

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

DISSERTATION

**Methodische Bewertung eines elektronischen
Zahlenverbindungstests (eNCT) bei Patienten mit
Leberzirrhose am Beispiel der minimalen hepatischen
Enzephalopathie**

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

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Tim Brötzmann
aus Magdeburg

Datum der Promotion: 18.09.2020

Vorwort

Die dieser Arbeit zugrundeliegenden Daten entstammen einer von d-LIVER (Projekt der EU-Kommission im Rahmen des 7 Framework Programms – Projektnummer: 287596) initiierten und finanzierten prospektiven Querschnittsstudie zur Charakterisierung eines neuartigen, elektronischen Zahlenverbindungstests (eNCT) bezüglich der Detektion einer minimalen hepatischen Enzephalopathie (mHE) bei Patienten mit manifester Leberzirrhose. Das d-LIVER Projekt zielt durch die Entwicklung eines patientenzentrierten telemedizinischen Ansatzes und durch die Entwicklung von speziellen Geräten auf eine Früherkennung von klassischen Krankheitsepisoden chronisch leberkranker Patienten ab.

Inhaltsverzeichnis

Abstrakt (Deutsch)	6
Abstract (English).....	8
Abkürzungsverzeichnis	9
1 Einleitung und Einführung in die Fragestellung	11
1.1 Leberzirrhose.....	11
1.2 Hepatische Enzephalopathie (HE)	14
1.2.1 Bedeutung der HE.....	14
1.2.2 Pathogenese der HE.....	14
1.2.3 Ätiologie, Diagnostik und Graduierung.....	15
1.3 Klinische Relevanz und Diagnostik der minimalen HE (mHE)	17
1.3.1 Wertigkeit des Zahlenverbindungstests (NCT) in der Diagnose der mHE .	20
1.4 Grundlagen methodischer Bewertung	22
1.4.1 Gütekriterien.....	22
1.4.2 Nebengütekriterien	24
1.5 Einführung in die Fragestellung.....	24
2 Methoden.....	26
2.1 Studienpopulation und Studiendesign	26
2.1.1 Ein- und Ausschlusskriterien.....	26
2.2 Untersuchungsablauf.....	27
2.3 Psychometrische Testverfahren	28
2.3.1 Portosystemischer Enzephalopathie (PSE)-Syndrom-Test	28
2.3.1.1 Testanweisung und -bearbeitung	29
2.3.1.2 Einzeltestverfahren.....	30
2.3.1.3 Auswertung und Interpretation	31
2.3.2 Elektronischer Zahlenverbindungstest (eNCT)	32
2.3.2.1 Konzept	32
2.3.2.2 Testanweisung und -bearbeitung	33
2.3.2.3 Auswertung und Interpretation	35
2.3.3 EncephallApp-Stroop Test (Stroop)	35
2.4 Fragebogen zum Feedback der Studienteilnehmer	35
2.5 Studienaufbau & statistische Methoden	35
2.5.1 Multiples Testen	38

3	Ergebnisse.....	39
3.1	Charakterisierung der Studienkohorte.....	39
3.1.1	Demographische Daten der Studienteilnehmer	40
3.1.2	Charakterisierung des Patientenkollektivs	40
3.2	Ergebnisse der im Studienprotokoll festgelegten Tests	42
3.2.1	Ergebnisse der Untertests im PSE-Syndrom-Test.....	42
3.2.2	Ergebnisse der Paralleltests im eNCT	43
3.3	Reliabilität.....	44
3.3.1	Retestreliabilität.....	44
3.3.2	Paralleltestreliabilität	45
3.3.3	Lerneffekte	48
3.4	Validität.....	49
3.4.1	Konstruktvalidität.....	49
3.4.2	Kriteriumsvalidität.....	50
3.5	Nebengütekriterien	51
3.6	Einflussfaktoren auf die eNCT-Performanz	54
3.6.1	Alter, Geschlecht und Bildungsgrad.....	54
3.6.2	Weitere eine Leberzirrhose charakterisierende Faktoren	57
3.6.3	Bildschirmdiagonale (Tabletgröße)	62
4	Diskussion	63
4.1	Zusammenfassung der Ergebnisse.....	63
4.2	Gütekriterien	64
4.2.1	Objektivität	64
4.2.2	Reliabilität.....	65
4.2.3	Validität	66
4.2.4	Nebengütekriterien.....	68
4.3	Einflussfaktoren auf die eNCT-Performanz	69
4.4	Limitationen	72
4.5	Klinischer Nutzen und Ausblick	74
5	Literaturverzeichnis.....	77
6	Abbildungsverzeichnis	86
7	Tabellenverzeichnis	87
8	Erklärung an Eides Statt.....	88
9	Curriculum Vitae	89

10	Danksagung.....	90
11	Anhang.....	91

Abstrakt (Deutsch)

Einleitung: Die Hepatische Enzephalopathie (HE) stellt eine häufige Komplikation der manifesten Leberzirrhose mit erheblichen Auswirkungen auf die Lebensqualität und Arbeitsfähigkeit der Patienten dar. Insbesondere die minimale Form der HE (mHE) unterliegt aufgrund der unterschweligen klinischen Symptomatik und ihrem episodenhaften Auftreten einer erschwerten Diagnosestellung. Die als Referenzstandard geltenden psychometrischen Testverfahren sind zeitaufwändig, erfordern geschultes Personal und eignen sich nicht für einen regelmäßigen, ambulant klinischen Einsatz. Ziel dieser Arbeit ist die methodische Bewertung eines elektronischen Zahlenverbindungstests (eNCT), welcher als ein schnell zu absolvierender Patientenselbsttest in der Diagnostik und Therapiekontrolle einer mHE entwickelt wurde und zukünftig regelmäßig im häuslichen Umfeld von chronisch Lebererkrankten eingesetzt werden könnte.

Methodik: Der eNCT wurde in Anlehnung an den Zahlenverbindungstest A (NCTA) konzipiert. 25 ungeordnete Zahlen werden auf einem Bildschirm dargestellt und sollen in aufsteigender Reihenfolge mit dem Finger berührt werden. Auf zwei unterschiedlichen Displaygrößen wurde die für die Testbearbeitung benötigte Zeit gemessen. Insgesamt 86 Patienten mit manifester Leberzirrhose und 61 gesunde Kontrollprobanden wurden in die Auswertung eingeschlossen. Eine methodische Bewertung des eNCT erfolgte, indem Testergebnisse (i) beider Gruppen verglichen, (ii) Parallel- und Retestreliaibilität mittels Cronbach alpha (rtt) und Intraklassenkorrelationen (IKK) bestimmt und (iii) mit Hilfe des als Referenzstandard geltenden portosystemischen Enzephalopathie (PSE)-Syndrom-Tests validiert wurden. Weiterhin identifizierten wir mit Hilfe einer multiplen Regressionsanalyse (iv) Einflussfaktoren auf die eNCT-Performanz.

Ergebnisse: Kernergebnisse der Untersuchung stellen sich wie folgt dar: (i) Patienten absolvieren den eNCT langsamer als Kontrollprobanden ($p < 0,001$). (ii) Der eNCT ist eine zuverlässige Messmethode mit hoher Retestreliaibilität (IKK=0,79-0,92) und sehr hoher Paralleltestreliaibilität (rtt=0,89-0,91). (iii) Hohe Korrelationen zwischen eNCT-Performanz und konstrukt-nahen Einzeltests des PSE-Syndrom-Tests ($r=0,43-0,80$) sowie die Korrelation zwischen eNCT-Performanz und dem Diagnosekriterium mHE ($r=0,66-0,69$) bescheinigen dem Testverfahren eine gute Validität. (iv) Die Merkmale Alter, Bildungsgrad, Gesamtpunktzahl in der Mini-Mental Status Examination (MMSE)

und eine bestehende medikamentöse HE wirksame Therapie konnten als unabhängige Prädiktoren der eNCT-Performanz identifiziert werden.

Schlussfolgerung: Der eNCT ist ein zuverlässiges Tablet-basiertes Testverfahren mit dem Potential eine mHE bei erwachsenen Patienten mit manifester Leberzirrhose zu quantifizieren. Der Gebrauch als telemedizinische Anwendung in Diagnostik und Therapiekontrolle einer mHE scheint in der bestehenden Form denkbar.

Abstract (English)

Introduction: Hepatic encephalopathy (HE) is known as a frequent complication of manifest liver cirrhosis with significant impact on quality of life and working ability as well as a high use of health system resources. In particular the minimal hepatic encephalopathy (mHE) underlies a difficult diagnosis due to missing clinical symptoms and episodic occurrence. The current reference standards are time-consuming, require trained staffing and seem not suitable for a regular outpatient use. Therefore, we aimed to investigate a first-time methodical review of the electronic number connection test (eNCT) which could be regularly used in the domestic environment as a quick self-test for patients with mHE in terms of diagnosis and therapy control.

Methods: The eNCT is designed based on the number connection test A (NCTA). 25 unordered numbers are represented on a screen and should be finger-touched in ascending order. The time required for processing was measured on two different sized screens. 86 patients with manifest hepatic cirrhosis and 61 control subjects were included in the analysis of this task. We aimed to investigate a first-time methodical review of the eNCT by (i) comparing test results of patients and healthy subjects, (ii) investigating data regarding retest reliability and parallel test reliability employing intraclasscorrelation (IKK) and Cronbach alpha (rtt) and (iii) by correlating findings to determine validity with results of the current reference standard PSE syndrome test. In addition, we examined (iv) influencing factors on eNCT performance with a multiple linear regression.

Results: Key findings of this study are (i) a slower test performance in patients compared to control subjects ($p < 0.001$). (ii) Within the methodical evaluation, a good to excellent retest reliability ($IKK = 0.79-0.92$) and parallel test reliability ($rtt = 0.89-0.91$) were shown, (iii) while high correlations between eNCT performance and construct close individual tests of the PSE syndrome tests ($r = 0.43-0.80$) as well as to diagnostic characteristics of mHE ($r = 0.66-0.69$) certify a good validity. (iv) We were able to identify age, education, scoring in mini-mental status examination (MMSE) and a present HE medication as independent predictors on eNCT performance.

Conclusion: The eNCT is a reliable tablet-based test procedure with the potential to quantify neuro psychological disorders in adult patients with liver cirrhosis. The use as a tele-medical application and screening process in diagnostics and therapy control for mHE seems conceivable in its existing form.

Abkürzungsverzeichnis

Abb.	Abbildung
ALAT	Alanin-Aminotransferase
AP	Alkalische Phosphatase
ASAT	Aspartat-Aminotransferase
ASH	alkoholische Steatohepatitis
AveNCT	eNCT im Average Konzept
bzw.	beziehungsweise
CFF	Kritische Flimmerfrequenz
CVK	Charité, Campus Virchow-Klinikum
EEG	Elektroenzephalographie
eNCT	elektronischer Zahlenverbindungstest
eNCT10Z	eNCT in der 10 Zoll Displaygröße
eNCT18Z	eNCT in der 18 Zoll Displaygröße
f ²	Effektstärke nach Cohen
G-GT	Gamma-Glutamyltransferase
ggf.	gegebenenfalls
Hb	Hämoglobin
HE	Hepatische Enzephalopathie
HESA	Hepatic Encephalopathy Scoring Algorithm
IKK	Intraklassenkorrelationen
INR	International Normalized Ratio
KON	Kontrollprobanden
KI	Konfidenzintervall
KP	Test Kreise-Punktieren
LNT	Linien-Nachfahr-Test
MELD-Score	Model of End Stage Liver Disease-Score
mHE	minimale Hepatische Enzephalopathie
MMSE	Mini Mental State Examination
n	Anzahl
NASH	nicht-alkoholische Steatohepatitis
NCTA	Zahlenverbindungstest A
NCTB	Zahlenverbindungstest B

oHE	offene Hepatische Enzephalopathie
p	p-Wert
PAT	Patientengruppe
Pearson- χ^2	Pearson-Chi-Quadrat Test
PHES	Portosystemic Hepatic Encephalopathy Score
PSE	portosystemische Enzephalopathie
r	Korrelationskoeffizient
r^2	Bestimmtheitsmaß des Regressionsmodells
rtt	Reliabilitätskoeffizienten mittels Methode nach Cronbach alpha
SBP	Spontan bakterielle Peritonitis
SD	Standardabweichung
SDT	Serial Dotting Test
Sig.	Signifikanz
SKK	Spearman Korrelationskoeffizient
Stroop	EncephalApp - Stroop Test
t	student's t-Test
Tab.	Tabelle
TIPPS	transjugulärer intrahepatischer portosystemischer Shunt
U	Wilcoxon-Vorzeichenrangtest
WHC	West Haven Kriterien
WP	Wertepunkte
Z.n.	Zustand nach
ZST	Zahlensymboltest

1 Einleitung und Einführung in die Fragestellung

1.1 Leberzirrhose

Die Zirrhose der Leber ist definiert als eine histologische Entwicklung von Regeneratknoten, die von fibrinösen Bindegewebssepten umgeben und durch eine chronische Lebererkrankung bedingt ist (1). Sie ist keine eigenständige Erkrankung sondern vielmehr die gemeinsame pathologische Endstrecke einer Vielzahl chronischer Lebererkrankungen, die sich durch einen langsam progredienten Verlauf über Jahre bis Jahrzehnte auszeichnet (2). Ferner stellt die Erkrankung eine präkanzeröse Kondition für die Entwicklung des hepatozellulären Karzinoms dar (3).

In Deutschland ist mit etwa 30 - 40% die alkoholische Steatohepatitis (ASH) führend. Virushepatitiden nehmen mit ungefähr 30% den zweiten Platz ein, während sich die nicht-alkoholische Steatohepatitis (NASH) und die kryptogene Leberzirrhose mit jeweils 10% den dritten Platz teilen (4). Die zahlreichen weiteren Ursachen machen mit etwa 10% nur einen kleinen Teil aus (siehe Abbildung (Abb.) 1).

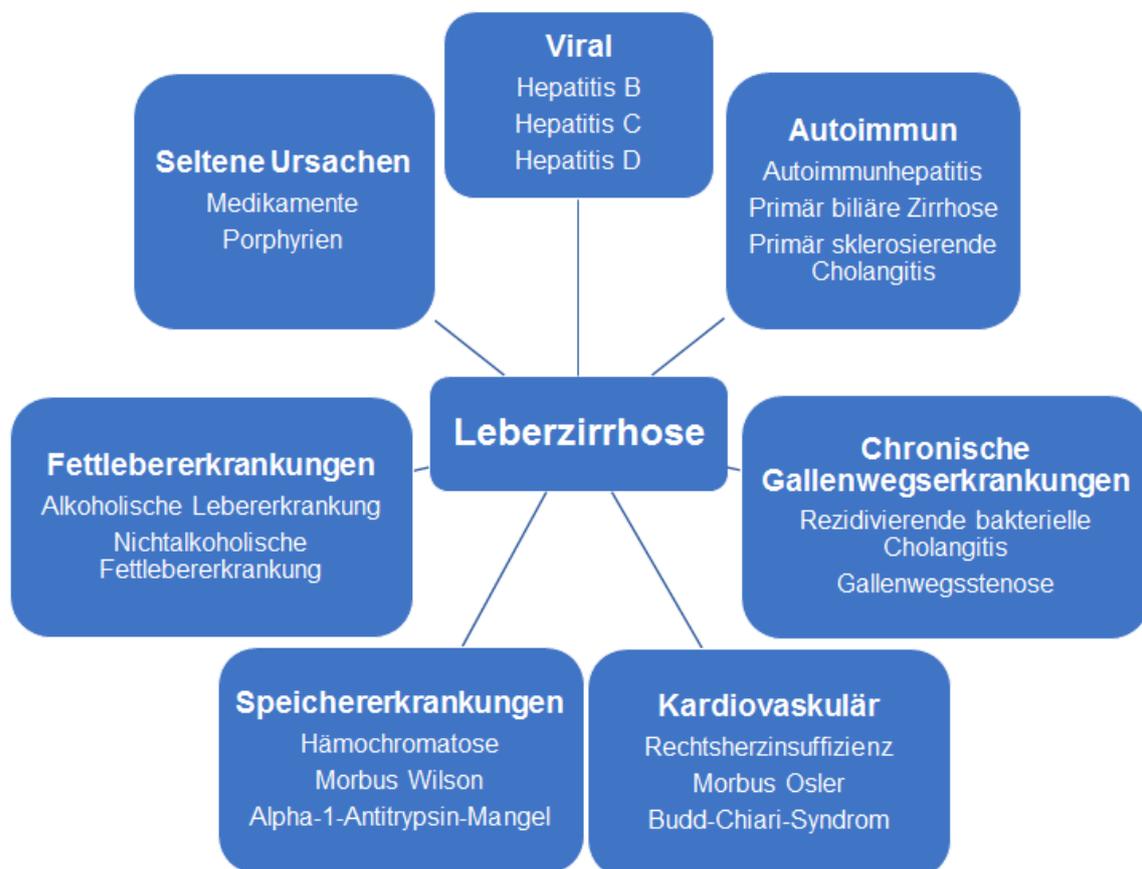


Abb. 1 Mögliche Ursachen einer Leberzirrhose. Modifizierte Grafik nach Wiegand et al.(2).

Die Studienergebnisse zur Epidemiologie dieser Erkrankung sind nicht einheitlich und stark abhängig vom untersuchten Kollektiv (4). Weltweit wird die Prävalenz auf 1% (1) und in Deutschland von einigen Autoren sogar auf bis zu 1,25% geschätzt (5). Im Jahr 2010 besaß Deutschland neben Finnland und Österreich die höchste zirrrosebedingte Mortalität in Westeuropa (6) mit einer Verdopplung der Todesfälle pro 100.000 Einwohner von 5 auf 9,9 zwischen 1980 und 2005 (7). Die Erkrankung und ihre Folgen stellen damit ein zunehmendes Problem für das Gesundheitssystem dar (2, 8). Infolge des bindegewebigen Umbaus und der Zerstörung der Läppchen –und Gefäßstruktur der Leber kommt es im Verlauf einer Leberzirrhose zu ausgeprägten funktionellen Folgen. Leberinsuffizienz, portale Hypertension auf Grundlage eines erhöhten intrahepatischen Gefäßwiderstandes und die Bildung intrahepatischer porto-systemischer Shunts zwischen Portalgefäßen und Lebervenen mit konsekutiver Minderperfusion der Leber können zu weitreichenden Komplikationen im Bereich der Syntheseleistung, Entgiftung und Zirkulation führen (5). Eine weit verbreitete Einteilung des Schweregrades einer Leberzirrhose ist die Stadieneinteilung anhand des CHILD-PUGH Scores (9, 10). Anhand von fünf Parametern wird eine Gesamtpunktzahl errechnet, wobei für jeden Parameter zwischen einem und drei Punkte vergeben werden. Der Score reicht von fünf bis 15 Punkten und ermöglicht eine Differenzierung in die Stadien CHILD A, B und C (9, 11) (siehe Abb. 2).

CHILD-PUGH Score und Interpretation			
Klassifikation	1	2	3
Serum Bilirubin (µmol/l)	<34	34-51	>51
Serum Albumin (g/L)	>35	28-35	<28
Aszites	nicht/gering	mäßig/therapierbar	massiv/therapierefraktär
Enzephalopathie	keine	minimal	schwer
Interpretation			
CHILD A = 5-6 Punkte (gute Leberfunktion)			
CHILD B = 7-9 Punkte (mäßige Leberfunktion)			
CHILD C = 10-15 Punkte (schlechte Leberfunktion)			

Abb. 2 CHILD-PUGH-Klassifikation. Grafik in Anlehnung an Child et al. und Pugh et al. (9, 11).

Mit Hilfe der Stadieneinteilung können Aussagen zum Langzeitüberleben und zur perioperativen Mortalität einer Leberzirrhose getroffen werden. Ein weiteres Instrument für die prognostische Einstufung der Schwere einer Lebererkrankung im Sinne eines disease severity scores ist der Model of End Stage Liver Disease-Score (MELD-Score) (12). Er dient zur Quantifizierung der Mortalität einer Leberzirrhose und wird vor allem in der Transplantationsmedizin bei der Priorisierung der Organzuteilung von Lebertransplantationen verwendet. Im Vergleich zu den CHILD-PUGH-Kriterien ist es das objektivere Verfahren (13). Die Berechnung des Punktwertes erfolgt mit Hilfe der folgenden Formel und der Laborparameter Serumkreatinin (Wert = 4 mg/dl bei Dialyse \leq 1 Woche vor Erhebung des MELD-Scores), Serumbilirubin und International Normalized Ratio (INR).

$$\text{MELD-Score} = 10 * (0,957 * \ln(\text{Kreatinin}) + 0,378 * \ln(\text{Bilirubin}) + 1,12 * \ln(\text{INR}) + 0,643)$$

Der errechnete Wert reicht von sechs bis 40 Punkten. Ein Punkte-Score unter neun geht mit einer niedrigen (1,9%) drei-Monatsmortalität, während ein Punkte-Score von 40 mit einer hohen (71,3%) drei-Monatsmortalität einhergeht (12).

In Abhängigkeit der funktionellen Störungen können sich eine Vielzahl an Komplikationen entwickeln. Neben Ikterus, Aszites, erhöhter Blutungsneigung, hepatorenalem Syndrom, Hyponatriämie, spontan bakterieller Peritonitis (SBP) oder Varizenblutung stellt die hepatische Enzephalopathie (HE) ein häufiges und schwerwiegendes Dekompensationszeichen einer Leberzirrhose dar (14, 15).

Der in der Regel asymptomatisch beginnende Verlauf chronischer Lebererkrankungen und der fließende Übergang in eine manifeste Zirrhose erschweren oftmals eine frühzeitige Diagnose und die sekundäre Prävention. Trotz verbesserter Screeningmaßnahmen und der Entwicklung nicht-invasiver diagnostischer Methoden zur Detektion einer Leberzirrhose bestehen sowohl in der Prävention der Erkrankung als auch in der Diagnostik und Behandlung der Komplikationen zahlreiche Verbesserungsmöglichkeiten (2, 5, 8, 15-17).

1.2 Hepatische Enzephalopathie (HE)

1.2.1 Bedeutung der HE

Die HE ist eine potentiell reversible, metabolisch bedingte Funktionsstörung des zentralen Nervensystems, die im Rahmen von akuten oder chronischen Lebererkrankungen auftritt (18-20). Das Spektrum dieser neuropsychiatrischen Störungen reicht von der Beeinträchtigung des Bewusstseins über Störungen der Persönlichkeit bis hin zur Minderung kognitiver Fähigkeiten unterschiedlicher Schweregrade (18, 19).

In der Literatur variieren die Angaben zur Prävalenz einer Leberzirrhose bedingten HE zwischen 20% und 80% (21-25). Gründe sind die oftmals schleichende Entwicklung HE bedingter Symptome, die Variabilität in den diagnostischen Kriterien zur Detektion sowie Unterschiede im Schweregrad der Leberzirrhose in den untersuchten Studienpopulationen der klinisch unauffälligen Form, der mHE (26, 27). Demgegenüber wird die Prävalenz einer klinisch manifesten HE zum Zeitpunkt der Diagnose Leberzirrhose ohne Komplikationen mit 10% - 14% und zum Zeitpunkt der Diagnose einer dekompensierten Leberzirrhose mit 16% - 21% angegeben (28). Es konnte gezeigt werden, dass das Auftreten einer HE nachteilige Effekte auf die Lebensqualität und die Ausführung komplexer Aufgaben hat. Insbesondere Alltagstätigkeiten wie die Teilnahme am Straßenverkehr sind deutlich beeinträchtigt (30, 31). Weiterhin konnten zahlreiche Studien einen negativen Effekt auf das Überleben darlegen (20, 29).

Ökonomisch stellt die HE ein zunehmendes Problem dar. Bei keiner anderen Komplikation der Leberzirrhose kommt es zu einer so hohen Inanspruchnahme der Ressourcen des Gesundheitssystems (8). Im Jahr 2003 betrug in den USA allein die Summe der HE bedingten Krankenhausaufenthalte annähernd 1 Milliarde US\$, indirekte Kosten für ambulante Behandlungen und ökonomischen Produktivitätsverlust nicht mit einberechnet (15).

1.2.2 Pathogenese der HE

"Er kann nicht gehalten werden, sagt unverständliche Dinge, ist boshaft und lässt sich nicht beruhigen"

Obwohl schon Hippokrates von Kos die HE vor mehr als 2000 Jahren an einem "gelblichen" Patienten beschrieb, sind die exakte Pathogenese und die molekularen Mechanismen nicht vollkommen verstanden (26, 32). Kircheis & Häussinger (33) beschreiben die HE als eine metabolisch induzierte primäre Störung der Gliazellfunktion, welche sekundär die neuronale Funktion beeinträchtigt. Von zentraler Bedeutung gilt die Akkumulation des Stoffwechselprodukts Ammoniak im Gehirn (34, 35). Den direkten Nachweis bei Patienten mit HE konnten Lockwood et al. 1991 mittels Positronen-Emissions-Tomographie erbringen (36). Da Astrozyten die einzigen Zellen im Zentralnervensystem sind, die überschüssiges Ammoniak durch Glutaminbildung fixieren können (37), bewirkt eine erhöhte Ammoniakkonzentration im Blut einen erhöhten intrazellulären Glutaminspiegel der Astroglia. Die osmotische Aktivität des Glutamins resultiert in einem Anschwellen der Zellen bis hin zur Entwicklung eines zerebralen Ödems. Auch präzipitierende Faktoren wie Medikamente, Elektrolytstörungen oder Infektionen können unabhängig vom Ammoniakspiegel des Blutes zu einer Störung der Astrozyten-Hydratation führen (33, 38). Durch Gliazellschwellung kommt es zur Aktivierung von N-Methyl-D-Aspartat-Rezeptoren und zur Bildung von reaktiven Sauerstoff- bzw. Stickstoffintermediaten, die ihrerseits die Schwellung der Astroglia begünstigen. Die konsekutive astrozytäre Dysfunktion mündet in einer Störung der glianeuronalen Kommunikation und der synaptischen Plastizität. Die vollständige Reversibilität einer HE bedingten Hirnfunktionsstörung wird infrage gestellt (33, 39). In einer 2010 veröffentlichten Studie von Bajaj und Kollegen (29) konnte erstmals gezeigt werden, dass neurologische Symptome bei Patienten mit manifesten HE-Episoden mit persistierenden und kumulativen Defiziten in den Bereichen Arbeitsgedächtnis, Antwortinhibition und Lernen assoziiert sind.

1.2.3 Ätiologie, Diagnostik und Graduierung

In Anlehnung an die Empfehlungen der von der American Association for the Study of Liver Disease und der European Association for the Study of the Liver aus dem Jahr 2014 beauftragten Arbeitsgruppe sollte die HE nomenklatorisch in vier Kriterien eingeordnet werden (26). Die Einteilung der Ätiologie erfolgt hierbei in drei Typen: (a) Enzephalopathie bei akutem Leberversagen, (b) Enzephalopathie bei portosystemischem Bypass ohne Lebererkrankung und (c) Enzephalopathie bei Leberzirrhose und portaler Hypertension. Die Einteilung nach klinischen Manifestationen basiert auf den West Haven Kriterien (WHC) (siehe Abb. 3) (26, 40).

Schweregrad	Bewusstseinslage	Neuropsychiatrische Symptome	Neurologische Symptome
Stadium 0 = mHE	Normal	Störungen nur durch psychometrische Tests zu erfassen	keine
Stadium I	Leichtgradige mentale Verlangsamung	Eu-/Dysphorie, Reizbarkeit und Angst, reduzierte Aufmerksamkeit	Gestörte Feinmotorik (Fingertremor, beeinträchtigt Schreibvermögen)
Stadium II	Verstärkte Müdigkeit, Apathie oder Lethargie	Leichte Persönlichkeitsstörung, minimale Desorientiertheit zu Zeit und Ort	Flapping-Tremor, Ataxie, verwaschene Sprache
Stadium III	Somnolenz	Aggressivität, ausgeprägte Desorientiertheit zu Zeit und Ort	Rigor, Krämpfe, Asterixis
Stadium IV	Koma	-	Hirndruckzeichen

Abb. 3 Semiquantitative Stadieneinteilung des mentalen Status bei Hepatischer Enzephalopathie (HE) nach den West Haven Kriterien (WHC). Modifizierte Grafik nach Conn & Bircher (45). mHE minimale hepatische Enzephalopathie.

Die Kriterien ermöglichen eine semiquantitative Einteilung der HE in Abhängigkeit klinischer Symptome wie Aufmerksamkeitsstörungen und Auffälligkeiten in Emotion, Affektion, Kognition, Motorik oder Psychomotorik. Die Stadien I bis IV werden von einigen Autoren auch als offene HE (oHE) zusammengefasst, während als allgemeiner Konsens gilt, das Stadium 0 als mHE zu bezeichnen (26, 41). Im Besonderen der Grad I ist vom Untersucher und dem Verhalten des Patienten abhängig und erschwert eine eindeutige Klassifizierung in Grad I und II, während Grad III und IV über Somnolenz und Koma eindeutiger definiert sind.

Ein neueres Graduierungssystem zur Charakterisierung einer HE ist der Hepatic Encephalopathy Scoring Algorithm (HESA), welcher den klinischen Eindruck mit einfachen Tests zur Überprüfung der neuropsychologischen Leistungsfähigkeit kombiniert und die Reliabilität der WHC verbessern kann (27, 42).

Von einer mHE, früher auch als subklinisch oder latent bezeichneten HE spricht man bei Fehlen von Symptomen in der konventionellen klinisch-neurologischen Untersuchung. In diesem Stadium können Defizite in kognitiven Teilbereichen von Aufmerksamkeit, visuell-räumlicher Wahrnehmung, Geschwindigkeit der Informationsbearbeitung, Feinmotorik und Kurzzeitgedächtnis auftreten (43). Im Unterschied zu den anderen Stadien kann die mHE nur präzise auf Grundlage psychometrischer Testungen oder neurophysiologischer Testverfahren erfasst werden (26, 40, 44). Hinsichtlich des zeitlichen Verlaufs wird die HE unterteilt in (1) episodische HE mit phasenhaftem Auftreten, (2) rezidivierende HE mit periodischem Auftreten in einem Zeitintervall von ≤ 6 Monaten und (3) persistierende HE. Als letztes Kriterium sollte das Vorhandensein von präzipitierenden Faktoren überprüft und differenziert werden in: (a) Fehlen von HE auslösenden Faktoren, (b) Spezifizierung von HE auslösenden Faktoren (u.a. Infektion, GI-Blutung, Elektrolytstörungen, Obstipation oder bestimmte Medikamente).

1.3 Klinische Relevanz und Diagnostik der minimalen HE (mHE)

Trotz fehlender klinischer Symptome sind die von einer mHE betroffenen Patienten in Bezug auf Arbeitsfähigkeit, Fahrtüchtigkeit und Lebensqualität merklich beeinträchtigt (27, 46, 47) und können, wenn korrekt diagnostiziert, von einer angepassten Therapie profitieren (26, 27). Zudem gilt die mHE als unabhängiger Prädiktor für die Entwicklung einer manifesten HE und für das Überleben von Patienten mit einer Leberzirrhose (26, 27, 48). Vor der eigentlichen Testung auf eine mHE sollten zunächst sichtbare neurologische Symptome und kognitive Defizite ausgeschlossen werden. Empfohlen wird die Verwendung der Mini-Mental State Examination (MMSE), um das kognitive Leistungsniveau zu erfassen (27, 40, 49). Ein Score von ≤ 25 impliziert eine kognitive Beeinträchtigung beziehungsweise das Vorhandensein einer oHE (49, 50). Der in der Differentialdiagnostik einer oHE oft verwendete venöse oder arterielle Ammoniakspiegel hat für die Diagnostik einer mHE keine Bedeutung. Obwohl Ammoniak in der Pathogenese der HE eine zentrale Rolle einnimmt, geben erhöhte Ammoniakspiegel im Blut keine ausreichenden Informationen zu Diagnose, Stadium oder Prognose einer HE in Patienten mit einer chronischen Lebererkrankung (26). Ferner sind wiederholte Bestimmungen zur Verlaufskontrolle meist entbehrlich und können die klinische Beurteilung keinesfalls ersetzen (22). Grund dafür ist, dass Ammoniak erst Einflüsse auf zerebrale Funktionen ausübt, sobald die Blut-Hirnschranke überwunden wird. In der

klinischen Praxis steht bisher kein laborchemischer Parameter zur Messung des Hirn-Ammoniak-Metabolismus zur Verfügung (51). Bildgebende Verfahren haben ihren Stellenwert in erster Linie in der Differentialdiagnostik bei klinischem Verdacht auf eine Pathologie anderer Genese, bei wissenschaftlichen Fragestellungen oder in Einzelfällen zur Verlaufskontrolle (26, 52).

Zur Diagnostik einer mHE wurden in den vergangenen Jahrzehnten eine Vielzahl von Testverfahren evaluiert (53-55). Heutzutage spielen vor allem psychometrische und neurophysiologische Testverfahren eine entscheidende Rolle. Zu den am häufigsten eingesetzten und am besten evaluierten Tests gehören die folgenden psychometrischen Testverfahren: der Zahlenverbindungstest A (number connection test A/NCTA), der Zahlenverbindungstest B (number connection test B/NCTB), der Block Design Test und der Zahlensymboltest (ZST). Für die Quantifizierung einer mHE wird empfohlen, dass mindestens zwei dieser Testverfahren miteinander kombiniert und alters –und bildungsadjustierte Normwerte in der zu untersuchenden Bevölkerungsgruppe vorliegen. Die etablierten psychometrischen- und neurophysiologischen Testverfahren werden im Folgenden erläutert:

Der **PSE-Syndrom-Test** ist eine Testbatterie bestehend aus fünf papierbasierten Einzeltests. Er ist relativ leicht anzuwenden, besitzt eine zuverlässige Reproduzierbarkeit sowie eine hohe externe Validität. In der englischsprachigen Literatur wird er häufig als Portosystemic Hepatic Encephalopathy Score (PHES) bezeichnet. Aufgrund der Überprüfung eines breiten Spektrums kognitiver Aspekte, normativen alterskorrigierten Referenzdaten und der geringen Kosten wird diese Testmethode in klinischen Studien häufig als Referenzstandard in der Diagnostik einer mHE eingesetzt (40, 56-58). Nachteile sind die verhältnismäßig lange Testdauer und Auswertung sowie die Notwendigkeit eines trainierten Untersuchers für Instruktion, Supervision und Analyse der Testergebnisse.

Die Bestimmung der „**Kritischen Flimmerfrequenz**“ (**CFF**) ist ein psychophysiologisches Testverfahren zur Erfassung zentralnervöser Aktiviertheitszustände. Basierend auf der Annahme, dass die Gliazellen der Retina den gleichen Funktionsstörungen unterliegen wie die Astrozyten im Gehirn wird hierbei mittels eines Flimmerfrequenzanalysators eine intrafoveale Lichtreizung erzeugt (18, 54). Der Lichtimpuls einer definierten Frequenz (25-60Hz) wird dem Patienten dabei in Schritten von 0,1 HZ in absteigender Richtung vorgegeben. Als Flimmerfrequenz wird die kritische Frequenz bezeichnet bei der der Eindruck des Gleichlichts in den des

Flimmerns übergeht (52). Die CFF stellt einen reproduzierbaren Parameter mit nur geringem Bias durch Trainingseffekte, Bildungsgrad oder untersucherbedingter Variabilität dar und wird als einfache, reproduzierbare und präzise Methode zur Objektivierung und Verlaufskontrolle einer mHE gesehen. Allerdings erfordert das CFF Binokularsehen und kann bei hohem Anforderungsanspruch an die Probanden mit einer Abbruchrate von bis zu 8% einhergehen (56). Die Werte zur Sensitivität schwanken in der Literatur stark. In einer im Jahr 2013 publizierten Metanalyse zur diagnostischen Genauigkeit dieser Methode in der Detektion einer mHE kommen die Autoren zu dem Ergebnis, dass die CFF eine hohe Spezifität (79%) bei moderater Sensitivität (61%) besitzt und empfehlen den Gebrauch des Testverfahrens als Ergänzung nicht aber als Ersatz zu psychometrischen Testungen (59).

Der **EncephalApp-Stroop Test (Stroop)** ist ein vielversprechendes Smartphone basiertes Testverfahren, welches mit Hilfe des Stroop-Effekts die psychomotorische Geschwindigkeit und kognitiv Flexibilität untersucht (53). Der Stroop-Effekt zeigt, dass trainierte Handlungen nahezu automatisch ablaufen, während ungewohnte Handlungen wie die Benennung der Farbe eines visuell dargebotenen Wortes, dessen Inhalt der dargebotenen Farbe widerspricht, eine höhere Aufmerksamkeit benötigen (57, 60) (siehe Abb. 4). Aktuelle Studienergebnisse bescheinigen dieser Testmethode eine gute Sensitivität für die Diagnose einer mHE sowie die Fähigkeit, die Entwicklung einer oHE vorherzusagen (58).



Abb. 4 Bildschirmfoto der EncephalApp-Stroop Test. Modifizierte Grafik nach Bajaj (61).

Der **Inhibitory Control Test** ist ein computerbasierter Daueraufmerksamkeitstest bei dem die Patienten instruiert werden, alternierend auf verschiedene Stimuli zu reagieren oder einen Zielreiz zu unterdrücken. Evaluiert werden Vigilanz, Aufmerksamkeit, Antwortunterdrückung und Arbeitsgedächtnis. Die Validität wird mit gut beurteilt, erfordert aber eine nicht eingeschränkte Kognition (26). Aufgrund schwacher Korrelationen mit anderen Tests in der mHE Diagnostik und fehlenden Normwerten bleibt der Test aktuell wissenschaftlichen Fragestellungen vorbehalten (57, 60, 62, 63). Eine **Elektroenzephalographie (EEG)** kann Veränderungen in der kortikalen Aktivität als Spektrum einer HE auch ohne Patientenkooperation und dem Risiko eines Lerneffekts detektieren (64). Allerdings sind die charakteristischen Veränderungen unspezifisch, da sie auch bei anderen metabolischen Enzephalopathien auftreten können und technische Expertise in der Interpretation erfordern. Innerhalb klinischer Fragestellungen und bei der Differenzialdiagnostik gestörter Hirnfunktionen besitzt das EEG allerdings nach wie vor seinen Stellenwert (60).

Akustisch evozierte Potentiale geben die Möglichkeit neurophysiologische Funktionen in Patienten mit Leberzirrhose zu evaluieren und neurologische oder neuropsychiatrische Untersuchungen sinnvoll zu ergänzen. Insbesondere Spätpotentiale (P300-Welle) lassen bereits eine geringfügige Hirnstörung bei Patienten mit einer mHE erkennen (62). Für die Routinediagnostik sind sie aufgrund ihrer inkonsistenten Ergebnisse jedoch entbehrlich (44, 52). Weiterhin werden für die Durchführung erfahrene Untersucher und spezielles Equipment benötigt.

Es bleibt zu betonen, dass bis heute kein Goldstandard in der Diagnostik der mHE existiert und von einem internationalen Expertengremium die Kombination von sowohl psychometrischen als auch neurophysiologischen Testverfahren empfohlen wird (26). Für die klinische Routine besteht die Möglichkeit, auf vertraute Einzeltestverfahren zurückzugreifen, wenn normative Referenzdaten zur Verfügung stehen und die jeweiligen Testverfahren für den Gebrauch in der Bevölkerungsgruppe validiert wurden (26, 27).

1.3.1 Wertigkeit des Zahlenverbindungstests (NCT) in der Diagnose der mHE

Erstmals wurde dieses Testverfahren als sogenannter „Trail Making Test“, bestehend aus dem Zahlenverbindungstest Version A (NCTA) und Version B (NCTB) zur Beurteilung von Hirnschäden untersucht und als ein einfaches, günstiges und schnelles neuropsychologisches Testverfahren in die klinische Diagnostik eingeführt:

"this short, inexpensive, and easily administered test may be a fairly valid indicator of certain effects of brain damage (65)."

Conn und Kollegen suchten nach einem geeigneten Diagnostikum zur quantitativen Erfassung des Schweregrades einer vorliegenden HE und führten 1977 den sogenannten PSE-Index ein (40). Dieser kombinierte die kognitiven Fähigkeiten mit dem arteriellen Ammoniakspiegel, dem Befund im EEG, dem Asterixis-Schweregrad und dem Testresultat im NCTA (66). In Abhängigkeit der Testzeit im NCTA wurden fünf Schweregrade der HE eingeteilt. Eine Korrektur der Normwerte hinsichtlich Alter, Geschlecht oder Bildungsstand wurde nicht vorgenommen.

In Deutschland verglichen Schäfer und Mitarbeiter 1981 zur Verlaufsbeurteilung der chronischen HE im Verlauf von acht Monaten verschiedene Messmethoden miteinander und prüften diese an Patienten mit Leberzirrhose. Zu den Messgrößen gehörten der NCTA, die zeitliche Orientierung, der Asterixisgrad, ein Rechentest und ein Schreibtest, die EEG-Untersuchung und der venöse Blutammoniakspiegel. Sie kamen zu der Schlussfolgerung, dass der NCTA leicht durchzuführen und gut reproduzierbar ist sowie eine vom Untersucher weitgehend unabhängige Aussage über den Schweregrad einer beginnenden chronischen HE geben kann (67). Weissenborn und Kollegen charakterisierten 1998 die Testverfahren NCT A & B als sensitive Testmethoden für die Diagnose einer frühen HE, insofern standardisierte Testversionen und alterskorrigierte Normwerte vorliegen. Für die zuverlässige Diagnostik einer mHE empfehlen sie die Benutzung des NCTA als Teil einer standardisierten Testbatterie, um möglichst verschiedene Teilbereiche der mentalen Leistungsfähigkeit zu erfassen (68). Als PSE-Syndrom-Test wurde eine papierbasierte, psychometrische und standardisierte Testbatterie von Schomerus und Mitarbeitern entwickelt und 1999 in Deutschland eingeführt (55). Dieses Testverfahren beinhaltet neben dem NCTA und B noch drei weitere psychometrische Einzeltests und erlaubt eine gute Quantifizierung der mHE mit hoher Sensitivität (43).

Zusammenfassend sollte erwähnt werden, dass der NCTA bis heute in vielen Studien zur Diagnostik einer mHE entweder als Einzeltest oder als Bestandteil eines diagnostischen Testverfahrens verwendet wurde, insbesondere da keine spezielle Expertise in der Durchführung des Testverfahrens benötigt wird und die Zeit für

Administration und Interpretation mit 30-120 Sekunden sehr kurz ist. (21, 23, 24, 28, 47, 66, 69, 70). Als allgemeiner Konsens für die Anwendung des NCTA in der Diagnostik einer mHE gilt, dass alters- und bildungskorrigierte Testwerte innerhalb der zu untersuchenden Bevölkerungsgruppe vorliegen und dass die Kombination des NCTA mit mindestens einem weiteren psychometrischen Einzeltest erfolgen sollte (27, 28, 40).

1.4 Grundlagen methodischer Bewertung

1.4.1 Gütekriterien

Gütekriterien, psychometrische Eigenschaften und Messeigenschaften sind synonyme Bezeichnungen für jene Kriterien, die für die Beurteilung eines Testverfahrens herangezogen werden. Sie sind Maß für Qualität und Vorhersagekraft eines spezifisch standardisierten Tests (71). Ferner ermöglichen sie einen Vergleich zwischen verschiedenen diagnostischen Methoden (72). Vor einem routinemäßigen Einsatz sollte jedes Testverfahren auf die Gütekriterien geprüft werden. Im deutschen Sprachraum werden in der klassischen Testtheorie drei Hauptgütekriterien beschrieben: Objektivität, Reliabilität und Validität (73). Es gilt zu beachten, dass diese Gütekriterien miteinander zusammenhängen. Reliabilität ist erst nach gegebener Objektivität zu erreichen, Validität erst nach gegebener Objektivität und Reliabilität.

Objektivität eines Tests ist ein wesentliches Gütekriterium, das die Vergleichbarkeit von Testleistungen verschiedener Testpersonen sicherstellt. Außerdem müssen klar und anwenderunabhängige Regeln für die Ergebnisinterpretation vorliegen (73). Sie lässt sich in drei Aspekte differenzieren: (a) Durchführungsobjektivität, (b) Auswertungsobjektivität und (c) Interpretationsobjektivität.

Reliabilität ist „*der Grad der Genauigkeit, mit dem das geprüfte Merkmal gemessen wird*“ (74). Zur Reliabilitätseinschätzung werden im Rahmen der klassischen Testtheorie mehrere Verfahren entwickelt (73). Zu ihnen zählen die (1) Retest-Reliabilität, die (2) Paralleltest-Reliabilität, die (3) Test-Halbierungs-Reliabilität und die Einschätzung auf Basis der (4) Inneren Konsistenz.

- (1) Bei der Retestreliabilität werden zwei Ergebnisse des gleichen Tests, die zu zwei unterschiedlichen Zeitpunkten erhoben wurden, miteinander korreliert (73). Je

stärker die Korrelationen der beiden Testergebnisse, desto besser die Retestrelia­bilität.

- (2) Die Paralleltestreliabilität beschreibt die Bestimmung der Korrelationen zwischen den beobachteten Testwerten in zwei parallelen Testformen, die inhaltlich möglichst aus ähnlichen Elementen bestehen und zeitnah an derselben Stichprobe erhoben wurden (73). Je stärker die Testergebnisse dabei korrelieren, desto größer ist die Paralleltestreliabilität.
- (3) Bei der Test-Halbierungs-Reliabilität wird ein Test in zwei Hälften geteilt, welche anschließend miteinander korreliert werden (72).
- (4) Die interne Konsistenz ist ein Maß für die Homogenität eines eindimensionalen Testverfahrens, bei dem mehrere Elemente dasselbe Konstrukt erheben (74).

Validität bezieht sich auf die Gültigkeit der Messung. Nur bei einem validen Test sind Messergebnisse interpretierbar. Die Validität eines Verfahrens wird bei Testentwicklung anhand mindestens eines der drei Validierungsmethoden geprüft (73, 74).

- (1) Unter inhaltlicher Validität versteht man, inwieweit ein Test das zu messende Merkmal repräsentativ erfasst (73). Sie wird in der Regel nicht numerisch anhand eines Kennwertes bestimmt, sondern argumentativ aufgrund logischer und fachlicher Überlegungen durch die Autorität von Experten beurteilt (72).
- (2) In der Konstruktvalidität wird überprüft, ob Hypothesen auf denen das Testverfahren beruht, bestätigt werden (71). Konvergente Validität hat ein Testverfahren, wenn die Testwerte zwischen zu überprüfenden und anderen Testverfahren, die das gleiche Konstrukt erheben, stark korrelieren (71). Eine geringe Korrelation des zu überprüfenden Testverfahrens zu Werten von Testverfahren, die ein anderes Konstrukt erheben, wird als diskriminative Validität bezeichnet (74). Entsprechende Zusammen­hangsstrukturen können durch bivariate und multivariate Verfahren überprüft werden (74).
- (3) Kriteriumsvalidität beschreibt die Korrelation der Werte eines neuen Testverfahrens mit dem Goldstandard (71). Eine kriterienbezogene Validierung ist daher nur möglich, wenn mindestens ein reliables und valides Außenkriterium, das für diagnostische Entscheidungen bedeutsam ist, vorliegt.

1.4.2 Nebengütekriterien

Neben den bereits beschriebenen Hauptgütekriterien existieren eine Reihe von sogenannten Nebengütekriterien von denen einige, für die vorliegende Arbeit wichtige, vorgestellt werden (72, 74):

Normierung bezieht sich auf die Referenzstichprobe (Eichstichprobe) des Tests. Diese stellt ein Bezugssystem dar, mit dessen Hilfe die Ergebnisse einer Testperson im Vergleich zu Merkmalsausprägungen anderer Personen eingeordnet und interpretiert werden können (73).

Testfairness ist gegeben wenn ein Test allen Personengruppen, für die er anwendbar sein soll, gleiche Chancen bietet bzw. wenn er keine Testpersonen systematisch aufgrund ihrer ethnischen, soziokulturellen oder geschlechtsspezifischen Gruppenzugehörigkeit benachteiligt (74).

Testökonomie ist erfüllt wenn ein Test relativ wenig Ressourcen wie Zeit, Kosten oder andere Formen (zum Beispiel Belastung des Testleiters) beansprucht, gemessen am diagnostischen Erkenntnisgewinn (73). Als besonders ökonomisch gelten computergestützte Testverfahren (74).

Zumutbarkeit umfasst die psychische, körperliche und zeitliche Belastung des Probanden und sollte die resultierende Nützlichkeit des Tests nicht übertreffen (72).

1.5 Einführung in die Fragestellung

Über den Verlauf einer manifesten Zirrhose hinweg kann die Leber den Funktionsverlust des Parenchyms nicht mehr ausreichend kompensieren. Neben Komplikationen wie Aszites, Magenfundus- und Ösophagusvarizen oder der Entwicklung einer SBP stellt die HE durch den Anstieg toxischer Stoffwechselprodukte ein häufiges und schwerwiegendes Problem für den Patienten dar. Das Spektrum klinischer Symptome ist weitreichend. Während sich die oHE durch quantitative Bewusstseinsstörungen bis hin zum Koma äußern kann, so entziehen sich die im Rahmen einer mHE auftretenden Symptome häufig der klinisch-neurologischen Untersuchung. Dennoch stellt Letztere eine hohe Bedeutung für die Arbeitsfähigkeit und Lebensqualität der Patienten dar. Defizite in kognitiven Teilbereichen wie Aufmerksamkeit, Konzentrationsvermögen und Wachheit (75) sind infolge mangelnder Fahrtüchtigkeit und Einschränkungen in der Navigation und Maschinenführung mit schweren Verletzungsfolgen und einem erhöhten Unfallrisiko assoziiert (30).

Die beschriebenen mit der mHE assoziierten Defizite stellen damit nicht zuletzt auch aus sozioökonomischer Sicht ein zunehmendes Problem unserer Gesellschaft dar. Eine frühzeitige Diagnose sowie eine an den Krankheitsverlauf angepasste Therapie würden die Lebensqualität und das Überleben der Patienten verbessern können.

Als Referenzstandards in der Diagnostik der mHE gelten psychometrische, papierbasierte Testverfahren, welche aufgrund ihres zeitlichen Aufwandes und dem erforderlichen Fachpersonal vor allem im Setting wissenschaftlicher Studien Verwendung finden. Darüber hinaus ist die Ausprägung der mHE einerseits episodenhaft, andererseits ist sie abhängig vom Einfluss verschiedener Faktoren wie Ernährung, medikamentöser Therapie oder gastrointestinalen Blutungen und erschwert damit eine zuverlässige Diagnosestellung. Ein schnell durchführbares und regelmäßig einsetzbares Testverfahren für die Detektion einer mHE in der ambulanten und stationären Versorgung würde eine Verbesserung in der Behandlung von chronischen Lebererkrankten ermöglichen. Zudem könnten die Anzahl der Krankenhausaufenthalte und damit die Kosten im stationären Bereich reduziert werden.

Der eNCT wurde in Anlehnung an den NCTA konzipiert und ist ein Tablet-basiertes, auf Selbsttestung ausgelegtes Testverfahren mit dem Ziel der quantitativen Erfassung und Objektivierung neuropsychischer Störungen. Als Teilbereich einer häuslich verwendeten Monitoring-Plattform für Patienten mit chronischer Lebererkrankung (d-LIVER Projekt) basiert das Testverfahren auf einem telemedizinischen Ansatz, bei dem die Ergebnisse kognitiver Testungen schnell und einfach übermittelt werden und einen stetigen Austausch zwischen Patient und Arzt ermöglichen. Durch eine regelmäßige Anwendung des eNCTs könnten Qualität der medizinischen Behandlung, sowie Prognose, Lebensqualität und Compliance von chronisch Lebererkrankten verbessert werden.

Das Ziel dieser Dissertationsschrift ist die erstmalige methodische Bewertung des eNCT hinsichtlich standardisierter Gütekriterien psychodiagnostischer Testverfahren. Diese erlauben eine Beurteilung von Qualität und Vorhersagekraft des eNCT und qualifizieren ihn für einen Vergleich mit bereits etablierten Testverfahren in der Diagnostik neuropsychischer Störungen bei erwachsenen Patienten mit einer manifesten Leberzirrhose. Weiterhin sollen Einflussfaktoren auf die eNCT-Performanz identifiziert, Testergebnisse bezüglich der Anwendung unterschiedlich großer Displays (10 Zoll versus 18 Zoll) verglichen, sowie Akzeptanz und Zumutbarkeit dieses neuen Testverfahrens evaluiert werden.

2 Methoden

2.1 Studienpopulation und Studiendesign

In einer prospektiven Querschnittstudie an der Charité, Campus Virchow-Klinikum (CVK) wurden von November 2014 bis Mai 2015 238 Studienteilnehmer untersucht. Patienten mit der Diagnose einer Leberzirrhose wurden ambulant in der Spezialsprechstunde für Lebererkrankungen oder im Rahmen eines stationären Aufenthaltes in der medizinischen Klinik mit Schwerpunkt Hepatologie und Gastroenterologie rekrutiert. Der Einschluss lebergesunder Kontrollprobanden aus dem Großraum Berlin-Brandenburg erfolgte im Rahmen einer klinikinternen Ausschreibung. Eine Subgruppe von 48 Probanden wurde innerhalb eines zeitlichen Abstandes von mindestens vier Tagen erneut untersucht. Alle Studienteilnehmer wurden mündlich und schriftlich über Art, Umfang und Ziel der Untersuchung aufgeklärt und gaben ihre schriftliche Einwilligung mit der Möglichkeit, diese jederzeit zurückzuziehen. Die Dokumentation der Untersuchungsdaten der Studienteilnehmer erfolgte einheitlich mit Hilfe eines Prüfbogens (Case Report Form) entsprechend dem von der Ethikkommission der Charité genehmigten Studienprotokolls (EA1/263/14).

2.1.1 Ein- und Ausschlusskriterien

Patienten und Kontrollen wurden nach folgenden Einschlusskriterien rekrutiert:

1. Alter der gesunden Kontrollprobanden ≥ 18 Jahre.
2. Alter der Patienten ≥ 18 Jahre.
3. Diagnosestellung einer Leberzirrhose jeglicher Genese bei bekannter Lebererkrankung (auch alkoholische Lebererkrankung) mittels gesicherter Histologie oder dem Verdacht auf Leberzirrhose wenn mindestens zwei der folgenden Kriterien zutreffen: (I) unregelmäßige Begrenzung der Leber in Computer- oder Magnetresonanztomografie, (II) Splenomegalie (Milz > 12 cm), (III) Fibroscan $> 17,6$ kPa, Vorliegen von Ösophagusvarizen, (IV) Thrombozytenzahl < 120 /nl, (V) Vorliegen von Aszites, (VI) HE nach den Westhaven-Kriterien.
4. Kein florider Alkohol- oder Drogenabusus innerhalb der letzten 6 Monate vor Studieneinschluss für sowohl Patienten als auch Kontrollprobanden.
5. Schriftliche Einwilligung zur Teilnahme an der Studie.

Der Ausschluss der Studienteilnehmer erfolgte anhand folgender Kriterien:

1. Score in der Mini Mental Status Examination (MMSE) <26 als Hinweis auf das Vorliegen einer Demenz (50).
2. Kontrollprobanden mit erhöhten laborchemischen Leberfunktionsparametern (ALAT und AP) außerhalb des Referenzbereichs (76).
3. Bekannte Rot-Grün Sehschwäche.
4. Bekannte neurologische oder psychiatrische Erkrankungen, welche zu einer eingeschränkten kognitiven Leistung führen (ausgenommen hierbei war die HE).
5. Studienteilnehmer mit aktueller ZNS-wirksamer Medikation. Davon ausgenommen: Patienten unter regelmäßiger Einnahme von Antidepressiva bei stabilem Krankheitsverlauf und Patienten mit verschriebener Schlafmedikation.

Der Ausschluss der für die Analyse der Retestreliabilität eingeschlossenen Studienteilnehmer erfolgte zusätzlich bei stationärem Patientenstatus und instabilem Krankheitsverlauf (notfallmäßige bzw. ungeplante Hospitalisierung) zwischen Test und Retest.

2.2 Untersuchungsablauf

Die Untersuchung der ambulant eingeschlossenen Probanden fand in den Räumlichkeiten der Spezialsprechstunde für Lebererkrankungen und der Ambulanz für Stoffwechselstörungen im CVK statt. Stationär rekrutierte Patienten wurden in den Patientenzimmern der Klinik für Gastroenterologie und Hepatologie des CVK untersucht. Um während der Testbearbeitung eine ruhige Atmosphäre ohne Störungen des Versuchsablaufs zu gewährleisten und eine ausreichende Konzentration zu ermöglichen wurden Krankenhauspersonal, Mitpatienten bzw. Besucher gebeten, das Zimmer für die Zeit der Testdurchführung nicht zu betreten. Die Testbearbeitung fand am Tisch in sitzender Position statt. Auf eine blendungsfreie und ausreichende Beleuchtung der Arbeitsmaterialien wurde geachtet. Ebenso erfolgte ein Hinweis an die Probanden, falls notwendig eine Lesebrille zu benutzen. Weiterhin wurden alle Untersuchungen durch speziell trainiertes medizinisches Personal durchgeführt. Vor Beginn der Untersuchung wurden Patienten zunächst über Ablauf, Zweck und Notwendigkeit der Untersuchung aufgeklärt und um ihr Einverständnis gebeten.

Es erfolgten die klinisch-internistische Untersuchung, sowie der Bestimmung des kognitiven Leistungsniveaus mittels (1) MMSE und die Detektion einer oHE durch (2)

HESA und (3) WHC. Bei Studieneinschluss erhielten alle Probanden eine venöse Blutentnahme von etwa 10 ml zur Bestimmung folgender Blutparameter: Hämoglobin (Hb), mittleres korpuskulares Volumen, mittlerer korpuskularer Hämoglobingehalt, Erythrozyten, Leukozyten, Thrombozyten, Natrium, Kalium, Kalzium, ASAT (Aspartat-Aminotransferase), ALAT (Alanin-Aminotransferase), AP (Alkalische Phosphatase), G-GT (Gamma-Glutamyltransferase), Bilirubin, Albumin, Quick, INR, Prothrombin und Ammoniak. Die Laboruntersuchung diente einerseits dem Ausschluss einer Lebererkrankung bei den Kontrollprobanden, andererseits zur laborchemischen Charakterisierung der Leberzirrhose innerhalb des Patientenkollektivs. Zur korrekten Bestimmung des Ammoniak-Spiegels erfolgte die Phlebotomie an nicht gestauten Venen mit anschließendem Transport (≤ 20 Minuten) der gekühlten Probe zur laboranalytischen Bestimmung des Serumspiegels.

Im Anschluss erfolgte die eigentliche psychometrische Testung der Studienteilnehmer in der folgenden Reihenfolge: (4) PSE-Syndrom-Test, (5) eNCT und (6) Stroop. Bei fehlendem Verständnis der Testanweisung wurde die gesamte Instruktion wiederholt. Weiterhin wurden vor Beginn der Einzeltests mit einem Übungsbeispiel begonnen um aufgetretene Fehler sofort korrigieren zu können und ausführungsbedingte Fehler zu vermeiden. Nach abgeschlossener Testprozedur wurden die Probanden gebeten, mit Hilfe eines (7) standardisierten Fragebogens ein Feedback hinsichtlich Anwendung und Akzeptanz von elektronischen Geräten zu geben sowie eine Aussage zur Anwendung der favorisierten Displaygröße im häuslichen Umfeld zu treffen.

2.3 Psychometrische Testverfahren

Alle Untersuchungsverfahren wurden standardisiert durchgeführt und werden im Folgenden beschrieben.

2.3.1 Portosystemischer Enzephalopathie (PSE)-Syndrom-Test

Der PSE-Syndrom-Test ist ein in Deutschland entwickeltes und normiertes psychodiagnostisches Verfahren zur quantitativen Erfassung und Objektivierung neuropsychischer Störungen bei erwachsenen Patienten mit Leberinsuffizienz (PSE-Syndrom-Test Manual), insbesondere zur quantitativen Erfassung der mHE (55). Der Test ist praktisch, einfach durchzuführen und hat sich in den durchgeführten Studien als eine präzise Methode zur Diagnostik und Verlaufskontrolle einer mHE erwiesen (26, 57, 60). Durch standardisierte Instruktionen, detaillierte Auswertungshinweise und

beigefügte normierte Schablonen erreicht das Testverfahren eine hohe Objektivität. Die Retestreliabilität in den Untertests beträgt für Patienten zwischen $r=0,56$ und $r=0,95$ (55). Daneben wird der Test dank hoher Werte für Sensitivität und Spezifität in der Literatur häufig als Referenzstandard bezeichnet (19, 57, 58). Bis zum heutigen Zeitpunkt konnte der Test in vielen Ländern (u.a. in Italien, Spanien, Großbritannien, Indien und den USA) validiert und standardisiert werden (56). Der PSE-Syndrom-Test wurde aus einer umfangreichen Testbatterie mit mehr als 30 verschiedenen Tests ausgearbeitet und besteht aus fünf verschiedenen Papier- und Bleistifttesten. Dazu gehören NCTA und NCTB, Zahlen-Symbol-Test (ZST), Linien-Nachfahr-Test (LNT) und der Test Kreise-Punktieren (KP), welcher in der Literatur auch als Serial Dotting Test (SDT) bezeichnet wird. Die Untertests sind in der Neurologie zur Beurteilung von psychomotorischen Fertigkeiten lange und gut etabliert (43, 68).

2.3.1.1 Testanweisung und -bearbeitung

Die Bearbeitung der Untertests erfolgt nach folgendem Prinzip: Anweisung für den Patienten und Überprüfung des Verständnisses; Bearbeitung der Übungsaufgaben; Kontrolle der Übungsaufgaben und ggf. Korrektur; Bearbeitung des eigentlichen Tests. Eine Ausnahme bildet der NCTB, dessen Bearbeitung schon im Vortest, dem NCTA geübt wird. Die Testvorlagen wurden freundlicherweise von Frau Prof. Dr. med. Karin Weissenborn (Klinik für Neurologie, Medizinische Hochschule Hannover) zu Verfügung gestellt, lagen aber bei Studienbeginn nicht rechtzeitig vor, so dass frei zugängliche, alternative Testversionen verwendet wurden. Modifikationen in der verwendeten Testvorlage werden bei den jeweiligen Untertests beschrieben und die eingesetzte Version visuell im Anhang dargestellt. Die Durchführung der Untertests wurde im Vergleich zum Manual des PSE-Syndrom-Test aus methodischen Gründen variiert und die Bearbeitung der Untertest erfolgte in folgender Reihenfolge:

1. NCTA I → 2. NCTA II → 3. NCTA III → 4. NCTB → 5. ZST → 6. LNT → 7. KP

Die Testbearbeitung des NCTA erfolgte hierbei wiederholt anhand von drei Parallelversionen, dargestellt mit römisch I, II und III.

2.3.1.2 Einzeltestverfahren

Zahlenverbindungstest A (NCTA)

Die Aufgabe im NCTA besteht darin, 25 ungeordnete auf einem Blatt Papier verteilte Zahlen so schnell wie möglich in aufsteigender Reihenfolge mit einem Stift zu verbinden (siehe Anhang 1). Dabei wird die Zeit vom Ansetzen des Stiftes auf die Zahl 1 bis zum Erreichen der Zahl 25 gestoppt und die Gesamtzeit in ganzen Sekunden auf dem Testbogen festgehalten. Bei Auslassen einer Zahl oder fehlerhafter Reihenfolge wird der Proband sofort unterbrochen und muss vor Weiterbearbeitung den Fehler korrigieren. Vor Testbearbeitung hat der Patient die Gelegenheit auf einem Demonstrationsblatt eine Übungsaufgabe zu bearbeiten und sich mit dem Verfahren vertraut zu machen (siehe Anhang 2).

Zahlenverbindungstest B (NCTB)

Im NCTB sind die Zahlen von 1 bis 13 und die Buchstaben von A bis L auf einem Blatt Papier dargestellt. Aufgabe ist es, die Zahlen und Buchstaben in alternierender Reihenfolge miteinander zu verbinden; d.h. von 1 nach A, von A nach 2 von 2 nach B usw. bis zum Ende (siehe Anhang 3). Die dafür erforderliche Zeit in Sekunden bildet das Testergebnis. Auf einen Probedurchgang vor Testanwendung wird verzichtet.

Zahlen-Symbol-Test (ZST)

Der ZST hat eine Zeitvorgabe von 90 Sekunden und beginnt nach einem Startzeichen. Ziel ist es so viele leere Kästchen wie möglich, welche sich unter einer Zahlenreihe befinden, mit dem richtigen Zeichensymbol zu vervollständigen. Die Zugehörigkeit zwischen Zahl und dem richtigen Symbol lässt sich aus einer Zahlen-Symbol-Reihe in der obersten Zeile der Testvorlage ablesen. Zur Übung des Testverfahrens bearbeitet der Proband vor Beginn der Messung eine Übungsreihe bestehend aus 8 Kästchen um etwaige Fehler sofort korrigieren zu können. Als Testergebnis gilt die Anzahl der korrekt zugeordneten Symbole innerhalb der Testzeit. Falsch zugeordnete, stark entstellte, räumlich gedrehte und gespiegelte Zeichensymbole werden nicht mitgezählt. In der verwendeten Testvorlage unterschieden sich einige Zeichensymbole wie auch die Zahlen-Symbolzuordnung im Vergleich zur Originalvorlage (siehe Anhang 4).

Linien-Nachfahr-Test (LNT)

Beim LNT wird gefordert schnellst möglich eine dritte Linie zwischen zwei auf einem Blatt Papier präsentierten Linien zu zeichnen, ohne die beiden vorgegebenen Linien zu berühren oder zu überschreiten. Vor Testdurchführung erfolgt die Bearbeitung einer

Übungsaufgabe (siehe Anhang 5). Beim LNT werden zwei Messergebnisse erhoben: Die benötigte Zeit für die Testbearbeitung in Sekunden, welche im Folgenden als LNT-Z bezeichnet wird und (b) die Fehlerauszählung und –gewichtung, nachfolgend als LNT-F bezeichnet. Zur Auswertung des LNT wird eine auf durchsichtiger Folie gedruckte Schablone mit dem dargestellten Raster verwendet (siehe Anhang 6). Diese Schablone wird auf den bearbeiteten Test gelegt und eine Fehlerauszählung mittels der Vergabe von Fehlerpunkten (FP) vorgenommen. Die Fehlerbestimmung erfolgt für jedes Rasterfeld separat. Vergeben werden die Fehlerpunkte beim Berühren (1FP) oder Überschreiten (2FP) der vorgezeichneten Linien, sowie beim Überschreiten der Rasterbegrenzung (3FP). Da sich ein Unterschied im Größenverhältnis zwischen verwendeter Testvorlage und der im PSE-Syndrom-Test bereitgestellten Auswertungsschablone zeigte, wurde im Rahmen der Studie eine eigene Schablone mitsamt Raster angefertigt. Diese basiert auf der verwendeten Testvorlage und ermöglicht die korrekte Bearbeitung von Fehlerzählung und –gewichtung.

Kreise-Punktieren (KP) / Serial Dotting Test (SDT)

Die Aufgabe beim KP besteht darin, mit einem Stift möglichst schnell und genau einen Punkt in die Mitte von Kreisen zu setzen, die in 10 Reihen bestehend aus 10 Kreisen angeordnet sind (siehe Anhang 7). Der Durchmesser der Kreise beträgt 10 mm und die Bearbeitung der Zeilen soll jeweils von links nach rechts erfolgen. Messergebnis ist die Testzeit in Sekunden für die Bearbeitung/Punktierung aller Kreise. Vor der eigentlichen Testung erfolgt zum Einüben in das Testverfahren die Bearbeitung einer Übungsaufgabe bestehend aus 2 Reihen zu je 10 Kreisen.

2.3.1.3 Auswertung und Interpretation

Neben der Ermittlung der Testergebnisse in Sekunden oder Fehlerzahl erfolgte auch eine Umwandlung der Rohtestwerte innerhalb der Untertests in alterskorrigierte Wertepunkte (WP). Die normativen Daten für diese Korrektur wurden bei der Entwicklung des PSE-Syndrom-Tests mittels einer Kontrollgruppe von 150 gesunden Probanden ermittelt und dienen als Referenz für die Bestimmung der WP. Für Leistungen im Bereich zwischen +/- einer Standardabweichung (SD) wird ein WP von 0 zugeordnet. Testwerte oberhalb der einfachen SD werden mit +1 WP gewichtet und Testwerte unterhalb der einfachen, zweifachen und dreifachen SD führen zu einem WP

von entweder -1, -2 oder -3. Durch anschließende Addition der WP wird ein Punktwert (der sogenannte PHES) errechnet, welcher schnell und einfach interpretiert werden kann. Insgesamt können WP zwischen +6 und – 18 erzielt werden. Die Diagnose einer mHE kann ab einem Score unterhalb von -4 WP gestellt werden.

2.3.2 Elektronischer Zahlenverbindungstest (eNCT)

Der eNCT wurde als Tablet-basiertes Testverfahren im Rahmen des FP7-EU-Projekts d-LIVER in Kooperation mit den Projektpartnern der Universität Newcastle und dem Fraunhofer-Institut für Biomedizinische Technik auf einem Windows Betriebssystem (Windows 8, Microsoft Corporation, Redmond, WA/USA) in Eigenherstellung zu Forschungszwecken entwickelt. Als digitale Variante des NCTA kann der eNCT im Sinne eines sprachfreien Intelligenz-Tests zur Messung der psychomotorischen Schnelligkeit angesehen werden. Das Testverfahren wurde bereits von medizinischem Fachpersonal geprüft und in verschiedenen Entwicklungsprozessen an die Patientenbedürfnisse im häuslichen Umfeld angepasst. Im Unterschied zur papierbasierten Version des NCTA sollen 25 ungeordnete Zahlen, dargestellt auf einem berührungsempfindlichen Bildschirm, in aufsteigender Reihenfolge so schnell wie möglich mit einem Finger berührt werden.

2.3.2.1 Konzept

Bei der Konzeption und Auswahl der Testvorlage wurden folgende Punkte berücksichtigt (siehe Abb. 5):

1. Das Bildschirmfenster beinhaltet 25 unterschiedlich angeordnete schwarze Ziffern auf grauem Hintergrund (1 bis 25).
2. Das Berühren der Ziffern erfolgt in aufsteigender Reihenfolge (1-2-3-4...) durch Antippen mit einem Finger.
3. Die Ziffern sind schwarz vor einem grauen, kreisförmigen Hintergrund und wechseln nur beim Antippen in richtiger Reihenfolge die Farbe des Zahlenhintergrundkreises von grau zu gelb.
4. Nur wenn der Test in korrekter Weise abgeschlossen ist und alle Ziffern einen gelben Hintergrund besitzen wird die Zeit gestoppt, intern gespeichert und auf dem Bildschirm dargestellt.

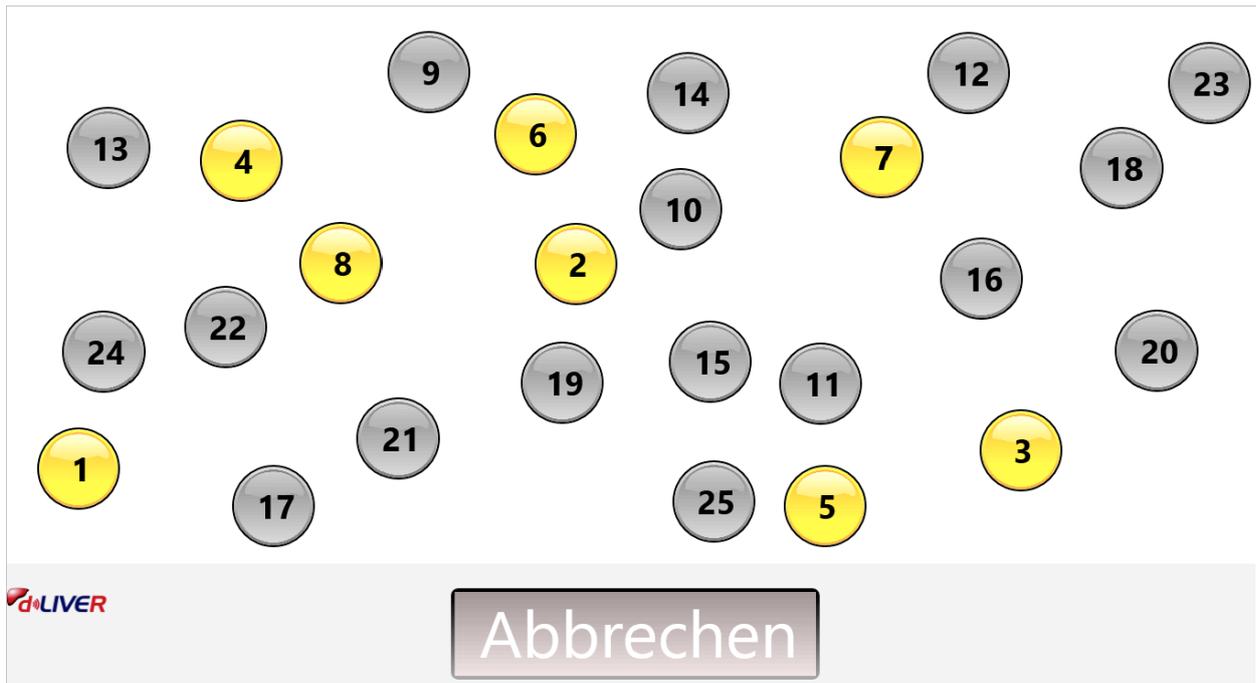


Abb. 5 Bildschirmfenster beim eNCT.

2.3.2.2 Testanweisung und -bearbeitung

Die Testbearbeitung des eNCT umfasst drei Testdurchgänge mit jeweils einer unterschiedlichen Testvorlage, d.h. drei verschiedene Parallelversionen (eNCT I, II und III) werden nacheinander absolviert. Die Entwicklung der Parallelversionen des eNCT erfolgte über eine versetzte Anordnung der Zahlen auf dem Bildschirm. Die Bearbeitung der Testvorlage lief folgendermaßen:

Anweisung für den Patienten und Überprüfung des Verständnisses; Bearbeitung der Übungsaufgabe; Kontrolle der Übungsaufgaben und ggf. Korrektur; Bearbeitung der Testvorlagen des eNCT I, II und III.

Um den Einfluss der Bildschirmgröße auf die Testzeit zu überprüfen und mögliche Benachteiligungen durch eine zu kleine Bildschirmgröße zu detektieren, erfolgte die Testbearbeitung des eNCT sowohl auf einem Tablet mit 10 Zoll Bildschirmdiagonale (Aspire Switch 10, Acer, Taipeh, Taiwan), nachfolgend als eNCT10Z bezeichnet, als auch auf einem Tablet mit 18 Zoll Bildschirmdiagonale (XPS One, Dell Inc., Round Rock, TX/USA), nachfolgend als eNCT18Z bezeichnet (siehe Abb. 6).

Nach Programmstart durch den Testleiter erfolgt die standardisierte Instruktion für den Übungsteil:

„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 so schnell wie möglich in aufsteigender Reihenfolge anzutippen. Tippen Sie die richtige Zahl an, ändert der umgebene 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.“

Die erforderliche Zeit wird innerhalb des elektronischen Gerätes automatisch gemessen und gespeichert. Die Testzeit wird vom Versuchsleiter auf dem Prüfbogen notiert und daran anschließend durch Berührung des Bildschirms die zweite (eNCT II) und dritte Testvorlage (eNCT III) in gleicher Weise gestartet.

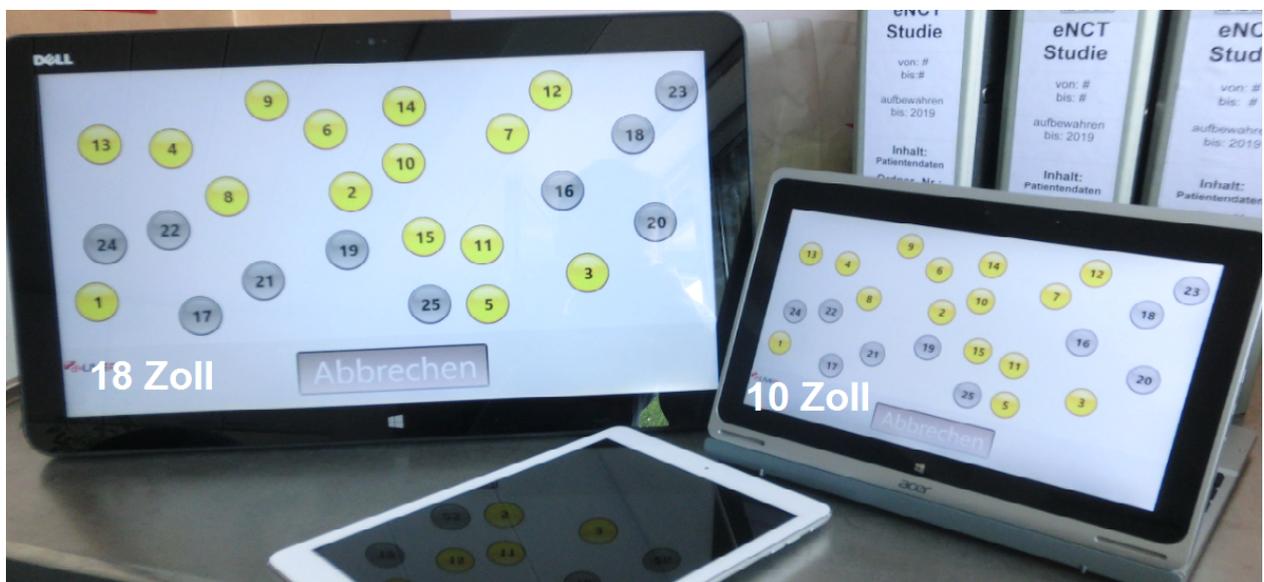


Abb. 6 Darstellung der unterschiedlichen Displaygrößen. Links: eNCT18Z; Rechts: eNCT10Z. eNCT10Z/18Z elektronischer Zahlenverbindungstest 10 Zoll/ 18 Zoll.

2.3.2.3 Auswertung und Interpretation

Als Testmaß gilt die Bearbeitungsgeschwindigkeit in Sekunden. Neben Ermittlung der rohen Testergebnisse für sowohl eNCT10Z I, II und III als auch eNCT18Z I, II, und III erfolgte die Berechnung der durchschnittlichen Bearbeitungszeit der drei Parallelversionen innerhalb einer Tabletgröße, im Folgenden als „Average Konzept“ bezeichnet (AveNCT10Z für die 10 Zoll Version bzw. AveNCT18Z für die 18 Zoll Version). Auf eine Interpretation der Testergebnisse und Festlegung von Cutpoint-Werten hinsichtlich der Diagnose einer mHE wird Rahmen der vorliegenden Arbeit verzichtet. Diese bildet die Grundlage einer weiteren Inauguraldissertation (77).

2.3.3 EncephallApp-Stroop Test (Stroop)

Dieses Smartphone-basierte Testverfahren wurde bereits kurz in Abschnitt 1.3 beschrieben. Da die Ergebnisse im Stroop auf die Fragestellungen der vorliegenden Arbeit keinen Einfluss haben und zudem Gegenstand einer weiteren Dissertation sind (77), wird diese Methode im Rahmen der vorliegenden Arbeit nicht weiter untersucht.

2.4 Fragebogen zum Feedback der Studienteilnehmer

Zur Evaluation der Patientenmeinung hinsichtlich Akzeptanz des elektronischen Testverfahrens gaben die Probanden im Anschluss der psychometrischen Testung ein kurzes Feedback. Mit Hilfe eines standardisierten Fragebogens wurden der regelmäßige Gebrauch von technischen Geräten, die Durchführbarkeit der Tests und die Displaygröße bewertet (siehe Anhang 1).

2.5 Studienaufbau & statistische Methoden

Grundlage für die statistischen Berechnungen bildeten das Statistical Package for Social Science Software (SPSS 22.0, SPSS Inc., Chicago, ILL/USA), das XLSTAT 2017 (Data Analysis and Statistical Solution for Microsoft Excel, Addinsoft, Paris, France) und das Graph Pad Prism 5.0 (Graphpad Software Inc., San Diego, CAL/USA). Anfänglich wurden alle Daten deskriptiv analysiert und auf fehlerhafte Eingaben sowie auf unschlüssige Werte untersucht. Für kategoriale Daten wurden absolute und relative Häufigkeiten einzelner Variablen tabellarisch dargestellt. Von metrisch skalierten Variablen wurden Mittelwert, Median und die Standardabweichung angegeben.

Hinsichtlich demographischer Daten, der Ergebnisse der im Studienprotokoll festgelegten Tests und der Auswertung des Fragebogens erfolgte ein Vergleich der

Untersuchungsgruppen. Sämtliche Gruppenvergleiche wurden für ungerichtete Hypothesen berechnet. Der kritische p-Wert wurde mit einem α -Level von 5% angesetzt. Für metrische Daten überprüften wir mittels Shapiro-Wilk Test sowie der graphischen Einschätzung anhand von Normalverteilungsdiagrammen, ob eine Normalverteilung angenommen werden kann. Bei gegebener Normalverteilung verwendeten wir den Levene-Test. War dieser geeignet, erfolgte der Gruppenvergleich mittels parametrischer Testung, dem Zweistichproben-t-Test für unabhängige Stichproben. War die Fehlervarianzhomogenität nicht gegeben, erfolgte der Welch-Test zur Gruppenvergleichbarkeit. Bei fehlender Normalverteilung nutzten wir den nichtparametrischen Mann-Whitney-U-Test zwischen zwei unverbundenen Stichproben. Kategoriale Daten wurden mittels Fishers Exakten Tests bzw. Chi-Quadrat miteinander verglichen.

Die Bewertung von Reliabilität und Validität erfolgte in der vorliegenden Arbeit auf Grundlage von Korrelationskoeffizienten. Diese können Werte von 0 bis 1 (bzw. -1) annehmen, wobei 1 bzw. -1 den größtmöglichen positiven bzw. negativen Zusammenhang bedeuten und 0 für das Fehlen eines Zusammenhangs zwischen den Variablen steht. Zur Beurteilung von Messfehlern und Ermittlung der Retestreliabilität des eNCT wurden Intra-Klassen-Korrelationen (IKK) einschließlich Konfidenzintervall verwendet. Das Auftreten von Lerneffekten bei wiederholter Testbearbeitung wurde mittels Wilcoxon-Vorzeichenrangtest untersucht. Voraussetzung für die Analyse von Retestreliabilität und Lerneffekt waren eine wiederholte Bearbeitung der im Studienprotokoll eingeschlossenen Testverfahren sowie ein stabiler Krankheitsverlauf. Ein Lerneffekt wurde bei einem Unterschied in der Testzeit zwischen Test und Retest angenommen. Mit Hilfe drei verschieden konzipierter Parallelversionen des eNCT wurde über eine Berechnung des Cronbach's alpha (rtt) die Paralleltestreliabilität bestimmt und mit der des NCTA verglichen. Weiterhin überprüften wir mittels Friedmanns zweifaktorieller Varianzanalyse nach Rang bei verbundenen Stichproben und Post-hoc Tests mit Dunn-Bonferroni korrigierten Gruppenvergleichen, ob die untersuchten Testversionen Unterschiede im Schwierigkeitsgrad besitzen.

Die methodische Bewertung von Konstrukt- und Kriteriumsvalidität erfolgte im Patientenkollektiv mittels Rangkorrelationskoeffizienten nach Spearman (SKK), um Effektstärken auch bei non-parametrischen und nicht linearen Zusammenhängen vergleichen zu können. Innerhalb der Konstruktvalidierung erfolgte die Überprüfung der divergenten Validität über Korrelationen zwischen dem Merkmal Testzeit im eNCT und

demographischen Daten, die im Wesentlichen andere Konstrukte abbilden. Dazu gehörten BMI, Art der postschulischen Ausbildung (Keine, Ausbildung, Studium), Beschäftigungsverhältnis (Vollzeit, Teilzeit, Arbeitslos, Arbeitsunfähig, Berentet und Student). Für die Untersuchung der konvergenten Validität wurde die Testzeit im eNCT mit den Einzeltests des PSE-Syndrom-Tests korreliert. Die Analyse der kriterienbezogenen Validität erfolgte hingegen mit Hilfe der Korrelation zwischen eNCT-Performanz und aus dem PHES abgeleiteten Diagnosekriterium mHE.

Die Beurteilung der Nebengütekriterien Testfairness und Zumutbarkeit sowie Unterschiede in der Handhabung zwischen den Displaygrößen wurde anhand des Fragebogens vorgenommen.

Um unabhängige Prädiktoren auf die eNCT-Performanz zu identifizieren, wurden zunächst potentielle Einflussfaktoren in einem linearen Regressionsmodell untersucht. Voraussetzung für die Aufnahme in das Modell war eine signifikante Korrelation zwischen potentielltem Einflussfaktor und dem Merkmal Testzeit im eNCT. In Anlehnung an die methodische Auswertung des PSE-Syndrom-Tests wurden in der Kontrollgruppe die Merkmale Alter, Geschlecht und Bildungsgrad analysiert, wobei hinsichtlich des Bildungsgrades gemäß des Deutschen Schul- und Bildungssystems drei Schulgrade definiert wurden: Grad 1: Hauptschule bzw. neun Jahre Schulunterricht; Grad 2: Realschule bzw. 10-12 Jahre Schulunterricht ohne Abitur; Grad 3: Abitur/Studium bzw. mindestens 12-13 Jahre Schulunterricht mit Abitur und ggf. Studium. Aufgrund einer ordinalen Skalierung erfolgte vor Modelleinschluss zudem eine Dummy-Codierung (Grad 1 ja=1/nein=0; Grad 2 ja=1/nein=0; Grad 3 ja=1/nein=0). Innerhalb der Patientenkohorte gehörten zu den untersuchten Variablen neben den bei Studieneinschluss erhobenen laborchemischen Blutparametern folgende Faktoren: MMSE-Score, MELD-Score, CHILD-PUGH-Score, Einnahme HE wirksamer oder diuretischer Medikamente, Anwesenheit von Ösophagusvarizen, stattgehabter SBP, Zustand nach transjugulärem intrahepatischer portosystemischer Shunt (TIPPS). Die multiple Regression erfolgte mittels schrittweiser Selektion. Dieses Verfahren ermöglicht die Aufnahme oder den Ausschluss von unabhängigen Variablen anhand ihres Signifikanzwertes und stellt den Einschluss der Variablen sicher, die den größten Anteil an der Erklärung der abhängigen Variablen haben. Das Regressionsmodell wurde mittels Varianzanalyse (F-Test) auf statistische Signifikanz hin überprüft und anhand seines Bestimmtheitsmaßes (r^2 -Wert) charakterisiert. Dieser gibt an, wieviel Prozent der Varianz durch die unabhängigen Variablen (Einflussfaktoren) von der abhängigen Variablen (eNCT-

Performanz) erklärt werden kann. In dieser Arbeit wurde zudem das angepasste r^2 verwendet, um bei steigender Variablenzahl im Regressionsmodell den Effekt eines Anstiegs des r^2 -Wertes zu korrigieren. Ein p -Wert $< 0,05$ galt hierbei als statistisch signifikant und bescheinigte der Variable die Eigenschaft als unabhängiger Prädiktor auf die eNCT-Performanz. Um den Zusammenhang zwischen Testzeit im eNCT und identifiziertem Einflussfaktor visuell zu demonstrieren, stellten wir die Wertepaare exemplarisch für den eNCT10Z I in einem Streudiagramm dar. Jedes Wertepaar wird dabei durch einen Punkt abgebildet und der Zusammenhang über eine Regressionsgerade beschrieben.

Um Aussagen zur Bedeutsamkeit der Ergebnisse treffen zu können, wurden zudem Effektstärken nach Cohen berechnet (78). Weiterhin wurden die Modelle hinsichtlich der Power in Abhängigkeit vom Stichprobenumfang untersucht.

Um den Einfluss der Displaygröße auf die Testzeit im eNCT zu konstatieren, wurden Bland-Altman-Diagramme erstellt. Ergänzend wurden Bias und zugehörige Übereinstimmungsgrenzen berechnet und in das Streudiagramm eingezeichnet.

2.5.1 Multiples Testen

Zum Nachweis von Effekten oder Zusammenhängen werden in der biomedizinischen Forschung oftmals statistische Signifikanztests verwendet und das Ergebnis in Form von p -Werten angegeben (79). Bei einem kleineren p -Wert als dem des vorgegebenen Signifikanzniveaus α (in der Regel 5%) wird die Nullhypothese verworfen. Man spricht von einem signifikanten Testergebnis. Hierbei begrenzt α die Wahrscheinlichkeit für den Fehler 1. Art, das heißt beim Ablehnen der Nullhypothese eine Fehlentscheidung zu treffen (79). Wenn nun mehrere Hypothesen an demselben Kollektiv gleichzeitig getestet werden spricht man von multiplen Testen. Hier steigt die Wahrscheinlichkeit bei Beibehaltung des Signifikanzniveaus α eine Falschaussage zu machen mit der Anzahl der durchgeführten Tests (80). Man spricht auch von einer Akkumulation des α -Fehlers, die eine Anpassung des Signifikanzniveaus nötig macht (81). Daher sollte zur Einhaltung des vorgegebenen Signifikanzniveaus bei Anwendung multipler Signifikanztests in konfirmatorischen Studien Korrekturverfahren für multiples Testen verwendet werden (79). Aufgrund des explorativen Studiendesigns wurde in der vorliegenden Arbeit in Anlehnung an Victor und Kollegen (80) auf eine Korrektur für multiples Testen verzichtet.

3 Ergebnisse

3.1 Charakterisierung der Studienkohorte

Die Studienpopulation bestand aus 238 Teilnehmern, von denen 28 Teilnehmer (oHE n=18, MMSE-Score < 26 Punkte n=7, Alkoholabusus n=1, Psychotrope Medikation n=1, Leberfibrose n=1) die Einschlusskriterien nicht erfüllten.

Um den Altersdurchschnitt zwischen Patienten und Kontrollen zu wahren, wurden weiterhin 63 junge Kontrollprobanden (Alter < 39 Jahre) aus den Analysen der vorliegenden Arbeit ausgeschlossen und im Rahmen einer gesonderten Dissertation untersucht (77).

Insgesamt 86 Patienten mit manifester Leberzirrhose konnten in unsere Arbeit eingeschlossen und mit 61 gesunden Kontrollprobanden verglichen werden (siehe Abb. 7).

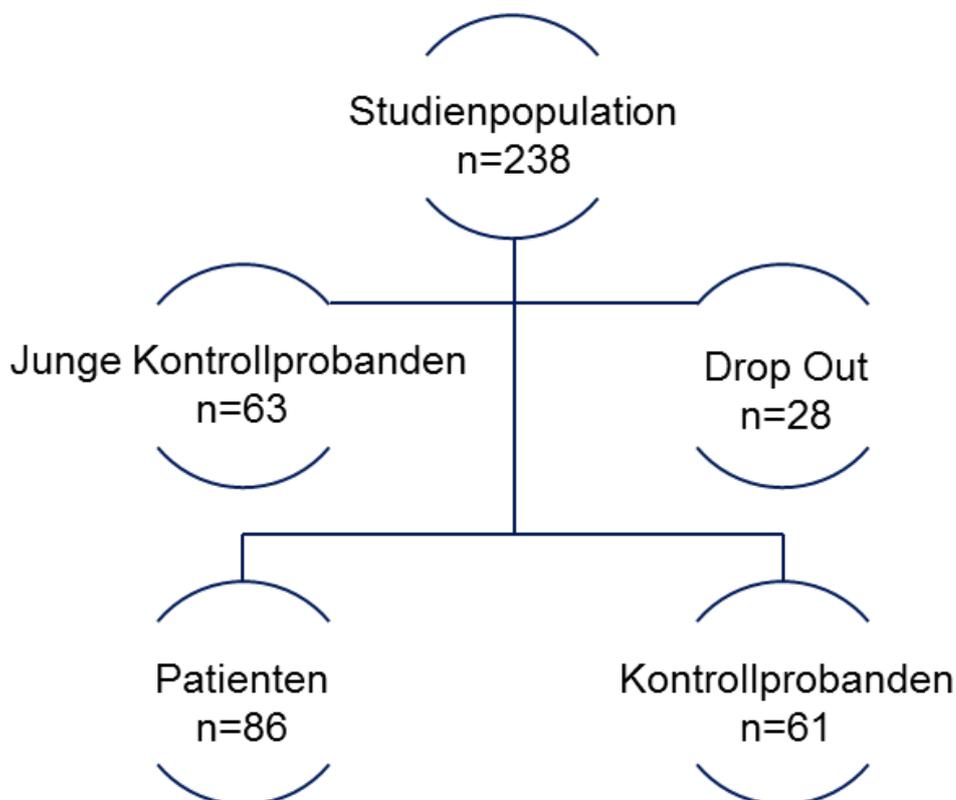


Abb. 7 Schematische Darstellung der eingeschlossenen Studienteilnehmer. n Anzahl.

3.1.1 Demographische Daten der Studienteilnehmer

In Bezug auf Alter, Geschlecht, Bildungsgrad und BMI konnten keine Unterschiede zwischen den beiden Gruppen festgestellt werden. Für den MMSE-Score ließen sich Unterschiede in den Mittelwerten für beide Gruppen aufzeigen. Die demographischen Daten von Patienten und Kontrollprobanden sind in Tabelle (Tab.) 1 dargestellt.

Tab. 1 Demographische Daten der Studienteilnehmer.			
	PAT	KON	p
Anzahl gesamt	86	61	
Alter [Jahre] (SD)	59,3(7.9)	58,5(11)	0,63 ¹
Altersbereich (Jahre)	39-76	39-74	
Geschlecht (Männer/Frauen)	51/35	30/31	0,22 ²
Bildungsgrad	Bis 9 Jahre	9	0,09 ³
	10-12 Jahre	31	
	Abitur/Studium	21	
BMI [kg/m ²] (SD)	28,0(6)	26,8(4)	0.19 ¹
MMSE-Score (SD)	28,9(1)	29,6(1)	<0,001¹

PAT Patientengruppe; KON Kontrollprobanden; SD Standardabweichung; MMSE Mini Mental State Examination; p p-Wert; ¹ t-Test; ² chi-Quadrat-Test; ³ Pearson-Chi-Quadrat-Test.

3.1.2 Charakterisierung des Patientenkollektivs

Die Ursache einer Leberzirrhose war in 34 Fällen eine ASH, in 15 Fällen eine Virushepatitis (Hepatitis B n=2; Hepatitis C n=13), in 8 Fällen eine NASH, in 18 Fällen eine andere Ursache (Autoimmunhepatitis, primär biliäre Zirrhose, primär sklerosierende Cholangitis, Hämochromatose), bei 11 Patienten konnten wir keine Ätiologie zuordnen (siehe Abb. 8).

Zur weiteren Charakterisierung der Leberzirrhose wurden neben der Berechnung des MELD-Scores und der Stadieneinteilung nach CHILD-PUGH auch eine etwaige HE-wirksame Medikation, eine medikamentöse diuretische Therapie, das Vorliegen von Ösophagusvarizen und der Zustand nach (Z.n.) SBP sowie die Versorgung mittels TIPPS innerhalb des Patientenkollektivs untersucht und in Tab. 2 dargestellt.

Der Mittelwert im MELD-Score betrug indes 13 bei einer Standardabweichung von +/- 5 und einem Median von 10. Im CHILD-PUGH-Score zeigte sich sowohl ein Mittelwert als auch ein Median von 7 (+/- 3).

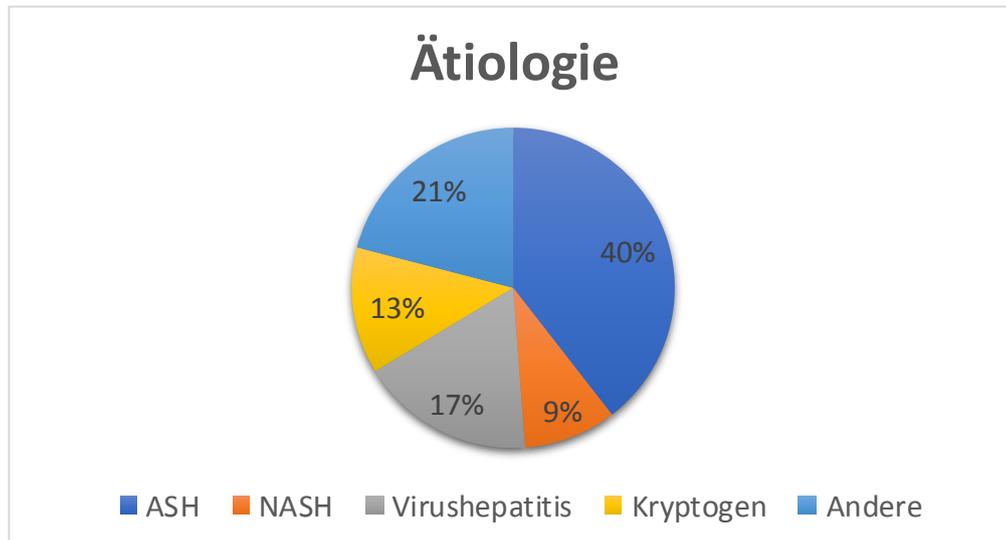


Abb. 8 Ätiologie der Leberzirrhose aller eingeschlossenen Patienten. ASH alkoholische Steatohepatitis; NASH nicht-alkoholische Steatohepatitis; Andere = Autoimmunhepatitis, primär biliäre Zirrhose, primär sklerosierende Cholangitis, Hämochromatose.

Tab. 2 Charakterisierung von Leberzirrhose, Therapie und Komplikationen.

	n	%		n	%
CHILD-PUGH-Stadium			Ösophagusvarizen		
Stadium A	32	40,5	ja	57	66,3
Stadium B	37	46,8	nein	9	10,5
Stadium C	10	12,7	unbekannt	20	23,3
HE wirksame Medikation			Z.n. SBP		
ja	28	32,6	ja	7	8,2
nein	58	67,4	nein	78	91,8
Diuretische Therapie			TIPPS		
ja	58	69,0	ja	10	16,1
nein	26	31,0	nein	52	83,9

n Anzahl; HE wirksame Medikation (Lactulose, Rifaximin oder Ornithinaspartat); Diuretische Therapie (Furosemid, Torasemid, Spironolacton); SBP Spontan bakterielle Peritonitis; TIPPS transjugulärer intrahepatischer portosystemischer Shunt.

3.2 Ergebnisse der im Studienprotokoll festgelegten Tests

3.2.1 Ergebnisse der Untertests im PSE-Syndrom-Test

Es ließen sich sowohl im NCTA, als auch in den weiteren Untertests des PSE-Syndrom-Tests, inklusive der finalen Punktezahl im PHES Unterschiede in den Testergebnissen zwischen Kontrollen und Patienten nachweisen. Die Patientengruppe generierte im Vergleich zu den gesunden Kontrollprobanden langsamere Testzeiten in den drei Parallelversionen des NCTA, im NCTB, im LNT-Z und im Test KP, während im ZST die Anzahl richtig zugeordneter Paare von Zahl und entsprechendem Symbol geringer ausfiel. Einzige Ausnahme bildete der Einzeltest LNT-F. Hier ließen sich zwischen den untersuchten Kollektiven keine Unterschiede detektieren. Die Testergebnisse in den papierbasierten Einzeltests sind in Tab. 3 für Patienten und Kontrollprobanden dargestellt.

Tab. 3 Einzeltests des PSE-Syndrom-Tests im Vergleich zwischen Kontrollprobanden und Patienten.

	PAT			KON			p
	n	M (SD)	Median (Q1-Q3)	n	M (SD)	Median (Q1-Q3)	
NCTA I [sek]	86	52,1(22)	49(35-63)	61	37,2(13)	35(28-46)	<0,001 ²
NCTA II [sek]	86	49,9(27)	41(41-60)	61	35,8(15)	31(26-41)	<0,001 ²
NCTA III [sek]	86	48,1(22)	44.5(32-57)	61	35,8(14)	32(27-43)	<0,001 ²
NCTB [sek]	83	141,4(99)	113(85-162)	59	89,9(40)	80(63-112)	<0,001 ²
ZST [Paare]	81	32,8(11)	33(24-41)	59	44,6(11)	44(38-53)	<0,001 ¹
LNT-Z [sek]	83	121(47)	109(83-109)	60	91,8 (34)	87(66-118)	<0,001 ²
LNT-F [Punkte]	83	28,4(39)	16(6-35)	60	24,6(32)	12.5(4-32)	0,29 ²
KP [sek]	86	52,9(19)	50(38-64)	60	42,5(14)	40(33-48)	<0,001 ²
PHES [Punkte]	81	-3,2(5)	-2(-6-0)	58	0,8(2.3)	1(0-2)	<0,001 ²

PAT Patientengruppe; KON Kontrollprobanden; n Anzahl; M Mittelwert; SD Standardabweichung; Q1-Q3 Perzentile 25-Perzentile 75; p p-Wert; NCTA/B Zahlenverbindungstest A/B; I,II,III Parallelversion I,II,III; ZST Zahlen-Symbol-Test; LNT-Z Zeit im Linien-Nachfahr-Test; LNT-F Fehler im Linien-Nachfahr-Test; KP Kreise-Punktieren; PHES Portosystemic Hepatic Encephalopathy Score; ¹ t-Test; ² Mann-Whitney-U-Test.

3.2.2 Ergebnisse der Paralleltests im eNCT

Innerhalb der statistischen Analyse zeigten sich im eNCT Unterschiede in den Testergebnissen der untersuchten Gruppen sowohl für die drei Parallelversionen als auch im Average Konzept nach der durchschnittlichen Testzeit für beide Tabletgrößen. In allen untersuchten Versionen absolvierte die Gruppe der Patienten den eNCT langsamer als die Gruppe der Kontrollprobanden. Die Testzeiten im eNCT sind in Tab. 4 für Patienten und Kontrollen getrennt dargestellt.

Tab. 4 Ergebnisse der Paralleltests im eNCT zwischen Kontrollprobanden und Patienten.

	PAT			KON			p
	n	M (SD)	Median (Q1-Q3)	n	M (SD)	Median (Q1-Q3)	
eNCT10Z I [sek]	86	45,3(17)	42(32-52)	61	34(8)	33(28-37)	<0,001 ²
eNCT10Z II [sek]	86	39,5(17)	37(30-44)	61	30,2(8)	28(24-35)	<0,001 ²
eNCT10Z III [sek]	86	40,7(15)	37(31-47)	61	30,3(8)	30(24-36)	<0,001 ²
AveNCT10Z [sek]	86	41,9(15)	38(32-47)	61	31,5(8)	30(26-36)	<0,001 ²
eNCT18Z I [sek]	85	47,1(15)	43(35-57)	59	34,3(8)	34(28-40)	<0,001 ²
eNCT18Z II [sek]	85	40,9(14)	38(30-49)	59	31,3(8)	30(24-36)	<0,001 ²
eNCT18Z III [sek]	85	43(16)	40(31-49)	59	32,1(9)	30(25-37)	<0,001 ²
AveNCT18Z [sek]	85	43,7(14)	41(33-52)	59	32,6(8)	30,7(26-38)	<0,001 ²

PAT Patientengruppe KON Kontrollprobanden; M Mittelwert; SD Standardabweichung; Q1-Q3 Perzentile 25-Perzentile 75; p p-Wert; eNCT10/18Z elektronischer Zahlenverbindungstest 10/18Zoll; AveNCT10/18Z Average Konzept 10/18 Zoll; I Parallelversion I; ¹t-Test; ² Mann-Whitney-U-Test.

3.3 Reliabilität

3.3.1 Retestreliaibilität

Zur Quantifizierung der Retestreliaibilität wurde ein Teil der Studienkohorte bestehend aus Kontrollprobanden (n=18) und ambulant getesteten Patienten (n=14) mit stabilem klinischen Verlauf zu einem zweiten Zeitpunkt erneut gemessen. Für beide elektronischen Testverfahren und dem NCTA ist die Retestreliaibilität in Tab. 5 dargestellt.

Tab. 5 Retestreliaibilität gemäß Intraklassenkorrelation (IKK).

	PAT		KON	
	IKK	95% KI	IKK	95% KI
eNCT10Z I	0,79**	0,37-0,93	0,82***	0,52-0,93
AveNCT10Z	0,92***	0,75-0,97	0,92***	0,75-0,97
eNCT18Z I	0,83**	0,48-0,94	0,88***	0,68-0,95
AveNCT18Z	0,85**	0,56-0,93	0,91***	0,61-0,97
NCTA I	0,82**	0,45-0,94	0,92***	0,78-0,97

PAT Patientengruppe KON Kontrollprobanden; 95% KI 95%iges Konfidenzintervall; eNCT10/18Z elektronischer Zahlenverbindungstest 10/18Zoll; AveNCT10/18Z Average Konzept 10/18 Zoll; I Parallelversion I; NCTA Zahlenverbindungstest A; *p≤0,05; **p≤0,01; ***p≤0,001.

In Abhängigkeit der Untersuchungsgruppe variierte die Retestreliaibilität im eNCT10Z I für beide Kohorten zwischen IKK-Werten von r=0,79 (p≤0,01) und r=0,82 (p≤0,001). Sie lag damit unter denen des AveNCT10Z (r=0,92; p≤0,001).

Ähnliche Ergebnisse ließen sich für den eNCT18Z detektieren. Hier variierten die IKK-Werte für Patienten und Kontrollprobanden in der Parallelversion I zwischen r=0,83 (p≤0,01) und r=0,88 (p≤0,001), sowie zwischen r=0,85 (p≤0,01) und r=0,91 (p≤0,001) im Average Konzept des eNCTA18Z.

Abhängig vom untersuchten Kollektiv generierten Studienteilnehmer im NCTA IKK-Werte von r=0,82 (p≤0,01) bzw. r=0,92 (p≤0,001).

3.3.2 Paralleltestreliabilität

Die Paralleltestreliabilität als Maß für Messgenauigkeit und Übereinstimmung der drei Testversionen ist separiert für Patienten und Kontrollen in Tab. 6 dargestellt.

Tab. 6 Paralleltestreliabilität gemäß Reliabilitätskoeffizient nach Cronbach alpha (rtt).

	PAT	KON
eNCT10Z	0,90	0,91
eNCT18Z	0,90	0,89
NCTA	0,94	0,88

PAT Patientengruppe; KON Kontrollprobanden; eNCT10/18Z elektronischer Zahlenverbindungstest 10/18Zoll; NCTA Zahlenverbindungstest A.

In Abhängigkeit des untersuchten Kollektivs variierten die rtt-Werte zwischen 0,90 und 0,91 im eNCT10Z und 0,90 bzw. 0,89 im eNCT18Z. Auch für den NCTA konnten bei rtt-Werten von 0,94 im Patientenkollektiv und 0,88 in der Gruppe der Kontrollprobanden hohe Reliabilitätskoeffizienten nachgewiesen werden.

Schwierigkeitsgrad der Parallelversionen

Zur Veranschaulichung sind die Testergebnisse im eNCT und NCTA mittels Fehlerbalken und 95%igem KI in Abb. 9 für Patienten und in Abb. 10 für Kontrollprobanden dargestellt. Statistische Unterschiede innerhalb den einzelnen Parallelversionen sind visuell gekennzeichnet.

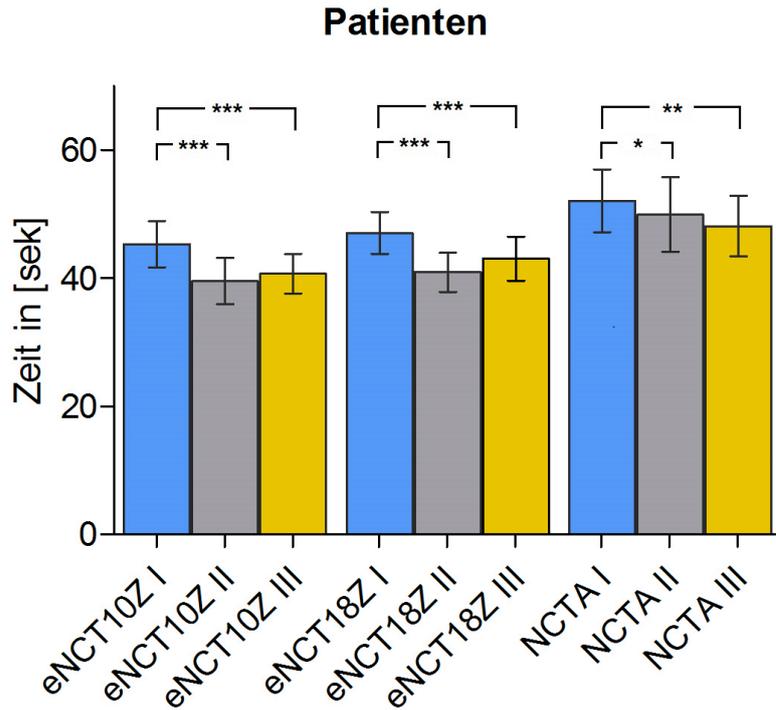


Abb. 9 Testzeiten der Parallelversionen für eNCT und NCTA in der Patientenkohorte dargestellt mit Hilfe von Fehlerbalken und 95%igem KI. eNCT10/18Z elektronischer Zahlenverbindungstest 10/18Zoll; AveNCT10/18Z Average Konzept 10/18 Zoll; I,II,III Parallelversion I,II,III; NCTA Zahlenverbindungstest A; KI Konfidenzintervall; * $p \leq 0,05$; ** $p \leq 0,01$; *** $p \leq 0,001$.

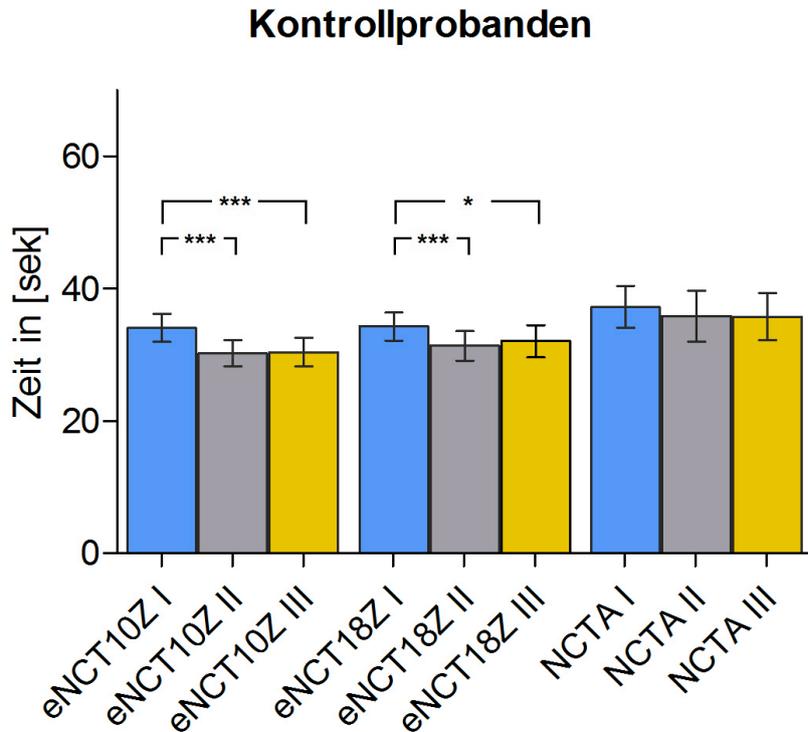


Abb. 10 Testzeiten in den Parallelversionen für eNCT und NCTA in der Kontrollkohorte dargestellt mit Hilfe von Fehlerbalken und 95%igem KI. eNCT10/18Z elektronischer Zahlenverbindungstest 10/18Zoll; AveNCT10/18Z Average Konzept 10/18 Zoll; I,II,III Parallelversion I,II,III; NCTA Zahlenverbindungstest A; KI Konfidenzintervall; * $p \leq 0,05$; ** $p \leq 0,01$; *** $p \leq 0,001$.

Im eNCT10Z zeigten sowohl Patienten ($\chi^2(2,n=86)=23,7$; $p<0,001$) als auch Kontrollprobanden ($\chi^2(2,n=61)=26,39$; $p<0,001$) einen Gesamtunterschied hinsichtlich der Testperformanz innerhalb der absolvierten Parallelversionen. Die daraufhin implementierten Post-hoc Tests zeigten für beide Kohorten Unterschiede in den Testzeiten zwischen eNCT10Z I und eNCT10Z II sowie zwischen eNCT10Z I und eNCT10Z III, während zwischen eNCT10Z II und eNCT10Z III keine Unterschiede aufgezeigt werden konnten.

Auch im eNCT18Z ergab sich ein Gesamtunterschied in den Testzeiten der absolvierten Parallelversionen für Patienten ($\chi^2(2,n=85)=31,76$; $p<0,001$) und Kontrollprobanden ($\chi^2(2,n=59)=15,39$; $p<0,001$). Hier zeigten Post-hoc Tests Unterschiede in der Testperformanz zwischen Parallelversion I und II sowie zwischen Parallelversion I und III, nicht aber zwischen Parallelversion II und III.

In der Analyse des NCTA zeigte sich ein Gesamtunterschied in den Testzeiten für sowohl Patienten ($\chi^2(2,n=86)=11,31$; $p<0,004$) als auch Kontrollprobanden ($\chi^2(2,n=61)=6,16$; $p<0,46$).

Im paarweisen Vergleich der Testzeiten konnte im Patientenkollektiv ein Unterschied zwischen den Parallelversionen NCTA I und NCTA II sowie zwischen NCTA I und NCTA III festgestellt werden. In den Parallelversionen NCTA II und NCTA III sahen wir keinen Unterschied. In der Gruppe gesunder Kontrollprobanden zeigten Implementierte Post-hoc Tests keine Unterschiede in der Testperformanz zwischen den einzelnen Parallelversionen des NCTA.

3.3.3 Lerneffekte

Um Lerneffekte im eNCT und NCTA zu untersuchen, wurden die Testzeiten von Test und Retest miteinander verglichen. Die Ergebnisse von Patienten (n=14) und Kontrollprobanden (n=18) sind in Tab. 7 dargestellt.

Tab. 7 Testergebnisse von Test und Retest im eNCT und NCTA.						
	PAT			KON		
	Test	Retest	U	Test	Retest	U
	M (SD)	M (SD)		M (SD)	M (SD)	
eNCT10Z I	37,2 (8)	39,1 (14)	-0,16	31,8 (8)	30,2 (9)	-0,92
AveNCT10Z	35,6 (8)	36,1 (10)	-0,14	29,2 (7)	27,4 (7)	-1,97*
eNCT18Z I	41,3 (11)	41,7 (12)	-0,21	32,8 (8)	31,1 (10)	-1,30
AveNCT18Z	39,7 (10)	37,8 (11)	-1,22	31,5 (8)	28,7 (8)	-2,58*
NCTA I	48,4 (20)	42,5 (15)	-1,19	35,4 (14)	33,3 (15)	-1,14

PAT Patientengruppe KON Kontrollprobanden; M Mittelwert; SD Standard-abweichung; U Wilcoxon-Vorzeichenrangtest; eNCT10/18Z elektronischer Zahlen-verbindungstest 10/18Zoll; AveNCT10/18Z Average Konzept 10/18 Zoll; I Parallelversion I; AveNCT18Z Average Konzept 18 Zoll; NCTA Zahlenverbindungstest A; *p≤0,05.

Im Vergleich der Testzeiten im eNCT10Z I ließen sich weder für Patienten noch für Kontrollen Unterschiede in der Testzeit zwischen beiden Messzeitpunkten nachweisen. Im AveNCT10Z fanden wir lediglich Unterschiede innerhalb der Kontrollkohorte (p=0,05) zwischen Test und Retest.

Im eNCT18Z I war weder für Patienten noch für Kontrollprobanden ein Unterschied in den Testzeiten zu beobachten. Der Vergleich der Testzeiten im AveNCT18Z ergab ausschließlich Unterschiede in den Testzeiten der Kontrollkohorte (p=0,01).

Für den NCTA I waren weder in der untersuchten Patientengruppe noch in der Kontrollgruppe Unterschiede im Vergleich der Testzeiten zwischen Test und Retest zu detektieren.

3.4 Validität

3.4.1 Konstruktvalidität

Korrelationen zwischen eNCT-Performanz und konstruktnahen bzw. konstruktfernen Merkmalen innerhalb der Patientenkohorte sind in Tab. 8 dargestellt.

Tab. 8 Konvergente und divergente Validität des eNCT gemäß Spearman Korrelationskoeffizient (SKK).					
		eNCT10Z I	AveNCTA10Z	eNCT18Z I	AveNCT18Z
Konvergente Validität					
NCTA	SKK	0,74***	0,76***	0,71***	0,74***
	n		86		85
NCTB	SKK	0,77***	0,80***	0,70***	0,73***
	n		83		82
ZST	SKK	-0,65***	-0,69***	-0,72***	-0,74***
	n		81		80
LNT-Z	SKK	0,25*	0,28*	0,32**	0,24*
	n		83		82
LNT-F	SKK	0,13	0,16	0,23	0,26
	n		86		85
KP	SKK	0,43***	0,50***	0,48***	0,49***
	n		83		82
Divergente Validität					
BMI	SKK	0,08	0,09	0,09	0,13
	n		86		85
Beschäftigungs- verhältnis	SKK	-0,08	-0,05	-0,10	-0,13
	n		86		86
Art der postschulischen Ausbildung	SKK	0,09	0,09	0,04	0,12
	n		86		85

n Anzahl; eNCT10/18Z elektronischer Zahlenverbindungstest 10/18Zoll; AveNCT10/18Z Average Konzept 10/18 Zoll; I Parallelversion I; NCTA Zahlenverbindungstest A; NCTB Zahlenverbindungstest B; ZST Zahlen-Symbol-Test; LNT-Z Zeit im Linien-Nachfahr-Test; LNT-F Fehler im Linien-Nachfahr-Test; KP Kreise-Punktieren; *p≤0,05; **p≤0,01; ***p≤0,001.

Die zur Analyse der konvergenten Validität verwendeten Einzeltests des PSE-Syndrom-Tests korrelierten mit Ausnahme des LNT gut bis sehr gut mit der eNCT-Performanz. Zwischen dem Merkmal Testzeit im eNCT und den Einzeltests NCTA, NCTB und ZST ergaben sich Wertebereiche von $r=0,65$ bis $r=0,80$ ($p \leq 0,001$).

Zwischen eNCT-Performanz und KP ließen sich zudem Korrelationen im Wertebereich von $r=0,43$ und $r=0,50$ ($p \leq 0,001$) nachweisen.

Hinsichtlich des Messergebnisses LNT-Z ließen sich zwischen eNCT und LNT allenfalls schwache Zusammenhänge in der Auswertung mit Werten zwischen $r=0,24$ und $r=0,32$ ($p \leq 0,05$) aufzeigen, während zum LNT-F keine Korrelationen bestanden.

Bei der Analyse der divergenten Validität zeigten die Variablen BMI, Beschäftigungsverhältnis und Art der postschulischen Ausbildung keine Zusammenhänge mit den eNCT-Testergebnissen. Für alle getesteten Variablen lagen die Spearman Korrelationskoeffizienten nahe Null.

3.4.2 Kriteriumsvalidität

Bivariate Korrelation zwischen dem Merkmal Testzeit im eNCT und der Diagnose mHE sind in Tab. 9 dargestellt.

Tab. 9 Kriteriumsvalidität des eNCT gemäß Spearman Korrelationskoeffizient (SKK).

		eNCT10Z I	AveNCT10Z	eNCTA18 I	AveNCT18Z
mHE-	SKK	0,66***	0,69***	0,67***	0,69***
Diagnose	n	82		81	

eNCT10/18Z elektronischer Zahlenverbindungstest 10/18Zoll; AveNCT10/18Z Average Konzept 10/18 Zoll; I Parallelversion I; SKK Spearman Korrelationskoeffizient; n Anzahl; mHE minimale Hepatische Enzephalopathie; * $p \leq 0,05$; ** $p \leq 0,01$; *** $p \leq 0,001$.

Das zur Beurteilung der kriterienbezogenen Validität verwendete Kriterium mHE-Diagnose zeigte eine gute Korrelation mit der eNCT-Performanz ($p \leq 0,001$). Die Werte variierten dabei zwischen $r=0,66$ und $r=0,69$.

3.5 Nebengütekriterien

143 Studienteilnehmer (Patienten n=84; Kontrollprobanden n=59) gaben ein schriftliches Feedback bezüglich der eigenen Erfahrung in der Benutzung elektronischer Geräte hinsichtlich Zumutbarkeit und favorisierter Bildschirmgröße (siehe Anhang 8.).

Testfairness

Die Gruppe der Kontrollprobanden (66%; n=39) besaß mehr Erfahrung in der Verwendung elektronischer Geräte (Smartphone/Tablet-Computer) als die Gruppe der Patienten (43%; n=36; siehe Abb. 11). Zwischen beiden Kohorten gab es signifikante Unterschiede hinsichtlich der Verteilung (Pearson- $\chi^2(1) = 7,51$; $p = 0,006$; $n = 143$).

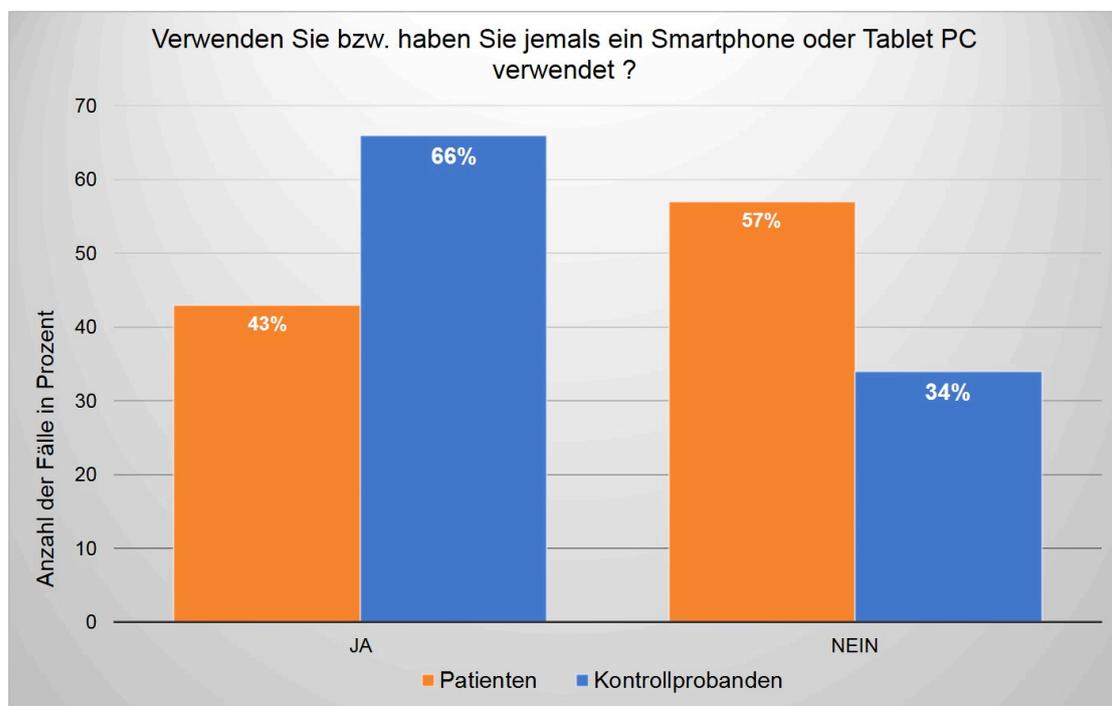


Abb. 11 Erfahrung in der Verwendung elektronischer Geräte.

Zumutbarkeit und favorisierte Displaygröße

Die Kohorte der Patienten bevorzugte mit 45% (n=38) die Anwendung des 10 Zoll Tablets gegenüber dem 18 Zoll Tablet mit 31% (n=26). Für 11% der Patienten spielte die Größe des Bildschirms keine Rolle (n=9). Auch die Mehrheit der Kontrollgruppe präferierte mit 53% (n=31) das 10 Zoll Tablet. Die Anwendung des 18 Zoll Tablets bevorzugten 22% der Kontrollprobanden (n=13), während für 20% (n=12) die Anwendung beider Größen vorstellbar wäre. 10% der Studienteilnehmer (Patienten n=11; Kontrollprobanden n=3) beurteilten eine tägliche Anwendung des elektronischen Testverfahrens als nicht zumutbar (siehe Abb. 12).

Zwischen den Antworten beider Studienkohorten zeigten sich keine signifikanten Unterschiede (Pearson- $\chi^2(3) = 5,85$ p=0,119; n=143).

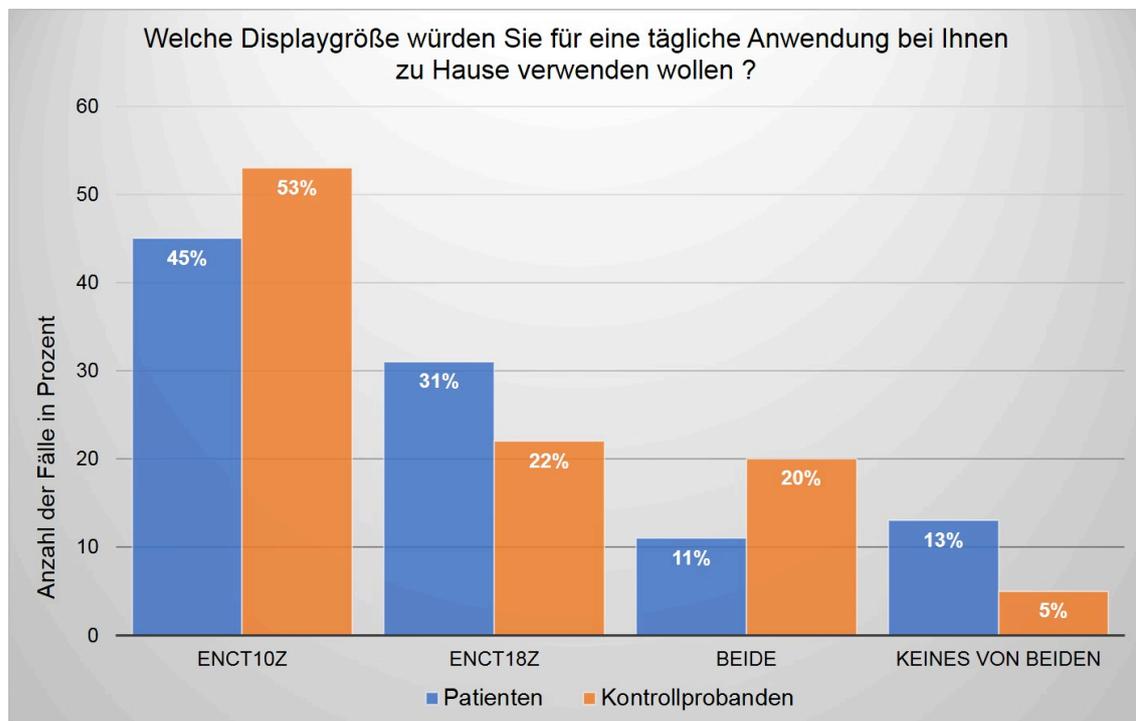


Abb. 12 Zumutbarkeit und favorisierte Displaygröße bei einer täglichen Anwendung des eNCT im heimischen Umfeld. eNCT10Z elektronischer Zahlenverbindungstest 10 Zoll; eNCT18Z elektronischer Zahlenverbindungstest 18 Zoll.

Unterschiede in der Handhabung zwischen den Displaygrößen

Für die Mehrzahl der Studienteilnehmer (57% der Kontrollprobanden n=33 und 55% der Patienten n=46) ergaben sich keine Unterschiede in der Handhabung unterschiedlich großer Displaygrößen. In der Patientenkohorte befanden 20% (n=17) den eNCT10Z sowie 24% (n=20) den eNCT18Z einfacher in der Handhabung. Diese Ergebnisse ähneln denen der Kontrollprobanden. Hier gaben 22% (n=13) an, auf dem 10 Zoll Tablet einfacher arbeiten zu können, während 21% (n=12) die Anwendung auf dem 18 Zoll als einfacher erachteten (siehe Abb. 13).

Zwischen beiden Kohorten gab es keine Unterschiede hinsichtlich der Verteilung (Pearson- $\chi^2(2)=0,25$; $p= 0,88$; $n=141$).

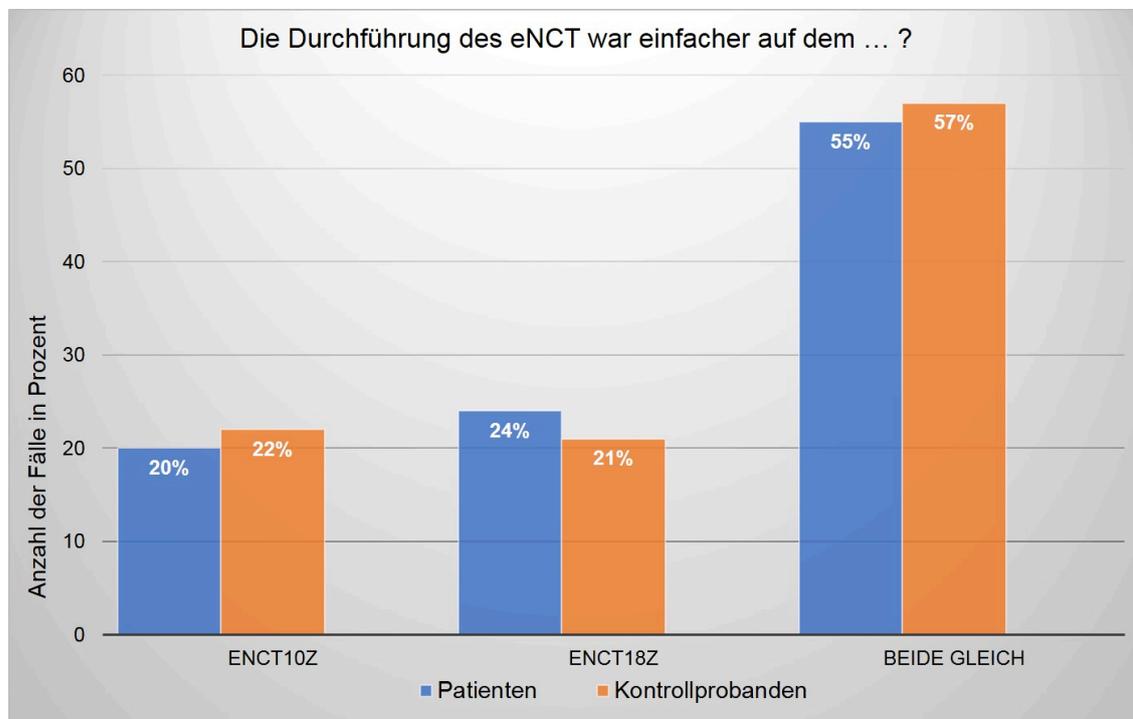


Abb. 13 Unterschiede in der Durchführung des eNCT hinsichtlich Displaygröße. eNCT10Z elektronischer Zahlenverbindungstest 10 Zoll; eNCT18Z elektronischer Zahlenverbindungstest 18 Zoll.

3.6 Einflussfaktoren auf die eNCT-Performanz

3.6.1 Alter, Geschlecht und Bildungsgrad

Korrelationsanalysen

Um mögliche Prädiktoren auf die Performanz im eNCT zu ermitteln, wurde der Zusammenhang zwischen dem Kriterium Testzeit im eNCT und den Merkmalen Alter, Geschlecht und Bildungsgrad innerhalb der Gruppe lebergesunder Kontrollprobanden untersucht. Die Ergebnisse der bivariaten Korrelationen sind in Tab. 10 dargestellt. Hier zeigte die abhängige Variable Testzeit im eNCT eine positive Korrelation zum Faktor Alter und eine negative Korrelation zum Faktor Bildungsgrad, wenngleich diese nicht über alle Auswertungskonzepte hinweg auftraten. Zwischen eNCT-Performanz und dem Merkmal Geschlecht konnte keinen Zusammenhang gefunden werden.

Tab. 10 Korrelationen zwischen eNCT-Performanz und den Faktoren Alter, Geschlecht und Bildungsgrad an gesunden Kontrollprobanden.

		eNCT10Z I	AveNCTA10Z	eNCT18Z I	AveNCT18Z
Alter	SKK	0,53***	0,57***	0,58***	0,56***
	n		61	59	
Geschlecht	SKK	0,07	0,14	0,08	0,16
	n		61	59	
Bildungsgrad	SKK	-0,28*	-0,28*	-0,16	-0,21
	n		61	59	

SKK Spearman Korrelationskoeffizient; n Anzahl; eNCT10/18Z elektronischer Zahlenverbindungstest 10/18Zoll; AveNCT10/18Z Average Konzept 10/18 Zoll; I Parallelversion I; * $p \leq 0,05$; ** $p \leq 0,01$; *** $p \leq 0,001$.

Lineare Regression

Um unabhängige Prädiktoren auf die eNCT-Performanz zu identifizieren, wurden die Merkmale Alter und Bildungsgrad in einem gemeinsamen Regressionsmodell analysiert. Die Ergebnisse dieser Analyse sind in Tab. 11 dargestellt.

Innerhalb der schrittweisen Regression wurden die Merkmale Schulgrade Grad 2 und 3 aus dem Modell ausgeschlossen. Im F-Test der ANOVA zeigten die Regressionsmodelle eine Signifikanz in ihrer Gesamtheit ($p \leq 0,05$).

Tab. 11 Ergebnisse der multiplen Regression nach Testzeiten im eNCT bei Kontrollprobanden. Angegeben sind die Betawerte der Prädiktoren im Regressionsmodell, das Bestimmtheitsmaß (r^2), die Effektstärke (f^2), die statistische Power und der F-Wert der ANOVA.

	n	r^2	Alter	Schulgrad 1	f^2	Power	F-Wert
eNCT10Z I	60	0,42	0,28**	9,63***	0,73	1	23,01
AveNCT10Z	60	0,44	0,29**	8,15**	0,79	1	24.57
eNCT18Z I	58	0,38	0,40***	4,91*	0,62	1	18.92
AveNCT18Z	58	0,42	0,31**	8,20**	0,71	1	21.59

n Anzahl der Patienten; eNCT10/18Z elektronischer Zahlenverbindungstest 10/18Zoll; AveNCT10/18Z Average Konzept 10/18 Zoll; I Parallelversion I; * $p \leq 0,05$; ** $p \leq 0,01$; *** $p \leq 0,001$.

Die Variablen Alter und der Schulgrad 1 konnten hingegen als unabhängige Prädiktoren auf die eNCT-Performanz identifiziert werden. Steigt das Lebensalter um ein Jahr so erhöht sich die Testzeit im eNCT je nach Auswertungskonzept zwischen 0,28 bis 0,40 Sekunden. Insofern generieren jüngere Probanden geringere Testzeiten als Ältere. Probanden mit Hauptschulabschluss bzw. neun Jahren Schulunterricht benötigen für die Absolvierung des eNCT zwischen 4,91 und 9,63 Sekunden länger als Kontrollprobanden mit Schulgrad 2 oder 3. Der lineare Zusammenhang zwischen eNCT-Performanz, Alter und Schulgrad 1 wird durch folgende Regressionsgleichung beschrieben. Hierbei stellt das Merkmal Schulgrad 1 eine binäre Variable dar (ja=1/nein=0).

$$\text{eNCT10Z I [sek]} = 16,25 + 0,28 \times \text{Alter} + 9,63 \times \text{Schulgrad 1}$$

$$\text{AveNCT10Z [sek]} = 13,58 + 0,29 \times \text{Alter} + 8,15 \times \text{Schulgrad 1}$$

$$\text{eNCT18Z I [sek]} = 10,31 + 0,40 \times \text{Alter} + 4,91 \times \text{Schulgrad 1}$$

$$\text{AveNCT18Z [sek]} = 13,33 + 0,31 \times \text{Alter} + 8,20 \times \text{Schulgrad 1}$$

Zusammenfassend konnten zwischen 38% und 44% der Gesamtstreuung in der eNCT-Performanz durch die beiden unabhängigen Variablen erklärt werden. Die Effektstärke f^2 nach Cohen lag je nach Auswertungskonzept zwischen 0,62 und 0,79 bei einer statistischen Power des Modells von 1.

Der lineare Zusammenhang zwischen der abhängigen Variable Testzeit im eNCT und den Einflussfaktoren Alter und Bildungsgrad ist exemplarisch am Konzept eNCT10Z I in einem Streudiagramm mit Regressionsgerade dargestellt (siehe Abb. 14 und 15).

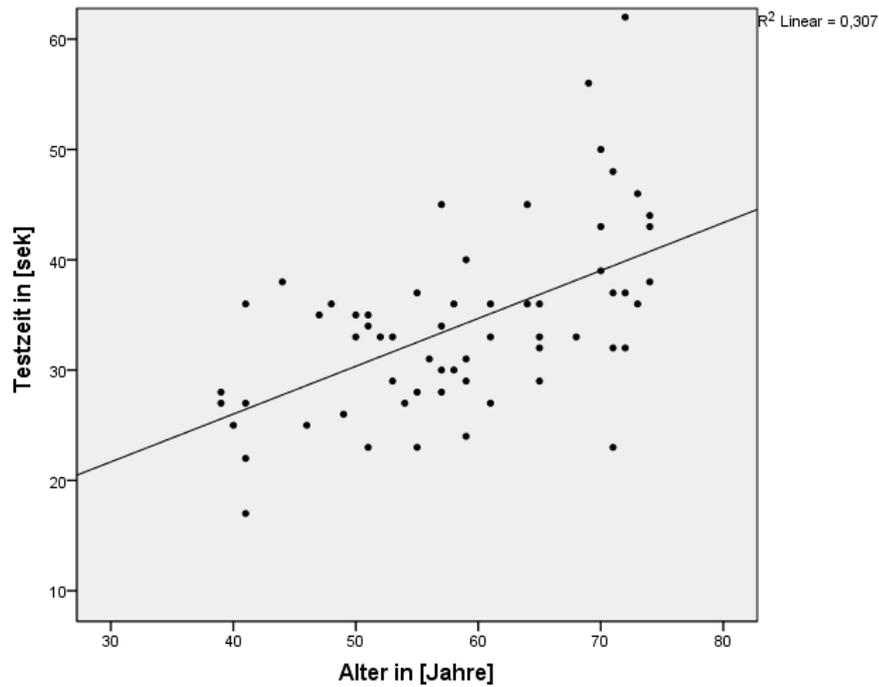


Abb. 14 Streudiagramm mit Regressionsgerade und Regressionsgleichung für den Zusammenhang zwischen Testzeit im eNCT10Z I und Alter. eNCT10Z elektronischer Zahlenverbindungstest 10 Zoll; I Parallelversion I; r^2 Bestimmtheitsmaß.

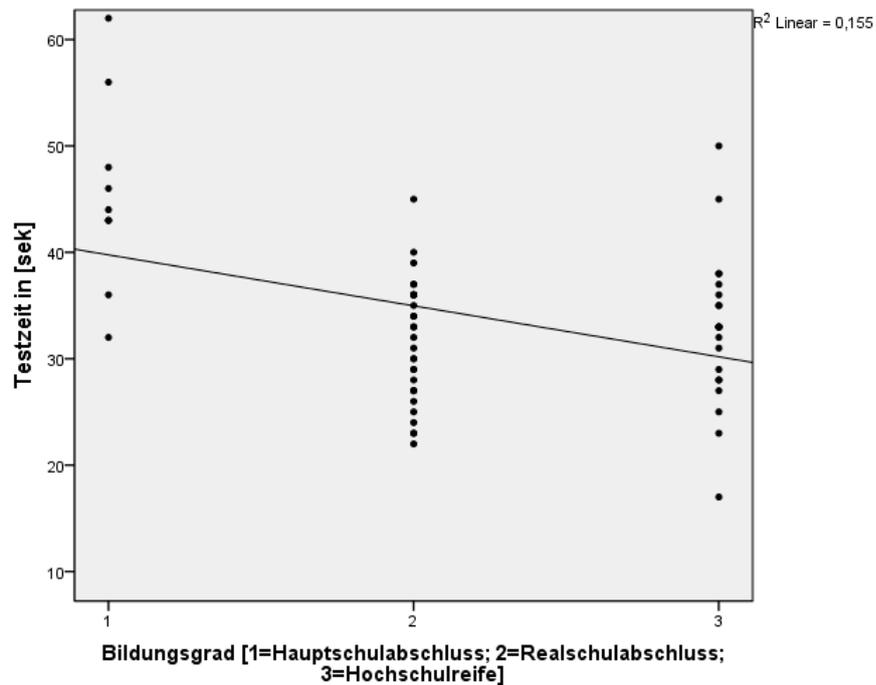


Abb. 15 Streudiagramm mit Regressionsgerade und Regressionsgleichung für den Zusammenhang zwischen Testzeit im eNCT10Z I und Bildungsgrad. eNCT10Z elektronischer Zahlenverbindungstest 10 Zoll; I Parallelversion I; r^2 Bestimmtheitsmaß.

3.6.2 Weitere eine Leberzirrhose charakterisierende Faktoren

Korrelationsanalysen

Innerhalb der Patientengruppe wurde die eNCT-Performanz auf mögliche Einflussfaktoren untersucht. Die Ergebnisse der bivariaten Korrelationen zwischen der abhängigen Variable Testzeit im eNCT und den untersuchten unabhängigen Variablen sind in Tab. 12 und 13 dargestellt.

Tab. 12 Korrelationen zwischen eNCT-Performanz und möglichen Einflussfaktoren innerhalb der Patientenkohorte.

		eNCT10Z I	AveNCTA10Z	eNCT18Z I	AveNCT18Z
BMI [kg/m ²]	SKK	-0,08	-0,05	-0,10	-0,13
	n		86	85	
MMSE-Score	SKK	-0,38***	-0,33**	-0,23*	-0,29**
	n		86	85	
CHILD-PUGH-Score	SKK	0,34**	0,28*	0,39***	0,31**
	n		79	78	
MELD-Score	SKK	0,35***	0,30**	0,35***	0,31**
	n		83	82	
HE wirksame Medikation [ja/nein]	SKK	0,26*	0,30**	0,23*	0,27*
	n		86	85	
Diuretische Therapie [ja/nein]	SKK	-0,19	-0,17	-0,13	-0,12
	n		84	83	
Ösophagusvarizen [ja/nein]	SKK	-0,10	-0,08	-0,11	-0,14
	n		86	85	
Z.n. SBP [ja/nein]	SKK	-0,24*	-0,24*	-0,11	-0,21
	n		85	84	
TIPPS [ja/nein]	SKK	0,09	0,01	0,04	-0,08
	n		62	61	

eNCT10/18Z elektronischer Zahlenverbindungstest 10/18Zoll; AveNCT10/18Z Average Konzept 10/18 Zoll; I Parallelversion I; MMSE Mini Mental State Examination; MELD Model of End Stage Liver Disease; SKK Spearman Korrelationskoeffizient; n Anzahl; HE Hepatische Enzephalopathie; HE wirksame Medikation (Lactulose, Rifaximin oder Ornithinaspartat); Diuretische Therapie (Furosemid, Torasemid, Spironolacton); SBP Spontan bakterielle Peritonitis; TIPPS transjugulärer intrahepatischer portosystemischer Shunt. *p≤0,05; **p≤0,01; ***p≤0,001.

Tab. 13 Korrelationen zwischen eNCT-Performanz und laborchemischen Blutparametern innerhalb der Patientenkohorte.

		eNCT10Z I	AveNCTA10Z	eNCT18Z I	AveNCT18Z
Natrium [mmol/l]	SKK	-0,16	-0,13	-0,22	-0,18
	n		83		82
Kalium [mmol/l]	SKK	0,16	0,17	0,06	0,16
	n		83		82
Hämoglobin [mg/dl]	SKK	-0,24*	-0,19	-0,19	-0,16
	n		85		84
Thrombozyten [μl]	SKK	-0,05	-0,05	-0,14	-0,08
	n		86		85
Kreatinin [mg/dl]	SKK	0,23	0,25	0,11	0,08
	n		85		84
Bilirubin [mg/dl]	SKK	0,17	0,15	0,26*	0,19
	n		85		84
ASAT [U/l]	SKK	0,06	0,02	0,15	0,08
	n		85		84
ALAT [U/l]	SKK	-0,11	0,13	0,01	0,01
	n		86		85
G-GT [U/l]	SKK	-0,12	-0,02	0,07	0,05
	n		86		85
Albumin [mg/dl]	SKK	-0,14	-0,11	-0,22*	-0,15
	n		81		80
AP [U/l]	SKK	0,13	0,13	0,12	0,08
	n		84		83
INR	SKK	0,32**	0,24*	0,33**	0,31**
	n		83		82
Ammoniak [mg/dl]	SKK	0,11	0,14	0,12	0,19
	n		76		75

eNCT10/18Z elektronischer Zahlenverbindungstest 10/18Zoll; AveNCT10/18Z Average Konzept 10/18 Zoll; I Parallelversion I; SKK Spearman Korrelationskoeffizient; n Anzahl; ASAT Aspartat-Aminotransferase; ALT Alanin-Aminotransferase; G-GT Gamma-glutamyltransferase; AP Alkalische Phosphatase; INR International Normalized Ratio; *p≤0,05; **p≤0,01; ***p≤0,001.

In der Analyse der laborchemischen Blutparameter konnte allein zwischen INR und eNCT-Performanz ein positiver Zusammenhang über alle Auswertungskonzepte hinweg detektiert werden. Weitere positive Korrelationen fanden sich zwischen Testzeit im eNCT und den Faktoren MELD-Score, sowie CHILD-PUGH-Score. Die Faktoren MMSE-Score, HE wirksame Medikation und Z.n. SBP zeigten hingegen einen negativen Zusammenhang mit der eNCT-Performanz.

Lineare Regression

Um unabhängige Prädiktoren auf die eNCT-Performanz zu identifizieren, wurden die Merkmale INR, MMSE-Score, MELD-Score, CHILD-PUGH-Score, HE wirksame Medikation und Z.n. SBP in einem gemeinsamen Regressionsmodell analysiert. Die Ergebnisse sind in Tab. 14 dargestellt. Innerhalb der schrittweisen Regression wurden die Merkmale INR, MELD-Score, CHILD-PUGH-Score, sowie Z.n. SBP aus dem Modell ausgeschlossen. Die Variablen MMSE-Score und HE wirksame Medikation konnten hingegen als unabhängige Prädiktoren auf die eNCT-Performanz identifiziert werden. Im F-Test der ANOVA zeigten die Regressionsmodelle eine Signifikanz in ihrer Gesamtheit ($p \leq 0,05$).

Tab. 14 Ergebnisse der multiplen Regression nach Testzeiten im eNCT bei Patienten. Angegeben sind die Betawerte der Prädiktoren im Regressionsmodell, das Bestimmtheitsmaß (r^2), die Effektstärke (f^2), die statistische Power und der F-Wert der ANOVA.

	n	r^2	MMSE Score	HE Medikation	f^2	Power	F-Wert
eNCT10Z I	86	0,24	-6,12***	10,53***	0,32	1	14,62
AveNCT10Z	86	0,23	-4,81***	10,38**	0,31	1	13,96
eNCT18Z I	85	0,12	-3,21*	7,92*	0,14	0,92	5,60
AveNCT18Z	85	0,16	-3,82**	8,52*	0,20	0,99	9,27

n Anzahl; HE Hepatische Enzephalopathie, HE wirksame Medikation (Lactulose, Rifaximin oder Ornithinaspartat); eNCT10/18Z elektronischer Zahlenverbindungstest 10/18Zoll; AveNCT10/18Z Average Konzept 10/18 Zoll; I Parallelversion I; * $p \leq 0,05$; ** $p \leq 0,01$; *** $p \leq 0,001$.

Patienten absolvieren den eNCT schneller, wenn sie einen höheren MMSE-Score generieren. Die Testzeit verringert sich im eNCT je nach Auswertungskonzept zwischen 3,2 und 6,1 Sekunden pro erreichtem Punkt im MMSE-Score. Weiterhin generieren Patienten mit HE wirksamer Medikation langsamere Testzeiten im eNCT als Patienten ohne diese. Hier erhöht sich die Testzeit im eNCT zwischen 7,9 bis 10,5 Sekunden bei Medikamenteneinnahme. Der lineare Zusammenhang zwischen Testzeit im eNCT, MMSE-Score und HE wirksamer Medikation wird durch folgende Regressionsgleichung beschrieben. Hierbei stellt das Merkmal HE Medikation eine binäre Variable dar (ja=1/nein=0).

$$\mathbf{eNCT10Z\ I\ [sek] = 239,43 - 6,12 \times MMSE\text{-}Score + 10,53 \times HE\ Medikation}$$

$$\mathbf{AveNCT10Z\ [sek] = 197,89 - 4,81 \times MMSE\text{-}Score + 10,38 \times HE\ Medikation}$$

$$\mathbf{eNCT18Z\ I\ [sek] = 152,85 - 3,21 \times MMSE\text{-}Score + 7,92 \times HE\ Medikation}$$

$$\mathbf{AveNCT18Z\ [sek] = 167,93 - 3,82 \times MMSE\text{-}Score + 8,52 \times HE\ Medikation}$$

Ferner konnten zwischen 12% und 24% der Gesamtstreuung in der eNCT-Performanz durch die beiden unabhängigen Variablen erklärt werden, während die Effektstärke f^2 nach Cohen zwischen 0,14 und 0,32. Die statistische Power des Modells betrug über alle Auswertungskonzepte hinweg mindestens 0,92.

Der lineare Zusammenhang zwischen der abhängigen Variable Testzeit im eNCT und den Einflussfaktoren MMSE-Score und HE wirksame Medikation ist exemplarisch am Konzept eNCT10Z I in einem Streudiagramm mit Regressionsgerade dargestellt (siehe Abb. 16 und 17).

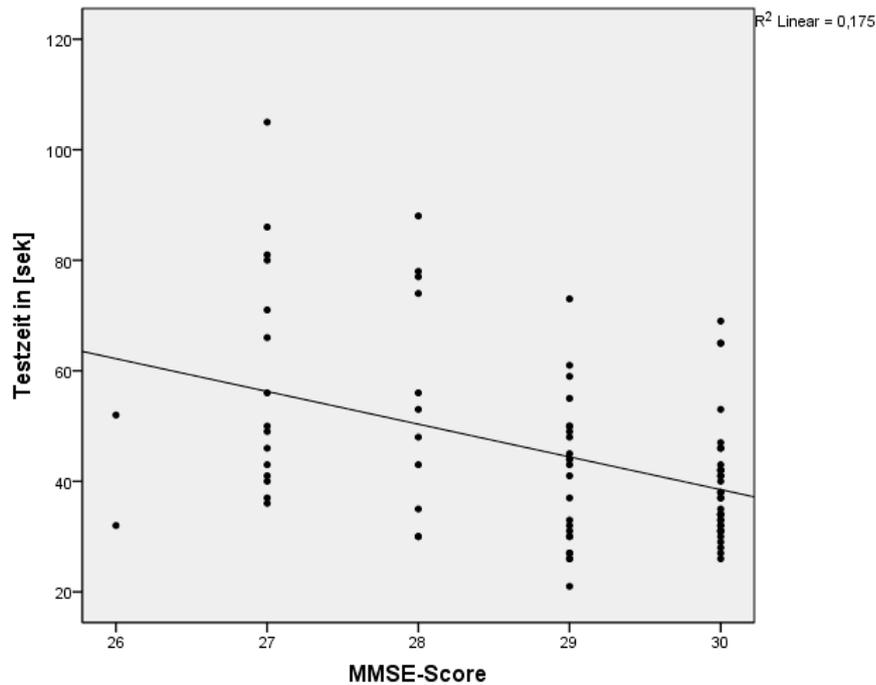


Abb. 16 Streudiagramm mit Regressionsgerade und Regressionsgleichung für den Zusammenhang zwischen Testzeit im eNCT10Z I und MMSE-Score. eNCT10Z elektronischer Zahlenverbindungstest 10 Zoll; I Parallelversion I; MMSE Mini Mental State Examination; r^2 Bestimmtheitsmaß.

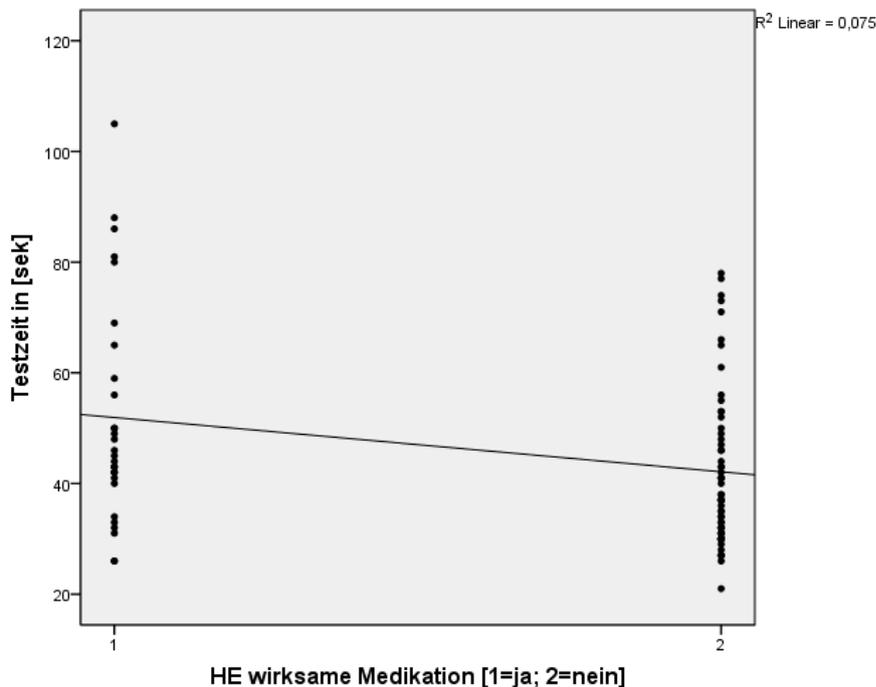


Abb. 17 Streudiagramm mit Regressionsgerade und Regressionsgleichung für den Zusammenhang zwischen Testzeit im eNCT10Z I und HE Medikation. eNCT10Z elektronischer Zahlenverbindungstest 10 Zoll; I Parallelversion I; HE Hepatische Enzephalopathie; HE wirksame Medikation (Lactulose, Rifaximin oder Ornithinaspartat); r^2 Bestimmtheitsmaß.

3.6.3 Bildschirmdiagonale (Tabletgröße)

Die durchschnittliche Übereinstimmung der Testzeit zwischen beiden Displaygrößen ist mittels Bland-Altman-Diagramm in Abb. 18 dargestellt.

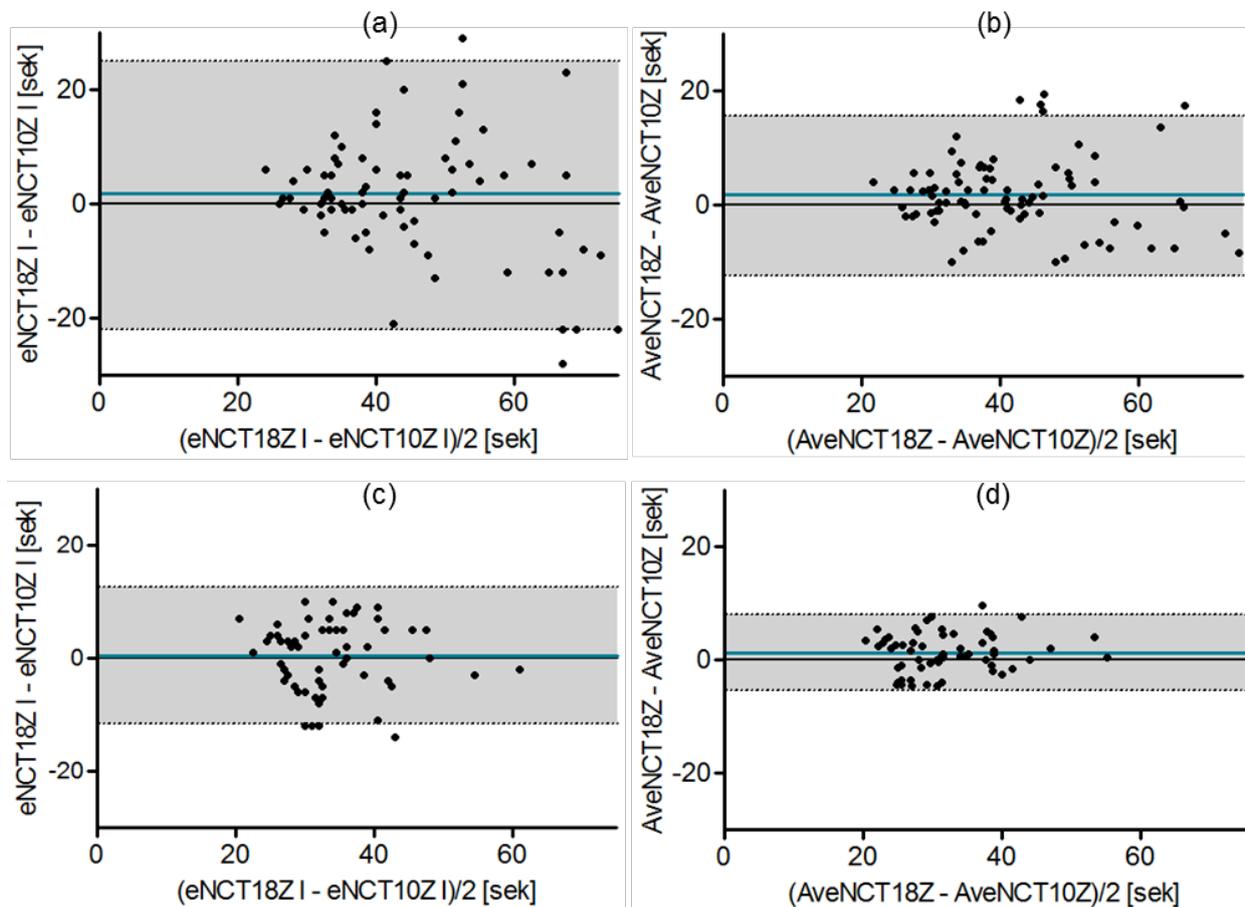


Abb. 18 Bland-Altman-Diagramme illustrieren den Grad der Übereinstimmung zwischen Testzeiten im eNCT innerhalb unterschiedlich großer Displays. Vergleich der Testwerte nach Parallelversion I (a) und Average Konzept (b) im Patientenkollektiv, sowie nach Parallelversion I (c) und Average Konzept (d) in der Gruppe der Kontrollprobanden. Bias (blaue Linie) und Grenzen der Übereinstimmung (graues Band). eNCT10/18Z elektronischer Zahlenverbindungstest 10/18Zoll; I Parallelversion; AveNCT10/18Z Average Konzept 10/18 Zoll.

Die Studienteilnehmer absolvierten den eNCT auf dem 10 Zoll Display schneller als auf dem 18 Zoll Display. Innerhalb der Patientenkohorte lag die durchschnittliche Übereinstimmung der Messergebnisse (Bias) für eNCT I und AveNCT bei 1,7 Sekunden, während diese im Kollektiv der Kontrollprobanden 0,4 Sekunden (eNCT I), respektive 1,3 Sekunden (AveNCT) betrug. Die Streuung der individuellen Messwertdifferenzen war im Average Konzept mit 14,3 Sekunden im Kollektiv der Patienten und mit 6,9 Sekunden in der Kontrollkohorte niedriger als im Konzept eNCT I. Hier betrug die zweifache Standardabweichung des Bias im Patientenkollektiv 24 Sekunden und in der Kontrollgruppe 12,4 Sekunden.

4 Diskussion

4.1 Zusammenfassung der Ergebnisse

In der vorliegenden Arbeit untersuchten wir Qualität und Vorhersagekraft des eNCT auf die Detektion einer kognitiven Beeinträchtigung am Beispiel der mHE bei Patienten mit Leberzirrhose. Die Betrachtung erfolgte dabei standardisiert anhand der Gütekriterien psychodiagnostischer Testverfahren: Objektivität, Reliabilität und Validität.

Das Vorhandensein einer kognitiven Beeinträchtigung wurde anhand der Testzeit ermittelt. Hierfür absolvierten Patienten mit manifester Leberzirrhose sowie lebergesunde Kontrollprobanden neben dem eNCT eine in der mHE Diagnostik als Referenzstandard geltende psychometrische Testbatterie (PSE-Syndrom-Test).

Um die methodische Bewertung zu verbessern und Benachteiligungen durch eine zu kleine Bildschirmgröße zu überprüfen, erfolgte die Testbearbeitung des eNCT wiederholt auf einem 10 Zoll Tablet (eNCT10Z) und einem 18 Zoll Tablet (eNCT18Z).

Mit Hilfe eines standardisierten Fragebogens wurde zudem die Beurteilung von Testfairness, Zumutbarkeit und favorisierter Bildschirmgröße ermöglicht.

Ergänzend überprüften wir, ob die demographischen Variablen Alter, Bildungsgrad und Geschlecht sowie darüberhinausgehende, eine Leberzirrhose charakterisierende, Faktoren Einfluss auf die Performanz im eNCT ausüben. Die Erhebung der genannten Parameter erfolgte in einem ausführlichen Anamnesegespräch, während laborchemische Leberfunktionsparameter durch eine zum Erstuntersuchungstermin durchgeführte Blutentnahme erhoben wurden.

Die Hauptergebnisse stellen sich wie folgt dar: (i) Patienten mit Leberzirrhose generieren im Unterschied zu gesunden Kontrollprobanden verlängerte Testzeiten. (ii) Hohe Reliabilitätkoeffizienten in Re- und Paralleltest bescheinigen dem eNCT eine gute Zuverlässigkeit. (iii) Durch hohe Korrelationen sowohl zu konstruktnahen Einzeltests des PSE-Syndrom-Tests als auch zum Diagnosekriterium mHE waren wir in der Lage, eine gute Validität des eNCT zu demonstrieren. (iv) In der Analyse möglicher Einflussfaktoren konnten die Faktoren Alter, Bildungsgrad, MMSE-Score, sowie eine medikamentöse HE wirksame Therapie als unabhängige Prädiktoren auf die Testzeit im eNCT identifiziert werden.

Zusammenfassend stellt der eNCT eine einfache, schnelle und zuverlässige Tablet-basierte Testmethode dar, mit dem Potential, neuropsychische Störungen bei erwachsenen Patienten mit manifester Leberzirrhose zu quantifizieren. Zudem ist für den Großteil der Studienteilnehmer eine regelmäßige Durchführung im häuslichen Umfeld vorstellbar. Durch die Möglichkeit einer telemedizinischen Anwendung bietet der eNCT optimale Voraussetzungen als unterstützende Funktion in der HE-Diagnostik und Therapieoptimierung für chronisch Lebererkrankte im ambulanten Setting.

4.2 Gütekriterien

4.2.1 Objektivität

Die Durchführungsobjektivität des eNCT kann als hoch angesehen werden, da dieses Testverfahren standardisierte Instruktionen sowie definierte Performanzkriterien ohne individuellen Spielraum zur Durchführung vorgibt. Nach einmaliger Einweisung durch geschultes Fachpersonal besteht die Möglichkeit der wiederholten Testbearbeitung auch ohne die Anwesenheit eines Versuchsleiters im Sinne der Selbsttestung (82). Äußere Störfaktoren lassen sich so auf ein Minimum reduzieren, da Motivation und Compliance der Testperson nicht selten vom Untersucher in unterschiedlichem Maße beeinflusst werden können (83, 84).

Die Auswertung der Testergebnisse erfolgte anhand der auf dem elektronischen Gerät gespeicherten Rohtestwerte in Sekunden. Da es sich bei den eNCT-Testergebnissen um reine Zeitmaße handelt, sind Fehler in der Auswertung als sehr gering zu erachten. Die Auswertung des Average Konzepts erfolgte in dieser Arbeit nachträglich durch eine manuelle Ermittlung der durchschnittlichen Testzeit aus allen Parallelversionen. Grundsätzlich kann die Auswertungsobjektivität des eNCT als hoch angesehen werden.

Die Testrohwerte bedürfen einer Normierung, um sie im Kontext eines diagnostischen Testverfahrens aussagekräftig interpretieren zu können. Nach Generierung von alterskorrigierten und bildungskorrigierten Normwerten an einem ausreichend großen Kollektiv könnten Testergebnisse in Zukunft einfach und personenunabhängig interpretiert werden.

4.2.2 Reliabilität

Nach Weissenborn (83) sollen für die adäquate Diagnostik einer mHE wiederholte Messungen bei klinisch stabilem Status ähnliche Testergebnisse generieren. Die Autorin fordert daher die Überprüfung der Retestreliabilität bei der Entwicklung neuer Testverfahren. Die Beurteilung der Reliabilitätskoeffizienten erfolgte in dieser Arbeit nach der Empfehlung von Lienert & Raatz (85), bei der standardisierte Tests eine Re- oder Paralleltestreliabilität von $r \geq 0,8$ aufweisen sollten.

Die erhobenen IKK bescheinigen dem eNCT eine gute Retestreliabilität und lagen innerhalb beider Studienkollektive im Bereich der IKK Werte des NCTA.

Für beide Displaygrößen ließ sich die Retestreliabilität mit Hilfe des Average Konzepts verbessern. Unterstützung findet dieses Ergebnis durch die Aussage von Held & Kollegen. Die Autoren propagieren, dass Mittelwerte wiederholter Messungen eine bessere Reliabilität als einzelne Messwerte besitzen (86).

Die Bereitstellung von Parallelversionen ist eine notwendige Voraussetzung, um Lerneffekte bei Verlaufsuntersuchungen und Therapiestudien ausschließen zu können (87). Im Vergleich der Reliabilitätskoeffizienten zwischen eNCT und NCTA mittels Methode nach Cronbach's alpha ergaben sich allenfalls diskrete Unterschiede zwischen den untersuchten Testmethoden. Die erhobenen rtt-Werte bescheinigen dem eNCT eine gute bis exzellente Paralleltestreliabilität.

Bezüglich der Reproduzierbarkeit von Testergebnissen beschreiben Schomerus et al. (55) in ihrem PSE-Syndrom-Test Manual mögliche Lern- oder Übungseffekte bei Verlaufsuntersuchungen mit der gleichen Testversion und somit eine Verzerrung des Testresultates. In der vorliegenden Arbeit ließen sich Übungseffekte im eNCT zwischen Test und Retest zwar nur innerhalb der Kontrollgruppe für das Average Konzept nachweisen, allerdings verbesserten sich die Studienteilnehmer in allen untersuchten Testverfahren in der Testzeit zwischen erster und zweiter Messung. Dieses Ergebnis unterstreicht die Bedeutung des Gebrauchs paralleler Testversionen mit ähnlichem Schwierigkeitsgrad als eine wichtige Voraussetzung für die Minderung von Lern- oder Übungseffekten innerhalb einer wiederholten Testbearbeitung. Erste Parallelversionen für den NCTA wurden bereits 1977 entwickelt (88) und nachfolgend auf Lerneffekte untersucht. Innerhalb der drei für den eNCT entwickelten Parallelversionen ließen sich Unterschiede im Schwierigkeitsgrad für sowohl Patienten als auch Kontrollen

detektieren. So absolvierten beide Kohorten die zweite und dritte Parallelversion schneller als die erste Parallelversion, während zwischen zweiter und dritter Parallelversion kein Unterschied in der Testzeit nachgewiesen werden konnte.

Hier stellt sich die Frage, ob dieses Resultat allein durch einen unterschiedlichen Schwierigkeitsgrad in den drei Parallelversionen zu erklären ist oder ob Übungseffekte das Ergebnis verzerren. Ein Blick auf die Analyse der Parallelversionen des NCTA liefert ähnliche Resultate. Sowohl für Patienten als auch für die Kontrollprobanden konnten wir einen Gesamtunterschied in der Testzeit der Parallelversionen nachweisen. Jedoch zeigten implementierte Post-hoc Tests nur im Kollektiv der Patienten Unterschiede zwischen den einzelnen Parallelversionen.

Für Schranz & Kollegen (87) ist die Kenntnis über wiederholungsbedingte Übungseffekte wesentlich zur Beurteilung von Testergebnissen psychometrischer Leistungstests, da Leistungssteigerungen bei wiederholter Testbearbeitung eher auf Übungseffekte zurückzuführen sind, als dass es sich um eine tatsächliche kognitive Besserung handelt. In ihrer Studie über Übungseffekte bei computergestützten psychologischen Leistungstests traten die größten Übungsgewinne bei der ersten Wiederholung auf, während die geringsten zwischen der dritten und vierten Wiederholung lagen. Vor diesem Hintergrund eignet sich das Average Konzept um künftige wiederholungsbedingte Übungseffekte zu begrenzen, wenn Parallelversionen mit ähnlichem Schwierigkeitsgrad vorliegen.

4.2.3 Validität

Nach Moosbrugger und Kelava (73) handelt es sich bei der Validität hinsichtlich der Testpraxis um das wichtigste Gütekriterium, da Objektivität und Reliabilität zwar eine hohe Messgenauigkeit ermöglichen, aber nur die günstigen Voraussetzungen für das Erreichen einer hohen Validität erbringen. Die Überprüfung der Validität des eNCT erfolgte in der vorliegenden Arbeit mit Hilfe von zwei Validierungsmethoden: der Konstrukt- und der Kriteriumsvalidität.

Die Beurteilung der Korrelationskoeffizienten erfolgte in Anlehnung an Nachtigall und Wirz (89). Die beiden Autoren empfehlen Korrelationen von $r \leq 0,20$ für konstruktfere Variablen und mindestens $r = 0,50$ bis $r = 0,70$ für konstruktnahe Variablen.

Innerhalb der Testvalidierung bescheinigen dem eNCT hohe Korrelationskoeffizienten zu den konstruktnahen Testverfahren NCTA, NCTB, ZST und KP eine hohe konvergente Validität.

Im Vergleich der Auswertungskonzepte bot das Average Konzept zu allen konstrukt-nahen Einzeltests (NCTA, NCTB, ZST, DST) bessere Korrelationen als die Auswertung der Testzeit nach Parallelversion I. Niedrige Korrelationskoeffizienten zu den untersuchten konstruktfernen Variablen BMI, Beschäftigungsverhältnis und Art der postschulischen Ausbildung hingegen sprechen für eine hohe divergente Validität des eNCT. Anzumerken gilt, dass der LNT hierbei eine Ausnahme bot und nur eine sehr schwache bivariate Korrelation zur eNCT-Performanz beobachtet werden konnte. Unterstützung findet dieses Ergebnis durch den Vergleich der im Studienprotokoll festgelegten Tests zwischen Patienten und Kontrollprobanden. Hier zeigte sich einzig im LNT-F ein fehlender Unterschied im Testergebnis zwischen gesunder und erkrankter Kohorte, während sich die Testergebnisse in allen anderen untersuchten Testverfahren zwischen beiden Gruppen unterschieden. Erwähnt sei, dass die Faktoren Alter, Geschlecht und Bildungsgrad als mögliche Einflussfaktoren auf den PSE-Syndrom-Test gelten und somit nicht als konstruktferne Kriterien sondern separat in Bezug auf die eNCT-Performanz untersucht wurden (siehe 3.6 Einflussfaktoren auf die eNCT-Performanz) (43).

Für die Prüfung der Kriteriumsvalidität korrelierten wir die Testzeit im eNCT mit dem Kriterium mHE. Die Diagnose einer mHE wurde innerhalb des Patientenkollektivs zuverlässig und valide anhand des als Referenzstandard geltenden PSE-Syndrom-Tests gestellt. Hochpositive Korrelationskoeffizienten sprechen hierbei für eine gute Kriteriumsvalidität des eNCT hinsichtlich der Detektion einer mHE. Im Vergleich der Auswertungskonzepte bot das Average Konzept dabei bessere Korrelationen zum Diagnosekriterium mHE als das Konzept eNCT I.

Zusammenfassend bescheinigen dem eNCT beide Validierungsmethoden eine hohe Validität und ermöglichen eine Interpretation der Messergebnisse hinsichtlich der Detektion von neuropsychischen Störungen bei erwachsenen Patienten mit manifester Leberzirrhose. Erwähnt sei zudem, dass sich sowohl die konvergente, als auch die kriterienbezogene Validität des eNCT mittels Average Konzept verbessern lässt.

4.2.4 Nebengütekriterien

Der eNCT konnte von Personen mit unterschiedlichstem Alters- und Bildungsgrad bearbeitet werden. Auf eine Übersetzung des Testverfahrens in eine andere Sprache konnte zudem, aufgrund der verwendeten arabischen Ziffern, verzichtet werden. Einzige Voraussetzung war die Beherrschung des Zählens von eins bis 25. Somit bot das Testverfahren allen Studienteilnehmern die gleichen Chancen, unabhängig der ethnischen, soziokulturellen oder geschlechterspezifischen Gruppenzugehörigkeit. Obwohl die Gruppe der Kontrollen mehr Erfahrung in der Verwendung von elektronischen Geräten besaß, stellte dies für die Patientenkohorte dank Übungsaufgabe und standardisierter Instruktion kein Hindernis in der Testbearbeitung dar. Insgesamt konnten alle Studienteilnehmer den eNCT erfolgreich absolvieren. Das Kriterium der Testfairness kann insofern als hoch angesehen werden.

Die Frage der favorisierten Displaygröße in der täglichen Anwendung zielte auf die Zumutbarkeit des eNCT im Sinne einer wiederholten Testbearbeitung des eNCT im häuslichen Umfeld ab. Hier beurteilten 90% der befragten Studienteilnehmer eine tägliche Heimanwendung des eNCT als akzeptabel. Im Kontext des klinischen Nutzens der frühzeitigen Detektion einer HE-Episode spricht dieses Resultat neben einer geringen psychischen, physischen und zeitlichen Belastung des Probanden für eine hohe Akzeptanz und Zumutbarkeit des Testverfahrens. Zudem müsste für eine klinische Nützlichkeit die Testung nicht zwangsläufig jeden Tag, sondern kann nach initialer Testung und Diagnostik in Abhängigkeit der klinischen Symptomatik intervallweise genutzt werden.

Weiterhin bietet der eNCT als digitales Testverfahren die Möglichkeit einer kostenlosen und einfachen Installation auf Tablet oder Notebook mit Touchfunktion. Nach erstmaliger Testbearbeitung mit einem dafür geschulten Testleiter besteht die Aussicht, den eNCT als Patientenselbsttest zu etablieren und folglich auf die Anwesenheit eines Testleiters zu verzichten. Die Bearbeitungszeit der drei Parallelversionen variiert je nach Leistungsfähigkeit der Testperson zwischen einer bis fünf Minuten. Im Vergleich zu den bereits etablierten Testverfahren in der Diagnostik einer mHE kann die Ressource Bearbeitungszeit als besonders günstig angesehen werden. Hier werden für den PSE-Syndrom-Test Zeiten von etwa 10-20 Minuten benötigt (55). In Zusammenschau der Ergebnisse kann dem eNCT eine respektable Testökonomie bescheinigt werden.

4.3 Einflussfaktoren auf die eNCT-Performanz

Eine zentrale Forderung an psychometrische Testverfahren in der Diagnostik einer mHE sind das Vorhandensein von Normwerten in der zu untersuchenden Bevölkerungsgruppe (27, 40). Um potentielle Einflussfaktoren auf den eNCT zu identifizieren und eine künftige Normierung zu ermöglichen, wurden verschiedene Parameter herangezogen und innerhalb des Kontroll- und Patientenkollektivs mittels einer linearen Mehrfachregression separiert untersucht. Schneider und Kollegen (90) beschreiben die lineare Regression als eine bedeutsame statistische Methode in der medizinischen Forschung, um Zusammenhänge zwischen verschiedenen Faktoren zu analysieren und prognostisch wichtigen Risikofaktoren zu identifizieren. Die Autoren fordern zugleich eine korrekte methodische Anwendung und Interpretation der gewonnenen Studienresultate. Um dieser Forderung gerecht zu werden, wurde auf eine Angabe und Interpretation der regressionsanalytischen Ergebnisse hinsichtlich der Faktoren Kausalität, fehlender Werte, Variablenselektion und Fallzahlplanung beziehungsweise Modellpower geachtet. Die Beurteilung der Bedeutsamkeit des Einflussfaktors erfolgte anhand der Effektstärke nach Cohen (78). Dabei entspricht $f^2=0,02$ einem schwachen, $f^2=0,15$ einem mittleren und $f^2=0,35$ einem starken Effekt.

In der Gruppe gesunder Kontrollprobanden konnte demonstriert werden, dass kein Zusammenhang zwischen Geschlecht und eNCT-Performanz besteht. Im Unterschied übten sowohl Alter als auch Bildungsgrad einen Einfluss auf die Testzeit im eNCT aus. Steigt das Alter um ein Jahr, so erhöht sich die Testzeit um 0,3 bis 0,4 Sekunden. Hinzu kommt, dass in Abhängigkeit des Auswertungskonzeptes Personen mit einem Hauptschulabschluss bzw. neun Jahren Schulunterricht den eNCT zwischen 4,9 und 9,6 Sekunden langsamer absolvieren als Personen mit einem höheren Bildungsgrad. Die beiden weiteren gemäß des Deutschen Schul- und Bildungssystems definierten Schulgrade besaßen keinen prädiktiven Wert für die Testzeit im eNCT. Das korrigierte r^2 als Maß der Güte des Regressionsmodells lag zwischen 0,38 und 0,44. Demzufolge können über 38% der Gesamtstreuung in der eNCT-Performanz durch die Variablen Alter und Bildungsgrad erklärt werden. Die Effektstärke f^2 nach Cohen lag über alle Auswertungskonzepte hinweg bei $\geq 0,62$. Dies entspricht einem starken Effekt. Die Analyse der Power der Regressionsmodelle deutet bei einem Wert von 1 für einen ausreichend großen Stichprobenumfang, während innerhalb der untersuchten Kohorte keine Beobachtung aufgrund fehlender Werte aus der Regressionsanalyse ausgeschlossen wurde.

Hinsichtlich der inhaltlichen Kausalität deckt sich unser Resultat mit denen vergangener Studienergebnisse. So beschreiben Boll und Reitan (91) eine Altersabhängigkeit der Performanz im Trail-Making-Test, dem Vorläufer des NCTA, während Weissenborn und Kollegen (68) in ihrer Studie aufzeigen konnten, dass bei Annahme eines festen Zeitlimits im NCTA 10% der Patienten (vor allem junge Patienten) falsch negativ klassifiziert werden, obwohl bei Studieneinschluss bereits manifeste Zeichen einer HE vorlagen. Im PSE-Syndrom-Test wurde interessanterweise auf eine Normwerterstellung für den Bildungsgrad verzichtet, obwohl ein signifikanter Einfluss der Variable Schulbildung auf die Performanz der in der Testbatterie enthaltenen Einzeltests dargestellt werden konnte (87). Grund dafür war zum einen die vernachlässigbar schwache Verschiebung der Testresultate in Abhängigkeit des Schulgrades und zum anderen die international unterschiedlichen Bildungssysteme, die einen Vergleich des Bildungsgrades erschwerten. In Anbetracht dieser Problematik empfehlen die Autoren eine individuelle Berücksichtigung des Einflusses der Schulbildung und eine Reanalyse der Ergebnisse im PSE-Syndrom-Test für Grenzfälle (87).

Unter den darüber hinaus untersuchten Faktoren zeigten ausschließlich der Merkmale medikamentöse HE wirksame Therapie und MMSE-Score einen Einfluss auf die Leistungsfähigkeit im eNCT. Erwartungsgemäß absolvieren Patienten den eNCT schneller, wenn sie einen höheren Punktwert in der MMSE erreichen. In Abhängigkeit des Auswertungskonzepts verringerte sich dabei die Testzeit im eNCT um 3,2 bis 6,1 Sekunden pro erreichtem Punkt im MMSE-Score.

Patienten mit medikamentöser HE-Prophylaxe absolvierten den eNCT langsamer als diejenigen ohne HE wirksame Medikation. Hier erhöht sich die Testzeit zwischen 7,9 und 10,5 Sekunden in Abhängigkeit von Displaygröße und Auswertungskonzept bei medikamentöser HE Prophylaxe. Zwischen 12% und 24% der Gesamtstreuung in der eNCT-Performanz werden durch die Variablen MMSE-Score und HE wirksame Medikation erklärt. Bei Effektstärken f^2 von 0,14-0,20 im eNCT18Z und 0,31-0,31 im eNCT10Z ist von einem mittleren bis starken Effekt der prädiktiven Variablen MMSE-Score und HE wirksame Medikation auf die Testzeit im eNCT auszugehen.

Innerhalb der Poweranalyse der linearen Regressionsmodelle sprechen Werte von $>0,9$ für einen ausreichend großen Stichprobenumfang. Auch in diesen Modellen wurde keine Beobachtung aufgrund fehlender Werte aus der Regressionsanalyse ausgeschlossen.

Die von uns demonstrierte Beziehung zwischen eNCT und MMSE findet indirekt Unterstützung durch die 2014 veröffentlichte Studie von Corrias & Kollegen (92), in welcher die Validität der MMSE bei der Detektion einer mHE untersucht wurde. Die Arbeitsgruppe kommt zu der Schlussfolgerung, dass ein MMSE-Score von < 26 Punkten ältere Patienten mit einer Leberzirrhose identifiziert, die zu einer Manifestation HE bedingter Symptome neigen. Die Autoren waren somit imstande eine inhaltliche Kausalität zwischen MMSE-Score und neuropsychischer Störung bei Lebererkrankten nachzuweisen. Die beschriebenen neuropsychischen Defizite lassen sich quantitativ mit dem eNCT erfassen. Auch die kausale Beziehung zwischen eNCT-Performanz und HE wirksamer Medikation erscheint plausibel. Innerhalb der Patientengruppe mit einer HE wirksamen Therapie ist aufgrund eines fortgeschrittenen Krankheitsstadiums die Wahrscheinlichkeit erwartungsgemäß höher, trotz medikamentöser Prophylaxe Defizite in kognitiven Teilbereichen zu entwickeln, als bei Patienten ohne Vergangenheit einer HE. Interessanterweise konnten wir keinen Zusammenhang zwischen eNCT-Performanz und dem zur Diagnose einer klinisch manifesten HE oft verwendeten Ammoniak-Blutspiegel demonstrieren. Gründe hierfür könnten neben einer überschätzten Bedeutung der Blutammoniumkonzentration in der Diagnostik einer HE auch die fehlende Berücksichtigung von Patienten mit einer oHE im Rahmen der vorliegenden Arbeit sein (40). Innerhalb dieser Patientenkohorte ist erwartungsgemäß mit höheren Ammoniakspiegeln zu rechnen, da nach aktuellen Empfehlungen die Diagnose einer oHE bei normaler Ammoniumkonzentration in Frage gestellt werden sollte (26).

Für die Merkmale, MELD-Score, INR, CHILD-PUGH-Score, sowie Z.n. SBP konnten wir eine signifikante Beziehung zur Testzeit im eNCT allein in den bivariaten Einzelanalysen nachweisen. Innerhalb der schrittweisen Selektion der multiplen Regressionsanalyse ließ sich kein signifikanter Einfluss dieser Merkmale auf die eNCT-Performanz darstellen, so dass ein Ausschluss aus der weiterführenden Modellanalyse erfolgte.

Die Übereinstimmung der Testergebnisse zwischen eNCT10Z und eNCT18Z reflektiert die Abhängigkeit der eNCT-Performanz von der Displaygröße. Hier ließen sich mithilfe der Bland-Altman-Methode beide Messmethoden visuell vergleichen und eine Verzerrung sowie Streuung der individuellen Messwertdifferenzen berücksichtigen.

Diese einfache grafische Methode berücksichtigt sowohl die durchschnittliche Übereinstimmung der Messergebnisse (Bias) als auch die Streuung der individuellen Messwertdifferenzen und hat sich als Verfahren der Wahl zum Vergleich von Messmethoden etabliert (93). Der im Diagramm grau hinterlegte Bereich stellt dabei die obere und untere Übereinstimmungsgrenze dar. In diesem Bereich befinden sich bei hinreichend symmetrischer Verteilung 95% der Messwerte (93). Vereinfacht formuliert ist die Übereinstimmung zweier Methoden entsprechend besser, je kleiner der Bias und desto enger die Übereinstimmungsgrenzen liegen.

Die dargestellten Bland-Altman-Diagramme zeigen eine gute Übereinstimmung der eNCT-Performanz auf den unterschiedlichen Bildschirmgrößen. Hier konnten wir zeitliche Vorteile für die kleinere Bildschirmgröße ermitteln, wenngleich der Unterschied innerhalb beider Kohorten mit 0,4 respektive 1,7 Sekunden für klein zu erachten ist. Bezüglich der Messwertdifferenzen zeigten Patienten im Vergleich zu Kontrollprobanden erwartungsgemäß eine größere Streuung der Daten. Interessanterweise konnte mittels Average Konzept die Streuung für beide Studienkollektive nahezu halbiert und somit die Übereinstimmung zwischen beiden Tabletgrößen deutlich verbessert werden.

Der Großteil der Studienteilnehmer beurteilte im Fragebogen zum Testfeedback ein Fehlen von Unterschieden in der Handhabung der unterschiedlichen Displaygrößen. Da die Mehrheit der Patienten und Kontrollprobanden zudem das 10 Zoll Tablet präferierte, wäre eine Veröffentlichung und klinische Etablierung des eNCT10Z nachvollziehbar.

4.4 Limitationen

Die dargestellten Ergebnisse müssen hinsichtlich verschiedener Einschränkungen interpretiert werden. Zunächst gilt es eine kritische Auseinandersetzung mit der Testmethodik und Statistik zu führen.

Die in der Arbeit verwendeten Testvorlagen des PSE-Syndrom-Tests waren frei zugänglich und zeigten geringfügige Unterschiede in Größe und Maßstab zu den Originalvorlagen des PSE-Syndrom-Test. Die Modifikationen in den verwendeten Testvorlagen wurden bei den jeweiligen Untertests genauer beschrieben und die eingesetzte Version visuell dargestellt. Ebenso wurde die Testreihenfolge im Vergleich zum Manuell des PSE-Syndrom-Tests aus methodischen Gründen variiert (siehe

Abschnitt 2.3.1.1 Testanweisung und -bearbeitung). Diese methodische Verzerrung muss bei der Interpretation der Ergebnisse berücksichtigt werden. Um die in dieser Arbeit gewonnenen Erkenntnisse zu stützen, sollte in künftigen konfirmatorischen Studien auf die Originalversion des PSE-Syndrom-Tests zurückgegriffen werden.

Auf eine Bestimmung der inhaltlichen Validität musste in dieser Dissertationsschrift verzichtet werden, da bisher keine Beurteilung des neu entwickelten Testverfahrens durch einen Expertenkonsens stattgefunden hat. Weiterhin ist eine grundsätzliche Bedingung für Validität die Existenz eines gültigen und anerkannten Goldstandards. Bis zum heutigen Zeitpunkt konnte kein einheitlicher Goldstandard in der mHE Diagnostik etabliert werden. Insofern erfolgte ein Vergleich der Testzeiten im eNCT mit den Testergebnissen des PSE-Syndrom-Tests. Dieser wird von vielen Autoren als Referenzstandard in der Diagnostik einer mHE angesehen (19, 27, 58, 60). Weiterhin muss der Umstand betrachtet werden, dass wir für die Bestimmung der kriterienbezogenen Validität den eNCT mit der aus dem PHES abgeleiteten Diagnose mHE (ja/nein) korrelierten. Der eNCT wurde in Anlehnung an den NCTA entworfen. Letzteres bedingt in Teilen die Punktezahl im finalen PHES. Insofern gilt es eine mögliche Verzerrung der Korrelation zwischen eNCT-Performanz und der Diagnose mHE zu berücksichtigen.

In der vorliegenden, hypothesengenerierenden Arbeit wurde der kritische p-Wert mit einem α -Level von 5% angesetzt. Diesbezüglich wurde bewusst sparsam mit dem Wort „signifikant“ umgegangen und die p-Werte bzw. p-Wertbereiche wurden gezielt angegeben, um eine fehlende Korrektur bei multiplem Testen zu berücksichtigen. Gleichwohl gilt es zu beachten, dass eine Kontrolle der Irrtumswahrscheinlichkeit fehlt und zufällige Ergebnisse nicht auszuschließen sind. Insofern fordern sowohl Bender und Kollegen (79) als auch die Arbeitsgruppe um Victor & Kollegen (80) eine Bestätigung der explorativen Testresultate in weiteren gezielten Studien.

Bezüglich des Studiendesigns gilt es zu bedenken, dass für die Ermittlung der Retestrelabilität nur ambulante getestete Studienteilnehmer mit stabilem Krankheitsverlauf eingeschlossen wurden. Die daraus resultierende geringe Fallzahl von 18 Kontrollprobanden und 14 Patienten ist bei der Bewertung der Reliabilität zu beachten und sollte innerhalb einer nachfolgenden Studienplanung mit einer höheren

Fallzahl berücksichtigt werden. Moosbrugger und Kelava (73) geben in ihrem Kapitel „Qualitätsanforderungen an einen psychologischen Test“ zu bedenken, dass die ermittelten Korrelationen abhängig vom Zeitintervall zwischen Test und Retest variieren können und folglich die Reliabilität beeinflussen. Als Gründe hierfür werden neben einer tatsächlichen Veränderung des zu testenden Merkmals insbesondere Übungs- und Erinnerungseffekte genannt. Leider war es aus organisatorischer Sicht nicht möglich, die zeitlichen Abstände zwischen beiden Testungen konstant zu halten. Hier lag das kürzeste Zeitintervall bei vier Tagen, während das längste Intervall bei drei Monaten lag.

4.5 Klinischer Nutzen und Ausblick

In Anbetracht der demographischen Entwicklung mit einer älter werdenden Bevölkerung bei erniedrigter Geburtenrate stehen wir vor einer großen gesundheitspolitischen Herausforderung (94). Die breitere Nutzung telemedizinischer Verfahren könnte neben einer Anhebung der therapeutischen Qualität zu einer voranschreitenden Enthospitalisierung führen. Indem Krankenhausaufenthalte reduziert, Versorgungslücken geschlossen und Patienten besser im häuslichen Umfeld versorgt werden, könnten langfristig Kosten und Ressourcen für das Gesundheitssystem geschont, sowie Compliance und Lebensqualität von chronisch Lebererkrankten verbessert werden (94, 95).

„Das perfekte Testverfahren, das in kurzer Zeit, mit geringem Aufwand und minimaler Einarbeitungszeit sehr differenzierte Ergebnisse liefert, gibt es leider nicht. Vielmehr müssen Kliniker und Forscher wohlüberlegt entscheiden, welches Testverfahren für die jeweilige Aufgabenstellung am besten geeignet ist“ (71).

Die vorliegende Dissertationsschrift stellt als sogenanntes „Proof of Concept“ eine Grundlage für künftige Studien bzw. Arbeiten zum eNCT dar und soll den Weg für eine Etablierung des eNCT in den klinischen Alltag ebnen. In dieser Arbeit konnte der Nachweis erbracht werden, dass der eNCT imstande ist, kognitive Dysfunktionen bei Patienten mit manifester Leberzirrhose aufzudecken und dabei die erforderlichen Qualitätsanforderungen an einen psychometrischen Test zu erfüllen. Dank des digitalen Designs, des einfach konzipierten Aufbaus sowie einer schnellen Bearbeitungs- und Auswertungszeit besteht die Möglichkeit der telemedizinischen Anwendung im Kontext einer häuslich verwendeten Monitoring-Plattform für Patienten mit einer chronischen Lebererkrankung.

Hinsichtlich der Diagnostik einer mHE gilt für die Anwendung des papierbasierten und bereits etablierten NCTA die allgemeine Expertenmeinung, dass alters- und bildungskorrigierte Testwerte innerhalb der zu untersuchenden Bevölkerungsgruppe vorliegen und dass die Kombination des NCTA mit mindestens einem weiteren psychometrischen Einzeltest erfolgen sollte (27, 28, 40). In Anlehnung an diesen Konsens stellen Wuensch und Kollegen (82) die Hypothese auf, dass der eNCT ähnlich dem KP eine starke motorische Komponente bezüglich der Geschwindigkeit besitzt. Insofern sollte der eNCT weniger als eine simple, digitale Kopie des NCTA, sondern vielmehr als eine Erweiterung des bestehenden Testverfahrens verstanden werden, welcher die untersuchten psychomotorischen Bereiche im NCTA um die Geschwindigkeitskomponente des KP ergänzt. Als möglichen Hinweis dafür beschreiben die Autoren eine höhere Sensitivität des eNCT gegenüber der separaten Testung mittels NCTA oder KP im Hinblick auf die Diagnose einer mHE (82).

Zusammenfassend ist eine klinische Anwendung des eNCT in Diagnostik, Therapiekontrolle und Verlaufsbeobachtung von chronisch Lebererkrankten denkbar, wenngleich diese Messmethode zur alleinigen Diagnosesicherung einer mHE nicht ausreicht, sondern bei Verdacht auf eine Erkrankung um die Testung mittels Referenzstandard ergänzt werden sollte (26, 82). Grundlage für eine künftige Interpretation der Messweltergebnisse bildet in erster Linie die Festlegung von geeigneten Cutpoint-Werten, um Spezifität bzw. Sensitivität zu ermitteln und den eNCT im Kontext mit etablierten Methoden vergleichen zu können. Ferner wird die Erstellung von alters- und bildungskorrigierten Normwerten an einem ausreichend großen und repräsentativen Kollektiv vorausgesetzt. Über die Berechnung der durchschnittlichen Bearbeitungszeit der drei Parallelversionen lassen sich sowohl Reliabilität, als auch Validität des eNCT verbessern. Eine Optimierung der Auswertungsobjektivität ließe sich erreichen, indem die Generierung der Testresultate des Average Konzepts anstatt durch die manuelle Berechnung des Testleiters automatisiert innerhalb des verwendeten Devices stattfinden würde. Um den Schwierigkeitsgrad der drei untersuchten Parallelversionen des eNCT besser vergleichen zu können wäre es hilfreich, die einzelnen Parallelversionen gegeneinander in einem ausreichend großen zeitlichen Abstand an gesunden Kontrollprobanden zu testen.

Schlussendlich gilt es, diese Bedingungen bei der Planung und Durchführung von prospektiven, confirmatorischen Studien zu berücksichtigen, um den eNCT langfristig in Forschung und Klinik konstituieren zu können.

5 Literaturverzeichnis

1. Schuppan D, Afdhal NH. Liver cirrhosis. *The Lancet*. 2008;371(9615):838-51.
2. Wiegand J, Berg T. The etiology, diagnosis and prevention of liver cirrhosis: part 1 of a series on liver cirrhosis. *Dtsch Arztebl Int*. 2013;110(6):85-91.
3. Forner A, Llovet JM, Bruix J. Hepatocellular carcinoma. *The Lancet*. 2012;379(9822):1245-55.
4. Niederau C. Epidemiologie der Leberzirrhose. *Hepatitis & more*. 2011;1:11-14.
5. 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.
6. Ascione A, Fontanella L, Imperato M, Rinaldi L, De Luca M. Mortality from liver cirrhosis and hepatocellular carcinoma in Western Europe over the last 40 years. *Liver Int*. 2017;37(8):1193-1201.
7. Statistisches Bundesamt. Todesursachenstatistik. Gesundheitsberichterstattung des Bundes 2011. (Accessed March 27, 2017, at https://www.destatis.de/GPStatistik/servlets/MCRfileNodeServlet/DEHeft_derivate_00010401/2120400117004.pdf;jsessionid=63E9248E8205002ABB17C3AAC75FE619)
8. 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.
9. Child C, Turcotte J. Surgery and portal hypertension. In: *The liver and portal hypertension*. Major Probl Clin Surg. 1964;1:1-85.
10. Kaplan DE, Dai F, Skanderson M, Aytaman A, Baytarian M, D'Addeo K, Fox R, Hunt K, Knott A, Mehta R, Pedrosa M, Pocha C, Valderrama A, Taddei T. Recalibrating the Child-Turcotte-Pugh Score to improve prediction of transplant-free survival in patients with cirrhosis. *Dig Dis Sci*. 2016;61(11):3309-20.
11. Pugh RNH, Murray-Lyon IM, Dawson JL, Pietroni MC, Williams R. Transection of the oesophagus for bleeding oesophageal varices. *British Journal of Surgery*. 1973;60(8):646-9.
12. Wiesner RH, McDiarmid SV, Kamath PS, Edwards EB, Malinchoc M, Kremers WK, Krom RAF, Kim WR. MELD and PELD: Application of survival models to liver allocation. *Liver Transplantation*. 2001;7(7):567-80.

13. Kamath PS, Wiesner RH, Malinchoc M, Kremers W, Therneau TM, Kosberg CL, D'Amico G, Dickson ER, Kim WR. A model to predict survival in patients with end-stage liver disease. *Hepatology*. 2001;33(2):464-70.
14. Das A, Dhiman RK, Saraswat VA, Verma M, Naik SR. Prevalence and natural history of subclinical hepatic encephalopathy in cirrhosis. *Journal of Gastroenterology and Hepatology*. 2001;16(5):531-35.
15. Poordad FF. Review article: the burden of hepatic encephalopathy. *Aliment Pharmacol Ther*. 2007;25 Suppl 1:3-9.
16. Jara M, Lock JF, Malinowski M, Jones D, Kiefer S, Neuhaus P, Stockmann M: D-Liver - a new approach for bridging therapy to transplantation of the chronic liver failure patient: 1661. *Transplantation*. 2012;94(10S):789.
17. Pascher A, Nebrig M, Neuhaus P. Irreversible liver failure: treatment by transplantation: part 3 of a series on liver cirrhosis. *Dtsch Arztebl Int*. 2013;110(10):167-73.
18. Zhan T, Stremmel W. The diagnosis and treatment of minimal hepatic encephalopathy. *Dtsch Arztebl Int*. 2012;109(10):180-87.
19. Weissenborn K. Portosystemic encephalopathy. *Handb Clin Neurol*. 2014;120:661-74.
20. Morgan MY, Blei A, Grungreiff K, Jalan R, Kircheis G, Marchesini G, Riggio O, Weissenborn K. The treatment of hepatic encephalopathy. *Metab Brain Dis*. 2007;22(3-4):389-405.
21. Li YY, Nie YQ, Sha WH, Zeng Z, Yang FY, Ping L, Jia L. Prevalence of subclinical hepatic encephalopathy in cirrhotic patients in China. *World Journal of Gastroenterology*. 2004;10(16):2397-401.
22. Kircheis G, Fleig WE, Görtelmeyer R, Grafe S, Häussinger D. Assessment of low-grade hepatic encephalopathy: A critical analysis. *Journal of Hepatology*. 2007;47(5):642-50.
23. Groeneweg M, Moerland W, Quero JC, Hop WCJ, Krabbe PF, Schalm SW. Screening of subclinical hepatic encephalopathy. *Journal of Hepatology*. 2000;32(5):748-53.
24. Saxena N, Bhatia M, Joshi YK, Tandon RK. Auditory P300 event-related potentials and number connection test for evaluation of subclinical hepatic encephalopathy in patients with cirrhosis of the liver: A follow-up study. *Journal of Gastroenterology and Hepatology*. 2001;16(3):322-7.

25. Sharma P, Sharma BC, Puri V, Sarin SK. Critical flicker frequency: Diagnostic tool for minimal hepatic encephalopathy. *Journal of Hepatology*. 2007;47(1):67-73.
26. 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.
27. Dhiman RK, Saraswat VA, Sharma BK, Sarin SK, Chawla YK, Butterworth R, Duseja A, Aggarwal R, Amarapurkar D, Sharma P, Madan K, Shah S, Seth AK, Gupta RK, Koshy A, Rai RR, Dilawari JB, Mishra SP, Acharya SK. Minimal hepatic encephalopathy: Consensus statement of a working party of the Indian National Association for Study of the Liver. *Journal of Gastroenterology and Hepatology*. 2010;25(6):1029-41.
28. 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.
29. 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.
30. Bajaj JS, Saeian K, Schubert CM, Hafeezullah M, Franco J, Varma RR, Gibson DP, Hoffmann RG, Stravitz RT, Heuman DM, Sterling RK, Shiffman M, Topaz A, Boyett S, Bell D, Sanyal AJ. Minimal hepatic encephalopathy is associated with motor vehicle crashes: The reality beyond the driving test. *Hepatology (Baltimore, Md)*. 2009;50(4):1175-83.
31. Malaguarnera G, Pennisi M, Bertino G, Motta M, Borzi AM, Vicari E, Bella R, Drago F, Malaguarnera M. Resveratrol in patients with minimal hepatic encephalopathy. *Nutrients*. 2018;10(3):329.
32. Grover VP, Tognarelli JM, Massie N, Crossey MM, Cook NA, Taylor-Robinson SD. The why and wherefore of hepatic encephalopathy. *Int J Gen Med*. 2015;8:381-90.
33. Kircheis G, Haussinger D. Hepatic encephalopathy. *Dtsch Med Wochenschr*. 2012;137(31-32):1582-85.

34. Weissenborn K, Ahl B, Fischer-Wasels D, van den Hoff J, Hecker H, Burchert W, Köstler H. Correlations between magnetic resonance spectroscopy alterations and cerebral ammonia and glucose metabolism in cirrhotic patients with and without hepatic encephalopathy. *Gut*. 2007;56(12):1736-42.
35. Ong JP, Aggarwal A, Krieger D, Easley KA, Karafa MT, Van Lente F, Arroliga AC, Mullen KD. Correlation between ammonia levels and the severity of hepatic encephalopathy. *The American Journal of Medicine*. 2003;114(3):188-93.
36. Lockwood AH, Yap EW, Wong WH. Cerebral ammonia metabolism in patients with severe liver disease and minimal hepatic encephalopathy. *J Cereb Blood Flow Metab*. 1991;11(2):337-41.
37. Cooper A, Plum F. Biochemistry and physiology of brain ammonia. *Physiol Rev* 1987;67:440–519.
38. Blei AT. Infection, inflammation and hepatic encephalopathy, synergism redefined. *Journal of Hepatology*. 2004;40(2):327-30.
39. Zhang G, Cheng Y, Shen W, Liu B, Huang L, Xie S. The short-term effect of liver transplantation on the low-frequency fluctuation of brain activity in cirrhotic patients with and without overt hepatic encephalopathy. *Brain Imaging Behav*. 2016.
40. 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.
41. Allampati S, Mullen KD. Nomenclature and definition of hepatic encephalopathy - An update. *Clinical Liver Disease*. 2015;5(3):68-70.
42. Hassanein TI, Hilsabeck RC, Perry W. Introduction to the Hepatic Encephalopathy Scoring Algorithm (HESA). *Dig Dis Sci*. 2008;53(2):529-38.
43. Weissenborn K, Ennen JC, Schomerus H, Rückert N, Hecker H. Neuropsychological characterization of hepatic encephalopathy. *Journal of Hepatology*. 2001;34(5):768-73.
44. Bajaj JS, Wade JB, Sanyal AJ. Spectrum of neurocognitive impairment in cirrhosis: Implications for the assessment of hepatic encephalopathy. *Hepatology*. 2009;50(6):2014-21.
45. Conn HO, Bircher J. Quantifying the severity of hepatic encephalopathy: syndromes and therapies. East Lansing MI: Medi Ed Press. 1993:13-26.

46. Schomerus H, Hamster W. Quality of life in cirrhotics with minimal hepatic encephalopathy. *Metabolic Brain Disease*. 2001;16(1):37-41.
47. Groeneweg M, Quero JC, De Bruijn I, Hartmann IJC, Essink-Bot M-I, Hop WCJ, Schalm SW. Subclinical hepatic encephalopathy impairs daily functioning. *Hepatology*. 1998;28(1):45-9.
48. Amodio P, Del Piccolo F, Marchetti P, Angeli P, Iemmolo R, Caregaro L, Merkel C, Gerunda G, Gatta, A. Clinical features and survival of cirrhotic patients with subclinical cognitive alterations detected by the number connection test and computerized psychometric tests. *Hepatology*. 1999;29(6):1662-7.
49. Folstein MF, Folstein SE, McHugh PR. Mini-mental state. *Journal of Psychiatric Research*. 1975;12(3):189-98.
50. Tsai L, Tsuang M. The mini-mental state test and computerized tomography. *Am J Psychiatry* 1979; 136 (4A):436–438.
51. Lockwood AH. blood ammonia levels and hepatic encephalopathy. *Metabolic Brain Disease*. 2004;19(3/4):345-9.
52. Kircheis G. Hepatische Enzephalopathie. *Der Gastroenterologe*. 2007;2(4):251-60.
53. Bajaj JS, Heuman DM, Sterling RK, Sanyal AJ, Siddiqui M, Matherly S, Luketic V, Stravitz RT, Fuchs M, Thacker LR, Gilles HC, White MB, Unser A, Hovermale J, Gravis E, Noble NA, Wade JB. Validation of EncephalApp, smartphone-based Stroop Test, for the diagnosis of covert hepatic encephalopathy. *Clinical Gastroenterology and Hepatology*. 2015;13(10):1828-35.e1.
54. Kircheis G, Wettstein M, Timmermann L, Schnitzler A, Häussinger D. Critical flicker frequency for quantification of low-grade hepatic encephalopathy. *Hepatology*. 2002;35(2):357-66.
55. Schomerus H, Weissenborn K, Hamster W, Rückert N, Hecker H. PSE-Syndrom-Test. Swets Test Services, Swets & Zeitlinger B.V., Frankfurt, 1999.
56. Romero-Gómez M, Córdoba J, Jover R, del Olmo JA, Ramírez M, Rey R, de Madaria E, Montoliu C, Nunez D, Flavia M, Company L, Rodrigo JM, Felipo V. Value of the critical flicker frequency in patients with minimal hepatic encephalopathy. *Hepatology*. 2007;45(4):879-85.
57. Morgan MY, Amodio P, Cook NA, Jackson CD, Kircheis G, Lauridsen MM, Montagnese S, Schiff S, Weissenborn K. Qualifying and quantifying minimal hepatic encephalopathy. *Metab Brain Dis*. 2016;31(6):1217-29.

58. 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 A, 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.
59. Torlot FJ, McPhail MJW, Taylor-Robinson SD. Meta-analysis: the diagnostic accuracy of critical flicker frequency in minimal hepatic encephalopathy. *Alimentary Pharmacology & Therapeutics*. 2013;37(5):527-36.
60. Kircheis G, Häussinger D. Hepatische Enzephalopathie. *Der Gastroenterologe*. 2016;11(1):4-15.
61. Bajaj JS, Thacker LR, Heuman DM, Fuchs M, Sterling RK, Sanyal AJ, Puri P, Siddiqui MS, Stravitz RT, Bourneva 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.
62. Ferenci P. Hepatische Enzephalopathie. *Hepatitis&more*. 2011;1:38-41.
63. Amodio P, Ridola L, Schiff S, Montagnese S, Pasquale C, Nardelli S, Pentassuglio I, Trezza M, Marzano C, Flaiban C, Angeli P, Cona G, Bisiacchi P, Gatta A, Riggio O. Improving the inhibitory control task to detect minimal hepatic encephalopathy. *Gastroenterology*. 2010;139(2):510-8.e2.
64. Guerit JM, Amantini A, Fischer C, Kaplan PW, Mecarelli O, Schnitzler A, Ubiali E, Amodio P. Neurophysiological investigations of hepatic encephalopathy: ISHEN practice guidelines. *Liver Int*. 2009;29(6):789-96.
65. Reitan RM. The relation of the Trail Making Test to organic brain damage. *Journal of Consulting Psychology*. 1955;19(5):393-4.
66. Conn HO, Leevy CM, Vlahcevic ZR, Rodgers JB, Maddrey WC, Seeff L, Levy LL. Comparison of lactulose and neomycin in the treatment of chronic portal-systemic encephalopathy. *Gastroenterology*. 1977;72(4):573-83.
67. Schäfer K, Pittner PM, Lutcke A, Wehr M, Bode JC. [Assessment of the course of chronic hepatic encephalopathy. Comparison of various measurements with special reference to the trail-making test]. *Dtsch Med Wochenschr*. 1981;106(28):904-9.

68. Weissenborn K, Rückert N, Hecker H, Manns MP. The number connection tests A and B: interindividual variability and use for the assessment of early hepatic encephalopathy. *Journal of Hepatology*. 1998;28(4):646-53.
69. Zeegen R, Drinkwater JE, Dawson M. Method for measuring cerebral dysfunction in patients with liver disease. *The British Medical Journal*. 1970;2(5710):633-6.
70. Dhiman RK, Saraswat VA, Verma M, Naik SR. Figure connection test: A universal test for assessment of mental state. *Journal of Gastroenterology and Hepatology*. 1995;10(1):14-23.
71. Ritschl V, Weigl R, Stamm T. *Wissenschaftliches Arbeiten und Schreiben: Verstehen, Anwenden, Nutzen für die Praxis*. Springer. 2016.
72. Rey GD. *Methoden der Entwicklungspsychologie: Datenerhebung und Datenauswertung* Books on Demand. 2012; Auflage 1.
73. Moosbrugger H, Kelava A. *Testtheorie und Fragenbogenkonstruktion*: Springer. 2007.
74. Döring N, Bortz J, Pöschel S. *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Springer. 2015.
75. Kappus MR, Bajaj JS. Covert Hepatic Encephalopathy: Not as Minimal as You Might Think. *Clinical Gastroenterology and Hepatology*. 2012;10(11):1208-19.
76. Lilford RJ, Bentham L, Girling A, Litchfield I, Lancashire R, Armstrong D, Jones R, Marteau T, Neuberger J, Gill P, Cramb R, Oliff S, Arnold D, Khan K, Armstrong MJ, Houlihan DD, Newsom PN, Chilton PJ, Moons K, Altman D. Birmingham and Lambeth Liver Evaluation Testing Strategies (BALLETS): a prospective cohort study. *Health Technol Assess*. 2013;17(28):1-307.
77. Rüter FD. *Diagnostische Wertigkeit eines elektronischen Tablet-Computer-basierten Zahlenverbindungstests zur Detektion der minimalen hepatischen Enzephalopathie bei Patienten mit Leberzirrhose*. Dissertation - Charité Universitätsmedizin Berlin. 2018.
78. Cohen J. A power primer. *Psychological Bulletin*. 1992;112(1):155-9.
79. Bender R, Lange S, Ziegler A. Multiple testing. *Dtsch Med Wochenschr*. 2007;132 Suppl 1:e26-9.
80. Victor A, Elsässer 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.

81. Webinger J, Keller D, Budrich B. Wie schreibe ich einer Doktorarbeit ?. Springer. 2014.
82. 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;00:000-000.
83. Huang M-H. Race of the interviewer and the black–white test score gap. *Social Science Research.* 2009;38(1):29-38.
84. Vormittag I, Ortner TM. In the eye of the examinee: likable examiners interfere with performance. *Social Psychology of Education.* 2014;17(3):401-17.
85. Lienert GA, Raatz U. Testaufbau und Testanalyse. Beltz, Psychologie Verlags Union. 1998.
86. Held L, Rufibach K, Seifert B. Medizinische Statistik: Konzepte, Methoden, Anwendungen. Pearson Studium. 2013.
87. Ennen JC. Der PSE-Syndrom-Test: Diagnosestandardisierung der latenten portosystemischen Enzephalopathie mittels psychometrischer Testverfahren. Dissertation - Medizinische Hochschule Hannover. 2000.
88. Conn HO. Trailmaking and number-connection tests in the assessment of mental state in portal systemic encephalopathy. *The American Journal of Digestive Diseases.* 1977;22(6):541-50.
89. Nachtigall C, Wirtz M. Wahrscheinlichkeitsrechnung und Inferenzstatistik Statistische Methoden für Psychologen Teil 2. Weinheim: Juventa Verlag. 2006.
90. Schneider A, Hommel G, Blettner M. Linear regression analysis: part 14 of a series on evaluation of scientific publications. *Dtsch Arztebl Int.* 2010;107(44):776-82.
91. Boll TJ, Reitan RM. Effect of age on performance of the Trail Making Test. *Percept Mot Skills.* 1973;36(3):691-4.
92. Corrias M, Turco M, Rui MD, Gatta A, Angeli P, Merkel C, Amodio P, Schiff S, Montagnese S. Covert hepatic encephalopathy: does the mini-mental state examination help? *J Clin Exp Hepatol.* 2014;4(2):89-93.
93. Grouven U, Bender R, Ziegler A, Lange S. [Comparing methods of measurement]. *Dtsch Med Wochenschr.* 2007;132 Suppl 1:e69-73.

94. Keidel M, Vauth F, Richter J, Hoffmann B, Soda H, Griewing B, Scibor M. Telerehabilitation nach Schlaganfall im häuslichen Umfeld. *Der Nervenarzt*. 2017;88(2):113-9.
95. Ganapathy D, Acharya C, Lachar J, Patidar K, Sterling RK, White MB, Ignudo C, Bommidi S, DeSoto J, Matherly S, Shaw J, Siddiqui MS, Puri P, Sanyal AJ, Luketic V, Lee H, Stravitz RT, Bajaj JS. The patient buddy app can potentially prevent hepatic encephalopathy-related readmissions. *Liver International*. 2017;00:1-9

6 Abbildungsverzeichnis

- Abb. 1:** Mögliche Ursachen einer Leberzirrhose.
- Abb. 2:** CHILD-PUGH-Klassifikation.
- Abb. 3:** Semiquantitative Stadieneinteilung des mentalen Status bei Hepatischer Enzephalopathie (HE) nach den West Haven Kriterien (WHC).
- Abb. 4:** Bildschirmfoto der EncephalApp–Stroop Test.
- Abb. 5:** Bildschirmfenster beim eNCT.
- Abb. 6:** Darstellung der unterschiedlichen Displaygrößen.
- Abb. 7:** Schematische Darstellung der eingeschlossenen Studienteilnehmer.
- Abb. 8:** Ätiologie der Leberzirrhose aller eingeschlossenen Patienten.
- Abb. 9:** Testzeiten der Parallelversionen für eNCT und NCTA in der Patientenkohorte dargestellt mit Hilfe von Fehlerbalken und 95%igem KI.
- Abb. 10:** Testzeiten in den Parallelversionen für eNCT und NCTA in der Gruppe der Kontrollprobanden dargestellt mit Hilfe von Fehlerbalken und 95%igem KI.
- Abb. 11:** Erfahrung in der Verwendung elektronischer Geräte.
- Abb. 12:** Zumutbarkeit und favorisierte Displaygröße bei einer täglichen Anwendung des eNCT im heimischen Umfeld.
- Abb. 13:** Unterschiede in der Durchführung des eNCT hinsichtlich der Displaygröße.
- Abb. 14:** Streudiagramm mit Regressionsgerade und Regressionsgleichung für den Zusammenhang zwischen Testzeit im eNCT10Z I und Alter.
- Abb. 15:** Streudiagramm mit Regressionsgerade und Regressionsgleichung für den Zusammenhang zwischen Testzeit im eNCT10Z I und Bildungsgrad.
- Abb. 16:** Streudiagramm mit Regressionsgerade und Regressionsgleichung für den Zusammenhang zwischen Testzeit im eNCT10Z I und MMSE-Score.
- Abb. 17:** Streudiagramm mit Regressionsgerade und Regressionsgleichung für den Zusammenhang zwischen Testzeit im eNCT10Z I und HE Medikation
- Abb. 18:** Bland-Altman-Diagramme illustrieren den Grad der Übereinstimmung zwischen Testzeiten im eNCT innerhalb unterschiedlich großer Displays.

7 Tabellenverzeichnis

- Tab. 1:** Demographische Daten der Studienteilnehmer.
- Tab. 2:** Charakterisierung von Leberzirrhose, Therapie und Komplikationen.
- Tab. 3:** Einzeltests des PSE-Syndrom-Tests im Vergleich zwischen Kontrollprobanden und Patienten.
- Tab. 4:** Ergebnisse der Paralleltests im eNCT zwischen Kontrollprobanden und Patienten.
- Tab. 5:** Retestreliabilität gemäß Intraklassenkorrelation (IKK).
- Tab. 6:** Paralleltestreliabilität gemäß Reliabilitätskoeffizient nach Cronbach alpha (rtt).
- Tab. 7:** Testergebnisse von Test und Retest im eNCT und NCTA.
- Tab. 8:** Konvergente und divergente Validität des eNCT gemäß Spearman Korrelationskoeffizient (SKK).
- Tab. 9:** Kriteriumsvalidität des eNCT gemäß Spearman Korrelationskoeffizient (SKK).
- Tab. 10:** Korrelationen zwischen eNCT-Performanz und den Faktoren Alter, Geschlecht und Bildungsgrad an gesunden Kontrollprobanden.
- Tab. 11:** Ergebnisse der multiplen Regression nach Testzeiten im eNCT bei Kontrollprobanden.
- Tab. 12:** Korrelationen zwischen eNCT-Performanz und möglichen Einflussfaktoren innerhalb der Patientenkohorte.
- Tab. 13:** Korrelationen zwischen eNCT-Performanz und laborchemischen Blutparametern innerhalb der Patientenkohorte.
- Tab. 14:** Ergebnisse der multiplen Regression nach Testzeit im eNCT bei Patienten.

8 Erklärung an Eides Statt

„Ich, Tim Brötzmann, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema „Methodische Bewertung eines elektronischen Zahlenverbindungstests (eNCT) bei Patienten mit Leberzirrhose am Beispiel der minimalen hepatischen Enzephalopathie“ selbstständig und ohne nicht offen gelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe. Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen 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, Graphiken und Tabellen) entsprechen den URM (siehe oben) 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.“

Tim Brötzmann
Promovend

9 Curriculum Vitae

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

10 Danksagung

Großen Dank möchte ich an sämtliche Studienteilnehmer aussprechen, ohne die das Gelingen der Arbeit nicht möglich gewesen wäre.

Ein besonderer Dank gilt meinem Doktorvater, Herrn Prof. Dr. med. Martin Stockmann der mir die theoretische Grundlage dieser Arbeit gelegt und es mir zur Bearbeitung überlassen hat.

Vielen Dank auch an meinen Betreuer Herrn Dr. med. Maximilian Jara für die umfangreiche Unterstützung und Förderung, die intelligenten Korrekturen und für sein Vertrauen in mich.

Bei Herrn Dr. rer. nat. Tilo Wunsch möchte ich mich für den konstruktiven Austausch und die Betreuung während der Studie bedanken.

Aufrichtigen Dank an Peter, Max, Basti, Stefan und Emma für das Gegenlesen der Arbeit oder die vielen positiven Impulse inner- und außerhalb der Dissertation.

Danke und herzliche Umarmungen an meine Eltern Korinna und Norbert für die beispiellose Unterstützung und Motivation über all die Jahre hinweg.

Mein größter Dank richtet sich an meinen wichtigsten Menschen.

Sarah, danke für dein Vertrauen und deine Geduld mit mir. Du hast mich während der Promotion nicht nur einzigartig unterstützt, motiviert und gefördert, sondern wenn es nötig war auch emotional abgeholt und meine Gedanken neu sortiert.

Für immer und Dich.

11 Anhang

Anhang 1: Zahlenverbindungstests A in Parallelversion I (NCTA I)

Number Connection Test 1

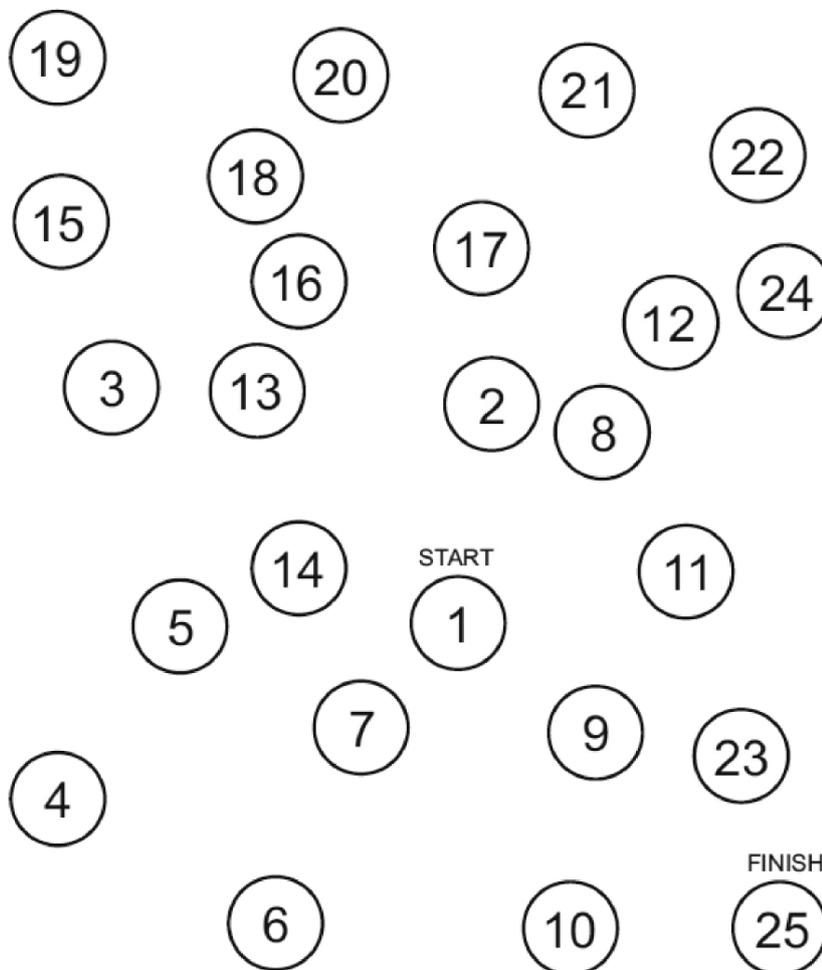


Pat-Code: □□_□□_□□□/□□

Klinik f. Allgemein-, Viszeral- und
Transplantationschirurgie

Datum: □□_□□_□□

Time of day: _____

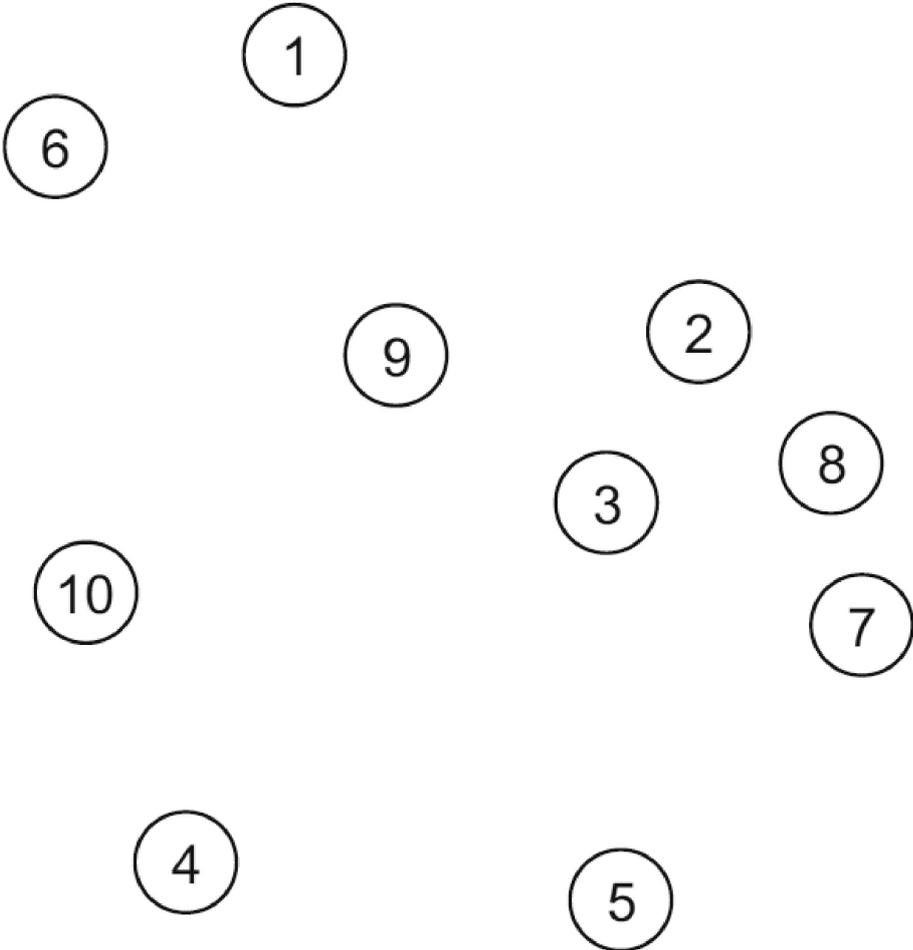




Klinik f. Allgemein-, Viszeral- und
Transplantationschirurgie

Zahlen-Verbindungs-Test

(Demonstrationsseite)



Anhang 3: Zahlenverbindungstest B (NCTB)

Number Connection Test Part B



Pat-Code: □□_□□_□□□/□□

Datum: □□_□□_□□

Time of day: _____



RESULTS: Time: _____ sec.

Verbinden Sie bitte die Zahlen und Buchstaben abwechselnd der Reihe nach.

The test grid contains 13 numbered circles and 12 lettered circles. The numbers are: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13. The letters are: A, B, C, D, E, F, G, H, I, J, K, L. The word "End" is written in blue at the top left, and "Begin" is written in blue in the center. The circles are scattered across the grid, and the task is to connect them in an alternating sequence starting from "Begin".

Anhang 4: Zahlen-Symbol-Test (ZST) in der verwendeten Vorlage

Zahlen-Symbol-Test



Pat-Code: □□_□□_□□□/□□

Datum: □□_□□_□□

Time of day: _____

RESULTS: _____ Points (max. 93)

10. Zahlen-Symbol-Test

1	2	3	4	5	6	7	8	9	Punkte
—	⊥	☐	L	U	O	^	X	=	90°

Beispiel

	2	1	3	7	2	4	8		2	1	3	2	1	4	2	3	5	2	3	1	4	5	6	3	1	4	
	1	5	4	2	7	6	3	5	7	2	8	5	4	6	3	7	2	8	1	9	5	8	4	7	3		
	6	2	5	1	9	2	8	3	7	4	6	5	9	4	8	3	7	2	6	1	5	4	6	3	7		
	9	2	8	1	7	9	4	6	8	5	9	7	1	8	5	2	9	4	8	6	3	7	9	8	6		

Anhang 5: Linien-Nachfahr-Test (LNT) inklusive Übungsaufgabe (oben)

Linien-Nachfahr-Test



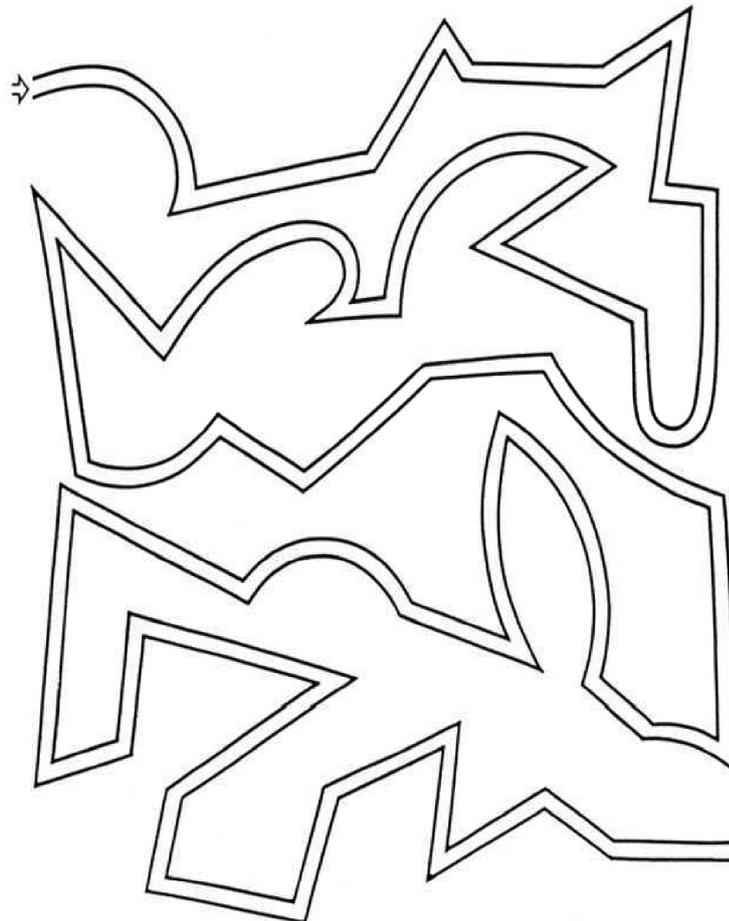
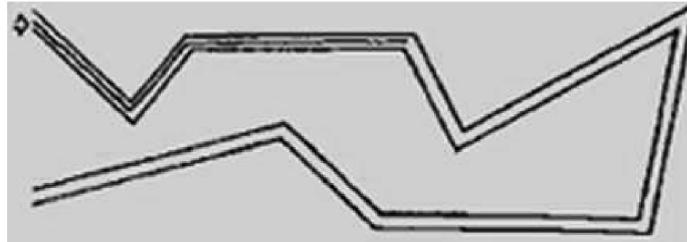
Pat-Code: □□_□□_□□□/□□

Datum: □□_□□_□□

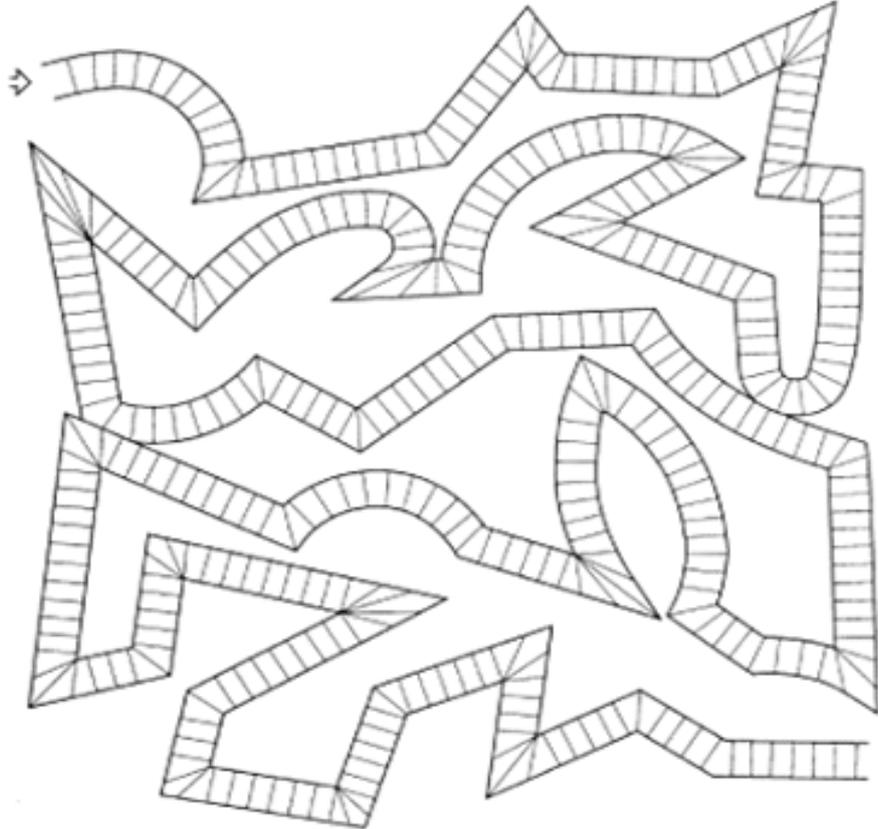
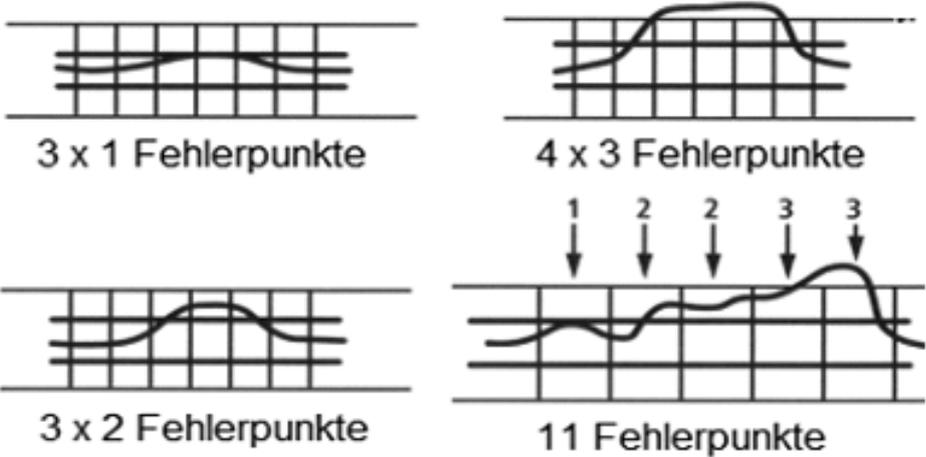
Time of day: _____

RESULTS: Time: _____sec.

Touching/ crossing the lines: ____ times



Anhang 6: Auswertungsschablone im Linien-Nachfahr-Test (LNT) inklusive Fehlerzählung und -gewichtung (oben)



Anhang 7: Test Kreise-Punktieren (KP) / Serial Dotting Test (SDT) inkl. Übungsreihe (oben)

Serial Dot Test

Pat-Code: □□_□□_□□□/□□

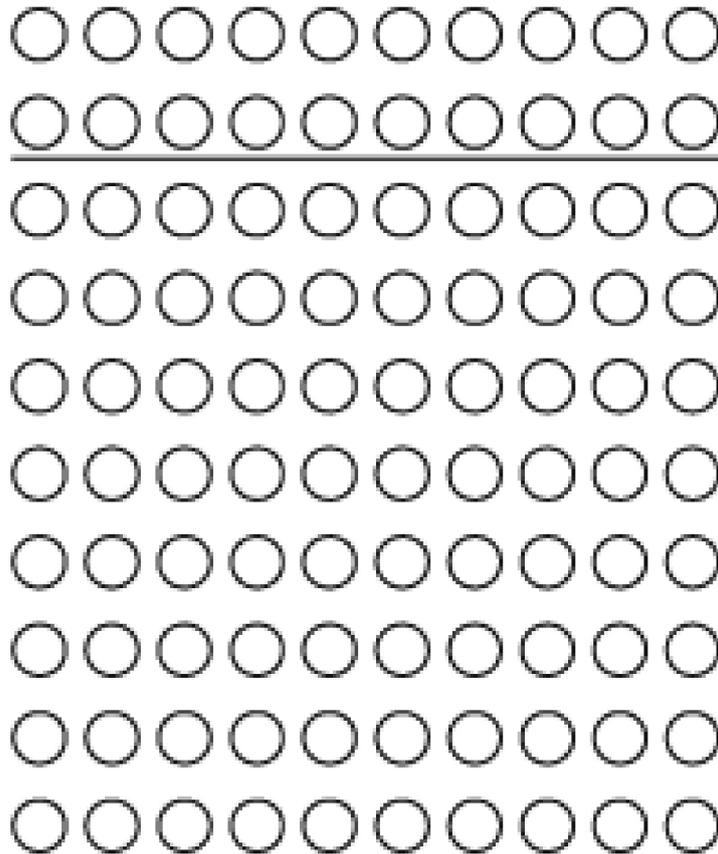
Datum: □□_□□_□□

Time of day: _____

RESULT: _____ sec.



Klinik f. Allgemein-, Viszeral- und
Transplantationschirurgie



Anhang 8: Fragebogen zum Patientenfeedback

Fragen zum Gebrauch von technischen Geräten

bitte schreiben Sie Ihre Antworten auf oder kreuzen Sie die zutreffende(n) Antwort(en) an

1. Welche der folgenden Geräte haben Sie zu Hause und benutzen es auch regelmäßig?
(bitte alle zutreffenden Antworten ankreuzen)?
 - a. **Einen DVD -Player / Recorder**
 - b. **Einen Computer (zu Hause oder bei der Arbeit)**
 - c. **Eine Digitalkamera**
 - d. **Ein Blutzuckermessgerät, ein automatisches Blutdruckmessgerät oder ähnliche medizinische Geräte**
 - e. **Ein Mobiltelefon**
2. Welche digitalen Angebote verwenden Sie, um mit anderen Mitmenschen in Kontakt zu treten (bitte alle zutreffenden Antworten ankreuzen)?
 - a. **E- Mail**
 - b. **SMS**
 - c. **Facebook oder andere soziale Netzwerke**
 - d. **Keine dieser Art**
3. Verwenden Sie bzw. haben Sie jemals ein Smartphone oder Tablet PC verwendet (Apple Iphone, Apple Ipad, Android phone, Blackberry etc.)?
 - a. **Ja**
 - b. **Nein**

Fragen zur Display-Größe und der Durchführbarkeit des kognitiven Tests

bitte schreiben Sie Ihre Antworten auf oder kreuzen Sie die zutreffende(n) Antwort(en) an

4. Welche der beiden Gerätegrößen würden Sie für die tägliche Anwendung bei Ihnen zu Hause verwenden wollen?
 - a. **Das 10“ Tablet**
 - b. **Das 18“ Tablet**
 - c. **Keines von beiden**
 - d. **Beide**
5. Sollten Sie eine der beiden Gerätegrößen favorisieren: Was ist der Grund dafür?
(mehrere Antworten möglich)
 - a. **Das Gerät ist in der Handhabung angenehmer**
 - b. **Das Gerät ist besser in mein häusliches Umfeld integrierbar**
 - c. **Die Anweisungen sind auf dem Gerät leichter ablesbar**
 - d. **Die Bedienung des Touchscreens ist für mich leichter**
 - e. **Andere Gründe (bitte schreiben Sie diese auf):**
6. Der Durchführung des Zahlenverbindungstests war meines Erachtens nach einfacher auf dem
 - a. **10“ Tablet**
 - b. **18“ Tablet**
 - c. **Auf beiden war die Durchführung gleichermäßen möglich**