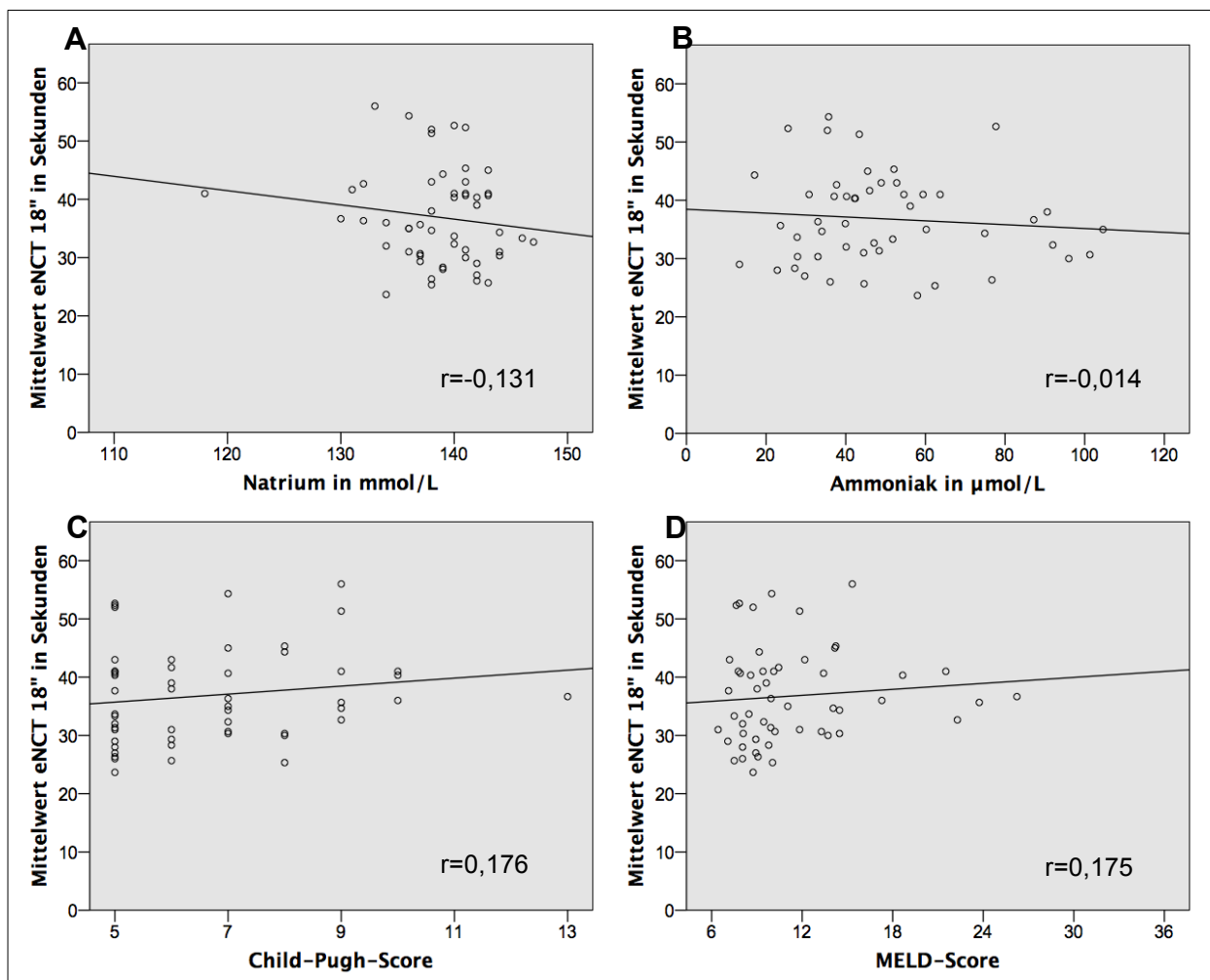


Abbildung 22: Streudiagramme der Korrelationsanalysen zwischen eNCT 18“ und Natrium (A), Ammoniak (B), Child-Pugh-Score (C) bzw. MELD-Score (D) – Jeder Kreis steht für die Testergebnisse eines Probanden, die diagonale Linie entspricht der Regressionsgeraden. (eNCT: elektronischer Zahlenverbindungstest; MELD: Model for End-Stage Liver Disease; r: Spearman-Korrelationskoeffizient)

Tabelle 15: Laborwerte der LZI-Untergruppe *keine mHE*

	N	Mittelwert ± SD	Median	Minimum	Maximum
MELD-Score	54	11,4 ± 4,5	9,9	6,4	26,2
Child-Pugh-Score	53	7 ± 2	6	5	13
Ammoniak (µmol/L)	51	49 ± 23	45	13	105
Natrium (mmol/L)	54	139 ± 5	139	118	147

Der MELD-Score wurde für die statistische Analyse nicht auf ganze Zahlen gerundet, sondern bei den exakten Werten belassen. N: Anzahl; SD: Standardabweichung; MELD: Model for End-Stage Liver Disease.

Die Überprüfung der Parameter als mögliche Einflussfaktoren auf die Testzeiten im eNCT führte mit Bestimmtheitsmaßen R^2 zwischen 0,002 und 0,015 (eNCT 10“) bzw. 0,005 und 0,024 (eNCT 18“) und zugehörigen p-Werten zwischen 0,390 und 0,696 (eNCT 10“) bzw. 0,272 und 0,517 (eNCT 18“) nicht zur Aufnahme weiterer Variablen in das Regressionsmodell.

6.7 Vergleich von Stroop-Test mit PHES und eNCT

6.7.1 Korrelationen zwischen Stroop-Test und eNCT bzw. PHES

Für eine Untersuchung auf mögliche Zusammenhänge zwischen den Resultaten im eNCT und im Stroop-Test sowie einen Vergleich der beiden Tests bezüglich ihrer Korrelation mit der PHES-Testbatterie bei Patienten mit Leberzirrhose und potenzieller mHE (also ohne oHE) wurde eine Korrelationsanalyse in der LZI-Kohorte durchgeführt.

Eine hohe ($r > 0,7$) bzw. annähernd hohe Korrelation konnte in der LZI-Untergruppe *mHE+keine mHE* zwischen den Ergebnissen im Stroop-Test und den Ergebnissen im eNCT 18“ ($r = 0,755$) bzw. eNCT 10“ ($r = 0,692$) gemessen werden (Abbildung 23).

Die stärksten Korrelationen zwischen Stroop-Test und PHES-Einzeltests wurden unabhängig von der untersuchten Kohorte (*mHE*, *keine mHE* oder *mHE+keine mHE*) für den NCT-A, NCT-B und DST errechnet (Ausnahme bei *keine mHE*: Korrelation mit NCT-B kleiner als mit SDT, Tabelle 16). Eine hohe Korrelation mit den Ergebnissen im Stroop-Test zeigte nur der DST für *mHE+keine mHE* ($r = -0,732$), gefolgt vom NCT-A ($r = 0,671$) und NCT-B ($r = 0,616$). Die Korrelationen (vgl. Abschnitt 6.3.1) vom NCT-A mit dem eNCT 10“ ($p = 0,04$) bzw. vom NCT-B mit dem eNCT 10“ ($p < 0,001$) und eNCT 18“ ($p = 0,01$) waren signifikant höher als mit dem Stroop-Test. Der Korrelationskoeffizient für

Stroop-Test und PHES-Gesamtwert, durch den die Validität des Stroop-Tests mit beurteilt werden kann, lag für *mHE+keine mHE* mit $r=-0,680$ im moderaten Bereich. Damit fällt die Korrelation des PHES mit dem Stroop-Test nicht-signifikant ($p=0,234$ bzw. $p=0,485$) niedriger aus als die Korrelation mit dem eNCT 10“ bzw. eNCT 18“.

Der Stroop-Test wiederum zeigte nur beim Vergleich mit dem LNT-Z-Wert eine signifikant bessere Korrelation als der eNCT 10“ ($p=0,012$) und der eNCT 18“ ($p=0,003$).

Tabelle 16: Spearman-Rangkorrelationskoeffizienten (Stroop-Test)

	mHE (N=26)	Keine mHE (N=57)	mHE+keine mHE (N=83)
NCT-A	$r=0,573$ ($p=0,051$)	$r=0,493$ ($p=0,001$)	$r=0,671$ ($p<0,001$)
NCT-B	$r=0,469$ ($p=0,124$)	$r=0,426$ ($p=0,007$)	$r=0,616$ ($p<0,001$)
DST	$r=-0,499$ ($p=0,099$)	$r=-0,596$ ($p<0,001$)	$r=-0,732$ ($p<0,001$)
LNT-F	$r=-0,109$ ($p=0,737$)	$r=0,363$ ($p=0,023$)	$r=0,253$ ($p=0,073$)
LNT-Z	$r=-0,161$ ($p=0,618$)	$r=0,262$ ($p=0,107$)	$r=0,511$ ($p<0,001$)
SDT	$r=0,326$ ($p=0,301$)	$r=0,461$ ($p=0,003$)	$r=0,551$ ($p<0,001$)
PHES	$r=-0,493$ ($p=0,103$)	$r=-0,509$ ($p=0,001$)	$r=-0,680$ ($p<0,001$)
eNCT 10“	$r=0,490$ ($p=0,106$)	$r=0,499$ ($p=0,001$)	$r=0,692$ ($p<0,001$)
eNCT 18“	$r=0,739$ ($p=0,006$)	$r=0,641$ ($p<0,001$)*	$r=0,755$ ($p<0,001$)**

*N=56; **N=82; NCT-A/B: Zahlenverbindungstest A/B; DST: Zahlensymboltest; LNT-F: Liniennachfahrttest Fehlerpunkte; LNT-Z: Liniennachfahrttest Zeit; SDT: Kreise-punktieren-Test; PHES: Psychometric Hepatic Encephalopathy Score; eNCT: elektronischer Zahlenverbindungstest; mHE: minimale hepatische Enzephalopathie; r: Spearman-Korrelationskoeffizient; N: Anzahl.

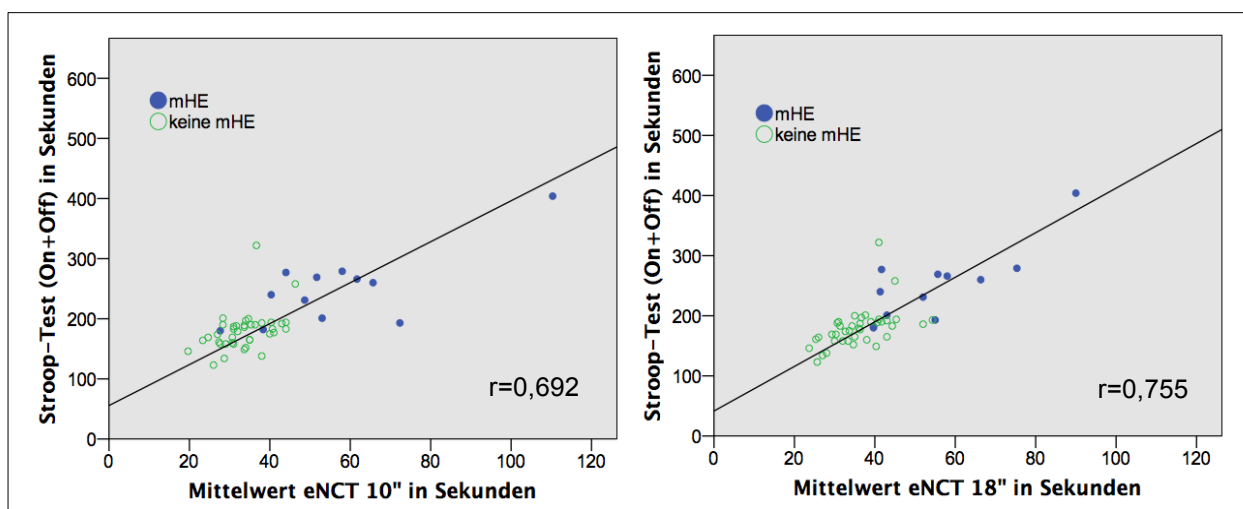


Abbildung 23: Streudiagramme der Korrelationsanalysen zwischen Stroop-Test und eNCT – Jeder Kreis steht für die Testergebnisse eines Probanden im Stroop-Test (y-Achse) und im jeweiligen eNCT (x-Achse), die diagonale Linie entspricht der Regressionsgeraden. (mHE: minimale hepatische Enzephalopathie; r: Spearman-Korrelationskoeffizient; eNCT: elektronischer Zahlenverbindungstest)

6.7.2 Testgenauigkeit des Stroop-Tests

Wie für den eNCT (vgl. Abschnitt 6.3.2) wurden die Grenzwerte des Stroop-Tests von 190 Sekunden für Probanden ≥ 45 Jahren (horizontale rote Linie in Abbildung 15) bzw. 145 Sekunden für Probanden < 45 Jahren an der LZI-Kohorte angewandt und die resultierende Aufteilung in *mHE* und *keine mHE* mithilfe einer Kreuztabelle mit der PHES-Klassifikation verglichen.

Mithilfe des Stroop-Tests konnten 78,4% (N=40/51) der Fälle richtig klassifiziert werden. Eine mHE wurde bei 21 von 51 Patienten diagnostiziert, was einer relativen Zunahme an mHE-Diagnosen im Vergleich zum PHES (N=12/51) von 17,7 Prozentpunkten entspricht. Eine falsch-negative Diagnose wurde in einem Fall gestellt (8,3%).

Insgesamt fielen die Sensitivität um 4,5, der NPW um 1,0 und der PPW um 10,1 Prozentpunkte im Stroop-Test niedriger, die Spezifität um 0,7 Prozentpunkte höher aus als im eNCT 10“. Im Vergleich mit dem eNCT 18“ waren die Sensitivität um 0,6, die Spezifität um 2,4 und der PPW um 12,5 Prozentpunkte niedriger, der NPW um 1,1 Prozentpunkte höher. Die Wahrscheinlichkeitsverhältnisse LR(+) bzw. LR(-) zeigten mit 3,61 bzw. 0,22 ebenso wie die resultierende DOR (16,7) schlechtere Werte für den Stroop-Test als für beide eNCT-Versionen. Die Interrater-Reliabilität war bei einem Kappa-Wert von 0,524 moderat und geringer als beim eNCT (Tabelle 17).

Tabelle 17: Kriterien der Testgenauigkeit (Stroop-Test) zur Detektion einer mHE

Stroop-Test (N=51)	
SENS [95%-KI]	91,7% [68,1-99,5%]
SPEZ [95%-KI]	74,4% [59,4-86,2%]
PPW [95%-KI]	52,4%* [31,7-72,5%]
NPW [95%-KI]	96,7%* [86,1-99,8%]
Präzision	78,4%
Kappa (p-Wert)	0,524 (p<0,001)
LR(+) [95%-KI]	3,58 [2,04-6,27]
LR(-) [95%-KI]	0,11 [0,02-0,74]
DOR [95%-KI]	31,9 [3,6-279,3]

*prävalenzkorrigierte Werte: PPW=55,8%, NPW=96,4%. SENS: Sensitivität; SPEZ: Spezifität; PPW: positiv-prädiktiver Wert; NPW: negativ-prädiktiver Wert; LR: Likelihood-Quotient; DOR: diagnostische Odds-Ratio; N: Anzahl; KI: Konfidenzintervall.

7 Diskussion

Ziele der vorliegenden Arbeit waren die Charakterisierung des neuentwickelten eNCTs und die Bestimmung seiner Wertigkeit in der mHE-Diagnostik bei Patienten mit Leberzirrhose. Damit sollten wichtige Voraussetzungen für einen Einsatz als Screeningtest geprüft werden, der - eingebunden in ein häusliches Überwachungssystem der Leberfunktion (vgl. Abschnitt 5.1) - der Kontrolle der kognitiven Leistungsfähigkeit dienen soll, um bei Verschlechterung frühzeitig und effektiv intervenieren zu können.

7.1 Diskussion der Methodik

Für einen aussagekräftigen Vergleich eines neuen Testverfahrens in der HE-Diagnostik mit einem Referenztest wurden durch die *internationale Gesellschaft für hepatische Enzephalopathie und Stickstoffmetabolismus* (ISHEN) Vorgaben zur Studienplanung und zum methodischen Aufbau einer Studie formuliert, deren Umsetzung in der eNCT-Studie nachfolgend diskutiert wird:

1. Die gewählte Kontrollgruppe sollte in den wesentlichen demografischen Variablen (Alter, Geschlecht und Bildung) mit der Patientenkohorte übereinstimmen (60).

Die Zusammensetzung der LZI-Kohorte unterschied sich nicht im Hinblick auf das Alter und die Bildung von der ≥ 39 J-Kohorte. Der Anteil männlicher Probanden in der LZI-Kohorte lag zwar 12,2 Prozentpunkte über dem Anteil in der ≥ 39 J-Kohorte, der Unterschied war allerdings nicht signifikant. Da das Geschlecht in vorigen Studien als Einflussfaktor auf etablierte psychometrische Tests ohnehin ausgeschlossen werden konnte, wurde auf Vorgaben zum Verhältnis der Geschlechter verzichtet (67, 99).

2. Zu vergleichende Tests sollten an einem Zeitpunkt nacheinander absolviert werden, von der Kontrollgruppe in mindestens einwöchigem Abstand ein zweites Mal (60).

In der eNCT-Studie sollte jeder Studienteilnehmer an einem Zeitpunkt alle zu vergleichenden Tests durchführen. Teilweise konnten lebergesunde Kontrollprobanden und Patienten mit Leberzirrhose auch für einen Wiederholungstest im Verlauf rekrutiert werden. Für das Zeitintervall zwischen Einschluss und Nachbeobachtung wurde bei der Planung der Studie auf die Festlegung von Grenzwerten verzichtet, da die Fachliteratur keine eindeutigen Vorgaben lieferte. Ein Mindestabstand von einem Tag wurde in einer vergleichbaren Studie von Gualtieri et al. akzeptiert, in den Leitlinien zur neuropsychologischen Diagnostik der HE von Randolph et al. wird ein Intervall von einer Woche gefordert und an anderer Stelle darf das Intervall sogar ein Jahr betra-

gen (60, 111, 118). So ergab sich in der eNCT-Studie für die hier untersuchten Kohorten im Median ein Zeitraum zwischen Einschluss und Nachbeobachtung von sieben Tagen. Unterschiedliche Zeitintervalle waren auch aufgrund des Studiendesigns unvermeidbar, da Probanden erst im Rahmen ihrer ärztlichen Kontrolltermine erneut eingeschlossen werden konnten. Aussagen zum Vergleich der Wiederholungsgenauigkeit von PHES und eNCT wurden davon nicht beeinflusst, da der Test-Retest-Zeitraum für den jeweiligen Probanden in beiden Testverfahren gleich war.

3. Von den Patienten mit Leberzirrhose sollten so viele klinisch relevante Daten wie möglich erhoben werden (60).

Mit Verweis auf den Methodenteil dieser Arbeit lässt sich eine ausführliche Anamnese und körperliche Untersuchung der Studienteilnehmer belegen, deren Ergebnisse für jeden Probanden in einem standardisierten CRF-Protokoll festgehalten wurden.

4. Mithilfe einer ROC-Analyse sollten optimale Grenzwerte für den neuen Test bestimmt werden (60).

Bei der Auswertung der eNCT-Studie wurde eine ROC-Analyse zur Grenzwertoptimierung durchgeführt, wobei geeignete Grenzwerte für eine Differenzierung zwischen normalen und pathologischen Testzeiten vom Schwerpunkt der jeweiligen Fragestellung abhängen (111). Da von Wuensch et al. bereits gezeigt werden konnte, dass der eNCT auf dem 10“-Bildschirm durchschnittlich um 1,5 Sekunden schneller zu absolvieren ist als auf dem 18“-Bildschirm, wurden für die beiden Testversionen zwei verschiedene Grenzwerte und Regressionsmodelle berechnet (98).

5. Die statistischen Eigenschaften der verschiedenen Testverfahren sollten gegen ihre praktische Anwendbarkeit abgewogen werden, um für jede Anwendungssituation den passenden Test bestimmen zu können (60).

Während in dieser Dissertation eine erste Einschätzung von Zeit- und Arbeitsaufwand erfolgte, müssen Normwerte und Daten zur Kostenschätzung und Eignung im häuslichen Umfeld durch weitere Studien gewonnen werden.

Das untersuchte Patientenkollektiv setzte sich aus 61 männlichen und 39 weiblichen Patienten mit Leberzirrhose zusammen. Ein solches Verhältnis wurde auch in anderen Studien berichtet (Tabelle 18) und spiegelt die geschlechtsspezifische Prävalenz der Leberzirrhose in der Gesamtbevölkerung wider, die durch Unterschiede in der Häufigkeit von Komorbiditäten der Leber und der Einnahme hepatogener Noxen begünstigt wird. So konnten Ferenci et al. in ihrer Kohortenstudie für ein männliches Kollektiv mit

HCV-Infektion eine höhere Inzidenz der Leberzirrhose nachweisen als Wiese et al. in einer ähnlichen Studie für ein weibliches Kollektiv mit HCV-Infektion (117, 119, 120).

Während neben dem Verhältnis männlicher zu weiblicher Patienten auch die durchschnittliche Ausbildungszeit der Patienten in der eNCT-Studie vergleichbar ist mit entsprechenden Kohorten anderer Studien zur Validierung von psychometrischen Testverfahren der HE, war der Altersdurchschnitt 2-16 Jahre höher als in diesen Studien (Tabelle 18). Eine mögliche Ursache ist die Rekrutierung von Patienten über die Leberambulanz, in der die Diagnose vieler Patienten zum Zeitpunkt des Einschlusses bereits langjährig bekannt war.

In der Validierungsstudie zum PHES wurden Patienten mit zurückliegendem Alkoholabusus ausgeschlossen (67). Zum Einschluss in die eNCT-Studie wurde eine äthyltoxische ebenso wie eine virale Genese der Leberzirrhose hingegen akzeptiert, weil gezeigt werden konnte, dass die Häufigkeit einer mHE unabhängig von der Ätiologie der Leberzirrhose ist (7, 59, 121). Dementsprechend ist der Anteil dieser Diagnosen bei den Patienten der eNCT-Studie wie auch bei vergleichbaren Studien führend (Tabelle 18). Der mittlere Schweregrad (MELD-Score) der Lebererkrankung eingeschlossener Patienten

Tabelle 18: Gegenüberstellung verschiedener Validierungsstudien zur HE-Diagnostik

	eNCT-Studie	Ennen (69)	Bajaj et al. (78)	Sharma et al. (122)	Allampati et al. (123)
LZI-Probanden (N)	100	88	167	50	437
Alter ($\bar{x} \pm$ SD Jahre)	59 \pm 8	45 \pm 14	55 \pm 7	43 \pm 10	57 \pm 8
Geschlecht (m:w)	61:39	59:41	71:29	42:8	64:36
Ausbildung ($\bar{x} \pm$ SD Jahre)	13 \pm 2	-	14 \pm 2	12 \pm 5	13 (12-14)
Genese: C2	37%	-	26%	40%	26%
HCV/HBV	21%	50%	35%	40%	43%
Sonstige	42%	50%	39%	20%	31%
Schweregrad: MELD	12,6	-	12	13,8	11,1
Child A	33%	34%	-	22%	-
Child B	46%	41%	-	50%	-
Child C	21%	25%	-	28%	-
HE-Status*: mHE	26%	22%	39%	53%	37%
oHE	12%	28%	38%	-	-

*: bezogen auf LZI-Kohorte + oHE-Kohorte; N: Anzahl; \bar{x} : Mittelwert; SD: Standardabweichung; m: männlich; w: weiblich; C2: Alkohol; HCV: Hepatitis-C-Virus; HBV: Hepatitis-B-Virus; MELD: Model for End-Stage Liver Disease; mHE/oHE: minimale / offensichtliche hepatische Enzephalopathie.

bzw. die Häufigkeit der Schweregrade A-C (Child-Pugh-Score) und die Prävalenz der mHE lagen ebenso im Bereich analoger Studien (69, 78, 122, 123).

Von einer Fallzahlplanung vor Studienbeginn wurde von dem beratenden Institut für Biometrie bei geringer Aussagekraft im Rahmen von Pilotstudien abgeraten.

Als Referenztest wurde in der eNCT-Studie die PHES-Testbatterie verwendet, da sie von den Leitlinien zur HE von Vilstrup et al. als einfach durchzuführender Bestandteil der mHE-Diagnostik mit guter externer Validität empfohlen wird, die im Vergleich zu anderen Testverfahren (Stroop-Test, CRT, ICT) bereits durch eine Vielzahl von Studien länderübergreifend für diesen Anwendungsbereich bestätigt werden konnte (5, 66, 124-128). Der PHES wurde gezielt zur mHE-Diagnostik entwickelt und ermöglicht eine Abgrenzung von anderen Ursachen kognitiver Dysfunktionen (67, 129).

Der CFF-Test als weiteres validiertes Testverfahren wurde nicht als Referenztest eingesetzt, da zum Zeitpunkt der eNCT-Studie keine Altersnormierung der Testwerte möglich war (73, 129). Die EEG ist als eine unspezifische Methode zur Diagnose kognitiver Dysfunktionen vorbeschrieben und stellte damit für die Verwendung in der eNCT-Studie ebenfalls keine Option dar (76). Der RBANS wurde in Abschnitt 3.5.4 als gleichwertige Alternative zum PHES vorgestellt. Das Testverfahren wurde allerdings nicht in die Leitlinien aufgenommen und aufgrund fehlender Normwerte für Deutschland auch nicht in die eNCT-Studie miteinbezogen (5, 71, 129).

Schlussfolgerungen:

Die eNCT-Studie, auf der diese Dissertation aufbaut, erfüllt die wesentlichen methodischen Vorgaben für einen Vergleich des eNCTs mit einem geeigneten Referenztest (PHES). Die Zusammensetzung des Patientenkollektivs der eNCT-Studie ist vergleichbar mit den Patientenkollektiven anderer Studien zur Evaluation von HE-Diagnoseverfahren. Die Studienergebnisse können daher miteinander verglichen werden.

7.2 Diskussion psychometrischer Testergebnisse

7.2.1 Psychometric Hepatic Encephalopathy Score (PHES)

Die signifikanten Unterschiede im PHES-Gesamtwert zwischen den Gruppen $\geq 39J$ und *LZI* bestätigten eine Beeinträchtigung der kognitiven Leistungsfähigkeit bei Patienten mit Leberzirrhose. Die Ergebnisse der Patienten mit Leberzirrhose, die bereits klinisch

mit einer oHE auffällig waren, waren dagegen überraschenderweise nicht signifikant schlechter als die Ergebnisse von Patienten ohne stattgehabte oHE. Eine Ursache dafür ist in der sehr geringen Anzahl an Patienten mit oHE zu sehen, die den PHES überhaupt absolvieren konnten (N=5/12). Jeder der fünf oHE-Patienten mit einem PHES-Resultat erzielte ein Ergebnis unterhalb des Grenzwerts. Die sieben übrigen Patienten mit oHE, die die Testbatterie nicht vollständig bearbeiten konnten, sind damit auch diejenigen, die am ehesten noch schlechtere Punktwerte erreichen und den Unterschied in den Testresultaten zu Patienten ohne oHE vergrößern würden.

Der PHES wurde (wie auch der eNCT) zur Detektion der mHE entwickelt. Die Diskriminierungsfähigkeit des PHES zwischen Patienten mit mHE und oHE ist von anderen Studien als eingeschränkt vorbeschrieben (130). Die Graduierung der oHE-Patienten erfolgte daher in dieser Studie wie bei Kircheis et al. und in den Leitlinien zur HE mithilfe der WHC und zusätzlich wie bei Hassanein et al. mithilfe der HESA-Kriterien (5, 63, 130). Eine mHE-Diagnose wurde hingegen durch das PHES-Resultat festgelegt, sodass alle 26 Patienten der LZI-Kohorte mit einem PHES unterhalb des Grenzwerts per definitionem eine mHE aufwiesen. Einer von 57 lebergesunden Probanden (1,8%) der ≥ 39 J-Kohorte erreichte ebenfalls einen PHES unterhalb des Grenzwerts und damit ein falsch-positives Ergebnis, während in der Studie von Weissenborn et al. zur Charakterisierung des PHES keiner von 120 Kontrollprobanden ein falsch-positives Ergebnis erzielte. Der Mittelwert des Alters lag hier allerdings bei nur 40 Jahren im Gegensatz zu 59 Jahren der ≥ 39 J-Kohorte (67). Da die eNCT- und Stroop-Test-Ergebnisse des Probanden mit falsch-positivem PHES-Resultat unauffällig waren und keine Hinweise auf kognitive Funktionsstörungen (MMST: >25 Punkte) oder Verletzung anderer Einschlusskriterien vorlagen, konnten die Daten in die statistischen Analysen miteinbezogen werden.

Die größten Unterschiede zwischen der mHE-Kohorte und der Gruppe ohne mHE waren in den PHES-Einzeltests DST, NCT-B und NCT-A messbar. Kein signifikanter Unterschied zeigte sich im LNT-Fehlerpunktwert. In der Validierungsstudie des PHES erwies sich der LNT im Hinblick auf eine zuverlässige Differenzierung zwischen LZI und oHE ebenso als der schwächste Test (69).

Schlussfolgerungen:

Der PHES zeigte sich in der eNCT-Studie als geeignetes Verfahren zur Unterscheidung zwischen Patienten mit Leberzirrhose *mit* mHE und *ohne* mHE, wobei die Einzeltests in

Übereinstimmung mit den Ergebnissen voriger Studien unterschiedlich stark zur Diskriminierung beitragen.

7.2.2 Elektronischer Zahlenverbindungstest (eNCT)

Signifikante Unterschiede in den Testzeiten konnten für beide eNCT-Versionen wie beim PHES zwischen den Gruppen $\geq 39J$ und *LZI*, nicht aber zwischen *oHE* und *LZI* gemessen werden. Auch hier gilt die Limitation für die *oHE*-Kohorte, dass Ergebnisse nur von den Patienten vorlagen, die den Test trotz *oHE* noch absolvieren konnten (eNCT 10“: N=7, eNCT 18“: N=9). Nach PHES-basierter Aufteilung der *LZI*-Kohorte in *mHE* und *keine mHE* konnte unter Anwendung des eNCTs ein signifikanter Unterschied zwischen diesen beiden Gruppen gemessen werden.

Schlussfolgerungen:

Der eNCT erfasst den Unterschied der kognitiven Leistungsfähigkeit zwischen leberge-sunden Probanden und Patienten mit Leberzirrhose sowie zwischen Patienten mit Leberzirrhose *mit* *mHE* und *ohne* *mHE* und ermöglicht damit die Detektion der *mHE* in einer Gruppe von Patienten mit Leberzirrhose.

7.2.2.1 Diagnostische Wertigkeit des eNCTs

Im folgenden Abschnitt soll im Vergleich zum Stroop-Test, zur CFF und zum ICT, die jeweils ebenso als Patientenselbsttests in Betracht gezogen werden, diskutiert werden, wie zuverlässig und valide Veränderungen der eNCT-Testzeiten zur *mHE*-Diagnostik beitragen können und welche weiteren Einflussfaktoren auf die Testresultate existieren.

Testgenauigkeit und Validität:

Da der eNCT bis zur Fertigstellung dieser Dissertation noch nicht abschließend durch Experten bewertet wurde, muss die Inhaltsvalidität in zukünftigen Arbeiten beurteilt werden. Die Bestimmung der Konstruktvalidität des eNCTs erfolgte anhand von Korrelationsanalysen. Obwohl für den eNCT noch keine Alterskorrektur der Testwerte durchgeführt werden konnte, wurden hohe Korrelationen mit dem PHES berechnet. Damit gelang der Nachweis einer guten studieninternen Validität des eNCTs (101, 111). Eine zukünftige Alterskorrektur der eNCT-Testzeiten könnte die Korrelation noch verbessern.

In vergleichbaren Validierungsstudien zum Stroop-Test, zur CFF und zum ICT wurden jeweils nur Korrelationen mit ausgewählten Einzeltests der PHES-Testbatterie berechnet. Dabei zeigte lediglich der Stroop-Test hohe Korrelationen, und zwar wie der eNCT mit dem NCT-A, dem NCT-B und dem DST (72, 78, 131).

Als Maß der Kriteriumsvalidität diente die bei der ROC-Analyse bestimmte Fläche unter der Kurve (AUC), die für alle drei Zahlenverbindungstests (NCT-A, eNCT 10“ und eNCT 18“) einen Wert über 0,9 annahm. Das heißt, mit einer Wahrscheinlichkeit von über 90% wird ein Proband ohne mHE schnellere Testzeiten erreichen als ein Proband mit mHE. Bei einem theoretischen Höchstwert von 1,0 entspricht dieser Wert damit einer sehr hohen Klassifizierungsfähigkeit des eNCTs (113, 132).

Während in der eNCT-Studie der gesamte PHES als Vergleichsmaß für die ROC-Analyse diente, definierten die Studien zu den anderen Testverfahren eine mHE bereits über das Vorliegen von pathologischen Resultaten in ein bis zwei Einzeltests. Insgesamt erreichten hier der Stroop-Test mit 0,91 und der ICT mit 0,902 eine mit dem eNCT vergleichbare AUC (72, 78, 131).

Die aus einer ROC-Analyse gewählten Grenzwerte für eine Differenzierung zwischen normalen und pathologischen Testzeiten ergeben sich aus den jeweiligen Fragestellungen, die daher bei Vergleichen der Genauigkeit unterschiedlicher Testverfahren zu beachten sind (111). In dieser Arbeit wurde die Bedeutung einer hohen Sensitivität und eines hohen negativ-prädiktiven Werts für den Einsatz des eNCTs als Screeningtest betont, da bei einem falsch-negativen Ergebnis mit eventuell ausbleibendem Arztkontakt im Gegensatz zu einem falsch-positiven Ergebnis mit eventuell folgendem Arztkontakt von gesundheitlichen Risiken für den Patienten auszugehen ist. Es überwiegt die Bedeutung der Möglichkeit, schwerwiegenden Komplikationen einer HE präventiv begegnen zu können. Außerdem könnten überflüssige Arztkonsultationen nach falsch-positiven Testergebnissen durch ein eskalierendes Stufenkonzept der Diagnostik (z. B. Wiederholungstests im zeitlichen Abstand mit Listung in einer Gruppe intensivierter Beobachtung, Integration weiterer von zu Hause aus messbarer Parameter des Allgemeinzustands und der Leberfunktion) auf ein geringes Maß begrenzt bleiben.

Unter Anwendung der Grenzwerte von 38 Sekunden (eNCT 10“) bzw. 41 Sekunden (eNCT 18“) erwies sich der eNCT im Vergleich mit dem Stroop-Test (89,1%), der CFF (55%) und dem ICT (88%) als das sensitivste Verfahren, während die Spezifität niedriger war als beim Stroop-Test (82,1%) oder der CFF (100%) und vergleichbar mit dem

ICT (77%). Auch hier ist zu beachten, dass als Vergleichsmaß bei der Validierung dieser Testverfahren jeweils nicht der gesamte PHES erhoben wurde (72, 78, 131).

Zusätzlich zur Sensitivität und zur Spezifität wurden in dieser Arbeit Vorhersagewerte, Likelihood-Quotienten und die diagnostische Odds-Ratio (DOR) berechnet. Hier konnte eine hohe Verlässlichkeit negativer Testergebnisse im Einklang mit den geforderten Eigenschaften für den eNCT als Screeningtest gezeigt werden. Ferner implizieren die Likelihood-Quotienten eine höhere Wahrscheinlichkeit auf eine mHE bei einem positiven Testergebnis im eNCT 18“ als im eNCT 10“ und eine niedrigere Wahrscheinlichkeit bei einem negativen Testergebnis im eNCT 10“ als im eNCT 18“. Insgesamt spricht die höhere DOR des eNCT 10“ im Vergleich mit dem eNCT 18“ für eine bessere diagnostische Trennfähigkeit des kleineren Tablet-Computers.

Wiederholungsgenauigkeit des eNCTs:

Für die Gesamtheit der Studienteilnehmer, die keine mHE oder oHE hatten und die Testbatterie ein zweites Mal absolvierten (N=50), wurden in der Publikation zur eNCT-Studie gute bis sehr gute IKK im NCT-A (IKK=0,94), eNCT 10“ (IKK=0,89) und eNCT 18“ (IKK=0,94) nachgewiesen, die jeweils größer waren als der entsprechende *Korrelationskoeffizient* r (98). Stärkere und insbesondere verschieden starke Lerneffekte zwischen den Testverfahren konnten damit ausgeschlossen werden, die laut Held et al. insbesondere dann wahrscheinlich sind, wenn der Wert des IKK kleiner ist als der *Korrelationskoeffizient* r (101). Die intraindividuellen Unterschiede der eNCT-Testzeiten zwischen Einschluss und Nachbeobachtung in der LZI-Kohorte repräsentieren also maßgeblich die Variabilität der kognitiven Leistungsfähigkeit, was für eine hohe Aussagekraft der eNCT-Testzeiten erwünscht ist. Entsprechend waren die Testzeitdifferenzen bei Patienten mit initialen Kognitionsdefiziten (mHE-Kohorte) größer (6,6 (eNCT 10“) bzw. 3,2 (eNCT 18“) Sekunden) als bei Patienten ohne mHE (0,9 bzw. 3,1 Sekunden). Für den NCT-A und andere Einzeltests der PHES-Testbatterie sind dagegen Lerneffekte bekannt, die zur Entwicklung von vier Parallelversionen des PHES und zur Erhebung eines Gesamtwerts führten (68, 129, 133). Der Stroop-Test und der ICT werden ebenso – und im Unterschied zur CFF – durch Lerneffekte beeinflusst (78, 122).

Der eNCT bietet im Gegensatz zu den beständigen Testbögen des PHES die Möglichkeit immer neuer Parallelversionen durch automatisch wechselnde Zahlenanordnungen.

Daher sind langfristige Lerneffekte, die über das Erlernen der Methodik hinausgehen, für den eNCT nicht zu erwarten.

Anders als in der Publikation von Wuensch et al. wurden zur Berechnung der IKK in dieser Dissertation lebergesunde Probanden ausgeschlossen und Patienten mit mHE eingeschlossen, um die Auswirkungen von Veränderungen der Therapie oder der Kognition sichtbar zu machen (98). Dennoch waren die innerhalb der LZI-Kohorte gemessenen IKK (N=25) hoch, bei Patienten mit mHE (N=12) niedriger als bei Patienten ohne mHE (N=13). Damit kann einerseits die hohe Wiederholungsgenauigkeit des eNCTs bestätigt und andererseits die erwünschte Variabilität der Testergebnisse bei Vorliegen einer mHE demonstriert werden, die sich (bemessen an den Differenzen der Testzeitunterschiede) deutlicher im eNCT 10“ zeigte. Die geringfügig höheren IKK bei Wuensch et al. sind also durch konstantere Bedingungen der untersuchten Kohorte zu erklären.

Veränderungssensitivität des eNCTs:

Probanden, die sich zum Zeitpunkt der Nachbeobachtung in ihrer HE-Diagnose verbessert hatten (N=8), konnten sich auch im eNCT 10“ mit durchschnittlich 10,9 Sekunden signifikant verbessern. Im eNCT 18“ war diese Verbesserung erst nach Ausschluss eines möglichen Ausreißers signifikant. Keiner der vier Patienten mit mHE, die bei Nachbeobachtung einen PHES oberhalb des Grenzwerts erreichten, absolvierte den eNCT bei Nachbeobachtung schneller als die jeweilige Grenzzeit. Dennoch könnte der eNCT auch in kurzfristigen Kontrolluntersuchungen aussagekräftige Ergebnisse liefern, da die Testzeiten als Tendenz unabhängig vom Grenzwert interpretierbar sind. Für eine bessere Bewertung dieser Hypothese ist eine größere Anzahl an Patienten notwendig.

Einflussfaktoren auf Testergebnisse des eNCTs:

Die durchgeführten Regressionsanalysen zeigten einen signifikanten Einfluss durch das Alter auf die Testwerte der Kohorte *keine mHE*, der mit Ausnahme des ICTs auch für den PHES, die CFF und den Stroop-Test nachgewiesen ist (66, 67, 72, 78, 131). Unabhängig von der Bildschirmgröße des Tablet-Computers werden dadurch aber nur etwas mehr als 10% der Varianz der eNCT-Resultate erklärt. Eine zuverlässige Vorhersage von Testzeiten auf Grundlage des Regressionsmodells ist daher nicht möglich. Da eine signifikante Beeinträchtigung der Testzeiten durch Geschlecht und Bildung sowie patientenspezifische Faktoren wie Aszites, HE-Medikation, vorige oHE oder Einschlusssta-

tus mithilfe der Regressionsanalysen ausgeschlossen wurde, sind die fehlenden 90% der Varianz am ehesten auf geringfügige Effekte einer Vielzahl verschiedener Variablen zurückzuführen. Andererseits traten einige Ausprägungen der Faktoren eventuell zu selten auf für einen stärkeren Einfluss auf das Regressionsmodell. Beispielsweise hatten von den hier berücksichtigten Studienteilnehmern (*keine mHE*, N=57) nur vier Patienten vor Einschluss eine oHE-Episode (bei insgesamt 19 von 95 getesteten Patienten mit Leberzirrhose). Für diesen Faktor konnte in Validierungsstudien zum Stroop-Test hingegen ein signifikanter Einfluss auf die Testresultate gezeigt werden (78).

Für die Kontrollprobanden konnten Wuensch et al. Bildung als zusätzlichen Einflussfaktor in einem Regressionsmodell nachweisen, das 40% (eNCT 10“) bzw. 44% (eNCT 18“) der Varianz erklärte (98). Dieser Anstieg der erklärten Varianz ist auch als Folge des Wegfalls patientenspezifischer Faktoren und damit einer geringeren Anzahl charakterisierender Variablen in der Kontrollkohorte zu deuten. Dennoch bleiben knapp 50% der Varianz ungeklärt, sodass zusätzliche unspezifische Ursachen, wie z.B. die Tageszeit, deren signifikanter Einfluss beispielsweise auf die CFF bekannt ist, die Motivation der Probanden und die ungewohnte Umgebung vermutet werden können (72, 134).

Außerdem ist zu hinterfragen, warum die Bildung im Gegensatz zu den lebergesunden Probanden keinen Einfluss auf die Testzeiten des Patientenkollektivs haben sollte. Eine alternative Erklärung für diese Beobachtung ist, dass die Varianz der Bildung in der LZI-Kohorte geringer war (13 ± 2 Ausbildungsjahre) als bei den lebergesunden Probanden (14 ± 3). Insgesamt hat die Bildung wie beim PHES, dessen Normwerte auch wegen der schwierigen Vergleichbarkeit unterschiedlicher Bildungssysteme nicht bildungskorrigiert wurden, nur eine geringe Auswirkung auf Testresultate im eNCT (69).

Die Regressionsanalyse auf der Suche nach weiteren Einflussfaktoren um beliebig viele Variablen zu erweitern, würde das Problem der Überanpassung von Regressionsanalysen verstärken und damit die Gültigkeit des Regressionsmodells weiter schwächen (135). Ohnehin darf das im Rahmen dieser Pilotstudie errechnete Regressionsmodell nicht außerhalb der Studienpopulation angewendet werden, da zunächst nur ein kleines Testkollektiv rekrutiert wurde. Zukünftige Studien mit größeren Kohorten sind daher wünschenswert, auch um altersabhängige Normwerttabellen zu ermöglichen.

Die untersuchten Blutwerte Ammoniak und Natrium und die Leberfunktionsscores MELD und Child-Pugh konnten als Einflussfaktoren auf die eNCT-Testzeiten von Patienten ohne mHE oder oHE bei nicht signifikanten *Bestimmtheitsmaßen* R^2 ausge-

geschlossen werden. Niedrige und nicht signifikante Korrelationskoeffizienten bestärkten dieses Fazit. Wie in Studien zum PHES und weiteren psychometrischen Testverfahren an anderer Stelle gezeigt, erwiesen sich die genannten Parameter damit nicht als geeignete Methoden für den Nachweis einer mHE, obwohl teilweise ein Einfluss auf psychometrische Testergebnisse (z. B. CFF und Stroop-Test) vorbeschrieben ist (5, 16, 69, 72, 78).

Schlussfolgerungen:

Der eNCT ist ein valides Testverfahren zur Beurteilung der kognitiven Leistungsfähigkeit bei Patienten mit Leberzirrhose. Der Zielbereich der Testgenauigkeit zur Anwendung des eNCTs als mHE-Screeninginstrument (hohe Sensitivität bei Spezifität über 70%) wurde erreicht. Der eNCT ist ein Test mit sehr hoher Wiederholungsgenauigkeit und zuverlässiger Abnahme der Testzeiten bei Reduktion der kognitiven Leistungsfähigkeit. Die Testzeiten werden signifikant durch das Alter und nicht signifikant durch eine Vielzahl unterschiedlicher Faktoren beeinflusst.

Insgesamt zeigte sich der eNCT gegenüber anderen mHE-Diagnostetests, für die gleichfalls eine Anwendung ohne Fachpersonal diskutiert wird (und soweit methodisch vergleichbar), für den Einsatz als Screeningtest überlegen.

7.2.2.2 Vergleich von PHES und eNCT

Aufgrund der Entwicklung des eNCTs als eine Kombination aus NCT und SDT wurde eine gute Korrelation mit den Ergebnissen im NCT-A, NCT-B und SDT erwartet.

Tatsächlich konnte eine hohe Korrelation zwischen Resultaten im eNCT und Resultaten im NCT-A, NCT-B und DST nachgewiesen werden. Für den SDT konnte dagegen nur eine mittlere Korrelation mit dem eNCT berechnet werden. Da auch die Effektstärke einer mHE auf die Testwerte der drei erstgenannten Tests vergleichbar hoch war wie die Effektstärke auf eNCT-Testzeiten, scheint der eNCT insbesondere durch die Eigenschaften dieser Tests charakterisierbar zu sein: Die papierbasierten Zahlenverbindungstests prüfen visuell-räumliche Orientierung, motorische Geschwindigkeit, Konzentration und Aufmerksamkeit. Der DST lässt sich als Test der geistigen Leistungsfähigkeit und des visuellen Kurzzeitgedächtnisses charakterisieren (136). Eine differenzierte Beurteilung, welche kognitiven Domänen durch den eNCT überprüft werden, bedarf einer abschließenden Einschätzung durch Fachexperten.

Insgesamt aber muss für den PHES durch die verschiedenen Einzeltests eine genauere Untersuchung kognitiver Teilbereiche im Vergleich zum singulären eNCT angenommen werden. Bei der Bewertung des eNCTs ist die Genauigkeit jedoch der Alltagstauglichkeit der Testverfahren gegenüberzustellen. Der eNCT war für die LZI-Kohorte im Durchschnitt knapp sieben Minuten schneller absolvierbar als der PHES und dürfte damit leichter in den Alltag integrierbar sein. Potenziell trägt das schnellere Testdesign zwar auch zu Einschränkungen der Testgenauigkeit bei, die Etablierung eines Goldstandards in der HE-Diagnostik sollte durch den eNCT aber ohnehin nicht realisiert werden. Vielmehr wurde als Ziel für den eNCT formuliert, dass eine mit etablierten Testverfahren vergleichbare Genauigkeit bei einfacherer Anwendung erreicht und damit perspektivisch ein Einsatz in der häuslichen Umgebung ermöglicht werden kann.

Dementsprechend konnte die Genauigkeit des eNCTs am PHES bemessen werden, obwohl dadurch a priori die Möglichkeit für den eNCT ausgeschlossen wurde, für die verschiedenen Kriterien der Testgenauigkeit bessere Werte als der PHES zu erzielen. In der LZI-Kohorte blieben mögliche Fehldiagnosen durch den PHES nämlich aufgrund seiner diagnosebestimmenden Eigenschaft immer unbemerkt und würden fälschlicherweise als richtig gewertet werden. Tatsächlich richtige Ergebnisse der jeweiligen Patienten im eNCT würden beim Vergleich mit dem PHES entsprechend irrtümlich als falsch gewertet werden. Dass es prinzipiell zu Fehldiagnosen im PHES kommen kann, zeigt die geringe falsch-positiv-Rate bei lebergesunden Probanden von 1,8%, die bei fehlender Lebererkrankung eindeutig keine mHE entwickeln können. Zusätzlich waren Patienten mit einer oHE zum Teil nicht in der Lage, die PHES-Testbatterie komplett durchzuführen. Dieser Anteil des oHE-Kollektivs (N=7/12) musste bei fehlendem PHES-Gesamtwert von den statistischen Analysen der Gütekriterien ausgeschlossen werden, obwohl der eNCT hier als schnelleres Testverfahren potenziell noch absolvierbar gewesen wäre (und in dieser Studie tatsächlich von zwei (eNCT 10“) bzw. vier (eNCT 18“) zusätzlichen Patienten mit oHE durchgeführt werden konnte). Alternativ wäre ein Vergleich der beiden Testverfahren über die Festlegung des eNCT-Grenzwerts als Mittelwert der gleichaltrigen Kontrollgruppe plus zwei Standardabweichungen möglich, eine von Weissenborn et al. durchgeführte Methode bei der Evaluation des PHES (67). Dieser Ansatz wurde verworfen, da im Gegensatz zum damaligen Zeitpunkt nun mit dem PHES ein etablierter Vergleichstest mit insgesamt niedriger Fehlerrate zur Verfügung stand und damit die Berechnung verschiedener Grenzwerte per ROC-Analyse als überlegene Methode für die hier bearbeitete Fragestellung möglich war (60, 109, 132).

Der PHES wurde in Abschnitt 3.5.4 mit einer Sensitivität von 96,0% und einer Spezifität von 100% vorgestellt. Zu berücksichtigen ist, dass diese Werte erhoben wurden, indem die Spezifität des PHES in einer Kontrollgruppe und die Sensitivität in einer Gruppe von Patienten mit Leberzirrhose und oHE berechnet wurde (67). Wird dieser Rechenweg für den eNCT angewendet und die Spezifität in der $\geq 39J$ -Kohorte bzw. die Sensitivität in der oHE-Kohorte berechnet, resultiert eine höhere Spezifität von 86,0% (eNCT 10“) bzw. 94,7% (eNCT 18“) bei nicht abschließend beurteilbarer Sensitivität (aufgrund der Gruppengröße von $N=7$ bzw. $N=9$). Die Evaluierung des eNCTs wurde in dieser Dissertation aber anhand der LZI-Kohorte vorgenommen, da es sich dabei um die Zielgruppe des Testverfahrens handelt. Zwischen den LZI-Untergruppen *mHE* und *keine mHE* sind allerdings diskretere Unterschiede der Testresultate zu erwarten als zwischen Kontrollprobanden und Patienten mit oHE, die zur Validierung des PHES verglichen wurden. Umso höher ist es zu bewerten, dass der eNCT dennoch mit dem PHES vergleichbar gute Werte für die Sensitivität erreichte.

Für den eNCT wurde mittels IKK eine hohe Wiederholungsgenauigkeit berechnet, welche zumeist im Bereich der Angaben aus dem PHES-Handbuch zu den Einzeltests des PHES lag (NCT-A und -B: 0,87; DST: 0,94; LNT-Z: 0,95; LNT-F: 0,56; SDT: 0,94). Dort fehlen allerdings Informationen über die verwendeten statistischen Tests, sodass die Werte nur eine eingeschränkte Vergleichbarkeit besitzen (68). Zwischen dem eNCT und dem PHES existiert bei Kappa-Werten von 0,609 (eNCT 10“) bzw. 0,621 (eNCT 18“) laut Landis und Koch außerdem eine sehr gute („substantial“) Übereinstimmung bzw. Interrater-Reliabilität (137). Als Maß der absoluten Übereinstimmung wäre ein Bland-Altman-Plot vorzuziehen (138). Hierbei wird der Mittelwert zweier Messungen gegen deren Differenz aufgetragen (139). Für die Charakterisierung des eNCTs konnte mit einer durchschnittlichen Abweichung von $4,0s \pm 8,0s$ bzw. $2,5s \pm 8,2s$ bereits eine gute Übereinstimmung zwischen NCT-A und eNCT 10“ bzw. eNCT 18“ gezeigt werden (98). Da PHES und eNCT in verschiedenen Dimensionen gemessen werden, war die Berechnung der absoluten Übereinstimmung sowie ein Vergleich der Wiederholungsgenauigkeit zwischen PHES-Gesamtwert und eNCT nicht sinnvoll möglich.

Weiterhin existiert ein relevanter Unterschied zwischen eNCT und PHES in der praktischen Anwendung. Während dem PHES die Möglichkeiten aktiver Rückmeldung und automatischer Auswertungen fehlen und sich der Papiertest damit bereits durch sein Testdesign als möglicher Suchtest für das häusliche Umfeld ausschließt, ist der eNCT

einzig in der richtigen Reihenfolge zu vervollständigen und ermöglicht perspektivisch die unmittelbare Reaktion auf das Testresultat durch das elektronische System und/oder Ärzte eines angeschlossenen Krankenhauses (vgl. Abschnitt 5.1).

Schlussfolgerungen:

PHES und eNCT stimmen in der Diagnosestellung der mHE sehr gut überein.

Die festgelegten Grenzwerte führten zu einer mit dem PHES vergleichbar hohen Sensitivität des eNCTs bei deutlich niedrigerer Spezifität.

Im Hinblick auf die Anwendung des Testverfahrens im häuslichen Umfeld erfüllt der eNCT im Gegensatz zum PHES nötige Voraussetzungen, was den eNCT zu einem optimalen Screeningtest macht, ohne die endgültige Diagnostik mittels PHES zu ersetzen.

7.2.3 Stroop-Test

Im Stroop-Test zeigten sich wie im eNCT signifikante Unterschiede in den Testzeiten zwischen den Gruppen $\geq 39J$ und *LZI*, nicht aber zwischen *LZI* und *oHE* (bei $N=3$). Nach PHES-basierter Aufteilung der LZI-Kohorte in *mHE* und *keine mHE* ergab sich ein signifikanter Unterschied der Stroop-Testresultate zwischen diesen beiden Gruppen.

Schlussfolgerungen:

Der Stroop-Test erfasst den Unterschied der kognitiven Leistungsfähigkeit zwischen lebergesunden Probanden und Patienten mit Leberzirrhose sowie zwischen Patienten mit Leberzirrhose *mit* mHE und *ohne* mHE und ermöglicht damit die Detektion der mHE in der Zielgruppe. Damit stimmen die Ergebnisse der eNCT-Studie mit den Ergebnissen der entsprechenden Validierungsstudien des Stroop-Tests überein (78, 123).

7.2.3.1 Vergleich von Stroop-Test mit PHES und eNCT

Hohe Korrelationen von Stroop-Testwerten und Resultaten in Papiertests wurden nur für den DST beobachtet, der auch in den Arbeiten von Bajaj et al. mit $r=-0,8$ am stärksten mit Stroop-Testwerten korrelierte (78). Die Korrelationen von NCT-A und NCT-B mit dem Stroop-Test waren signifikant schlechter als mit dem eNCT. Als Maß der Validität wurde zwischen Stroop-Testwerten und PHES-Gesamtwerten eine nicht-signifikant schwächere Korrelation gemessen als zwischen eNCT-Testzeiten und PHES-Gesamtwerten. Die Messung der Effektstärke zeigte, dass der Stroop-Test eine gerin-

gere diskriminatorische Fähigkeit erreicht als der eNCT. Hohe Korrelationen zwischen Stroop-Test und eNCT sprechen dennoch für ähnliche Testeigenschaften mit Betonung der kognitiven Verarbeitungsgeschwindigkeit und visuell-räumlicher Leistungen.

Für den Stroop-Test war eine Sensitivität der mHE-Diagnostik in einer Gruppe von Patienten mit Leberzirrhose von 89,1%, eine Spezifität von 82,1%, eine AUC von 0,91 und ein IKK von 0,832 vorbeschrieben (78). Demnach weisen die Ergebnisse der eNCT-Studie auf eine höhere Reliabilität (IKK) und höhere Sensitivität des eNCTs bei vergleichbarer Testgenauigkeit (AUC) und niedrigerer Spezifität hin. In der Arbeit von Bajaj et al. wurde der PHES allerdings nicht vollständig durchgeführt. Eine mHE-Diagnose wurde gestellt, wenn die Ergebnisse von zwei Papiertests um über zwei Standardabweichungen vom Mittelwert abwichen (78). Wird der Stroop-Test mit dem kompletten PHES verglichen, so zeigten die Resultate der eNCT-Studie eine Sensitivität von 91,7% und eine Spezifität von 74,4%. Insgesamt fielen im Vergleich mit dem eNCT 10“ mit Ausnahme der Spezifität und im Vergleich mit dem eNCT 18“ mit Ausnahme des NPWs alle Testgenauigkeitskriterien für den Stroop-Test niedriger aus.

Die Daten der eNCT-Studie sprechen für eine höhere Validität und Genauigkeit des eNCTs. Dieser Unterschied ist einerseits auf die geringere Anzahl vollständiger Stroop-Test-Resultate (N=51, mit 12x mHE) zurückzuführen, könnte andererseits aber auch auf eine geringere diagnostische Wertigkeit des Stroop-Tests in der mHE-Diagnostik hinweisen. So ist die Leseaufgabe im Stroop-On-Modus beispielsweise assoziiert mit verbaler Intelligenz, die im Rahmen der mHE weniger stark betroffen ist (17, 140).

Schlussfolgerungen:

Der eNCT war dem Stroop-Test in dieser ersten Anwendungsstudie in seiner diagnostischen Wertigkeit leicht überlegen. Bei einem zukünftigen Einsatz als Screeningtest wäre der eNCT dem Stroop-Test demnach vorzuziehen. Größere Kohorten sind allerdings notwendig, um diesen Unterschied zu validieren.

7.3 Einschränkungen und Limitationen der Testverfahren

7.3.1 Methodische Limitationen

- Die eingeschränkte Konzentrationsfähigkeit von Patienten mit oHE verlängerte die Durchführung der Testreihe von durchschnittlich 30 bis 40 Minuten auf eine Dauer

von bis zu drei Stunden. Dieser Umstand dürfte sich zusätzlich negativ auf die Leistungsfähigkeit und damit die Testwerte ausgewirkt haben, was die Anwendung möglichst simpler Testverfahren in einer häuslichen Umgebung bestärkt.

- Die persönliche Konzentrations- und Leistungsfähigkeit ist abhängig vom Umfeld (Beleuchtung, Lärm, Arbeitsfläche, etc.) und von der Tageszeit bzw. der individuellen zirkadianen Rhythmik (141, 142). Diese Faktoren ließen sich beim Einschluss der Studienteilnehmer in Abhängigkeit vom klinischen Betrieb nur begrenzt berücksichtigen und sind daher als mögliche Einflussfaktoren auf die Testwerte zu betrachten.

- Trotz gewissenhafter Prüfung der Ein- und Ausschlusskriterien mussten beispielsweise bezüglich des Alkoholkonsums oder der Einnahme von Medikamenten wahrheitsgemäße Angaben der Probanden angenommen werden.

Es wird geschätzt, dass bei der Hälfte aller Patienten mit einer kognitiven Dysfunktion und vermuteter HE stattdessen andere Ursachen (z. B. Infektionen, metabolische Erkrankungen, strukturelle Läsionen) vorliegen (143). Daher muss trotz sorgfältiger Prüfung jedes einzelnen Probanden eine geringe Fehlerrate angenommen werden.

- Die Hilfsinstrumente WHC, MMST und HESA zur Einschätzung der kognitiven Leistungsfähigkeit sind in abnehmender Stärke von subjektiven Bewertungen des Prüfers beeinflusst. Um diese Tatsache in ihrer Bedeutung gering zu halten, wurden die prüfenden Doktoranden in der Erhebung der verschiedenen Wertungsbögen mithilfe von objektivierbaren Kriterien oder standardisierten Fragebögen trainiert.

- Abnahme und Auswertung von Blutproben sind trotz hoher Qualitätsstandards von dem Risiko (prä- und post-)analytischer Fehler begleitet, deren Häufigkeit auf 0,01 bis 0,6% geschätzt wird (144).

- Bei der Interpretation signifikanter Ergebnisse ist zu beachten, dass mit steigender Anzahl durchgeführter Signifikanztests auch die Wahrscheinlichkeit steigt, trotz eines p -Werts $<0,5$ einen *Fehler 1. Art* zu begehen und damit irrtümlicherweise eine Signifikanz anzunehmen. Auch bei der Analyse mehrerer Hypothesen an derselben Stichprobe tritt das Problem des multiplen Testens auf. Es existieren verschiedene Korrekturverfahren (z. B. Bonferroni-Korrektur), die abhängig sind von der Anzahl durchgeführter Tests. Auf die Anwendung dieser Verfahren wurde in der vorliegenden Dissertation weitgehend verzichtet, da die eNCT-Studie als eine explorative Pilotstudie zu betrachten ist, deren Resultate unter anderem zur Entwicklung enger gefasster Hypothesen in Folgestudien beitragen sollten. Eine genaue Aussage über die Anzahl

durchgeführter Signifikanztests ist in diesem Rahmen kaum möglich, weshalb laut Victor et al. auf Korrekturverfahren verzichtet werden darf. Dementsprechend ist bei der Deutung der Ergebnisse aber ein relevantes Risiko zu berücksichtigen, dass unter allen durchgeführten Tests wenigstens ein Ergebnis fälschlicherweise als signifikant akzeptiert wurde (145).

- Während wesentliche Kriterien der Testgenauigkeit berechnet werden konnten, ist eine Generalisierung und Übertragbarkeit der hier gewonnenen Grenzwerte auf die Gesamtheit der Patienten mit Leberzirrhose sowie die Berechnung adäquater alterskorrigierter Normwerte bei der geringen Anzahl rekrutierter Leberzirrhosepatienten mit mHE (N=26) oder oHE (N=12) nicht möglich.
- Obwohl der Anteil der Patienten mit mHE in der LZI-Kohorte mit 31,3% im Vergleich zu anderen Studien nicht hoch war, dürfte dieser Wert den wahren Anteil in der Gesamtheit der Patienten mit Leberzirrhose aufgrund eines systematischen Selektionsfehlers noch überschätzt haben: Stationär eingeschlossene Patienten waren häufiger von Komplikationen der Leberzirrhose (z. B. HE) betroffen als die ambulant eingeschlossene Patientengruppe, welche die Gesamtheit der Patienten mit Leberzirrhose besser repräsentiert (78, 122, 123). In der Tat wurde für eine ambulante deutsche LZI-Kohorte mit 25% eine niedrigere Prävalenz der mHE publiziert (117). Daher wurden vergleichsweise prävalenzkorrigierte Vorhersagewerte berechnet, die zu einer diskreten Steigerung des NPWs bzw. Verminderung des PPWs führten.
- Die Werte der Testgenauigkeit sind für die Population errechnet worden, die zuvor bereits die Grenzwerte definierte. Um diesen so genannten Resubstitutionsschätzer, der häufig zu einer verzerrt hohen Genauigkeit des Tests führt, zu umgehen, müssten die Grenzwerte an einer Kohorte berechnet und an einer zweiten Kohorte mit ähnlicher Zusammensetzung kreuzvalidiert werden. Für dieses Vorgehen sind Folgestudien mit größeren Kohorten notwendig (108).

Schlussfolgerungen:

Insbesondere aufgrund der geringen Anzahl rekrutierter Patienten mit Leberzirrhose erfordern Aussagen, die für die Gesamtheit der Patienten mit Leberzirrhose gelten sollen, die Validierung der Ergebnisse in größeren Kohorten. Dennoch liefert die eNCT-Studie verlässliche und für ähnliche Kollektive repräsentative Daten, da weitere methodische Limitationen größtenteils als unvermeidbare Einflussfaktoren auch bei vergleichbaren Studien zu berücksichtigen wären.

7.3.2 Limitationen des PHES

Der PHES wurde als Vergleichstest verwendet, weil er speziell zur mHE-Diagnostik entwickelt und häufig validiert wurde (66, 67, 126, 127). Da die Gütekriterien des eNCTs in Bezug auf den PHES bestimmt wurden, ist die Genauigkeit des PHES für die Aussagekraft der Messungen relevant. Fehlerhafte Diagnosen durch den PHES würden ansonsten unbemerkt die Charakterisierung der eNCT-Eigenschaft verzerren (vgl. Abschnitt 7.2.2.2). Hinweise auf Ungenauigkeiten des PHES lieferte die Arbeitsgruppe um Kircheis et al., die zum einen erhebliche Fehler in der Durchführung der PHES-Testbatterie und damit Fehldiagnosen in knapp 50% der Fälle feststellte und zum anderen Zweifel an der Diskriminierungsfähigkeit des PHES zwischen mHE und oHE Grad 1 formulierte (130). Durchführungsfehler wurden demnach durch das Wechseln des Stifts während eines Tests, die mehrfache Anwendung gleicher NCT-A-Versionen im Laufe eines Testdurchlaufs, ausbleibendes sofortiges Unterbrechen des Probanden durch den Prüfer bei Fehlern und fehlendes oder fehlerhaftes Ausfüllen des Übungsblatts gemacht (130, 146). Obwohl es für diese Sichtweise eine ausführliche Gegendarstellung von Weissenborn et al. gibt, die wiederum die Studienqualität von Kircheis et al. kritisiert, wurde versucht, die genannten Fehlerquellen in der eNCT-Studie zu vermeiden (147).

Der PHES wird auch deswegen weiterhin als geeigneter Vergleichstest angesehen, da für die Testbatterie und den eNCT die Unterscheidung zwischen Patienten *ohne* mHE und *mit* mHE die entscheidende Indikation darstellt und nicht die Unterscheidung zwischen mHE und oHE. Letztere ist durch das Auftreten klinischer Symptome definiert, weshalb psychometrische Tests zur oHE-Diagnose nicht notwendig sind. Daher ist die von Kircheis et al. geäußerte Kritik am PHES nicht in vollem Umfang nachvollziehbar.

Der PHES könne laut Kircheis et al. außerdem im Verlauf nicht sicher bei Patienten mit einer oHE zwischen Verbesserungen und stabiler HE unterscheiden (130). Da in der eNCT-Studie nur von vier Patienten mit oHE und klinischer Besserung im Verlauf Daten in einem Wiederholungstest erhoben werden konnten, kann diese Behauptung nicht abschließend bewertet werden. Es erzielten aber immerhin drei der vier Patienten im PHES bei Nachbeobachtung ein verbessertes Ergebnis.

Bezüglich der Anwendung des PHES in der eNCT-Studie ist ein systematischer Messfehler des PHES aufgrund der leichten Modifikationen und der veränderten Reihenfolge der PHES-Einzeltests nicht auszuschließen, betreffe aber alle Probanden gleichermaßen.

ßen. Außerdem konnte in der Arbeit zur Charakterisierung des PHES gezeigt werden, dass die Anwendung unterschiedlicher Paralleltestversionen nur im DST signifikanten Einfluss auf die Testergebnisse hatte, der allerdings wiederum das Gesamtergebnis des PHES nicht veränderte (69).

Unter den objektiven Testkriterien des PHES ist der Fehlerpunktwert des LNTs der einzige mithilfe einer Auswertungsschablone subjektiv zu bestimmende Wert. Diese Auswertung ist teilweise zeitintensiv und kann trotz Schablone zwischen verschiedenen Prüfern unterschiedlich ausfallen. Eine Validierung von 120 der hier erhobenen LNT-Fehlerpunktwerte durch einen externen Prüfer bestätigte eine hohe Konsistenz der Ergebnisse mit maximalen Abweichungen im PHES-Gesamtwert von zwei Punkten. Bei zwei Patienten unterschied sich die HE-Diagnose aufgrund eines abweichenden LNT-Fehlerpunktwerts.

Schlussfolgerungen:

Es gibt bisher keinen definierten Goldstandard zur mHE-Diagnose. So wird die Aussagekraft des PHES und anderer Tests in der mHE-Diagnostik bis heute kontrovers diskutiert, wodurch die großen Schwankungen in den Angaben zur mHE-Prävalenz begünstigt werden (8-12). Die Vielzahl der Studien, die dem PHES seine hohe diagnostische Wertigkeit bestätigten, die Vorgaben der Leitlinien zur HE, die geringen Beschaffungskosten und das für den eNCT beabsichtigte Einsatzgebiet rechtfertigen dennoch die Anwendung des PHES als Vergleichstest.

7.3.3 Limitationen des eNCTs

Ein computerbasiertes Testverfahren ist in der Anwendung nur so stabil und sicher wie der verwendete Computer. Während der Rekrutierungsphase gab es in einigen Fällen Probleme mit der Reaktion des Bildschirms auf Berührungen oder notwendige Aktualisierungen des Systems machten einen Einsatz des Tablet-Computers unmöglich. Diese Situationen traten zwar sehr selten auf, könnten aber in der praktischen Durchführung weitreichende Konsequenzen haben. Es ist also unbedingt erforderlich, das zur Anwendung kommende technische System bestmöglich abzusichern. Das betrifft neben der Systemstabilität auch eine biometrische Identitätsprüfung, z. B. per Fingerabdruck und/oder Kamera, und den Datenschutz, da ein alleiniger Zugriff durch den Patienten

sichergestellt werden muss und ein Transfer sensibler patientenbezogener Daten an unbeteiligte Dritte verhindert werden muss.

Eine Limitation psychometrischer Tests und damit auch des eNCTs ist trotz objektiver Bewertungskriterien die Abhängigkeit von der Mitarbeit der Probanden. In der Anwendungssituation sollten Patienten daher initial trainiert werden und regelmäßig über die Relevanz der Testresultate aufgeklärt werden, so wie die Probanden auch im Studienumfeld auf die Notwendigkeit einer bestmöglichen Absolvierung der Tests hingewiesen wurden. Am Ende des Tests sollte der Proband eine Rückmeldung über sein Testresultat erhalten können. Das war in dieser Studie bei fehlenden Referenzwerten noch nicht möglich.

Bei Patienten mit einer oHE war die Konzentrationsfähigkeit teilweise so stark eingeschränkt, dass sie während eines Testdurchlaufs an ihre Aufgabe erinnert werden mussten. Es ist also notwendig, eine automatische Hilfestellung einzurichten oder einen zweiten Grenzwert zu bestimmen, um Patienten mit hochgradiger kognitiver Einschränkung frühzeitig zu identifizieren. Diese Funktion ist für die selbstständige Anwendung des Testverfahrens in Abwesenheit von medizinischem Fachpersonal wünschenswert.

Da der NCT-A als Vorlage für den eNCT diente und gleichzeitig Bestandteil des PHES ist, war ein systematischer Messfehler bei Vergleichen mit dem PHES zu erwarten. Alternativ stand die Berechnung eines PHES ohne NCT-A zur Diskussion. Durch die hohen Effektstärken und Korrelationskoeffizienten verschiedener Papiertests konnte aber gezeigt werden, dass der NCT-A nicht als einziger Papiertest ähnliche Eigenschaften wie der eNCT hat. Somit konnte auf die Berechnung eines modifizierten PHES ohne NCT-A verzichtet werden.

Den vorliegenden Daten folgend könnten sich die durchschnittlichen Testzeiten eines Probanden bei regelmäßiger Anwendung durch einen geringen Lerneffekt leicht verbessern. Daher könnten die aus der einzeitigen Stichprobe berechneten Grenzwerte zu langsam ausfallen und zu einer erhöhten Rate falsch-negativer Diagnosen führen. Eine erneute Untersuchung der Lerneffekte nach regelmäßiger Anwendung des eNCTs ist zu empfehlen. Durch ständig neue Zahlenanordnungen beim eNCT sind die Lerneffekte im Gegensatz zum PHES allerdings auf das Erlernen der Methodik begrenzt.

Die Ergebnisse der Pilotstudie zum eNCT konnten die Frage, ob Verschlechterungen der kognitiven Leistungsfähigkeit mithilfe des eNCTs im Verlauf zuverlässig und frühzei-

tig diagnostiziert werden können, aufgrund des Studiendesigns nicht abschließend klären. Hierfür sind prospektive Studien notwendig.

Hingegen wurde die Verbesserung der kognitiven Leistungsfähigkeit bei den vier mHE-Patienten, die bei Nachbeobachtung einen PHES oberhalb des Grenzwerts erzielten, vom eNCT nicht festgestellt. Zusätzlich wurde ein oHE-Patient nicht detektiert. Diese Ergebnisse sind als Hinweise auf Einschränkungen in der Testgenauigkeit des eNCTs zu interpretieren, die ebenfalls durch weitere Studien überprüft werden müssen und möglicherweise dadurch begünstigt werden, dass es sich beim eNCT entgegen der Vorgaben der *ISHEN* und im Gegensatz zum PHES nicht um eine Testbatterie handelt. Dennoch belegen die erhobenen Daten, dass der eNCT als Screeningtest eine hohe Aussagekraft hat und als sinnvolle Ergänzung in der mHE-Diagnostik eingesetzt werden könnte. Auch andere Studien zeigten, dass eine geringere Anzahl verwendeter Tests, beispielsweise eine Reduktion auf NCT-A und LNT, zielführend sein kann (8).

Schlussfolgerungen:

Trotz des frühen Entwicklungsstands der getesteten eNCT-Version konnten valide Ergebnisse erzielt werden. Gleichzeitig wurde an einigen Stellen Optimierungspotenzial für notwendige Folgestudien festgestellt, durch die genauere Aussagen über die diagnostische Wertigkeit getroffen und altersgerechte Normwerte berechnet werden können.

7.3.4 Limitationen des Stroop-Tests

Nur zwei Drittel aller Probanden waren dazu bereit, den Stroop-Test am Ende der aufwendigen Studien-Testserie durchzuführen. Bei neun von zwölf Patienten mit oHE musste der Test abgebrochen werden, da die Patienten von der Aufgabenstellung überfordert waren und sich die Fehlerrückmeldung des Tablet-Computers negativ auf die bereits niedrige Motivation auswirkte. Zusätzlich ist limitierend anzumerken, dass der Stroop-Test bei Legasthenie, Rot-Grün-Sehschwäche oder eingeschränkten Sprachkenntnissen keine aussagekräftigen Ergebnisse liefern kann.

Schlussfolgerungen:

Der Stroop-Test empfiehlt sich durch sein elektronisches Testdesign prinzipiell für die Anwendung in einem häuslichen Umfeld. Der hohe Schwierigkeitsgrad stellt für die Pa-

tienten allerdings ein Hindernis für eine zuverlässige und selbstständige Bearbeitung des Tests dar.

7.4 Neue Erkenntnisse und möglicher Anwendungsbereich

In Abschnitt 3.5.4 wurden Vorgaben an ein geeignetes mHE-Testverfahren präsentiert, von denen die kurze Anwendungsdauer, die leichte Übersetzbarkeit und die Möglichkeit alternativer Testanordnungen sowie (zukünftig) alterskorrigierter Normwerte für den eNCT zutreffen. Allerdings wurde ein computerbasierter Test nicht als geeignete Grundlage für einen Goldstandard erachtet, da er nicht transportierbar, unhandlich und teuer wäre (60). Einerseits wurde der eNCT nicht als möglicher Goldstandard entwickelt, andererseits ist die soeben zitierte Expertenmeinung von 2009 in dieser Form in Frage zu stellen. Der eNCT empfiehlt sich ganz im Gegenteil zu dieser Ansicht nämlich durch seine Praktikabilität und seine einmaligen, verhältnismäßig geringen Beschaffungskosten als mHE-Screeningtest. Ressourcenverbrauchendes Druckerpapier wird im Gegensatz zum PHES nicht benötigt, da vom Computer automatisch immer neue Testvarianten erzeugt werden. Außerdem ist die Durchführung des eNCTs simpel genug für ein schnelles Verständnis, sodass der Test nicht nur automatisch und objektiv auswertbar, sondern auch wenig fehleranfällig ist. Insbesondere entfallen die meisten der von Kircheis et al. für die Durchführung des PHES genannten Fehlerquellen (130).

Der eNCT ist nach einer initialen Trainingsphase selbstständig durchführbar, perspektivisch kann daher auf betreuendes Fachpersonal verzichtet werden. Durch eine regelmäßige Anwendung des Tests wäre nicht nur eine schnellere Detektion pathologischer Zustände sowie die Überwachung von Therapieeffekten möglich, die Patienten würden über einen unmittelbaren Kontakt zur Klinik auch langfristig und ohne Unterbrechungen an eine medizinische Betreuung gebunden werden. Um eine hohe Adhärenz der Patienten sicherzustellen, sollte ein Screeningtest möglichst wenig belastend für den Patienten, einfach und schnell durchzuführen sein. Die komplexen psychometrischen Testbatterien PHES und RBANS erfassen zwar diverse kognitive Domänen, die Dauer der Tests charakterisiert den eNCT aber als das einfachste bzw. schnellste Testverfahren.

Im Laufe der Studie äußerten einige Probanden Zweifel an der Notwendigkeit einer regelmäßigen Testdurchführung, da ihnen ihre abnehmende kognitive Leistungsfähigkeit auch ohne Tests auffiele. Hier müssten die Patienten ausführlich und mehrmals über die Notwendigkeit des häuslichen Kontrollsystems aufgeklärt werden, da eine diskret

beginnende mHE für den Patienten unbemerkt sein kann und selbst das Registrieren von Symptomen nur selten zu einem direkten Arztkontakt führt (148).

Der eNCT wurde von den Studienteilnehmern dennoch überwiegend positiv bewertet. Fast 90% der eingeschlossenen Patienten mit Leberzirrhose wären trotz der geäußerten Zweifel zur regelmäßigen Durchführung eines elektronischen Testverfahrens bereit. Zusammengefasst ist das Testdesign des eNCTs für einen Einsatz als mHE-Screeningtest und für eine Untersuchung der Anwendbarkeit im häuslichen Umfeld geeignet.

Außerdem erfüllte der eNCT in dieser ersten Studie zur diagnostischen Wertigkeit mit einer hohen Sensitivität, Reliabilität und Validität auch die qualitativen Voraussetzungen für einen Einsatz als mHE-Screeningtest und eine Implementierung in den Alltag von Patienten mit Leberzirrhose. Die für die Berechnung der Gütekriterien notwendigen Grenzzeiten wurden in der Publikation von Wuensch et al. höher gewählt (eNCT 10“: 42,8s; eNCT 18“: 44,5s) als in dieser Arbeit, in der die Fähigkeit des eNCTs als sensibler Screeningtest betont werden sollte (98). Damit wird ein Spektrum der Testgenauigkeit abgebildet, die der eNCT je nach beabsichtigter Strenge der Testzeitbewertung annehmen kann und die dem Testverfahren eine individualisierte Diagnostik ermöglicht.

Der eNCT kann also als zeitgerechtes Diagnoseverfahren zu einer modernen und ökonomischen Therapie von Lebererkrankungen beitragen. Folgestudien der Phasen drei und vier nach Jensen und Abel sind aber zur Evaluation einer selbstständigen Durchführbarkeit im beabsichtigten Einsatzgebiet zwingend notwendig (108). Ein Ziel muss nach Optimierung der eNCT-Software und der Benutzeroberfläche die Bestimmung altersabhängiger Norm- oder Grenzwerte in der klinischen Anwendungssituation sein. Dafür muss eine größere Stichprobe rekrutiert werden, deren Kontrollprobanden eine ausgewogene Alters- und Bildungsstruktur haben und die idealerweise den eNCT zum Vergleich mit dem PHES zur gleichen Tageszeit an einem anderen Tag durchführen (Phase 3). Mithilfe der dabei erhobenen Daten könnten Altersklassen ermittelt werden, zwischen denen in einer Kontrollkohorte signifikante Unterschiede in den Testzeiten messbar sind. Für jede Gruppe könnte dann ein Grenzwert gebildet werden oder es könnten Testwertbereiche definiert werden, die eine Bewertung des mHE-Risikos z. B. in gering, mittel und hoch ermöglichen. Für diesen Fall wären dann auch Multilevel-Likelihood-Quotienten berechenbar, die zur weiteren Einschätzung der Testgenauigkeit dienen könnten (108, 113).

Eine weitere Studie müsste zur „Schaden-Nutzen-Abwägung“ die Resultate einer Langzeitanwendung bei Patienten mit Leberzirrhose im häuslichen Umfeld überprüfen (Phase 4). Insbesondere fehlen bisher Daten, die dem eNCT die zuverlässige Detektion einer Abnahme kognitiver Leistungsfähigkeit im Verlauf und die hypothetisch beschriebenen Vorteile in der Prävention bestätigen. Außerdem sollte eine Prüfung der externen Validität in vergleichbaren Patientengruppen ohne Zieldiagnose stattfinden.

Deutlich geworden ist auch, dass ein einmaliges Ergebnis im eNCT nicht als alleiniger Indikator über die Therapie eines Patienten entscheiden kann. Das ist nicht nur mit der eingeschränkten Spezifität des eNCTs zu begründen, sondern ebenso durch die unklaren Folgen einer mHE-Diagnose. Zum einen gibt es bisher keine klaren Therapieempfehlungen für Patienten mit einer mHE, zum anderen lässt sich nur aus der Diagnose keine sichere prognostische Bedeutung ableiten (5, 11). Auch daher besteht das Ziel in der Implementierung eines häuslichen Kontrollsystems, dessen zusätzliche Patientendaten (z. B. Blutwerte) und mehrfachen eNCT-Resultate in den Entscheidungsalgorithmus über eine Therapie integriert werden sollen.

Die eingangs formulierten Fragestellungen können insgesamt positiv beantwortet werden: Der eNCT stellte sich in dieser Pilotstudie als valides und reliables Instrument zur Einschätzung der kognitiven Leistungsfähigkeit von Patienten mit Leberzirrhose dar, das eine mit etablierten Testverfahren vergleichbare Genauigkeit bei einfacherer Anwendung erreicht. Bezüglich der weiteren in Abschnitt 5.5.1 genannten Gütekriterien lässt sich der eNCT darüber hinaus als objektives, ökonomisches und zumutbares Testverfahren charakterisieren, das durch Parallelversionen vergleichbare Ergebnisse ermöglicht. Nützlichkeit und Nicht-Verfälschbarkeit müssen in weiteren Studien, durch die eine Normierung angestrebt werden kann, genauer untersucht werden. Das Alter wurde als Einflussfaktor auf die Testzeiten von Leberzirrhosepatienten bestätigt und es konnten erste Grenzwerte mit hoher diagnostischer Trennfähigkeit errechnet werden. Dabei fielen die Werte der Testgenauigkeit für Spezifität und positive Vorhersagewerte zwar schwächer aus als die des diagnosegebenden Vergleichstests (PHES), dafür empfiehlt sich der eNCT aber aufgrund seiner hohen Sensitivität und negativen Vorhersagewerte sowie seines einfachen Testdesigns im Gegensatz zu anderen untersuchten Testverfahren (Stroop-Test, CFF, ICT) für eine selbstständige Anwendung als mHE-Screeningtest (129).

7.5 Ausblick

Der eNCT und das damit verbundene *d-LIVER*-Projekt erschließen einen ganz neuen Ansatz in der Betreuung von Patienten mit Lebererkrankungen, dessen Verwirklichung zur Reduktion von Therapiekosten, Arztkonsultationen und Krankenhauseinweisungen führen soll. Dafür wurde ein häusliches Kontrollsystem untersucht, dessen Bestandteile (z. B. eNCT) zur Erkennung und Frühintervention von Dekompensationen bei Patienten mit Leberzirrhose beitragen sollen. Als zentrale Schnittstelle dieses Systems kann ein Tablet-Computer dienen (Abbildung 6D), der die sensiblen Patientendaten verwaltet und über eine schnelle Kommunikation mit dem betreuenden medizinischen Personal eine semiautomatische Therapieänderung zulässt.

Die Ergebnisse der vorliegenden Dissertation sind ein erster Hinweis darauf, dass mittels eNCT die Notwendigkeit eines Arztbesuchs zur Untersuchung und Behandlung einer potenziellen HE von zu Hause aus evaluiert werden kann und in vielen Fällen ein überflüssiger Weg zum Arzt erspart bleibt, ohne sich nachteilig auf den Gesundheitszustand des jeweiligen Patienten auszuwirken. Als Alternative zu Grenzwerten könnten durch Multilevel-Likelihood-Verfahren auch Testwertbereiche eines geringen, mittleren und hohen mHE-Risikos definiert werden. So könnte der eNCT als eine Art Frühwarnsystem insbesondere in infrastrukturell schwächer versorgten Regionen sein präventives Potenzial ausschöpfen und die Angst von Patienten vor akuten Komplikationen und damit verbundenen Krankenhauseinweisungen reduzieren sowie die Kosten für das Gesundheitssystem senken. Da auf diesem Weg auch die ärztlichen Ressourcen effektiver genutzt würden, verspricht der eNCT Vorteile für alle Akteure des Gesundheitswesens.

Der Test kann zukünftig eine wichtige Rolle in der primären, sekundären und sogar tertiären Prävention einnehmen und damit das Konzept der HE-Diagnostik gezielt ergänzen.

8 Literaturverzeichnis

1. Gesundheit - Todesursachen in Deutschland. Wiesbaden: Statistisches Bundesamt; 2015 [updated January 6, 2016. accessed March 27, 2016. Available from: http://www.destatis.de/DE/Publikationen/Thematisch/Gesundheit/Todesursachen/Todesursachen2120400147004.pdf?__blob=publicationFile.]
2. Sauerbruch T, Appenrodt B, Schmitz V, Spengler U. The conservative and interventional treatment of the complications of liver cirrhosis: Part 2 of a series on liver cirrhosis. *Dtsch Arztebl Int.* 2013;110(8):126-32, I.
3. Rakoski MO, McCammon RJ, Piette JD, Iwashyna TJ, Marrero JA, Lok AS, Langa KM, Volk ML. Burden of cirrhosis on older Americans and their families: analysis of the health and retirement study. *Hepatology.* 2012;55(1):184-91.
4. Zhan T, Stremmel W. The diagnosis and treatment of minimal hepatic encephalopathy. *Dtsch Arztebl Int.* 2012;109(10):180-7.
5. Vilstrup H, Amodio P, Bajaj J, Cordoba J, Ferenci P, Mullen KD, Weissenborn K, Wong P. Hepatic encephalopathy in chronic liver disease: 2014 Practice Guideline by the American Association for the Study of Liver Diseases and the European Association for the Study of the Liver. *Hepatology.* 2014;60(2):715-35.
6. Ferenci P, Lockwood A, Mullen K, Tarter R, Weissenborn K, Blei AT. Hepatic encephalopathy--definition, nomenclature, diagnosis, and quantification: final report of the working party at the 11th World Congresses of Gastroenterology, Vienna, 1998. *Hepatology.* 2002;35(3):716-21.
7. Koch H, Schauder P, Schafer G, Dahme B, Ebel W, Vahldiek B, Konig F, Henning H. [Diagnosis and prevalence of latent hepatic encephalopathy]. *Z Gastroenterol.* 1990;28(11):610-5.
8. Schomerus H, Schreiegg J. Prevalence of latent portasystemic encephalopathy in an unselected population of patients with liver cirrhosis in general practice. *Z Gastroenterol.* 1993;31(4):231-4.
9. Groeneweg M, Quero JC, De Bruijn I, Hartmann IJ, Essink-bot ML, Hop WC, Schalm SW. Subclinical hepatic encephalopathy impairs daily functioning. *Hepatology.* 1998;28(1):45-9.
10. Bajaj JS, Schubert CM, Heuman DM, Wade JB, Gibson DP, Topaz A, Saeian K, Hafeezullah M, Bell DE, Sterling RK, Stravitz RT, Luketic V, White MB, Sanyal AJ. Persistence of cognitive impairment after resolution of overt hepatic encephalopathy. *Gastroenterology.* 2010;138(7):2332-40.
11. Hartmann IJ, Groeneweg M, Quero JC, Beijeman SJ, de Man RA, Hop WC, Schalm SW. The prognostic significance of subclinical hepatic encephalopathy. *Am J Gastroenterol.* 2000;95(8):2029-34.
12. Sharma P, Sharma BC, Puri V, Sarin SK. Critical flicker frequency: diagnostic tool for minimal hepatic encephalopathy. *J Hepatol.* 2007;47(1):67-73.
13. El Khoury AC, Klimack WK, Wallace C, Razavi H. Economic burden of hepatitis C-associated diseases in the United States. *J Viral Hepat.* 2012;19(3):153-60.
14. El Khoury AC, Wallace C, Klimack WK, Razavi H. Economic burden of hepatitis C-associated diseases: Europe, Asia Pacific, and the Americas. *J Med Econ.* 2012;15(5):887-96.
15. Poordad FF. Review article: the burden of hepatic encephalopathy. *Aliment Pharmacol Ther.* 2007;25 Suppl 1:3-9.
16. Bajaj JS, Riggio O, Allampati S, Prakash R, Gioia S, Onori E, Piazza N, Noble NA, White MB, Mullen KD. Cognitive dysfunction is associated with poor socioeconomic status in patients with cirrhosis: an international multicenter study. *Clin Gastroenterol Hepatol.* 2013;11(11):1511-6.

17. Schomerus H, Hamster W, Blunck H, Reinhard U, Mayer K, Dolle W. Latent portasystemic encephalopathy. I. Nature of cerebral functional defects and their effect on fitness to drive. *Dig Dis Sci.* 1981;26(7):622-30.
18. Kircheis G, Knoche A, Hilger N, Manhart F, Schnitzler A, Schulze H, Haussinger D. Hepatic encephalopathy and fitness to drive. *Gastroenterology.* 2009;137(5):1706-15 e1-9.
19. Bustamante J, Rimola A, Ventura PJ, Navasa M, Cirera I, Reggiardo V, Rodes J. Prognostic significance of hepatic encephalopathy in patients with cirrhosis. *J Hepatol.* 1999;30(5):890-5.
20. Martini GA. Acute Liver Failure: History and Epidemiology. In: Brunner G, Mito M, editors. *Artificial Liver Support: Concepts, Methods, Results.* 2nd ed. Berlin & Heidelberg, Deutschland: Springer-Verlag; 1992. p. 3-20.
21. Amodio P. Hepatic encephalopathy: historical remarks. *J Clin Exp Hepatol.* 2015;5(Suppl 1):S4-6.
22. Sherlock S, Summerskill WH, White LP, Phear EA. Portal-systemic encephalopathy; neurological complications of liver disease. *Lancet.* 1954;267(6836):454-7.
23. Summerskill WH, Davidson EA, Sherlock S, Steiner RE. The neuropsychiatric syndrome associated with hepatic cirrhosis and an extensive portal collateral circulation. *Q J Med.* 1956;25(98):245-66.
24. Zeegen R, Drinkwater JE, Dawson AM. Method for measuring cerebral dysfunction in patients with liver disease. *Br Med J.* 1970;2(5710):633-6.
25. Conn HO. Quantifying the severity of hepatic encephalopathy: syndromes and therapies. In: Conn HO, Bircher J, editors. *Hepatic encephalopathy: syndromes and therapies.* Bloomington, Illinois, USA: Medi-Ed Press; 1993. p. 13-26.
26. Haussinger D, Schliess F. Pathogenetic mechanisms of hepatic encephalopathy. *Gut.* 2008;57(8):1156-65.
27. Albrecht J, Jones EA. Hepatic encephalopathy: molecular mechanisms underlying the clinical syndrome. *J Neurol Sci.* 1999;170(2):138-46.
28. Gorg B, Schliess F, Haussinger D. Osmotic and oxidative/nitrosative stress in ammonia toxicity and hepatic encephalopathy. *Arch Biochem Biophys.* 2013;536(2):158-63.
29. Haussinger D, Gorg B. Interaction of oxidative stress, astrocyte swelling and cerebral ammonia toxicity. *Curr Opin Clin Nutr Metab Care.* 2010;13(1):87-92.
30. Rama Rao KV, Norenberg MD. Glutamine in the pathogenesis of hepatic encephalopathy: the trojan horse hypothesis revisited. *Neurochem Res.* 2014;39(3):593-8.
31. Haussinger D, Sies H, Gerok W. Functional hepatocyte heterogeneity in ammonia metabolism. The intercellular glutamine cycle. *J Hepatol.* 1985;1(1):3-14.
32. Haussinger D. Low grade cerebral edema and the pathogenesis of hepatic encephalopathy in cirrhosis. *Hepatology.* 2006;43(6):1187-90.
33. Haussinger D, Kircheis G, Fischer R, Schliess F, vom Dahl S. Hepatic encephalopathy in chronic liver disease: a clinical manifestation of astrocyte swelling and low-grade cerebral edema? *J Hepatol.* 2000;32(6):1035-8.
34. Norenberg MD, Martinez-Hernandez A. Fine structural localization of glutamine synthetase in astrocytes of rat brain. *Brain Res.* 1979;161(2):303-10.

35. Lange SC, Bak LK, Waagepetersen HS, Schousboe A, Norenberg MD. Primary cultures of astrocytes: their value in understanding astrocytes in health and disease. *Neurochem Res.* 2012;37(11):2569-88.
36. Parpura V, Heneka MT, Montana V, Oliek SH, Schousboe A, Haydon PG, Stout RF, Jr., Spray DC, Reichenbach A, Pannicke T, Pekny M, Pekna M, Zorec R, Verkhratsky A. Glial cells in (patho)physiology. *J Neurochem.* 2012;121(1):4-27.
37. Miese F, Kircheis G, Wittsack HJ, Wenserski F, Hemker J, Modder U, Haussinger D, Cohnen M. 1H-MR spectroscopy, magnetization transfer, and diffusion-weighted imaging in alcoholic and nonalcoholic patients with cirrhosis with hepatic encephalopathy. *AJNR Am J Neuroradiol.* 2006;27(5):1019-26.
38. Minguez B, Garcia-Pagan JC, Bosch J, Turnes J, Alonso J, Rovira A, Cordoba J. Noncirrhotic portal vein thrombosis exhibits neuropsychological and MR changes consistent with minimal hepatic encephalopathy. *Hepatology.* 2006;43(4):707-14.
39. Tanigami H, Rebel A, Martin LJ, Chen TY, Brusilow SW, Traystman RJ, Koehler RC. Effect of glutamine synthetase inhibition on astrocyte swelling and altered astroglial protein expression during hyperammonemia in rats. *Neuroscience.* 2005;131(2):437-49.
40. Laubenberger J, Haussinger D, Bayer S, Gufler H, Hennig J, Langer M. Proton magnetic resonance spectroscopy of the brain in symptomatic and asymptomatic patients with liver cirrhosis. *Gastroenterology.* 1997;112(5):1610-6.
41. Kale RA, Gupta RK, Saraswat VA, Hasan KM, Trivedi R, Mishra AM, Ranjan P, Pandey CM, Narayana PA. Demonstration of interstitial cerebral edema with diffusion tensor MR imaging in type C hepatic encephalopathy. *Hepatology.* 2006;43(4):698-706.
42. Scott TR, Kronsten VT, Hughes RD, Shawcross DL. Pathophysiology of cerebral oedema in acute liver failure. *World J Gastroenterol.* 2013;19(48):9240-55.
43. Albrecht J, Norenberg MD. Glutamine: a Trojan horse in ammonia neurotoxicity. *Hepatology.* 2006;44(4):788-94.
44. Stewart VC, Sharpe MA, Clark JB, Heales SJ. Astrocyte-derived nitric oxide causes both reversible and irreversible damage to the neuronal mitochondrial respiratory chain. *J Neurochem.* 2000;75(2):694-700.
45. Celentano JJ, Gyenes M, Gibbs TT, Farb DH. Negative modulation of the gamma-aminobutyric acid response by extracellular zinc. *Mol Pharmacol.* 1991;40(5):766-73.
46. Kruczek C, Gorg B, Keitel V, Bidmon HJ, Schliess F, Haussinger D. Ammonia increases nitric oxide, free Zn(2+), and metallothionein mRNA expression in cultured rat astrocytes. *Biol Chem.* 2011;392(12):1155-65.
47. Gorg B, Foster N, Reinehr R, Bidmon HJ, Hongen A, Haussinger D, Schliess F. Benzodiazepine-induced protein tyrosine nitration in rat astrocytes. *Hepatology.* 2003;37(2):334-42.
48. Shan X, Chang Y, Lin CL. Messenger RNA oxidation is an early event preceding cell death and causes reduced protein expression. *FASEB J.* 2007;21(11):2753-64.
49. Oenarto J, Karababa A, Castoldi M, Bidmon HJ, Gorg B, Haussinger D. Ammonia-induced miRNA expression changes in cultured rat astrocytes. *Sci Rep.* 2016;6:18493.
50. Schliess F, Gorg B, Haussinger D. Pathogenetic interplay between osmotic and oxidative stress: the hepatic encephalopathy paradigm. *Biol Chem.* 2006;387(10-11):1363-70.
51. Murthy CR, Rama Rao KV, Bai G, Norenberg MD. Ammonia-induced production of free radicals in primary cultures of rat astrocytes. *J Neurosci Res.* 2001;66(2):282-8.

52. Reinehr R, Gorg B, Becker S, Qvartskhava N, Bidmon HJ, Selbach O, Haas HL, Schliess F, Haussinger D. Hypoosmotic swelling and ammonia increase oxidative stress by NADPH oxidase in cultured astrocytes and vital brain slices. *Glia*. 2007;55(7):758-71.
53. Shawcross DL, Balata S, Olde Damink SW, Hayes PC, Wardlaw J, Marshall I, Deutz NE, Williams R, Jalan R. Low myo-inositol and high glutamine levels in brain are associated with neuropsychological deterioration after induced hyperammonemia. *Am J Physiol Gastrointest Liver Physiol*. 2004;287(3):G503-9.
54. Haussinger D, Laubenberger J, vom Dahl S, Ernst T, Bayer S, Langer M, Gerok W, Hennig J. Proton magnetic resonance spectroscopy studies on human brain myo-inositol in hypo-osmolarity and hepatic encephalopathy. *Gastroenterology*. 1994;107(5):1475-80.
55. Donovan JP, Schafer DF, Shaw BW, Jr., Sorrell MF. Cerebral oedema and increased intracranial pressure in chronic liver disease. *Lancet*. 1998;351(9104):719-21.
56. Dejong CH, van de Poll MC, Soeters PB, Jalan R, Olde Damink SW. Aromatic amino acid metabolism during liver failure. *J Nutr*. 2007;137(6 Suppl 1):1579S-85S; discussion 97S-98S.
57. Saxena N, Bhatia M, Joshi YK, Garg PK, Dwivedi SN, Tandon RK. Electrophysiological and neuropsychological tests for the diagnosis of subclinical hepatic encephalopathy and prediction of overt encephalopathy. *Liver*. 2002;22(3):190-7.
58. Timmermann L, Butz M, Gross J, Ploner M, Sudmeyer M, Kircheis G, Haussinger D, Schnitzler A. Impaired cerebral oscillatory processing in hepatic encephalopathy. *Clin Neurophysiol*. 2008;119(2):265-72.
59. Gitlin N, Lewis DC, Hinkley L. The diagnosis and prevalence of subclinical hepatic encephalopathy in apparently healthy, ambulant, non-shunted patients with cirrhosis. *J Hepatol*. 1986;3(1):75-82.
60. Randolph C, Hilsabeck R, Kato A, Kharbanda P, Li YY, Mapelli D, Ravdin LD, Romero-Gomez M, Stracciari A, Weissenborn K, International Society for Hepatic E, Nitrogen M. Neuropsychological assessment of hepatic encephalopathy: ISHEN practice guidelines. *Liver Int*. 2009;29(5):629-35.
61. Bajaj JS, Cordoba J, Mullen KD, Amodio P, Shawcross DL, Butterworth RF, Morgan MY, International Society for Hepatic E, Nitrogen M. Review article: the design of clinical trials in hepatic encephalopathy--an International Society for Hepatic Encephalopathy and Nitrogen Metabolism (ISHEN) consensus statement. *Aliment Pharmacol Ther*. 2011;33(7):739-47.
62. Hassanein TI, Hilsabeck RC, Perry W. Introduction to the Hepatic Encephalopathy Scoring Algorithm (HESA). *Dig Dis Sci*. 2008;53(2):529-38.
63. Hassanein T, Blei AT, Perry W, Hilsabeck R, Stange J, Larsen FS, Brown RS, Jr., Caldwell S, McGuire B, Nevens F, Fontana R. Performance of the hepatic encephalopathy scoring algorithm in a clinical trial of patients with cirrhosis and severe hepatic encephalopathy. *Am J Gastroenterol*. 2009;104(6):1392-400.
64. Singh ID, Barclay JA, Cooke WT. Blood-ammonia levels in relation to hepatic coma and the administration of glutamic acid. *Lancet*. 1954;266(6820):1004-7.
65. Lockwood AH. Blood ammonia levels and hepatic encephalopathy. *Metab Brain Dis*. 2004;19(3-4):345-9.
66. Duarte-Rojo A, Estradas J, Hernandez-Ramos R, Ponce-de-Leon S, Cordoba J, Torre A. Validation of the psychometric hepatic encephalopathy score (PHES) for identifying patients with minimal hepatic encephalopathy. *Dig Dis Sci*. 2011;56(10):3014-23.
67. Weissenborn K, Ennen JC, Schomerus H, Ruckert N, Hecker H. Neuropsychological characterization of hepatic encephalopathy. *J Hepatol*. 2001;34(5):768-73.

68. Schomerus H, Weissenborn K, Hecker H, Hamster W, Rückert N. PSE-Syndrom-Test. Psychodiagnostisches Verfahren zur quantitativen Erfassung der (minimalen) portosystemischen Enzephalopathie (PSE). 2nd ed. Hannover, Deutschland: Medizinische Hochschule Hannover, Digitale Medien; 2013.
69. Ennen J. Der PSE-Syndrom-Test. Diagnosestandardisierung der latenten portosystemischen Enzephalopathie mittels psychometrischer Testverfahren. Hannover, Deutschland: Medizinische Hochschule Hannover; 2000.
70. Duff K, Hobson VL, Beglinger LJ, O'Bryant SE. Diagnostic accuracy of the RBANS in mild cognitive impairment: limitations on assessing milder impairments. *Arch Clin Neuropsychol*. 2010;25(5):429-41.
71. Randolph C, Tierney MC, Mohr E, Chase TN. The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): preliminary clinical validity. *J Clin Exp Neuropsychol*. 1998;20(3):310-9.
72. Kircheis G, Wettstein M, Timmermann L, Schnitzler A, Haussinger D. Critical flicker frequency for quantification of low-grade hepatic encephalopathy. *Hepatology*. 2002;35(2):357-66.
73. Romero-Gomez M, Cordoba J, Jover R, del Olmo JA, Ramirez M, Rey R, de Madaria E, Montoliu C, Nunez D, Flavia M, Company L, Rodrigo JM, Felipe V. Value of the critical flicker frequency in patients with minimal hepatic encephalopathy. *Hepatology*. 2007;45(4):879-85.
74. Lauridsen MM, Thiele M, Kimer N, Vilstrup H. The continuous reaction times method for diagnosing, grading, and monitoring minimal/covert hepatic encephalopathy. *Metab Brain Dis*. 2013;28(2):231-4.
75. Bajaj JS, Saeian K, Verber MD, Hischke D, Hoffmann RG, Franco J, Varma RR, Rao SM. Inhibitory control test is a simple method to diagnose minimal hepatic encephalopathy and predict development of overt hepatic encephalopathy. *Am J Gastroenterol*. 2007;102(4):754-60.
76. Guerit JM, Amantini A, Fischer C, Kaplan PW, Mecarelli O, Schnitzler A, Ubiali E, Amodio P, members of the IcoNI. Neurophysiological investigations of hepatic encephalopathy: ISHEN practice guidelines. *Liver Int*. 2009;29(6):789-96.
77. Stroop JR. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*. 1935;18(6):643-62.
78. Bajaj JS, Heuman DM, Sterling RK, Sanyal AJ, Siddiqui M, Matherly S, Luketic V, Stravitz RT, Fuchs M, Thacker LR, Gilles H, White MB, Unser A, Hovermale J, Gavis E, Noble NA, Wade JB. Validation of EncephalApp, Smartphone-Based Stroop Test, for the Diagnosis of Covert Hepatic Encephalopathy. *Clin Gastroenterol Hepatol*. 2014;13:1828-35.
79. Bajaj JS, Thacker LR, Heuman DM, Fuchs M, Sterling RK, Sanyal AJ, Puri P, Siddiqui MS, Stravitz RT, Bouneva I, Luketic V, Noble N, White MB, Monteith P, Unser A, Wade JB. The Stroop smartphone application is a short and valid method to screen for minimal hepatic encephalopathy. *Hepatology*. 2013;58(3):1122-32.
80. Gluud LL, Dam G, Borre M, Les I, Cordoba J, Marchesini G, Aagaard NK, Risum N, Vilstrup H. Oral branched-chain amino acids have a beneficial effect on manifestations of hepatic encephalopathy in a systematic review with meta-analyses of randomized controlled trials. *J Nutr*. 2013;143(8):1263-8.
81. Holecek M. Branched-chain amino acids and ammonia metabolism in liver disease: therapeutic implications. *Nutrition*. 2013;29(10):1186-91.
82. Morgan MY, Alonso M, Stanger LC. Lactitol and lactulose for the treatment of subclinical hepatic encephalopathy in cirrhotic patients. A randomised, cross-over study. *J Hepatol*. 1989;8(2):208-17.

83. Strauss E, Tramote R, Silva EP, Caly WR, Honain NZ, Maffei RA, de Sa MF. Double-blind randomized clinical trial comparing neomycin and placebo in the treatment of exogenous hepatic encephalopathy. *Hepatogastroenterology*. 1992;39(6):542-5.
84. Bajaj JS, Etemadian A, Hafeezullah M, Saeian K. Testing for minimal hepatic encephalopathy in the United States: An AASLD survey. *Hepatology*. 2007;45(3):833-4.
85. Romero-Gomez M, Boza F, Garcia-Valdecasas MS, Garcia E, Aguilar-Reina J. Subclinical hepatic encephalopathy predicts the development of overt hepatic encephalopathy. *Am J Gastroenterol*. 2001;96(9):2718-23.
86. WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects Ferney-Voltaire, France [updated 2016. accessed June 16, 2016. Available from: <http://www.wma.net/en/30publications/10policies/b3/index.html>.]
87. Biecker E, Hausdorfer I, Grunhage F, Strunk H, Sauerbruch T. Critical flicker frequency as a marker of hepatic encephalopathy in patients before and after transjugular intrahepatic portosystemic shunt. *Digestion*. 2011;83(1-2):24-31.
88. Merola J, Chaudhary N, Qian M, Jow A, Barboza K, Charles H, Teperman L, Sigal S. Hyponatremia: A Risk Factor for Early Overt Encephalopathy after Transjugular Intrahepatic Portosystemic Shunt Creation. *J Clin Med*. 2014;3(2):359-72.
89. Padmanabhan U, Leslie K, Eer AS, Maruff P, Silbert BS. Early cognitive impairment after sedation for colonoscopy: the effect of adding midazolam and/or fentanyl to propofol. *Anesth Analg*. 2009;109(5):1448-55.
90. Campagna F, Montagnese S, Schiff S, Biancardi A, Mapelli D, Angeli P, Poci C, Cillo U, Merkel C, Gatta A, Amodio P. Cognitive impairment and electroencephalographic alterations before and after liver transplantation: what is reversible? *Liver Transpl*. 2014;20(8):977-86.
91. Philips BJ, Lane K, Dixon J, Macphee I. The effects of acute renal failure on drug metabolism. *Expert Opin Drug Metab Toxicol*. 2014;10(1):11-23.
92. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12(3):189-98.
93. Folstein MF, Robins LN, Helzer JE. The Mini-Mental State Examination. *Arch Gen Psychiatry*. 1983;40(7):812.
94. da Fonseca-Wollheim F. Preanalytical increase of ammonia in blood specimens from healthy subjects. *Clin Chem*. 1990;36(8 Pt 1):1483-7.
95. Wiesner R, Edwards E, Freeman R, Harper A, Kim R, Kamath P, Kremers W, Lake J, Howard T, Merion RM, Wolfe RA, Krom R, United Network for Organ Sharing Liver Disease Severity Score C. Model for end-stage liver disease (MELD) and allocation of donor livers. *Gastroenterology*. 2003;124(1):91-6.
96. Wiesner RH, McDiarmid SV, Kamath PS, Edwards EB, Malinchoc M, Kremers WK, Krom RA, Kim WR. MELD and PELD: application of survival models to liver allocation. *Liver Transpl*. 2001;7(7):567-80.
97. Pugh RN, Murray-Lyon IM, Dawson JL, Pietroni MC, Williams R. Transection of the oesophagus for bleeding oesophageal varices. *Br J Surg*. 1973;60(8):646-9.
98. Wuensch T, Ruether DF, Zollner C, Mueller T, Jung T, Kaffarnik M, Kassner U, Schott E, Kiefer S, Pratschke J, Stockmann M, Jara M. Performance characterization of a novel electronic number connection test to detect minimal hepatic encephalopathy in cirrhotic patients. *Eur J Gastroenterol Hepatol*. 2017;29(4):456-63.

99. Weissenborn K, Ruckert N, Hecker H, Manns MP. The number connection tests A and B: interindividual variability and use for the assessment of early hepatic encephalopathy. *J Hepatol.* 1998;28(4):646-53.
100. Joebges EM, Heidemann M, Schimke N, Hecker H, Ennen JC, Weissenborn K. Bradykinesia in minimal hepatic encephalopathy is due to disturbances in movement initiation. *J Hepatol.* 2003;38(3):273-80.
101. Held L, Rufibach K, Seifert B. *Medizinische Statistik: Konzepte, Methoden, Anwendungen.* 1st ed. München, Deutschland: Pearson Studium; 2013.
102. Henderson AR. Testing experimental data for univariate normality. *Clin Chim Acta.* 2006;366(1-2):112-29.
103. Kim HY. Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restor Dent Endod.* 2013;38(1):52-4.
104. Backhaus K, Erichson B, Plinke W, Weiber R. *Multivariate Analysemethoden.* 14th ed. Berlin & Heidelberg, Deutschland: Springer-Verlag; 2016. p. 63-134; 73-213.
105. Field A. *Discovering Statistics Using SPSS.* 4th ed. London, UK: SAGE Publications Ltd; 2013. p. 347-94; 539-83.
106. Wasserstein RL, Lazar NA. The ASA's Statement on p-Values: Context, Process, and Purpose. *The American Statistician.* 2016;70(2):129-33.
107. Bühner M. *Einführung in die Test- und Fragebogenkonstruktion.* 3rd ed. München, Deutschland: Pearson Studium; 2010. p. 33-44.
108. Jensen K, Abel U. [Methodology of diagnostic validation studies. Errors in study planning and evaluation]. *Med Klin (Munich).* 2000;95(1 Spec No):54-60.
109. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, Lijmer JG, Moher D, Rennie D, de Vet HC, Kressel HY, Rifai N, Golub RM, Altman DG, Hooft L, Korevaar DA, Cohen JF, Group S. STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies. *Clin Chem.* 2015;61(12):1446-52.
110. Durlak JA. How to select, calculate, and interpret effect sizes. *J Pediatr Psychol.* 2009;34(9):917-28.
111. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales - a practical guide to their development and use.* 5th ed. Oxford, UK: Oxford University Press; 2015.
112. McGraw K, Wong S. Forming Inferences About Some Intraclass Correlation Coefficients. *Psychological Methods.* 1996;1(1):30-46.
113. Fischer JE, Bachmann LM, Jaeschke R. A readers' guide to the interpretation of diagnostic test properties: clinical example of sepsis. *Intensive Care Med.* 2003;29(7):1043-51.
114. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA.* 1994;271(9):703-7.
115. Irwig L, Tosteson AN, Gatsonis C, Lau J, Colditz G, Chalmers TC, Mosteller F. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med.* 1994;120(8):667-76.
116. Arroyo V, Gines P, Gerbes AL, Dudley FJ, Gentilini P, Laffi G, Reynolds TB, Ring-Larsen H, Scholmerich J. Definition and diagnostic criteria of refractory ascites and hepatorenal syndrome in cirrhosis. International Ascites Club. *Hepatology.* 1996;23(1):164-76.

117. Sivanathan V, Kittner JM, Sprinzl MF, Weinmann A, Koch S, Wiltink J, Nguyen-Tat M, Marquardt JU, Worns MA, Zimmermann T, Lang H, Galle PR, Schattenberg JM. [Etiology and complications of liver cirrhosis: data from a German centre]. *Dtsch Med Wochenschr.* 2014;139(36):1758-62.
118. Gualtieri CT, Johnson LG. Reliability and validity of a computerized neurocognitive test battery, CNS Vital Signs. *Arch Clin Neuropsychol.* 2006;21(7):623-43.
119. Ferenci P, Ferenci S, Datz C, Rezman I, Oberaigner W, Strauss R. Morbidity and mortality in paid Austrian plasma donors infected with hepatitis C at plasma donation in the 1970s. *J Hepatol.* 2007;47(1):31-6.
120. Wiese M, Grungreiff K, Guthoff W, Lafrenz M, Oesen U, Porst H, East German Hepatitis CSG. Outcome in a hepatitis C (genotype 1b) single source outbreak in Germany--a 25-year multicenter study. *J Hepatol.* 2005;43(4):590-8.
121. Pomier-Layrargues G, Nguyen NH, Faucher C, Giguère JF, Butterworth RF. Subclinical Hepatic Encephalopathy in Cirrhotic Patients: Prevalence and Relationship to Liver Function. *Canadian Journal of Gastroenterology.* 1991;5(4).
122. Sharma P, Kumar A, Singh S, Tyagi P, Kumar A. Inhibitory control test, critical flicker frequency, and psychometric tests in the diagnosis of minimal hepatic encephalopathy in cirrhosis. *Saudi J Gastroenterol.* 2013;19(1):40-4.
123. Allampati S, Duarte-Rojo A, Thacker LR, Patidar KR, White MB, Klair JS, John B, Heuman DM, Wade JB, Flud C, O'Shea R, Gavis EA, Unser AB, Bajaj JS. Diagnosis of Minimal Hepatic Encephalopathy Using Stroop EncephalApp: A Multicenter US-Based, Norm-Based Study. *Am J Gastroenterol.* 2016;111(1):78-86.
124. Amodio P, Campagna F, Olianias S, Iannizzi P, Mapelli D, Penzo M, Angeli P, Gatta A. Detection of minimal hepatic encephalopathy: normalization and optimization of the Psychometric Hepatic Encephalopathy Score. A neuropsychological and quantified EEG study. *J Hepatol.* 2008;49(3):346-53.
125. Dhiman RK, Saraswat VA, Verma M, Naik SR. Figure connection test: a universal test for assessment of mental state. *J Gastroenterol Hepatol.* 1995;10(1):14-23.
126. Romero Gomez M, Cordoba J, Jover R, del Olmo J, Fernandez A, Flavia M, Company L, Poveda MJ, Felipe V, Red Nacional de Investigacion en Encefalopatia H. [Normality tables in the Spanish population for psychometric tests used in the diagnosis of minimal hepatic encephalopathy]. *Med Clin (Barc).* 2006;127(7):246-9.
127. Seo YS, Yim SY, Jung JY, Kim CH, Kim JD, Keum B, An H, Yim HJ, Lee HS, Kim CD, Ryu HS, Um SH. Psychometric hepatic encephalopathy score for the detection of minimal hepatic encephalopathy in Korean patients with liver cirrhosis. *J Gastroenterol Hepatol.* 2012;27(11):1695-704.
128. Wunsch E, Koziarska D, Kotarska K, Nowacki P, Milkiewicz P. Normalization of the psychometric hepatic encephalopathy score in Polish population. A prospective, quantified electroencephalography study. *Liver Int.* 2013;33(9):1332-40.
129. Goldbecker A, Weissenborn K, Hamidi Shahrezaei G, Afshar K, Rumke S, Barg-Hock H, Strassburg CP, Hecker H, Tryc AB. Comparison of the most favoured methods for the diagnosis of hepatic encephalopathy in liver transplantation candidates. *Gut.* 2013;62(10):1497-504.
130. Kircheis G, Fleig WE, Gortelmeyer R, Grafe S, Haussinger D. Assessment of low-grade hepatic encephalopathy: a critical analysis. *J Hepatol.* 2007;47(5):642-50.
131. Bajaj JS, Hafeezullah M, Franco J, Varma RR, Hoffmann RG, Knox JF, Hirschke D, Hammeke TA, Pinkerton SD, Saeian K. Inhibitory control test for the diagnosis of minimal hepatic encephalopathy. *Gastroenterology.* 2008;135(5):1591-600 e1.
132. Swets JA. Measuring the accuracy of diagnostic systems. *Science.* 1988;240(4857):1285-93.

133. Conn HO. Trailmaking and number-connection tests in the assessment of mental state in portal systemic encephalopathy. *Am J Dig Dis.* 1977;22(6):541-50.
134. Tarter RE, Hegedus AM, Van Thiel DH, Schade RR, Gavaler JS, Starzl TE. Nonalcoholic cirrhosis associated with neuropsychological dysfunction in the absence of overt evidence of hepatic encephalopathy. *Gastroenterology.* 1984;86(6):1421-7.
135. Babyak MA. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosom Med.* 2004;66(3):411-21.
136. Schomerus H, Hamster W. Neuropsychological aspects of portal-systemic encephalopathy. *Metab Brain Dis.* 1998;13(4):361-77.
137. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-74.
138. Bewick V, Cheek L, Ball J. Statistics review 7: Correlation and regression. *Crit Care.* 2003;7(6):451-9.
139. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1(8476):307-10.
140. MacLeod CM. Half a century of research on the Stroop effect: an integrative review. *Psychol Bull.* 1991;109(2):163-203.
141. Bermudez EB, Klerman EB, Czeisler CA, Cohen DA, Wyatt JK, Phillips AJ. Prediction of Vigilant Attention and Cognitive Performance Using Self-Reported Alertness, Circadian Phase, Hours since Awakening, and Accumulated Sleep Loss. *PLoS One.* 2016;11(3):e0151770.
142. Dijk DJ, Duffy JF, Czeisler CA. Circadian and sleep/wake dependent aspects of subjective alertness and cognitive performance. *J Sleep Res.* 1992;1(2):112-7.
143. Rahimi RS, Elliott AC, Rockey DC. Altered mental status in cirrhosis: etiologies and outcomes. *J Investig Med.* 2013;61(4):695-700.
144. O'Kane M. The reporting, classification and grading of quality failures in the medical laboratory. *Clin Chim Acta.* 2009;404(1):28-31.
145. Victor A, Elsasser A, Hommel G, Blettner M. Judging a plethora of p-values: how to contend with the problem of multiple testing--part 10 of a series on evaluation of scientific publications. *Dtsch Arztebl Int.* 2010;107(4):50-6.
146. Kircheis G, Hilger N, Haussinger D. Value of critical flicker frequency and psychometric hepatic encephalopathy score in diagnosis of low-grade hepatic encephalopathy. *Gastroenterology.* 2014;146(4):961-9.
147. Weissenborn K, Hecker H, Ruckert N. Pitfalls in psychometric testing. *Gastroenterology.* 2015;148(3):664-5.
148. Shawcross DL, Dunk AA, Jalan R, Kircheis G, de Knegt RJ, Laleman W, Ramage JK, Wedemeyer H, Morgan IE, New Insights Steering C. How to diagnose and manage hepatic encephalopathy: a consensus statement on roles and responsibilities beyond the liver specialist. *Eur J Gastroenterol Hepatol.* 2016;28(2):146-52.

9 Eidesstattliche Versicherung

„Ich, Ferenc Darius R ther, versichere an Eides statt durch meine eigenh ndige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema *„Diagnostische Wertigkeit eines elektronischen Tablet-Computerbasierten Zahlenverbindungstests zur Detektion der minimalen hepatischen Enzephalopathie bei Patienten mit Leberzirrhose“* selbstst ndig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die w rtlich oder dem Sinne nach auf Publikationen oder Vortr gen anderer Autoren beruhen, sind als solche in korrekter Zitierung (siehe „Uniform Requirements for Manuscripts (URM)“ des ICMJE – www.icmje.org) kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Grafiken und Tabellen) entsprechen den URM (s.o.) und werden von mir verantwortet.

Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erkl rung mit dem Betreuer angegeben sind. S mtliche Publikationen, die aus dieser Dissertation hervorgegangen sind und bei denen ich Autor bin, entsprechen den URM (s.o.) und werden von mir verantwortet.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§156,161 des Strafgesetzbuches) sind mir bekannt und bewusst.“

Datum

Unterschrift

Anteilserkl rung an erfolgten Publikationen

Ferenc Darius R ther hatte folgenden Anteil an der folgenden Publikation:

Publikation 1: Wuensch T, **Ruether DF**, Zollner C, Mueller T, Jung T, Kaffarnik M, Kassner U, Schott E, Kiefer S, Pratschke J, Stockmann M, Jara M. Performance characterization of a novel electronic number connection test to detect minimal hepatic encephalopathy in cirrhotic patients. *European Journal of Gastroenterology & Hepatology*. 2017;29(4):456-63.

Beitrag im Einzelnen:

Ferenc Darius R ther leistete einen gro en Beitrag zur Durchf hrung der Studie durch Optimierung der Studienplanung, Datenerhebung von der H lfte der Studienteilnehmer (Befragung, Untersuchung, Testdurchf hrung und ggf. Studium der elektronischen Patientenakten),  bertragung der Daten in eine SPSS-Datenbank und Verwaltung der SPSS-Datenbank. Au erdem trug er durch statistische Analysen, Interpretation erhobener Daten sowie schriftliche Entw rfe und der kritischen Durchsicht des finalen Manuskripts wesentlich zur erfolgten Publikation bei.

Unterschrift, Datum und Stempel des betreuenden Hochschullehrers

Unterschrift des Doktoranden
(Ferenc Darius R ther)

10 Lebenslauf

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

11 Publikationsliste

Publikation 1:

Wuensch T, **Ruether DF**, Zollner C, Mueller T, Jung T, Kaffarnik M, Kassner U, Schott E, Kiefer S, Pratschke J, Stockmann M, Jara M. Performance characterization of a novel electronic number connection test to detect minimal hepatic encephalopathy in cirrhotic patients. *European Journal of Gastroenterology & Hepatology*. 2017;29(4):456-63.

12 Danksagung

Der erfolgreiche Abschluss der vorliegenden Dissertation wäre mir ohne die Hilfe folgender Menschen nicht möglich gewesen. Mein besonderer Dank gilt daher:

Herrn Professor Dr. med. Martin Stockmann
für die Überlassung des interessanten Themas und die hilfreichen Anregungen zur Durchführung und Auswertung der Studie.

Meinem Betreuer, Herrn Dr. rer. nat. Tilo Wunsch,
der als Ansprechpartner jederzeit für mich erreichbar war und mich mit seinem großen Arbeitseinsatz in der Betreuung der Studie fortlaufend motivierte.

Herrn Andreas Böttner und Herrn Klaus Lenz
vom Institut für Biometrie und Klinische Epidemiologie der Charité – Universitätsmedizin Berlin für ihre fachkundigen Impulse zur Herangehensweise an eine statistische Auswertung wissenschaftlicher Daten.

Frau Professor Dr. med. Karin Weissenborn
für die Bereitstellung von PHES-Handbuch und Testunterlagen.

Den Mitarbeitern der verschiedenen Krankenhausabteilungen
für ihre bereitwillige Kooperation zur Rekrutierung von Probanden.

Allen Patienten und Probanden,
die mit ihrer investierten Zeit zum Erfolg der Studie beigetragen haben.

Meiner Familie und meinen Freunden
für ihre Unterstützung und Rücksichtnahme während arbeitsintensiver Phasen.

13 Anhang

Anhang 1: Laborwertetabelle

		N	Mittelwert ± SD	Median	Minimum	Maximum
Natrium (mmol/L)	LZI	85	138 ± 5	138	118	147
	oHE	12	135 ± 5	135	124	147
	≥39J	55	141 ± 3	141	131	146
Kalium (mmol/L)	LZI	85	4,0 ± 0,5	4,0	2,9	5,1
	oHE	12	4,2 ± 0,6	4,1	3,4	5,5
	≥39J	54	4,4 ± 0,5	4,3	3,5	5,9
Kreatinin (mg/dl)	LZI	88	1,03 ± 0,70	0,83	0,49	5,39
	oHE	12	1,32 ± 0,77	1,13	0,59	2,70
	≥39J	54	0,90 ± 0,27	0,83	0,56	2,07
INR-Wert	LZI	85	1,31 ± 0,25	1,27	0,96	2,53
	oHE	11	1,53 ± 0,27	1,45	1,13	2,13
	≥39J	56	1,10 ± 0,33	1,02	0,89	3,05
Quick (%)	LZI	86	70 ± 15	68	30	107
	oHE	12	57 ± 12	58	36	81
	≥39J	56	93 ± 20	97	25	122
Albumin (g/L)	LZI	83	37 ± 7	37	23	50
	oHE	9	29 ± 5	31	20	35
	≥39J	52	46 ± 2	46	40	50
AP (U/L)	LZI	86	162 ± 159	125	31	1355
	oHE	12	160 ± 124	110	53	421
	≥39J	51	65 ± 14	62	36	104
GGT (U/L)	LZI	88	168 ± 200*	100	13	1278
	oHE	12	154 ± 131	105	13	434
	≥39J	54	33 ± 25	24	10	134



*: bei Einheiten, die keine negativen Werte annehmen können, ist eine Standardabweichung größer als der Mittelwert ein rein mathematisches Konstrukt. Im Weiteren wurde hier der Median verwendet. N: Anzahl; SD: Standardabweichung; LZI: Patienten mit Leberzirrhose ohne klinische Zeichen einer HE; oHE: Patienten mit Leberzirrhose und klinischen Zeichen einer HE; ≥39J: lebergesunde Probanden ab 39 Jahren; INR: International Normalized Ratio; AP: Alkalische Phosphatase; GGT: Gammaglutamyltransferase.

Anhang 1: Laborwertetabelle (Fortsetzung)

		N	Mittelwert ± SD	Median	Minimum	Maximum
AST (U/L)	LZI	87	70 ± 82*	48	13	688
	oHE	12	87 ± 76	45	29	233
	≥39J	52	29 ± 15	25	7	112
ALT (U/L)	LZI	88	48 ± 51*	34	10	374
	oHE	12	56 ± 55	25	14	172
	≥39J	54	26 ± 10	24	8	51
Hb (g/dl)	LZI	87	12,4 ± 2,4	12,4	7,0	17,1
	oHE	12	10,5 ± 2,7	9,9	6,6	15,6
	≥39J	55	14,7 ± 1,2	14,8	11,5	17,1
Faktor II (%)	LZI	40	68 ± 20	64	32	121
	oHE	3	46 ± 8	46	38	53
	≥39J	31	100 ± 20	103	35	126
Thrombozytenanzahl / nl	LZI	87	118 ± 63	100	32	343
	oHE	12	105 ± 66	102	29	212
	≥39J	55	234 ± 54	231	141	364
Bilirubin (mg/dl)	LZI	87	2,7 ± 5,9	1,2	0,2	49,2
	oHE	12	6,5 ± 8,0	1,8	0,9	22,0
	≥39J	52	0,5 ± 0,3	0,5	0,2	1,3
Ammoniak (µmol/L)	LZI	78	53 ± 25	49	13	122
	oHE	10	65 ± 28	73	32	104
	≥39J	45	27 ± 8	28	11	47
MELD-Score	LZI	85	12,6 ± 5,3	10,4	6,4	31,6
	oHE	11	18,9 ± 8,6	16,9	7,8	36,1
	≥39J	50	7,9 ± 3,3	6,9	6,4	22,7
Child-Pugh-Score	LZI	81	7 ± 2	7	5	13
	oHE	9	10 ± 2	11	7	13
	≥39J	50	5 ± 1	5	5	7

*: bei Einheiten, die keine negativen Werte annehmen können, ist eine Standardabweichung größer als der Mittelwert ein rein mathematisches Konstrukt. Im Weiteren wurde hier der Median verwendet. N: Anzahl; SD: Standardabweichung; LZI: Patienten mit Leberzirrhose ohne klinische Zeichen einer HE; oHE: Patienten mit Leberzirrhose und klinischen Zeichen einer HE; ≥39J: leberge-sunde Probanden ab 39 Jahren; AST: Aspartat-Aminotransferase; ALT: Alanin-Aminotransferase; Hb: Hämoglobin; MELD: Model for End-Stage Liver Disease.



Anhang 2: CRF-Dokumentationsbogen

	RE-TEST: <input type="checkbox"/> JA <input type="checkbox"/> NEIN; DATUM nächster Charité-Termin:	 <small>CharitéCentrum für Chirurgische Medizin</small>
d-LIVER eNCT Evaluation - Einschluss		
Name: _____	Datum: <input type="text"/> / <input type="text"/> / <input type="text"/>	
Vorname: _____	Teilnehmer-ID: <input type="text"/> _ <input type="text"/> _ <input type="text"/> <input type="text"/> / <input type="text"/>	
Geburtsdatum: <input type="text"/> / <input type="text"/> / <input type="text"/>	Tel./Mobil: _____	
Hausarzt: _____	Adresse: _____	
Tel./Mobil: _____		_____

Patient: <input type="checkbox"/>		Proband: <input type="checkbox"/>
Stammdaten		
Alter: <input type="text"/> <input type="text"/>	Beschäftigung:	Höchster Abschluss:
Geschlecht: m <input type="checkbox"/> w <input type="checkbox"/>	Vollzeit <input type="checkbox"/>	Hochschule <input type="checkbox"/>
Gewicht: <input type="text"/> <input type="text"/> <input type="text"/> kg	Teilzeit <input type="checkbox"/>	Gymnasium <input type="checkbox"/>
Größe: <input type="text"/> <input type="text"/> <input type="text"/> cm	Arbeitslos <input type="checkbox"/>	Mittelschule <input type="checkbox"/>
Anzahl der Schuljahre: <input type="text"/> <input type="text"/>	Arbeitsunfähig <input type="checkbox"/>	Kein Abschluss <input type="checkbox"/>
	Berentet/im Ruhestand <input type="checkbox"/>	
	Student <input type="checkbox"/>	

Ein- und Ausschlusskriterien	
Sehhilfe	Ja <input type="checkbox"/> Nein <input type="checkbox"/>
Rot-Grün Sehschwäche	Ja <input type="checkbox"/> Nein <input type="checkbox"/>
Bekannte Lebererkrankung	Ja <input type="checkbox"/> Nein <input type="checkbox"/>
Bei bekannter Lebererkrankung	
Bestätigte Leberzirrhose durch Biopsie	Ja <input type="checkbox"/> Nein <input type="checkbox"/>
Datum Biopsie	<input type="text"/> / <input type="text"/> / <input type="text"/>
ODER	
V.a. Leberzirrhose – wenn 2 der folgenden Kriterien zutreffen	
(I) Unregelmäßige Begrenzung der Leber im CT oder MRT	Ja <input type="checkbox"/> Nein <input type="checkbox"/> unbekannt <input type="checkbox"/>
(II) Milz >12cm	Ja <input type="checkbox"/> Nein <input type="checkbox"/> unbekannt <input type="checkbox"/>
(III) Fibroscan >17,6 kPa	Ja <input type="checkbox"/> Nein <input type="checkbox"/> unbekannt <input type="checkbox"/>
(IV) Varizen/PHG	Ja <input type="checkbox"/> Nein <input type="checkbox"/> unbekannt <input type="checkbox"/>
(V) Thrombozyten <120/nl	Ja <input type="checkbox"/> Nein <input type="checkbox"/> unbekannt <input type="checkbox"/>
(VI) Aszites	Ja <input type="checkbox"/> Nein <input type="checkbox"/> Grad mild <input type="checkbox"/> moderate <input type="checkbox"/> schwer <input type="checkbox"/>
(VII) Hepatische Enzephalopathie nach Westhaven	Ja <input type="checkbox"/> Nein <input type="checkbox"/> Grad 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/>
Neuro-/Psychiatrische Einschränkung	
Bekannte Neuro-/psychiatrische Erkrankung	Ja <input type="checkbox"/> Nein <input type="checkbox"/>
Aktuelle Medikation mit Benzodiazepinen oder Zolpidem	Ja <input type="checkbox"/> Nein <input type="checkbox"/>
Bei regelmäßige Einnahme von Psychopharmaka:	
(I) Vom Arzt verschriebene Schlafmedikation (Quetiapine oder Melperon)	Ja <input type="checkbox"/> Nein <input type="checkbox"/>
(II) Regelmäßige Einnahme von Antidepressive und stabiler	

Anhang 2: CRF-Dokumentationsbogen (Fortsetzung)

		 CharitéCentrum für Chirurgische Medizin	
Krankheitsverlauf der Depression			
Mini Mental Status Test (MMST) >25 Punkte		Ja <input type="checkbox"/> Nein <input type="checkbox"/>	
Erreichte Punkte MMST: <input type="text"/> <input type="text"/>			
Details zur Lebererkrankung			
Ätiologie: ALD <input type="checkbox"/> NAFLD <input type="checkbox"/> HBV <input type="checkbox"/> HCV <input type="checkbox"/> PBC <input type="checkbox"/> PSC <input type="checkbox"/> AIH <input type="checkbox"/> Hämochromatose <input type="checkbox"/> kryptogen <input type="checkbox"/> Andere:	Bildgebung: Ultraschall Abdomen CT Abdomen/Leber MRT Abdomen/Leber	Erfolgt? <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	Zirrhose? <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Milz >12cm? <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>			
Fibroscan: Ja <input type="checkbox"/> Nein <input type="checkbox"/> Datum Fibroscan: <input type="text"/> / <input type="text"/> / <input type="text"/>			
Fibroscan-Stiffness: <input type="text"/> <input type="text"/> , <input type="text"/> kPa			
TIPSS <input type="checkbox"/> Nein, <input type="checkbox"/> Ja, wann _____			
Aszites			
Aktueller Aszites: Ja <input type="checkbox"/> Nein <input type="checkbox"/> Falls ja, Schweregrad: mild <input type="checkbox"/> moderate <input type="checkbox"/> schwer <input type="checkbox"/> Aktuelle periphere Ödeme: Ja <input type="checkbox"/> Nein <input type="checkbox"/> Therapierefraktärer Aszites: Ja <input type="checkbox"/> Nein <input type="checkbox"/> Frühere Aszitespunktionen: Ja <input type="checkbox"/> Nein <input type="checkbox"/> Anzahl der Aszitespunktionen in den letzten 12 Monaten: <input type="text"/> <input type="text"/> Datum letzte Aszitespunktion: <input type="text"/> / <input type="text"/> / <input type="text"/> Frühere Spontanperitonitis: Ja <input type="checkbox"/> Nein <input type="checkbox"/> Antibiotika-Prophylaxe: Ja <input type="checkbox"/> Nein <input type="checkbox"/> Wenn ja welches:.....	KG 12/12 Max <input type="text"/> <input type="text"/> <input type="text"/> kg Min <input type="text"/> <input type="text"/> <input type="text"/> kg Diuretika: Ja <input type="checkbox"/> Nein <input type="checkbox"/> Diuretika-Unverträglichkeit: Ja <input type="checkbox"/> Nein <input type="checkbox"/> Furosemid-Dosis: <input type="text"/> <input type="text"/> <input type="text"/> mg Regime: Spironolacton-Dosis: <input type="text"/> <input type="text"/> <input type="text"/> mg Regime: Andere Diuretika & Dosis:		
Enzephalopathie			
Jemals stat. Aufnahmen wegen Enzephalopathie Frühere stat. Aufnahmen wegen Enzephalopathie 12M Wenn ja, Anzahl Wenn ja, Datum letzter Aufnahme Frühere medikament. Enzephalopathie-Prophylaxe ITS-Aufenthalt wegen Enzephalopathie Aktuell Laktulose zur Enzephalopathie-Prophylaxe Aktuell Rifaximin zur Enzephalopathie-Prophylaxe	Ja <input type="checkbox"/> Nein <input type="checkbox"/> Ja <input type="checkbox"/> Nein <input type="checkbox"/> <input type="text"/> <input type="text"/> / <input type="text"/> <input type="text"/> / <input type="text"/> Ja <input type="checkbox"/> Nein <input type="checkbox"/> Ja <input type="checkbox"/> Nein <input type="checkbox"/> Ja <input type="checkbox"/> Nein <input type="checkbox"/> Ja <input type="checkbox"/> Nein <input type="checkbox"/>		