

Aus der Klinik für Anästhesiologie
mit Schwerpunkt operative Intensivmedizin
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

Das postoperative Delirium.
Vergleich der
Nursing Delirium Screening Scale
und der
Confusion Assessment Method
im Screening
auf der peripheren chirurgischen Station.

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

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Sabine Schust

aus Berlin

Gutachter: 1. Prof. Dr. med. C. Spies
2. Prof. Dr. med. J. Gallinat
3. Prof. Dr. med. W. Wilhelm

Datum der Promotion: 04.02.2011

Abkürzungsverzeichnis

ACHI	Allgemeinchirurgie
ASA	American Society of Anesthesiologists
ASA-PS	American Society of Anesthesiologists Physical Status Classification
AUC	Area Under the Curve
CAM	Confusion Assessment Method
CI	Konfidenzintervall
CIWA-Ar	Clinical Withdrawal Assessment for Alcohol
CRS	Confusion Rating Scale
DDS	Delirium Detection Scale
DR	Delirium Room
DRS	Delirium Rating Scale
DSM-III	Diagnostic and Statistical Manual of Mental Disorders, 3 rd Edition
DSM-III-R	Diagnostic and Statistical Manual of Mental Disorders, 3 rd Edition, revised
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders, 4 th Edition
FPR	Falsch-Positiv-Rate
FNR	Falsch-Negativ-Rate
GABA	Gamma-Amino-Buttersäure
GYN	Gynäkologie
HELP	Hospital Elder Life Program
HNO	Hals-Nasen-Ohren-Heilkunde
ICD-10	International Statistical Classification of Diseases and Related Health Problems
ICU	Intensive Care Unit
ISPOR PRO	International Society for Pharmacoeconomics and Outcome Research Task Force for Translation and Cultural Adaptation of Patient Related Outcomes
LR-	Likelihood Ratio für einen negativen Test
LR+	Likelihood Ratio für einen positiven Test
MMSE	Mini Mental State Examination
NPW	Negativer Prädiktiver Wert

Abkürzungsverzeichnis

Nu-DESC	Nursing Delirium Screening Scale
P	Wahrscheinlichkeit des Fehlers 1.Art (α -Fehler)
PPW	Positiver Prädiktiver Wert
PPPD	pyloruserhaltende Pancreatico-Duodenostomie
RASS	Richmond Agitation Sedation Scale
ROC	Receiver Operating Characteristics
SOP	Standard Operating Procedure
SPSS	Statistic Package of the Social Science Software
STARD	Standards for the Reporting of Diagnostic Accuracy Studies
TEP	Totale Endoprothese
TIVA	Total intravenöse Anästhesie
WHO	World Health Organisation

Inhaltsverzeichnis

	Seite
1 Einleitung	4
1.1 Bedeutung	4
1.2 Definition	4
1.3 Inzidenz und Prävalenz	4
1.4 Pathogenese und Risikofaktoren	5
1.4.1 Prädisponierende Faktoren.....	6
1.4.2 Präzipitierende Faktoren.....	6
1.5 Outcome.....	7
1.6 Diagnose	7
1.7 Therapie und Management	9
1.8 Prävention	10
2 Fragestellung	12
3 Methodik	13
3.1 Studiendesign	13
3.2 Studienpopulation	13
3.2.1 Einschlusskriterien	13
3.2.2 Ausschlusskriterien	14
3.3 Erhobene Daten	15
3.3.1 Basisdaten	15
3.3.2 Delirdiagnostik	15
3.3.2.1 DSM-IV	16
3.3.2.2 Nu-DESC.....	17
3.3.2.3 CAM	18
3.4 Statistik.....	18
4 Ergebnisse	20
4.1 Studienpopulation	20
4.2 Deskriptive Statistik	21
4.2.1 Basisdaten der Gesamtpopulation	21
4.2.2 Basisdaten deliranter und nicht deliranter Patienten im Vergleich	22
4.3 Delir-Testung.....	23

Inhaltsverzeichnis

4.3.1 Inzidenz und Dauer des postoperativen Delirs auf der peripheren Station	24
4.3.2 Inzidenz von Einzelfaktoren	25
4.3.2.1 Diagnostic and Statistical Manual of Mental Disorders – DSM-IV	26
4.3.2.2 Confusion Assessment Method – CAM	27
4.3.2.3 Nursing Delirium Screening Scale – Nu-DESC	28
4.4 Testgüte	30
4.4.1 Sensitivität	31
4.4.2 Spezifität	32
4.4.3 Falsch-Positiv-Rate	33
4.4.4 Falsch-Negativ-Rate	34
4.4.5 Positiver Prädiktiver Wert (PPW)	34
4.4.6 Negativer Prädiktiver Wert (NPW)	34
4.4.7 Likelihood Ratio für einen positiven Test (LR+)	34
4.4.8 Likelihood Ratio für einen negativen Test (LR-)	34
4.4.9 Prä-Test-Wahrscheinlichkeit (Prävalenz)	35
4.4.10 Prä-Test-Chance (pre-test-odds)	35
4.4.11 Post-Test-Chance (post-test-odds)	35
4.4.12 Post-Test-Wahrscheinlichkeit	35
4.4.13 Genauigkeit (Accuracy)	35
4.4.14 ROC-Kurve	36
4.4.15 Interrater-Reliabilität	38
4.4.16 Durchführungszeit	38
5 Diskussion	39
5.1 Inzidenz	39
5.2 Risikofaktoren	39
5.3 Outcome	40
5.4 Testgüte	40
5.4.1 Nu-DESC	40
5.4.2 CAM	42
5.5 Methodenkritik	44
5.5.1 Bias	44
5.5.1.1 Selection Bias	44

Inhaltsverzeichnis

5.5.1.2 Information Bias.....	44
5.5.2 Confounder	45
5.5.3 Zufall	46
5.5.4 Fallzahl	46
5.5.5 Reproduzierbarkeit.....	46
5.5.6 Qualität des Goldstandards	47
5.7 Praxisrelevanz der gefundenen Ergebnisse.....	48
6 Zusammenfassung	50
6.1 Hintergrund	50
6.2 Methodik.....	50
6.3 Ergebnisse	50
6.4 Schlussfolgerung.....	51
7 Literaturverzeichnis	52
8 Publikationsliste	60
9 Abbildungsverzeichnis	61
10 Tabellenverzeichnis.....	62
11 Danksagung	64
12 Lebenslauf.....	65
13 Erklärung an Eides Statt	66

1 Einleitung

1.1 Bedeutung

Ein sich postoperativ entwickelndes Delir ist eine häufige Komplikation bei chirurgischen Patienten (siehe 1.3), die mit negativen Auswirkungen auf Morbidität und Mortalität einhergeht (siehe 1.2 und 1.5). Durch eine effektive Prävention und Behandlung ließen sich diese negativen Folgen verhindern beziehungsweise verringern (siehe 1.7 und 1.8). Um vom Delir betroffene Patienten schneller identifizieren zu können, ist eine Implementierung geeigneter Screening-Instrumente als Voraussetzung für eine frühzeitige und effektive Therapie nötig. Allerdings fehlen bisher Validierungs- und Vergleichsstudien, die auf chirurgischen Stationen durchgeführt wurden.

1.2 Definition

Das postoperative Delir ist als ein Zustand definiert, der durch Störungen von Aufmerksamkeit, Gedächtnis, Wahrnehmung, kognitiver Funktion, Orientierung, Verhalten, Psychomotorik und Schlaf gekennzeichnet ist. Diese Merkmale entwickeln sich akut und zeigen einen fluktuierenden Verlauf¹⁻⁵. In der Literatur werden derzeit drei Subtypen unterschieden: Das hyperaktive Delir, in dem Symptome wie Agitation bis hin zur Aggression im Vordergrund stehen, das hypoaktive Delir, das eher durch fehlende psychomotorische Aktivität imponiert, sowie eine gemischte Form, die Züge beider Subtypen vereint^{4,6}.

Fälle, in denen nicht alle Eigenschaften eines vollen Deliriums ausgeprägt sind, werden als subsyndromales Delir bezeichnet^{7,8}.

1.3 Inzidenz und Prävalenz

Das postoperative Delir ist eine vergleichsweise häufige und im Verlauf potentiell tödliche Komplikation chirurgischer Eingriffe^{9,10}. Seine Prävalenz wird häufig unterschätzt, die Rate an Fehldiagnosen ist hoch^{11,12}. Olin und Kollegen konnten eine Prävalenz von ca. 51% (26 von 51 Patienten) nach großen abdominal-chirurgischen Eingriffen nachweisen¹³. Die in der Literatur angegebenen Inzidenzen für das Delir im Allgemeinen variieren stark - je nach Studiendesign zwischen 10% und 40%¹¹. Für das postoperative Delir wurden Inzidenzen von 7-52% angegeben¹⁴. Nach Cole et al. ist ein Delir bei 10-15% der älteren allgemein-chirurgischen Patienten zu finden¹⁵.

Die meisten bisher erhobenen Daten beziehen sich auf eine Patientenpopulation spezifischer medizinischer Richtungen oder eng gefasster Altersstrukturen; eine Untersuchung der Prävalenz am gemischten Hochrisiko-Patientenkollektiv der peripheren chirurgischen Station liegt bisher noch nicht vor.

1.4 Pathogenese und Risikofaktoren

Die Pathophysiologie des postoperativen Delirs ist bis heute nicht gänzlich geklärt. Eine Hypothese ist die des gestörten Neurotransmitter-Stoffwechsels. Diskutiert wird hier vor allem die Bedeutung eines Mangels an Acetylcholin sowie eines Dopaminüberschusses, während Serotonin und Gamma-Amino-Buttersäure (GABA) eher mit Delirien aufgrund schwerer internistischer Grunderkrankungen in Zusammenhang gebracht werden¹⁶⁻¹⁸. Hála stellte die Hypothese auf, dass oben genannte Störungen des Acetylcholin-Haushaltes durch Mikrozirkulationsstörungen hervorgerufen werden, welche Folge einer systemischen Entzündungsreaktion als Antwort auf das chirurgische Trauma sein könnten¹⁹.

Ein anderer Erklärungsansatz ist der einer krankheitsbedingten Hyperreagibilität der hypothalamo-hypophyseo-adrenalen Achse - ein Konzept, das Young und Inouye jedoch als wenig befriedigend zur Erklärung der Pathogenese des postoperativen Delirs bewerten²⁰. White et al. dagegen fanden eine erniedrigte Aktivität von Plasmaesterasen im Blut deliranter Patienten²¹. Diese sind wichtige Enzyme des Medikamentenstoffwechsels und könnten daher eine mögliche Erklärung dafür sein, dass ein Delir durch viele Arzneimittel auslösbar ist. Allen Erklärungsversuchen gemeinsam ist jedoch die Erkenntnis, dass die Entstehung eines Delirs eher als ein Syndrom, ausgelöst durch viele verschiedene sich gegenseitig beeinflussende Faktoren, zu verstehen ist²⁰.

Inouye et al. entwickelten dabei ein Konzept, bei dem einerseits dem Patienten individuell innewohnende Eigenschaften (prädisponierende Faktoren) und andererseits von außen einwirkende Größen (präzipitierende Faktoren) miteinander interagieren und ein Delir auslösen können²². So kann beispielsweise ein schweres Trauma in Form eines großen chirurgischen Eingriffs oder aber eine schwere Infektion ein Delir bei einem bisher gesunden Patienten verursachen. Andererseits ist nur ein geringes Trauma nötig, um Gleiches bei einem Patienten hervorzurufen, der viele Risikofaktoren auf sich vereint und daher sehr vulnerabel für die Entwicklung eines Delirs ist.

Auf der Basis dieser Risikofaktoren lassen sich Modelle entwickeln, um Patienten mit einem erhöhten Risiko für ein Delir zu identifizieren und den entsprechenden Präventionsstrategien zuzuführen²²⁻²⁵.

1.4.1 prädisponierende Faktoren

Zwei systematische Reviews identifizierten 25 bzw. 27 relevante Artikel, die insgesamt 61 Risikofaktoren untersuchten. Die wichtigsten hierbei waren: eine vorbestehende Demenz, hohes Alter^{9, 26}, die Einnahme einer hohen Anzahl von Medikamenten, schwere Krankheit, männliches Geschlecht⁹, eine vorbestehende Einschränkung der kognitiven, funktionellen und sensorischen Fähigkeiten, Depressionen, die präoperative Einnahme psychotroper Medikamente, vorbestehende psychopathologische Symptome, Wohnen in einem Pflegeheim und stärkere Komorbidität²⁶. Litaker et al. nennen darüber hinaus Drogenkonsum und Delirien in der Vorgeschichte als Risikofaktoren²⁷.

1.4.2 präzipitierende Faktoren

Vor allem nicht elektive Eingriffe/Notoperationen stellen chirurgische Risikofaktoren dar⁹. Auch scheint die Lokalisation des operativen Eingriffs eine Rolle zu spielen. So fand Marcantonio signifikant häufiger Delirien nach Operationen in der nicht kardialen Thoraxchirurgie und der Aortenaneurysma-Chirurgie²⁸.

Ferner ist es scheinbar unerheblich, ob die Patienten eine Regionalanästhesie oder eine Allgemeinnarkose erhalten^{9, 29}. Dies wiederum könnte ein Hinweis darauf sein, dass direkte Gewebstraumen durch die Operation einen größeren Einfluss auf die Entwicklung eines Delirs haben als die oftmals angeschuldigte Allgemeinanästhesie. Auch gibt es bisher noch keine eindeutigen Hinweise, ob ganz bestimmte Anästhetika ein Delir begünstigen.

Perioperativ tragen vor allem ein starker Blutverlust, viele Bluttransfusionen und ein niedriger Hämatokrit (<30%) zur Entstehung eines Delirs bei^{9, 30}.

Postoperativ können unerkannte Infektionen, Stoffwechselstörungen, Myokardinfarkte oder respiratorisches Versagen zu einem Delir führen^{9, 26}. Auch das postoperative Fehlen einer Brille oder eines Hörgerätes, eine verspätete oder fehlende postoperative Mobilisation sowie Manipulationen wie das Legen eines Blasenkatheters sind präzipitierende Faktoren^{9, 31}.

Patienten, die einen Drogenentzug, insbesondere einen Alkohol- und Nikotinentzug erleben, sind besonders gefährdet^{9, 32}.

Alagiakrishnan et al. schätzen, dass bis zu 39% aller Fälle von Delirien allein durch Medikamente ausgelöst werden. Hierzu gehören besonders Benzodiazepine, Opioiden und anticholinerg wirkende Substanzen^{15, 16, 33}.

1.5 Outcome

Bereits nach drei Monaten zeigt sich unter den ehemals deliranten Patienten eine höhere Rate von Einweisungen in Pflegeheime³⁴, eine Verschlechterung der funktionellen Fähigkeiten sowie eine erhöhte Mortalitätsrate³⁵. Edlund et al. beobachteten zudem eine erhöhte Bettlägerigkeit nach 6 Monaten³⁶. Diese Ergebnisse werden durch Langzeit-Studien von ein bis drei Jahren Dauer weitgehend bestätigt³⁷⁻⁴⁰. Besonders auffällig waren die Resultate, wenn das Delir zum Zeitpunkt der Entlassung noch bestand⁴¹.

Darüber hinaus werden Folgen wie eine längere Verweildauer im Krankenhaus und auf der Intensivstation sowie häufigere Komplikationen beschrieben^{13, 34}. Dabei scheint es auch von Bedeutung zu sein, wie lange ein Delir besteht. Pisani und Kollegen zeigten in einer Studie mit 304 Patienten einer Intensivstation, dass die Anzahl der deliranten Tage direkt mit der 1-Jahres-Mortalitätsrate korrelierte, die umso höher war, je länger das Delir bestand⁴².

Außerdem wird diskutiert, ob ein Delir mit einem höheren Risiko der Entwicklung einer kognitiven Dysfunktion bis hin zur Demenz einhergeht⁴³. Die Ergebnisse hierzu sind uneinheitlich. Während Bickel et al. einen Zusammenhang zwischen postoperativem Delir und Verschlechterung der kognitiven Fähigkeiten besonders im Langzeit-Follow-Up über drei Jahre nachweisen⁴⁰, finden Rudolph et al. diesen nur nach 7 Tagen, aber nicht mehr nach drei Monaten⁴⁴.

Auch Patienten, die nicht das Vollbild eines Delirs entwickeln, sondern nur einzelne Symptome zeigen (so genanntes subsyndromales Delir, siehe 1.2), haben zum Teil ein ähnlich schlechtes oder gar schlechteres Outcome als Patienten mit einem milden Delir^{7, 45}.

1.6 Diagnose

Die eindeutige Diagnose eines Delirs ist nach wie vor schwierig, da gerade in der häufig betroffenen Gruppe der alten und multimorbiden Patienten eine Abgrenzung zur Demenz mitunter kompliziert sein kann. Ein Delir wird heutzutage ausschließlich auf der Basis der klinischen Untersuchung und Einschätzung diagnostiziert. Um die Fest-

Einleitung

stellung im klinischen Alltag zu erleichtern, wurden verschiedene Diagnoseschemata entwickelt.

Als der Goldstandard wird hierbei aktuell die DSM-IV-Klassifikation (*Diagnostic and Statistical Manual of Mental Disorders, 4th Edition*) angesehen ⁴⁶. Mit ihrer Hilfe wurden mehr Fälle eines Delirs diagnostiziert als mit der ICD-10 Klassifikation (*International Classification of Diseases*), und dies sowohl in der Allgemeinbevölkerung als auch in selektierten Populationen akut kranker und hospitalisierter Patienten ^{47, 48}. Die Anwendung der DSM-IV-Klassifikation ist jedoch zeitintensiv und bedarf einer umfassenden Ausbildung, so dass sie als Screening-Methode im Alltag auf der peripheren Station nicht geeignet scheint.

Zum Screening auf ein Delir wurden verschiedene andere Testmethoden entwickelt, hierzu gehören die *Confusion Assessment Method (CAM)* ⁴⁹, die *Nursing Delirium Screening Scale (Nu-DESC)* ⁵⁰ und die *Delirium Detection Scale (DDS)* ⁵¹.

Die CAM erwies sich als ähnlich valide wie die *Delirium Rating Scale (DRS)* ⁵². Zudem bewies sie die höchste Übereinstimmung mit der DSM-IV-Klassifikation verglichen mit den Kriterien der DSM-III, DSM-III-R und der ICD-10 ⁵³. Eine entsprechende Ausbildung in der Durchführung vorausgesetzt, können Sensitivitäten und Spezifitäten von über 90% erreicht werden ^{49, 54}. Die CAM wurde für die Durchführung durch nicht psychiatrisches Personal entwickelt und benötigt weniger als fünf Minuten zur Anwendung ⁴⁹. Es gibt eine gekürzte Fassung, die aus vier Fragen besteht.

Die Nu-DESC ist eine vergleichsweise neue Testmethode. Sie wurde 2005 als eine Erweiterung der *Confusion Rating Scale (CRS)* von Gaudreau et al. entworfen und durch die Zuhilfenahme des Kriteriums „psychomotorische Retardierung“ weiter an die DSM-IV Kriterien angenähert ⁵⁰. Die Nu-DESC zeigte eine ähnliche Diskriminationsstärke auf der Basis der *area under the curve (AUC Test)* wie die CAM und kann innerhalb von ca. einer Minute durchgeführt werden ⁵⁰. In ihrer Validierungsstudie bewies sie außerdem eine hohe Sensitivität und Spezifität mit Werten von rund 86% respektive 87% ⁵⁰.

Die DDS wurde in der Abteilung für Anästhesie und Intensivmedizin der Charité zur Quantifizierung der Stärke eines Delirs entwickelt und ist eine modifizierte Version des *Clinical Withdrawal Assessment for Alcohol (CIWA-Ar)* ⁵¹.

1.7 Therapie und Management

Nur eine rechtzeitige Früherkennung und Therapie des postoperativen Delirs können dessen Dauer und Schwere verringern und damit negative Folgen mildern⁵⁵⁻⁵⁷. Eine Intervention sollte außerdem so schnell wie möglich erfolgen, da bereits gezeigt wurde, dass das Outcome von betroffenen Patienten umso schlechter ist, je länger das Delir besteht⁴².

Bei Naughton et al. resultierte die Einführung eines Früherkennungs- und Managementprogramms in einer Erniedrigung der Delirprävalenz, einer geringeren Verweildauer im Krankenhaus und einer reflektierteren Verabreichung psychotroper Medikamente⁵⁸. Die frühzeitige Erkennung präzipitierender Faktoren und deren Beseitigung (z.B. Absetzen anticholinergischer Medikamente) sowie eine supportive Therapie (Rehydratation, Sauerstoffgabe, Kalorienbilanzierung, frühe Mobilisierung) sind auch heute noch die Hauptkomponenten einer effektiven Delirtherapie¹⁵.

Eine Betreuung durch ein spezialisiertes Team scheint dagegen keine klaren Vorteile gegenüber der gewöhnlichen Behandlung zu haben^{59, 60}. Flaherty et al. entwickelten eine Delir-Einheit (so genannter *Delirium Room – DR*) mit 4 Betten, die unter anderem eine 24-stündige Intensivüberwachung durch Krankenschwestern gewährleisten soll, jedoch fehlen weitere Daten, um die Effektivität dieses Pilotprojektes zu beurteilen⁶¹. In dieser Abteilung sollte u.a. vollständig auf körperliche Fixierung verzichtet werden, da diskutiert wird, ob diese beim hyperaktiven Delir Agitationen und daraus resultierende Verletzungen des Patienten eher noch verstärkt⁶².

Bezüglich der pharmakologischen Therapie des Delirs ist die Datenlage wenig aussagekräftig. Im Wesentlichen beschränkt sich die medikamentöse Therapie des Delirs nach wie vor auf die Gabe von Antipsychotika oder gemäß der Acetylcholin-Mangel-Hypothese auf die Therapie mit Acetylcholinesterase-Inhibitoren. 2007 und 2008 wurden drei Metaanalysen durchgeführt, deren Einschlusskriterien nur eine bzw. drei Studien erfüllten⁶³⁻⁶⁵. Darin wurden Haloperidol, Chlorpromazin und Lorazepam sowie Haloperidol, Risperidon, Olanzapin und Placebo miteinander verglichen. Dabei erwiesen sich Haloperidol und Chlorpromazin als ähnlich wirksam, wobei Chlorpromazin jedoch leicht (statistisch nicht signifikant) die kognitive Funktion beeinträchtigte⁶⁵. Haloperidol in niedrigen Dosen (<3mg pro Tag) hatte einen positiven Einfluss auf die Schwere und Dauer des Delirs, unterschied sich in dieser Wirkung jedoch nicht von den oben genannten atypischen Antipsychotika. Hoch dosiertes Haloperidol (>4.5 mg pro Tag)

dagegen war in einer Studie mit dem vermehrten Auftreten extrapyramidal-motorischer Störungen assoziiert⁶³. Overshott et al. untersuchten den Effekt des Acetylcholinesterasehemmers Donepezil auf die Dauer eines Delirs, hier konnte keine positive Beeinflussung festgestellt werden⁶⁴.

1.8 Prävention

Studien zur primären Prävention eines Delirs sind relativ rar.

Inouye et al. konnten zeigen, dass ein Präventionsprogramm, welches ein Managementprotokoll für sechs Risikofaktoren (Einschränkungen der Kognition, des Hörens und des Sehen, Immobilität, Schlafentzug und Dehydratation) bereitstellt, die Anzahl und Dauer deliranter Episoden unter den Patienten senkt⁶⁶. Es hatte jedoch keinen Einfluss auf die Schwere oder die Rezidivraten einmal aufgetretener Delirien, was wiederum ein Indikator dafür sein könnte, dass die Primärprävention die effektivste Strategie zur Bekämpfung des postoperativen Delirs ist.

In einer neueren Metaanalyse wurden sechs Studien eingeschlossen, von denen allerdings nur eine die nötige Power besaß, um signifikante Unterschiede zum primären Endpunkt der Studie nachzuweisen. Dabei wurde eine Reduktion der Delir-Inzidenz in der Interventionsgruppe belegt. Die Intervention (in Form forcierter gerontologischer Konsile) war besonders erfolgreich in der Verhinderung schwerer Fälle, während kein Effekt auf die Dauer der einzelnen Episoden, die Länge des Krankenhausaufenthaltes, die Einweisung in ein Pflegeheim oder die kognitiven Einschränkungen nachgewiesen werden konnte⁶⁷.

Eine andere Studie untersuchte die Prophylaxe mit niedrig dosiertem Haloperidol, welche die Delir-Inzidenz nicht senken konnte, wohl aber die Schwere und Dauer der Episoden sowie die Länge des Krankenhausaufenthaltes⁶⁷.

Eine nicht randomisierte Studie wies darauf hin, dass eine Reduktion der unter 1.4.1 und 1.4.2 genannten Risikofaktoren z.B. durch eine Verbesserung der Umgebungsbedingungen auf Station, eine individuelle Rehabilitation oder eine aktive Ernährung bei Patienten nach Oberschenkelhalsbruch sowohl die Inzidenz von Delirien als auch die Komplikationsrate verringern konnte⁶⁸. Allerdings reicht eine bloße Einführung so genannter *Good Practice Guidelines* nicht aus⁶⁹, nur die nötige Therapietreue führt zur Senkung von Inzidenz und Dauer postoperativer Delirien⁷⁰. Ein Beispiel hierfür ist das *Yale Delirium Prevention Model of Care*⁶⁷, welches als *Hospital Elder Life Program*

Einleitung

(HELP) ⁷¹ in drei Ländern in die klinische Routine integriert wurde und sowohl die Delir-Inzidenz als auch die Kosten senken konnte ⁷².

Allgemein gilt es als wünschenswert, die unter 1.4.2 genannten krankenhausspezifischen Faktoren (Langzeitaufenthalte, Aufenthalt auf Intensivstationen, Anzahl der Verlegungen, Fehlen von Lesehilfen, Fehlen von Familienanschluss) weitestgehend zu vermeiden ⁷³. Der Umstand, dass eine frühe Verlegung in eine ambulante Rehabilitation die Inzidenz von Delirien signifikant senken konnte ⁷⁴, unterstreicht nochmals den Einfluss krankenhausspezifischer Prozesse auf die Entstehung eines Delirs.

2 Fragestellung

Die Prävalenz- und Inzidenzraten für ein Delir sind bei chirurgischen Patienten sehr hoch.

Dabei ist ein sich postoperativ entwickelndes Delir oft das erste erkennbare Zeichen einer ernsthaften Erkrankung oder Medikamentenintoxikation und sollte daher schnellstmöglich diagnostiziert werden, um die zugrunde liegende Ursache rascher behandeln zu können⁴. Dennoch wird in bis zu einem Drittel der Fälle das Delir von den behandelnden Ärzten nicht erkannt⁷⁵. In der Konsequenz erhöhen sich die Komplikationsraten sowie die Länge des Krankenhausaufenthaltes um das bis zu 3-Fache und die Mortalität steigt auf bis zu 65%^{12, 38, 76}. Interventionsprogramme konnten zeigen, dass die Implementierung eines Bewertungs- und Management-Protokolls die Prävalenz von Delirien sowie die Länge des Krankenhausaufenthaltes signifikant senken konnte^{55, 58}.

Die Einführung eines verlässlichen Delirium-Screening-Instruments für die routinemäßige postoperative Überwachung auf der chirurgischen Station wäre demnach wünschenswert, da sie eine deutlich höhere Effektivität verspricht als eine zufällige und somit wenig sensitive Diagnosestellung.

Der derzeitige Goldstandard zur Diagnose eines Delirs sind die DSM-IV-Kriterien, allerdings eignen diese sich aufgrund ihrer Komplexität nicht zum Einsatz im klinischen Alltag. In der Vergangenheit wurden daher verschiedene Testmethoden entwickelt, die leichter anzuwenden sind und sich überdies in unterschiedlichen Patientenkollektiven als valide erwiesen haben. Das Ziel dieser Studie war es, die Validität dreier dieser Screening-Methoden (Nu-DESC, DDS und CAM) unter den Bedingungen der chirurgischen Station zu vergleichen. Davon werden in der vorliegenden Dissertation nur die Ergebnisse für die Nu-DESC und die CAM beschrieben.

Die Nu-DESC ist speziell für die Anwendung durch das Pflegepersonal konzipiert worden und könnte sich daher besonders zur Integration in den Stationsalltag eignen. Bisher wurde ihre Validität jedoch nur im klinischen Setting der onkologischen bzw. geriatrischen Station überprüft, eine Validierung für die Chirurgie steht bisher noch aus.

Die CAM ist eine vielfach validierte Methode, die allerdings noch nicht auf ihre tägliche Reproduzierbarkeit und Eignung zur wiederholten Anwendung in der klinischen Routine der peripheren chirurgischen Station hin getestet wurde.

3 Methodik

3.1 Studiendesign

Dieser prospektiven Observationsstudie lag das Ziel zugrunde, die Effektivität verschiedener Delirium-Screening-Scores im Verlauf der postoperativen Überwachung auf der peripheren chirurgischen Station zu vergleichen. Sie wurde durch den Ethikrat der *Charité - Universitätsmedizin Berlin* genehmigt. Die Studie wurde zwischen Ende Mai 2007 und Anfang Mai 2008 am Campus des Virchow Klinikums Berlin durchgeführt.

Die Patienten wurden nach der präoperativen Anästhesievisite anhand des präoperativen Anästhesieprotokolls auf ihre Eignung zum Einschluss in die Studie hin überprüft. Geeignete Patienten wurden mindestens zwölf Stunden vor der geplanten Operation hinsichtlich ihrer Bereitschaft zur Teilnahme befragt und aufgeklärt. Eine schriftliche Teilnahmebestätigung wurde eingeholt.

Die Daten wurden von geschulten Mitarbeitern der Anästhesie erhoben. In der Anwendung der verschiedenen Diagnostikinstrumente wurden diese von einem Psychiater angeleitet und betreut.

Pflegekräfte und ärztliches Personal der chirurgischen Stationen waren gegenüber den Ergebnissen des Delir-Scorings geblindet.

3.2 Studienpopulation

Die Patienten wurden in die Studie aufgenommen, sofern alle Einschluss- und keine der Ausschlusskriterien erfüllt waren. Ausreichende Sprachkenntnisse wurden während des Aufklärungsgesprächs überprüft.

3.2.1 Einschlusskriterien

Die Einschlusskriterien lauteten:

- Aufnahme ins Virchow Klinikum Berlin
- Aufnahme im Zeitraum Ende Mai 2007 bis Anfang Mai 2008
- Aufnahme zur elektiven Operation in den Bereichen
 1. Allgemeinchirurgie (ACHI)
 2. Gynäkologie (GYN)
 3. Hals-Nasen-Ohren-Heilkunde (HNO)

4. Orthopädie (ORTHO)

- Präoperative Anästhesievisite bis spätestens 18 Uhr des Vortages
- American Society of Anesthesiologists Physical Status (ASA-PS) von mindestens III (entspricht schwerer, aber kompensierter Komorbidität)
- Geplante Operationsdauer >90 Minuten
- Stark invasiver Charakter der Operation:
 1. ACHI: z.B. Leberteilresektionen, Pyloruserhaltende Pankreatico-Duodenostomie (PPPD), Hemicolektomien, große Gefäßchirurgie oder Vergleichbares
 2. HNO: Neck Dissection
 3. GYN: Tumorresektion mittels Laparotomie
 4. ORTHO: Hüft-TEPs
- Alter über 18 Jahre
- Schriftliche Zustimmung zur Teilnahme
- Ausreichende Deutschkenntnisse

3.2.2 Ausschlusskriterien

Die Ausschlusskriterien waren:

- neurologische Vorerkrankungen
- psychiatrische Vorerkrankungen
- stattgehabter zerebraler Insult in der Vorgeschichte
- Substanzmissbrauch in der Vorgeschichte, inklusive
 1. Alkohol
 2. Medikamenten
 3. anderen Drogen
- Mangelnde Deutschkenntnisse
- Teilnahme an anderen Studien
- Mitarbeiter der Charité.

3.3 Erhobene Daten

3.3.1 Basisdaten

Folgende Daten wurden aus dem präoperativen Anästhesieprotokoll, dem Operationsprotokoll, dem Aufwachraumprotokoll sowie der Patientenakte entnommen:

- Alter
- Geschlecht
- ASA-PS Klassifikation
- Richmond Agitation Sedation Scale (RASS)
- Vorliegende Grunderkrankungen
- Dauer der Operation
- Dauer der Narkose
- Art der Narkose
- Operationsgebiet
- Länge des Aufenthaltes auf der Intensivstation
- Postoperative Liegedauer.

3.3.2 Delirdiagnostik

Zum Ausschluss grober kognitiver Einschränkungen wurde mit allen teilnehmenden Patienten zunächst einmalig am präoperativen Tag eine Untersuchung mit der *Mini Mental State Examination* (MMSE) durchgeführt. Diese galt bis zu einem Cut-Off von 24 Punkten als unauffällig.

Nach bestandener MMSE wurden die Patienten vom präoperativen Tag an täglich ein Mal zwischen 16 und 20 Uhr auf das Auftreten eines Delirs hin untersucht. Die Datenerhebung endete am sechsten postoperativen Tag, bei Verlegung in andere Krankenhäuser oder bei erneuter Operation wurde die Datenerhebung vorzeitig abgebrochen.

Die Diagnose eines Delirs wurde mit Hilfe der DSM-IV Kriterien gestellt. Außerdem wurden die Patienten mit der Nu-DESC, der CAM und der DDS getestet, wovon in der vorliegenden Arbeit nur die Ergebnisse der Nu-DESC und der CAM dargelegt werden. Die Ärzte und das Pflegepersonal der jeweiligen Stationen hatten keinen Zugriff auf die Resultate der Screeningtests.

3.3.2.1 DSM-IV

Die DSM-IV Kriterien fassen die Hauptmerkmale eines Delirs in vier Punkten zusammen (Abbildung 1). Die einzelnen Eigenschaften werden nicht wie beim DDS oder Nu-DESC bepunktet, sondern müssen alle vorhanden sein. Ein Delir liegt also nur vor, wenn alle vier Kriterien erfüllt sind. Unter 4.3.2.1 ist aufgeschlüsselt, wie die Untersucher dieser Studie über das Vorliegen oder Fehlen der einzelnen Eigenschaften entschieden haben.

A	Bewusstseinsstörung (d.h. reduzierte Klarheit der Umgebungswahrnehmung) mit eingeschränkter Fähigkeit, die Aufmerksamkeit zu richten, aufrecht zu erhalten oder zu verlagern.
B	Veränderung kognitiver Funktionen (Störungen des Gedächtnisses, der Orientierung und Sprache) oder Entwicklung einer Wahrnehmungsstörung, die nicht besser durch eine bereits bestehende oder sich entwickelnde Demenz erklärt werden kann.
C	Entwicklung des Störungsbildes innerhalb einer kurzen Zeitspanne (Stunden oder Tage) und Fluktuation der Symptomatik im Tagesverlauf.
D	Hinweise aus der Anamnese, der körperlichen Untersuchung oder Laborbefunden, dass das Krankheitsbild durch einen medizinischen Krankheitsfaktor verursacht ist.

Abbildung 1: Kriterien zur Diagnose nach DSM-IV

Eine reduzierte Klarheit der Umgebungswahrnehmung kann dabei auch eine Einschränkung der Erweckbarkeit beinhalten, die allerdings nicht bis zum Stupor oder Koma reichen darf. Die Konzentrationsfähigkeit kann gestört sein und damit dazu führen, dass der Patient leicht ablenkbar ist.

Die Störungen des Gedächtnisses betreffen vorwiegend das Kurzzeitgedächtnis. Die Patienten sind häufig zu Zeit und/oder Ort sowie anderen Personen desorientiert, selten bezieht sich die Desorientiertheit auf die eigene Person. Gerade letzteren Punkt zu erfassen kann oft schwierig sein, da die Patienten meist nur unzusammenhängend kommunizieren und Fragen zur Person nicht logisch kohärent beantworten können. Daher muss bei der Diagnosestellung nach DSM-IV Kriterien zum Teil auf Angaben aus den Krankenakten, von Angehörigen oder betreuendem Personal zurückgegriffen werden.

Wahrnehmungsstörungen umfassen Fehlinterpretationen, Wahnwahrnehmungen, Wahnvorstellungen und Halluzinationen. Am häufigsten sind hierbei visuelle Illusionen und Halluzinationen, seltener leiden die Patienten auch unter olfaktorischen, gustatorischen, taktilen oder auditorischen Sinnestäuschungen⁵.

3.3.2.2 Nu-DESC

Entsprechend den Vorgaben der *International Society for Pharmacoeconomics and Outcome Research Task Force for Translation and Cultural Adaptation of Patient Related Outcomes (ISPOR PRO)*⁷⁷ wurde die Nu-DESC von Lütz et al. ins Deutsche übersetzt⁷⁸. Sie besteht aus fünf Merkmalen, die jeweils mit einem Punktwert von 0 bis 2 bewertet werden können. Die 0 steht hierbei für das Fehlen der beschriebenen Eigenschaft, die 1 für das Vorhandensein derselben und die 2 für eine besonders starke Ausprägung des Merkmals. Die Punktwerte der einzelnen Dimensionen werden aufaddiert, ab einem Wert größer 1 fällt der Test positiv aus (Abbildung 2).

	Symptome	Symptom Bewertung	
1	Desorientierung Manifestierung einer Desorientierung zu Zeit oder Ort durch Worte oder Verhalten oder Nicht-Erkennen der umgebenden Personen.	<input type="checkbox"/> 0	<input type="checkbox"/> 1 <input type="checkbox"/> 2
2	Unangemessenes Verhalten Unangemessenes Verhalten zu Ort und/oder Person: z.B. Ziehen an Kathetern oder Verbänden, Versuch aus dem Bett zu steigen, wenn es kontraindiziert ist und so weiter.	<input type="checkbox"/> 0	<input type="checkbox"/> 1 <input type="checkbox"/> 2
3	Unangemessene Kommunikation Unpassende Kommunikation zu Ort und/oder Person, zum Beispiel zusammenhanglose oder gar keine Kommunikation, unsinnige oder unverständliche sprachliche Äußerungen.	<input type="checkbox"/> 0	<input type="checkbox"/> 1 <input type="checkbox"/> 2
4	Illusionen / Halluzinationen Sehen oder Hören nicht vorhandener Dinge, Verzerrung optischer Eindrücke.	<input type="checkbox"/> 0	<input type="checkbox"/> 1 <input type="checkbox"/> 2
5	Psychomotorische Retardierung Verlangsamte Ansprechbarkeit, wenige oder keine spontane Aktivität/Äußerung, z.B. wenn der Patient angestupst wird, ist die Reaktion verzögert und/oder der Patient ist nicht richtig erweckbar.	<input type="checkbox"/> 0	<input type="checkbox"/> 1 <input type="checkbox"/> 2
Summe			
Delirium		≥ 2 <input type="checkbox"/> ja	< 2 <input type="checkbox"/> nein

Abbildung 2: Nu-DESC nach Gaudreau et al. 2005 und Lütz et al. 2008

3.3.2.3 CAM

Die zur einfacheren Anwendung verkürzte CAM (*CAM Diagnostic Algorithm*) beinhaltet vier Merkmale. Davon müssen Merkmal eins, zwei und drei zwingend vorhanden sein, um ein Delir zu diagnostizieren, das vierte Merkmal kann fakultativ vorhanden sein (Abbildung 3). Unorganisiertes Denken wird bei Auffälligkeiten mit Hilfe eines speziellen Fragenkatalogs eruiert, zu dem für gesunde Probanden normalerweise leicht zu beantwortende Fragen wie beispielsweise: „Schwimmt ein Stein auf Wasser?“ gehören.

	Symptome	Symptom Bewertung
1	Akuter Beginn oder fluktuierender Verlauf (Hat sich der mentale Status verändert?)	ja / nein
	UND	
2	Unaufmerksamkeit (Fähigkeit, Aufmerksamkeit zu richten, aufrecht zu erhalten oder zu verlagern vorhanden?)	ja / nein
	PLUS	
3	Unorganisiertes Denken	ja / nein
	ODER / UND	
4	Veränderte Bewusstseinslage (gesteigerte oder verminderte Aktivität?)	ja / nein

Abbildung 3: Die CAM nach Inouye et al. 1990 ⁴⁹

3.4 Statistik

Für alle Studienvariablen wurden deskriptive Statistiken berechnet. Kategorische Variablen wurden als Zahlen (Prozentzahlen) angegeben, kontinuierliche Variablen als Mittelwerte mit 95%-Konfidenzintervall (95%-CI) im Falle einer vorliegenden Normalverteilung oder als Median mit 25- und 75%-Perzentilen bei nicht normal verteilten Werten. Bezüglich der erfassten klinischen Parameter wurden Unterschiede zwischen den Gruppen mit dem *exakten Test nach Fisher* für Häufigkeiten bzw. dem *t-Test* für normalverteilte kontinuierliche und dem *Mann-Whitney U Test* für nicht normalverteilte kontinuierliche Werte beurteilt.

Die diagnostische Leistungsfähigkeit der Nu-DESC und der CAM wurde mit Hilfe der *Receiver Operating Characteristics (ROC)* Analyse bewertet. Hierbei wurden die Ergebnisse nach DSM-IV Kriterien als Goldstandard herangezogen.

Sensitivitäten und Spezifitäten der genannten Scores wurden mit dem *McNemar's Test* untereinander verglichen. Als Signifikanzniveau wurde dabei $p < 0.05$ festgelegt. Die ermittelten p-Werte sind als exploratorisch zu verstehen.

Methodik

Bei wiederholten Messungen an ein und demselben Patienten (sogenannte *clustered binary data* - gruppierte binäre Daten) wurde ein spezielles Testverfahren angewendet, das den Zusammenhang zwischen diesen wiederholten Messungen berücksichtigt. Die Kalkulationen von kumulativer Sensitivität und Spezifität sowie die dazugehörigen Vergleichstestungen wurden daher mit dem *weighted estimator*⁷⁹ durchgeführt.

Die Interrater-Übereinstimmung für die verschiedenen Testmethoden (Nu-DESC und CAM) wurde mit Cohens Kappa überprüft (inklusive 95%-CI) und mit Altmans Beschreibungen bewertet.

Alle Daten wurden mit dem *Statistical Package of the Social Science Software* (SPSS, Version 15.0) für *Microsoft Windows* berechnet.

4 Ergebnisse

4.1 Studienpopulation

Es wurden 116 Patienten, die gemäß Ein- und Ausschlusskriterien geeignet waren, in die Studie aufgenommen. Von diesen mussten 28 im weiteren Verlauf ausgeschlossen werden. 2 Probanden hatten entgegen vorher zugänglichen Informationen doch einen zerebralen Insult in der Vorgeschichte, ein Patient erwies sich als neurochirurgisch und bei 25 Patienten waren die Daten unvollständig oder nicht verfügbar. Der Ablauf des Screenings inklusive der jeweiligen Patientenzahlen ist im Folgenden in einem Flussdiagramm entsprechend den *Standards for the Reporting of Diagnostic Accuracy Studies (STARD)* dargestellt (Abbildung 4).

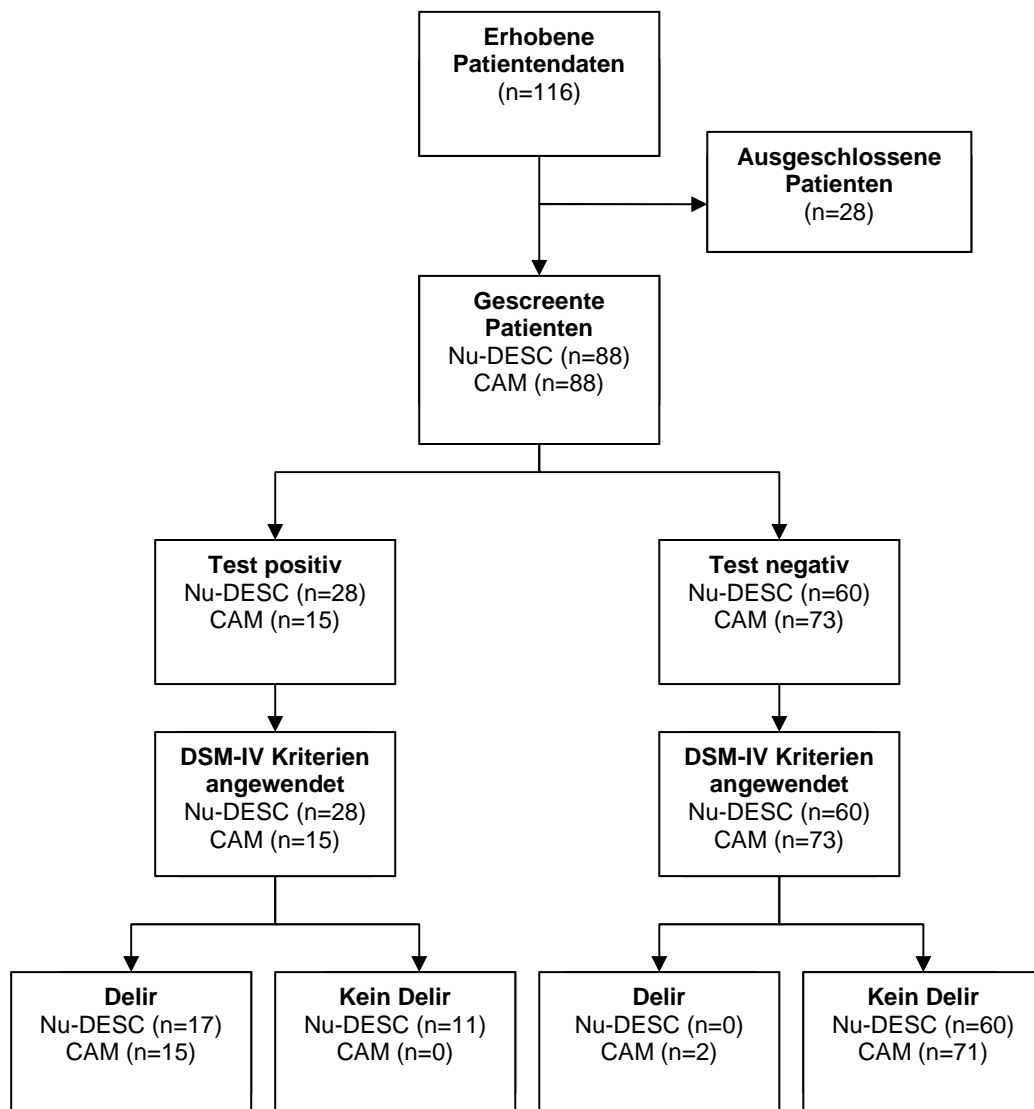


Abbildung 4: Flussdiagramm des Studienverlaufs gemäß STARD (Bossuyt et al. 2003)

4.2 Deskriptive Statistik

4.2.1 Basisdaten der Gesamtpopulation

Insgesamt wurden mehr Männer als Frauen in die Studie eingeschlossen. Alle gehörten der ASA-PS Gruppe III an. Das Alter der Probanden rangierte dabei zwischen 30 und 88 Jahren. Bei dem Großteil der Patienten erfolgte ein intra-abdomineller Eingriff auf der Allgemein Chirurgie. Die Mehrzahl der Narkosen wurde hierbei mit volatilen Anästhetika aufrechterhalten (Tabellen 1-3).

Tabelle 1: präoperative Basisdaten

<u>Geschlecht</u>		
Anzahl Männer (% in Klammern)	57	(64,8)
Anzahl Frauen (% in Klammern)	31	(35,2)
Mittleres Alter [Jahre], (95%-CI in Klammern)	65,4	(63,4 - 67,5)
Anzahl ASA-PS III (% in Klammern)	88	(100)
<u>Behandelnde Fachabteilung</u>		
Anzahl Allgemein Chirurgie (% in Klammern)	77	(87,5)
Anzahl Gynäkologie	5	(5,7)
Anzahl andere Fachabteilungen	6	(6,8)

Tabelle 2: Basisdaten zu Operation und Narkose

Mittelwert Operationsdauer [min], (95%-CI in Klammern)	251	(228 - 275)
Mittelwert Narkosedauer [min], (95%-CI in Klammern)	355	(329 - 380)
<u>Aufrechterhaltung der Narkose</u>		
Anzahl volatile Anästhetika (% in Klammern)	76	(86,4)
Anzahl Propofol (% in Klammern)	12	(13,6)
<u>Operationsgebiet</u>		
Intraabdominell (% in Klammern)	80	(90,9)
Gefäße	3	(3,4)
Urogenital	3	(3,4)
Hals/Kopf (extrakraniell)	1	(1,1)
Muskulo-skelettales System	1	(1,1)

Ergebnisse

Tabelle 3: postoperative Basisdaten

Mediane postoperative Verweildauer im Krankenhaus [Tage], (25%-und 75% Perzentile in Klammern)	14	(11 - 23)
Mediane postoperative Verweildauer auf der <i>Intensive Care Unit</i> (ICU) [Tage], (25%-und 75% Perzentile in Klammern)	2	(1 - 3)

4.2.2 Basisdaten deliranter und nicht deliranter Patienten im Vergleich

Sowohl beim Alter als auch beim Geschlecht sowie bezüglich der Art der Narkose und der Dauer der Operation konnten keine signifikanten Unterschiede zwischen deliranten und nicht deliranten Patienten gefunden werden. Allerdings wiesen Probanden, die nach DSM-IV Kriterien im Laufe der sechs beobachteten postoperativen Tage ein Delir entwickelten, eine signifikant längere Liegedauer auf der ICU und insgesamt postoperativ auf. Außerdem dauerte bei ihnen die Narkose signifikant länger (Tabelle 4).

Tabelle 4: Vergleich Basisdaten Patienten mit und ohne Delir nach DSM-IV

	Delirium – DSM IV (n=17)	kein Delirium – DSM IV (n=71)	p-Wert
Alter [Jahre] ²	68,7 (63,4 – 74,0)	64,7 (62,4 – 66,9)	0,127
Anzahl weibliches Geschlecht (%) ³	5 (5,7)	26 (29,5)	0,778
Operationsdauer [min] ²	292 (247 – 337)	240 (213 – 267)	0,078
Narkosedauer [min] ²	429 (365 – 493)	341 (313 – 368)	0,006
Art der Narkose ³			
Volatile (%)	14 (15,9)	62 (70,5)	0,332
Total intravenöse Anästhesie (TIVA)	2 (2,3)	4 (4,5)	
Postoperative Verweildauer ICU [Tage] ¹	2,5 (1,0 – 5,0)	2,0 (0,25 – 2,75)	0,017
Postoperative Verweildauer Krankenhaus [Tage] ¹	20,0 (14,0 – 46,0)	14,0 (10,25 – 21,0)	0,006

¹ Median, 25- und 75%-Perzentile in Klammern, p-Werte für Mann-Whitney-U-Test

² Mittelwerte, 95%-CI in Klammern, p-Werte für t-Test

³ Anzahl, Prozentualer Anteil in Klammern, p-Werte für Exakten Test nach Fisher

Ergebnisse

Um die Vorhersagekraft der verschiedenen Testmethoden bezüglich des Outcomes in den Vergleich zueinander zu setzen, wurden die postoperative Verweildauern auf der ICU sowie im Krankenhaus insgesamt für die verschiedenen Testmethoden gegenüber gestellt (Tabelle 5).

Tabelle 5: Delirium-Score und Outcome, Median, 25- und 75% Perzentile in Klammern, p-Werte für Mann-Whitney-U-Test

Test	Aufenthaltsdauer ICU (Tage)			Postoperative Krankenhausverweildauer (Tage)		
	Delirium	Kein Delirium	p	Delirium	Kein Delirium	p
DSM-IV	2,5 (1,0 - 5,0)	2,0 (0,25 - 2,75)	0,017	20,0 (14,0 - 46,0)	14,0 (10,25 - 21,0)	0,006
CAM	3,0 (1,5 - 5,5)	2,0 (0,75 - 2,25)	0,009	20,0 (14,0 - 43,0)	14,0 (10,75 - 21,25)	0,020
Nu-DESC	3,0 (1,0 - 5,0)	1,0 (0,0 - 3,0)	0,001	20,5 (14,0 - 41,5)	13,0 (10,0 - 19,0)	0,001

Dabei bestätigten sich die die Ergebnisse, die Patienten mit einem Delir nach Goldstandard zeigten. Auch Probanden, die nach CAM oder Nu-DESC delirant waren, wiesen eine längere Verweildauer auf der ICU sowie insgesamt postoperativ auf. Die stärksten Signifikanzen ließen sich hierbei bei der Nu-DESC nachweisen.

4.3 Delir-Testung

Im Folgenden werden die Ergebnisse der Testungen mit den DSM-IV Kriterien, der CAM und der Nu-DESC sowohl in Einzelfaktoren als auch im Gesamtergebnis wiedergegeben. Inzidenz und Dauer eines Delirs wurden dabei sowohl für die Anzahl der Patienten als auch für die Anzahl der beobachteten Patiententage insgesamt berechnet.

4.3.1 Inzidenz und Dauer des postoperativen Delirs auf der peripheren Station

Nach den Kriterien des Goldstandards (DSM-IV Kriterien) entwickelten 17 Patienten innerhalb des beobachteten Zeitraums ein Delir. Die CAM detektierte 15 Patienten, die Nu-DESC 28. Die durchschnittliche Dauer eines einmal aufgetretenen Delirs betrug zwei Tage nach DSM-IV Kriterien und Nu-DESC sowie einen Tag nach CAM (Tabellen 6-8).

Tabelle 6: Anzahl der Patienten CAM versus DSM-IV Kriterien

CAM	DSM-IV		Gesamt
	Delir	Kein Delir	
Delir	15	0	15
Kein Delir	2	71	73
Gesamt	17	71	88

Tabelle 7: Anzahl der Patienten Nu-DESC versus DSM-IV Kriterien

Nu-DESC	DSM-IV		Gesamt
	Delir	Kein Delir	
Delir	17	11	28
Kein Delir	0	60	60
Gesamt	17	71	88

Tabelle 8: mittlere Dauer des Delirs nach DSM-IV, CAM und Nu-DESC in Tagen, Median, 25%- und 75%-Perzentile in Klammern

	DSM-IV	CAM	Nu-DESC
Dauer Delir	2,0 (1,0 - 4,0)	1,0 (1,0 - 3,0)	2,0 (1,0 - 3,0)

Stellt man die Ergebnisse nach Patiententagen dar, hätten 88 Patienten an maximal 528 postoperativen Tagen gescreent werden können. Durch vorzeitige Entlassung von 8 Patienten wurden 503 Tage erfasst. An zwei Tagen waren Patienten aufgrund von

Ergebnisse

komatösen Zuständen (RASS < -2, entspricht fehlender gezielter Kontaktaufnahme auf Ansprache) nicht beurteilbar. Daher waren 501 Messungen der DSM-IV Kriterien, der CAM und der Nu-DESC auswertbar. Von diesen 501 Tagen wurden 39 Tage (7,8%) als Delir-positiv gemäß dem Goldstandard klassifiziert. Die CAM detektierte 27 delirante Tage (5,4%) während die Nu-DESC an insgesamt 65 Tagen (13,0%) positiv zeichnete (Tabellen 9+10).

Tabelle 9: Anzahl der Patiententage CAM versus DSM-IV Kriterien

CAM	DSM-IV		Gesamt
	Delir	Kein Delir	
Delir	27	0	27
Kein Delir	12	462	474
Gesamt	39	462	501

Tabelle 10: Anzahl der Patiententage Nu-DESC versus DSM-IV Kriterien

Nu-DESC	DSM-IV		Gesamt
	Delir	Kein Delir	
Delir	37	28	65
Kein Delir	2	434	436
Gesamt	39	462	501

4.3.2 Inzidenz von Einzelfaktoren

Nachstehend wird angegeben, wie viele der 88 Patienten bzw. der 501 gewerteten Patiententage einzelne Kriterien der DSM-IV, der CAM und der Nu-DESC an mindestens einem der sechs beobachteten postoperativen Tage erfüllten. Außerdem werden für jeden Test die Gesamtpunktzahlen bzw. die Gesamtzahl der erfüllten Kriterien zusammengefasst.

4.3.2.1 Diagnostic and Statistical Manual of Mental Disorders – DSM-IV

In der Aufstellung der Anzahl der erfüllten Kriterien wird ersichtlich, dass etwa genau so viele Patienten wie die, die ein Delir entwickelten, also alle vier Kriterien nach DSM-IV erfüllten (siehe 4.3.1: 17 Patienten), innerhalb der sechs postoperativen Tage zumindest an einem Tag mit drei der unten genannten Kriterien des DSM-IV Katalogs auffällig waren. Am häufigsten wurden Patienten mit der akuten Entwicklung von Bewusstseinsstörungen beobachtet (Tabelle 11).

Die stattgehabte Operation wurde aufgrund potenzieller Störungen u.a. des Elektrolyt-Haushalts, Säurebasen-Haushalts oder des neuronalen Transmitterstoffwechsels als medizinischer Krankheitsfaktor gewertet. Da daher definitionsgemäß alle Patienten das Merkmal „medizinischer Krankheitsfaktor vorhanden“ erfüllten, wird auf die gesonderte Darstellung dieses vierten Kriteriums in den Tabellen im weiteren Verlauf dieser Arbeit verzichtet.

Tabelle 11: DSM-IV – Erfüllungsgrad der Einzelkriterien Bewusstseinsstörung, Veränderung kognitiver Funktionen und akute Entwicklung und Fluktuation in Anzahl der Patienten, Prozentangaben in Klammern. 0 = nicht erfüllt, 1 = erfüllt

	Anzahl der Patienten (% in Klammern)		
	Bewusstseinsstörung	Veränderung kognitiver Funktionen	akute Entwicklung und Fluktuation
0	63 (71,6)	61 (69,3)	53 (60,2)
1	25 (28,4)	27 (30,7)	35 (39,8)
Gesamt	88 (100)	88 (100)	88 (100)

In Tabelle 12 wurde die im Verlauf der sechs postoperativen Tage an mindestens einem Tag maximal erreichte Anzahl der Faktoren zusammengefasst.

Tabelle 12: An mindestens einem Tag maximal erreichte Anzahl der Faktoren im DSM-IV, Anzahl der Patienten (% in Klammern)

Maximale Punktzahl im DSM-IV	Anzahl der Patienten (% in Klammern)
0	0 (0,0)
1	53 (60,2)
2	0 (0,0)
3	18 (20,5)
4	17 (19,3)

Ergebnisse

Dargestellt nach einzelnen Patiententagen war der akute Beginn von Symptomen mit 84 positiv gemessenen Tagen das am häufigsten auffällige DSM-IV Kriterium, gefolgt von den Veränderungen der kognitiven Funktion. Bewusstseinsstörungen wurden am seltensten gesehen. Auch hier glich die Anzahl der Tage, an denen drei Kriterien der DSM-IV erfüllt waren, denen, an denen ein volles Delir vorlag (Tabellen 13+14).

Tabelle 13 DSM-IV – Erfüllungsgrad der Einzelkriterien Bewusstseinsstörung, Veränderung kognitiver Funktionen und akute Entwicklung und Fluktuation in Anzahl der Patiententage, Prozentangaben in Klammern. 0 = nicht erfüllt, 1 = erfüllt

	Anzahl der Patiententage (% in Klammern)		
	Bewusstseinsstörung	Veränderung kognitiver Funktionen	akute Entwicklung und Fluktuation
0	444 (88,6)	434 (86,6)	418 (83,4)
1	57 (11,4)	67 (13,4)	83 (16,6)
Gesamt	501 (100)	501 (100)	501 (100)

Tabelle 14: An mindestens einem Tag maximal erfüllte Anzahl der Faktoren im DSM-IV, Anzahl der Patiententage (% in Klammern)

1	415 (82,8)
2	4 (0,8)
3	43 (8,6)
4	39 (7,8)
Gesamt	501 (100)

4.3.2.2 Confusion Assessment Method - CAM

In der Einzelauswertung der CAM ist ersichtlich, dass die weitaus meisten Patienten Veränderungen des Bewusstseins zeigten und einen akuten Beginn der erfassten Symptome aufwiesen. Unaufmerksamkeit und unorganisiertes Denken waren weniger häufig anzutreffen (Tabelle 15).

Ergebnisse

Tabelle 15: CAM – Erfüllungsgrad der Einzelkriterien akuter Beginn, Unaufmerksamkeit, unorganisiertes Denken und veränderte Bewusstseinslage in Anzahl der Patienten, Prozentangaben in Klammern. 0 = nicht erfüllt, 1 = erfüllt

	Anzahl der Patienten (% in Klammern)			
	Akuter Beginn	Unaufmerksamkeit	Unorganisiertes Denken	Veränderte Bewusstseinslage
0	48 (54,5)	73 (82,9)	73 (82,9)	47 (53,4)
1	40 (45,5)	15 (17,1)	15 (17,1)	41 (46,6)
Gesamt	88 (100)	88 (100)	88 (100)	88 (100)

Auch in der Aufschlüsselung der einzelnen Tage zeigt sich, dass eine Veränderung der Bewusstseinslage das häufigste durch die CAM erfasste Symptom war. Ebenso oft wurde ein akuter Beginn festgestellt (Tabelle 16).

Tabelle 16: CAM – Erfüllungsgrad der Einzelkriterien akuter Beginn, Unaufmerksamkeit, unorganisiertes Denken und veränderte Bewusstseinslage in Anzahl der Patiententage, Prozentangaben in Klammern. 0 = nicht erfüllt, 1 = erfüllt

	Anzahl der Patiententage (% in Klammern)			
	Akuter Beginn	Unaufmerksamkeit	Unorganisiertes Denken	Veränderte Bewusstseinslage
0	392 (78,2)	473 (94,4)	475 (94,8)	390 (77,8)
1	109 (21,8)	28 (5,6)	26 (5,2)	111 (22,2)
Gesamt	501 (100)	501 (100)	501 (100)	501 (100)

4.3.2.3 Nursing Delirium Screening Scale - Nu-DESC

Mehr als die Hälfte aller Patienten zeigte im postoperativen Verlauf eine mindestens milde psychomotorische Retardierung, bei fast 10% war diese sogar stark ausgeprägt. Das zweithäufigste durch die Nu-DESC detektierte Symptom war eine moderate Desorientierung (Tabelle 17).

Ergebnisse

Tabelle 17: Nu-DESC – Erfüllungsgrad der Einzelkriterien Desorientierung, unangemessenes Verhalten, unangemessene Kommunikation, Illusion/Halluzination und psychomotorische Retardierung in Anzahl der Patienten, Prozentangaben in Klammern. 0 = nicht erfüllt, 1 = leichte Ausprägung, 2 = schwere Ausprägung

	Anzahl der Patienten (% in Klammern)				
	Desorientierung	Unangemessenes Verhalten	Unangemessene Kommunikation	Illusion / Halluzination	Psycho-motorische Retardierung
0	65 (73,9)	77 (87,5)	67 (76,1)	72 (81,8)	33 (37,5)
1	20 (22,7)	10 (11,4)	14 (15,9)	15 (17,1)	47 (53,4)
2	3 (3,4)	1 (1,1)	7 (8,0)	1 (1,1)	8 (9,1)
Gesamt	88 (100)	88 (100)	88 (100)	88 (100)	88 (100)

Tabelle 18 fasst die im Verlauf der sechs postoperativen Tage an mindestens einem Tag maximal erreichte Punktzahl zusammen.

Tabelle 18: An mindestens einem Tag maximal erreichte Punktzahl im Nu-DESC, Anzahl der Patienten (% in Klammern)

Maximale Punktzahl im Nu-DESC	Anzahl der Patienten (% in Klammern)
0	27 (30,7)
1	34 (38,6)
2	11 (12,5)
3	5 (5,7)
4	4 (4,6)
5	2 (2,3)
6	3 (3,4)
7	1 (1,1)
8	1 (1,1)
9	0 (0,0)
10	0 (0,0)

Bezogen auf die 501 Einzelmessungen wurden die Patienten an mehr als 30% der postoperativen Tage mit einer leichten bis schweren psychomotorischen Retardierung auffällig. Mit großem Abstand war das in der Häufigkeit des Auftretens folgende Symptom eine milde Desorientierung. Am seltensten wurde ein unangemessenes Verhalten beobachtet (Tabellen 19+20).

Ergebnisse

Tabelle 19: Nu-DESC – Erfüllungsgrad der Einzelkriterien Desorientierung, unangemessenes Verhalten, unangemessene Kommunikation, Illusion/Halluzination und psychomotorische Retardierung in Anzahl der Patiententage, Prozentangaben in Klammern. 0 = nicht erfüllt, 1 = leichte Ausprägung, 2 = schwere Ausprägung

	Anzahl der Patiententage (% in Klammern)				
	Desorientierung	Unangemessenes Verhalten	Unangemessene Kommunikation	Illusion / Halluzination	Psycho-motorische Retardierung
0	457 (91,2)	488 (97,4)	457 (91,2)	466 (93,0)	347 (69,4)
1	41 (8,2)	12 (2,4)	37 (7,4)	34 (6,8)	141 (28,2)
2	3 (0,6)	1 (0,2)	7 (1,4)	1 (0,2)	12 (2,4)
Gesamt	501 (100)	501 (100)	501 (100)	501 (100)	501 (100)

Tabelle 20 fasst die erreichte Punktzahl der Nu-DESC an den einzelnen Patiententagen zusammen.

Tabelle 20: Erreichte Punktzahl im Nu-DESC, Anzahl der Patiententage (% in Klammern)

Maximale Punktzahl im Nu-DESC	Anzahl der Patiententage (% in Klammern)
0	321 (64,1)
1	115 (23,0)
2	33 (6,6)
3	13 (2,6)
4	10 (2,0)
5	4 (0,8)
6	3 (0,6)
7	1 (0,2)
8	1 (0,2)
9	0 (0,0)
10	0 (0,0)

4.4 Testgüte

Zur Beschreibung der Güte eines Tests werden Begriffe wie Sensitivität, Spezifität, Falschpositiv- und Falschnegativrate sowie Relevanz und Segreganz genutzt. Deren Bedeutung sowie die jeweiligen Ergebnisse werden nachfolgend dargestellt. Dabei werden Sensitivität und Spezifität wegen ihrer besonderen Bedeutung für die Testgüte in ihrem Verlauf getrennt nach Patiententagen betrachtet. Der p-Wert für McNemar in Tabelle 23 sagt aus, ob sich die Werte für die durchschnittliche Sensitivität und Spezifität von Nu-DESC und CAM signifikant unterscheiden. Für alle anderen Kenngrößen wurden die Ergebnisse zusammengefasst über alle Patiententage beschrieben.

Ergebnisse

Tabelle 21: Berechnung der Hauptmerkmale von Screening-Tests; a, b, c und d entsprechen den jeweiligen Zellen in der Kontingenztafel mit a – DSM-IV positiv und Test positiv, b – DSM-IV negativ und Test positiv, c – DSM-IV positiv und Test negativ, d – DSM-IV negativ und Test negativ (siehe Tab. 9+10)

Sensitivität	$a / (a + c)$
Spezifität	$d / (b + d)$
Falsch-Positiv-Rate (FPR)	1 - Spezifität
Falsch-Negativ-Rate (FNR)	1 - Sensitivität
Positiver Prädiktiver Wert (PPW)	$a / (a + b)$
Negativer Prädiktiver Wert (NPW)	$d / (c + d)$
Likelihood Ratio für positiven Test (LR+)	Sensitivität / (1 - Spezifität)
Likelihood Ration für negativen Test (LR-)	(1 - Sensitivität) / Spezifität
Prä-Test-Wahrscheinlichkeit (Prävalenz)	$(a + c) / (a + b + c + d)$
Prä-Test-Chance (pre-test-odds)	Prävalenz / (1- Prävalenz)
Post-Test-Chance (post-test-odds)	Prä-Test-Chance * LR+
Post-Test-Wahrscheinlichkeit	Post-Test-Chance / (Post-Test-Chance - 1)
Genauigkeit (Accuracy)	$a + d / (a + b + c + d)$

4.4.1 Sensitivität

Die Sensitivität beschreibt, wie viele der nach einem Goldstandard in einer Population als krank definierten Individuen ein bestimmter Test erkennt.

Die CAM zeigte in dieser Studie eine durchschnittliche Sensitivität von 74.9% (95%-CI: 57.7% - 92.2%) (Tabelle 23). Sie detektierte 27 der 39 nach Goldstandard deliranten Patiententage und erreichte ihre maximale Sensitivität von 90% am dritten postoperativen Tag. Am sechsten postoperativen Tag betrug die Sensitivität der CAM 0%. Gemäß DSM-IV Kriterien war an diesem Tag noch ein Patient delirant.

Verglichen mit der CAM war die Nu-DESC der sensitivere Test ($p=0.013$), sie bewies mit 37 von 39 korrekt klassifizierten Patiententagen eine mittlere Sensitivität von 97.7% (95%-CI: 93.0% - 102.3%) (Tabelle 23). Abbildung 5 veranschaulicht den Verlauf der Sensitivitäten beider Testmethoden über den Zeitraum der beobachteten sechs postoperativen Tage.

Ergebnisse

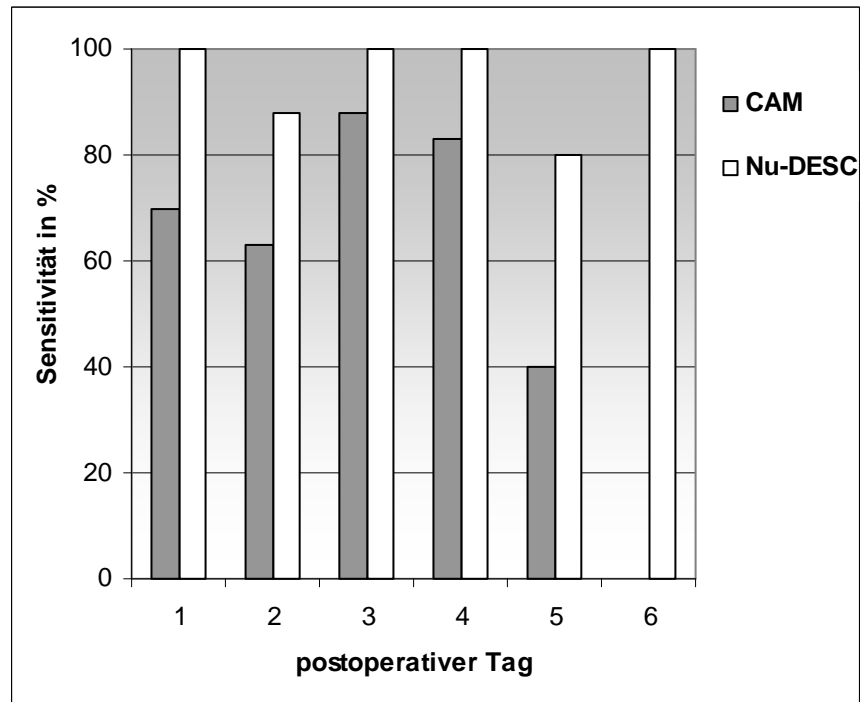


Abbildung 5: Sensitivitäten über sechs postoperative Tage

4.4.2 Spezifität

Die Spezifität ist ein Maß dafür, wie viele der in einer Population als gesund definierten Individuen ein Test richtig als gesund klassifiziert.

In der vorliegenden Studie bewies die CAM verglichen zur Nu-DESC die höchste Spezifität von 100% ($p < 0.001$) (Tabelle 23). Sie erkannte alle nach DSM-IV Kriterien gesunden Patienten an allen Tagen richtig als gesund.

Die Nu-DESC dagegen variierte in ihrer Spezifität im Verlauf zwischen 90% und 97%, insgesamt lag sie durchschnittlich bei 92.3% (Tabelle 23). Abbildung 6 stellt die Spezifität getrennt nach beobachteten Tagen dar.

Ergebnisse

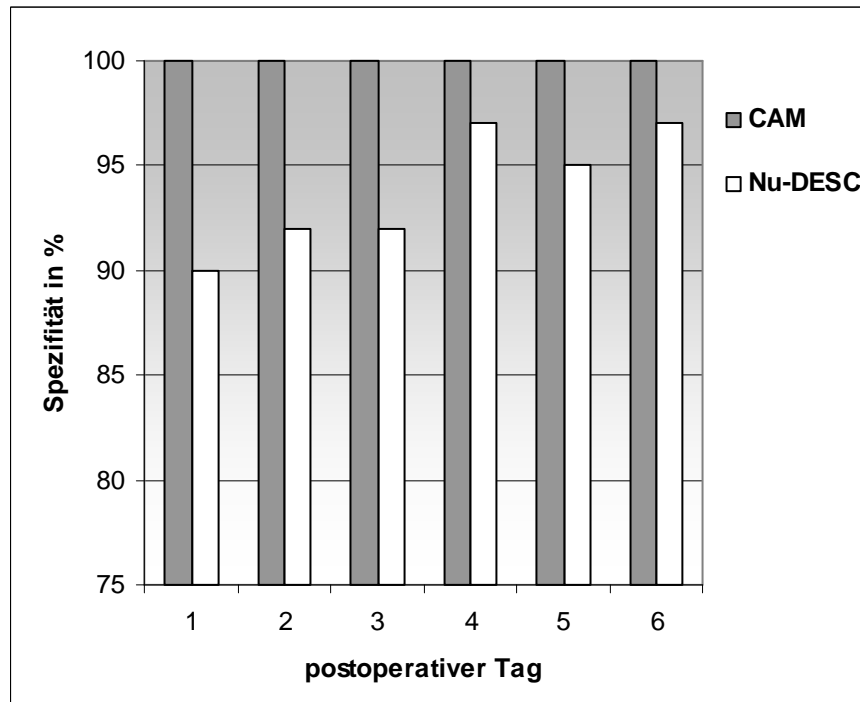


Abbildung 6: Spezifitäten über sechs postoperative Tage

4.4.3 Falsch-Positiv-Rate

Die Falsch-Positiv-Rate gibt an, wie häufig ein positiv getesteter Patient nach dem Goldstandard nicht krank ist. Die Rate lag für die CAM bei 0%, bei der Nu-DESC dagegen bei 7,7% (Tabelle 23).

Allerdings gingen 21,4% der 28 nach Nu-DESC falsch positiv detektierten Tage nach ein oder zwei Tagen in ein Delir nach DSM-IV Kriterien über. Weiteren 25% der falsch positiven Nu-DESC Tage ging ein Delir nach DSM-IV voraus (Tabelle 22).

Tabelle 22: Anzahl der Patiententage mit Delir (DSM-IV) zeitlich vor oder nach einem falsch positiven Testergebnis bei Anwendung der Nu-DESC

	vorausgegangenes Delir vor		nachfolgendes Delir nach		kein Übergang	Gesamt
	einem Tag	zwei Tagen	einem Tag	zwei Tagen		
Anzahl Patiententage	6	1	3	3	15	28

4.4.4 Falsch-Negativ-Rate

Die Falsch-Negativ-Rate gibt die Prozentzahl der Patienten an, die trotz negativen Testergebnisses nach Goldstandard krank sind. Die Falsch-Negativ-Rate der CAM war 26,1%, die der Nu-DESC 2,3% (Tabelle 23).

4.4.5 Positiver Prädiktiver Wert (PPW)

Der Positive Prädiktive Wert gibt an, mit welcher Wahrscheinlichkeit ein Test-Positiver auch tatsächlich krank ist. Der CAM wies einen PPW von 100% auf, die Nu-DESC einen von 56,9% (Tabelle 23).

4.4.6 Negativer Prädiktiver Wert (NPW)

Der Negative Prädiktive Wert, auch Segreganz genannt, ist ein Maß für die Wahrscheinlichkeit, dass ein negativ getesteter Proband auch wirklich gesund ist. Der NPW der CAM war 97,5%, der NPW der Nu-DESC betrug 99,5% (Tabelle 23).

4.4.7 Likelihood Ratio für einen positiven Test (LR+)

Sie gibt an, wie viel Mal häufiger ein positives Testresultat bei Personen mit Erkrankung im Vergleich zu Personen ohne Erkrankung vorkommt. Ein Wert von 1 bedeutet, dass die Wahrscheinlichkeit für ein positives Testergebnis bei Gesunden und Kranken gleich groß ist und der Test somit keine Aussagekraft hat. Ein Wert größer 3 gilt dagegen als akzeptabel, ein Wert größer 10 als gut. Je größer die LR+ ist, desto besser ist der Test geeignet, eine erkrankte Person zu erkennen.

Die LR+ für die Nu-DESC betrug 12,25. Die LR+ für CAM ließ sich mathematisch nicht berechnen, weil die Falsch-Positiv-Rate 0 % betrug und somit durch Null hätte geteilt werden müssen. Auf jeden Fall war die LR+ für CAM wesentlich größer als 10 und damit als „gut“ anzunehmen (Tabelle 23).

4.4.8 Likelihood Ratio für einen negativen Test (LR-)

Die LR- gibt an, wie viel wahrscheinlicher ein negatives Testergebnis bei Kranken als bei Gesunden auftritt. Ein Wert von Eins würde bedeuten, dass die Wahrscheinlichkeit für ein negatives Testergebnis bei Gesunden und Kranken gleich groß ist. Werte unter 0,3 sind akzeptabel, Werte unter 0,1 gelten als gut. Je kleiner also die LR-, umso besser ist der Test geeignet, eine gesunde Person zu erkennen.

Die LR- der CAM betrug 0,25, die LR- der Nu-DESC war 0,03 (Tabelle 23).

4.4.9 Prä-Test-Wahrscheinlichkeit (Prävalenz)

Die Prä-Test-Wahrscheinlichkeit entspricht der Prävalenz und damit der Wahrscheinlichkeit, ob ein Patient an einer Erkrankung leidet, bevor zusätzliche Information, z.B. in Form von weiteren Testergebnissen, vorliegen.

Sie lag in dieser Studie bei 0,08 (8%) (Tabelle 23).

4.4.10 Prä-Test-Chance (pre-test-odds)

Prä- und Post-Test-Chance sind Größen, die zur Berechnung der Post-Test-Wahrscheinlichkeit benötigt werden. Die Prä-Test-Chance betrug in dieser Studie 0,09 (9%) (Tabelle 23).

4.4.11 Post-Test-Chance (post-test-odds)

Die Post-Test-Chance lag für die Nu-DESC bei 1,10. Für die CAM ließ sie sich wegen der Spezifität von 100% nicht berechnen. Mathematisch ging sie gegen unendlich, praktisch war sie mit sehr viel größer als 1 anzunehmen (Tabelle 23).

4.4.12 Post-Test-Wahrscheinlichkeit

Sie drückt aus, wie wahrscheinlich die Erkrankung nach dem Vorliegen von Testergebnissen ist. Sie berücksichtigt im Gegensatz zum Positiv Prädiktiven Wert die Prävalenz der Erkrankung im Testkollektiv.

Für die Nu-DESC betrug die Post-Test-Wahrscheinlichkeit in dieser Studie 0,52, d.h. mit einer Chance von 52% war ein positiv getesteter Patient unter Berücksichtigung der relativ niedrigen Prävalenz auch tatsächlich krank.

Für die CAM ließ sie sich wiederum wegen der Spezifität von 100% nicht berechnen. Sie ging mathematisch gegen unendlich, war praktisch aber mit 100% anzunehmen (Tabelle 23).

4.4.13 Genauigkeit (Accuracy)

Die Genauigkeit gibt den Anteil der korrekten Testergebnisse an allen Ergebnissen eines Tests wieder.

Die Accuracy der CAM betrug in dieser Studie 97,6%, die der Nu-DESC 94,1% (Tabelle 23).

Ergebnisse

Tabelle 23: Testgütekriterien von CAM und Nu-DESC im Vergleich

	CAM	Nu-DESC	p-Wert für McNemar
Sensitivität (95%-CI in Klammern)	74,9 % (57,7 - 92,2)	97,7% (93,0 - 102,3)	0,013
Spezifität (95%-CI in Klammern)	100,0% (100,0-100,0)	92,3% (88,4 - 96,2)	<0,001
Falsch-Positiv-Rate	0,0%	7,7%	
Falsch-Negativ-Rate	26,1%	2,3%	
Positiver Prädiktiver Wert	100,0%	56,9%	
Negativer Prädiktiver Wert	97,5%	99,5%	
Likelihood Ratio für positiven Test	>10	12,25	
Likelihood Ratio für negativen Test	0,25	0,03	
Prä-Test-Wahrscheinlichkeit	0,08	0,08	
Prä-Test-Chance	0,09	0,09	
Post-Test-Chance	>1	1,10	
Post-Test-Wahrscheinlichkeit	1,00	0,52	
Genauigkeit	97,6%	94,1%	

4.4.14 ROC-Kurve

Die *Receiver Operating Characteristics Curve* (ROC-Analyse) liefert eine graphische Darstellung der diagnostischen Leistungsfähigkeit eines Tests im Vergleich zu einem Referenzwert und ermöglicht darüber hinaus die Ermittlung von Grenzwerten (Cut-Off-Werten). Sie stellt die Abhängigkeit der Effizienz von der Fehlerrate dar, indem die relative Häufigkeit aller richtig positiven Testergebnisse (Sensitivität) gegen die relative Häufigkeit aller falsch positiven Testergebnisse (1-Spezifität) aufgetragen wird.

In diesem Fall wurden die CAM und die Nu-DESC im Vergleich dargestellt, als Goldstandard dienten die DSM-IV Kriterien (Abbildung 7).

Ergebnisse

Die *Area under the Curve* (AUC) ist dabei ein Maß für den Zugewinn an Information, den der Test durch seine Anwendung liefert. Sie kann Werte zwischen 0,5 und 1 annehmen, wobei 0,5 einer geringen Testgüte und 1 der maximalen Übereinstimmung mit dem Referenzwert entspricht.

Der in der vorliegenden Studie für die Nu-DESC ermittelte Grenzwert, ab dem der Test als positiv gewertet werden sollte, lag bei >1 . Für die CAM konnte aufgrund ihrer dichotomen Einteilung kein cut-off ermittelt werden.

Die CAM zeigte dabei eine AUC von 0,85, die AUC der Nu-DESC war 0,99. Damit wiesen beide einen statistisch signifikanten Unterschied zur Indifferenzfläche von 0,5 (siehe oben) auf (beide $p < 0,001$) (Tabelle 24).

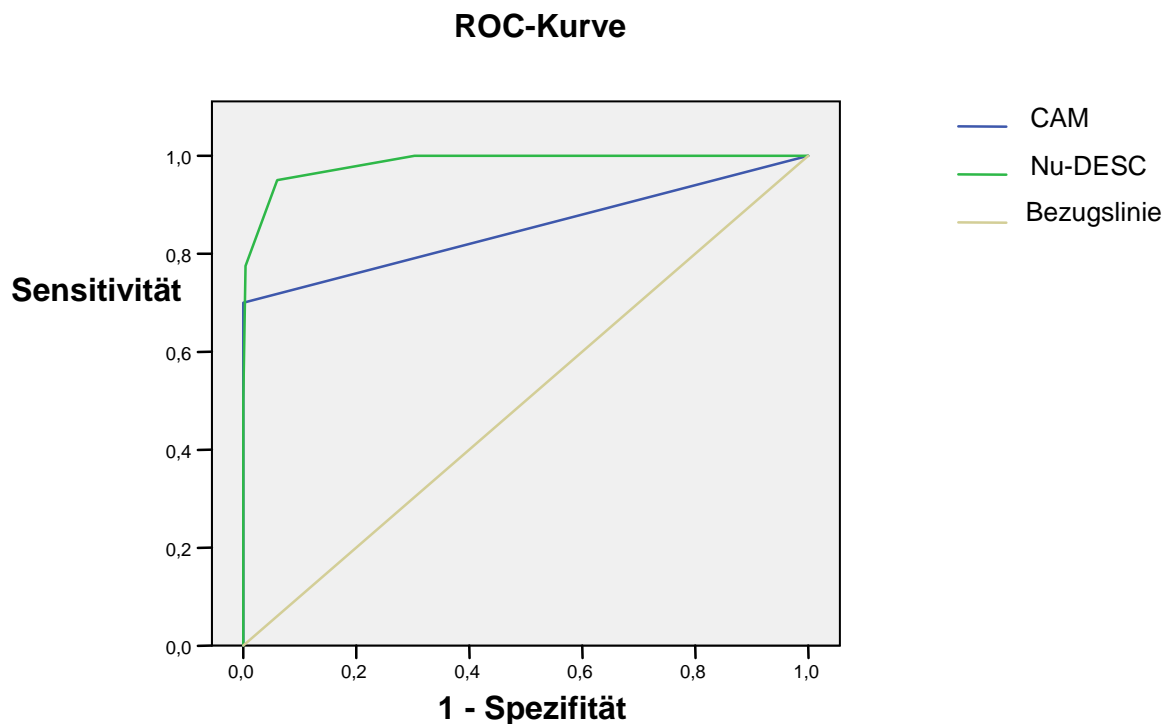


Abbildung 7: ROC-Kurve für CAM und Nu-DESC, Goldstandard = DSM-IV Kriterien

Tabelle 24: Area under the Curve für CAM und Nu-DESC im Vergleich, p-Wert für Unterschied zur Indifferenzfläche von 0,5

	AUC	CI	p
CAM	0,85	0,76-0,94	$<0,001$
Nu-DESC	0,98	0,97-1,00	$<0,001$

4.4.15 Interrater-Reliabilität

Die Interrater-Reliabilität ist ein Maß für die Übereinstimmung der Testergebnisse, wenn zwei Untersucher unabhängig voneinander den gleichen Test an demselben Patienten durchführen. Idealerweise sollte zwischen beiden Testungen keine große Zeitspanne liegen, um bei veränderlichen Symptomen die Ergebnisse nicht zu verfälschen. In der vorliegenden Studie wurde eine Stichprobe von 19 der 88 Patienten durch zwei Interviewer in einem Zeitintervall von nicht mehr als einer Stunde gescreent. Es wurde die Reliabilität bei Betrachtung des Endergebnisses des jeweiligen Tests, also die Unterscheidung der Patienten in den Kategorien Delir/kein Delir untersucht.

Dabei zeigte die CAM eine „sehr gute“ Übereinstimmung ($\kappa=1,00$), die Interrater-Reliabilität der Nu-DESC war „gut“ ($\kappa=0,83$). Die Übereinstimmung unterschied sich zwischen beiden Testmethoden somit signifikant. ($p<0,001$) (Tabelle 25).

Tabelle 25: Interrater-Reliabilität für CAM und Nu-DESC im Vergleich, p-Wert für McNemar

	kappa	95%-CI	p
Nu-DESC	0,83	0,49-1,16	<0,001
CAM	1,00	1,00-1,00	<0,001

4.4.16 Durchführungszeit

Um den Aufwand zur Durchführung von CAM und Nu-DESC zu objektivieren, wurde bei 10 separaten Testungen die benötigte Zeit erfasst. Hier zeigte sich, dass die Nu-DESC im Schnitt innerhalb von weniger als anderthalb Minuten vervollständigt werden konnte, während für die Komplettierung der CAM durchschnittlich mehr als fünf Minuten nötig waren, was einen signifikanten Zeitunterschied ($p<0,001$) darstellt (Tabelle 26).

Tabelle 26: Durchschnittlicher Zeitaufwand zur Durchführung von CAM und Nu-DESC in Minuten, 95%-CI in Klammern, p-Wert für t-Test

Zeit Nu-DESC [min]	Zeit CAM [min]	p-Wert
1,16 (1,04 - 1,28)	5,31 (4,90 - 5,72)	<0,001

5 Diskussion

Im folgenden Teil werden die wesentlichen Ergebnisse dieser Arbeit diskutiert, in Bezug zur aktuellen Literatur gesetzt sowie die Limitationen ihrer Verallgemeinerungswürdigkeit betrachtet.

5.1 Inzidenz

In der vorliegenden Studie entwickelten 19% (17 von 88) der über sechs postoperative Tage beobachteten Patienten ein Delir nach DSM-IV Kriterien. Erschreckend hierbei war, dass keiner dieser Fälle vom Krankenhauspersonal erkannt oder behandelt wurde. Die in dieser Studie gefundene Inzidenz stimmt mit Ergebnissen anderer Arbeiten überein, die allerdings vorwiegend am selektierten Patientengut (kardiale Operationen)⁸⁰ oder an Patienten eines anderen ASA-Physical Status^{81, 82} durchgeführt wurden.

Die CAM zeichnete für 17% der Patienten dieser Studie positiv. Olin et al. dagegen definierten 26 von 51 (51%) Patienten mit großen abdominal-chirurgischen Eingriffen als delirant nach der CAM. Diese Diskrepanz kann zum einen an der unterschiedlichen Größe der Studienpopulation liegen, zum anderen auch daran, dass Olin und Kollegen zur Klassifizierung deliranter Patienten nicht nur die Ergebnisse der CAM betrachteten, sondern auch Informationen aus der Krankenakte über auffälliges Verhalten mit hinzuzogen und so die Anzahl der gewerteten Delirien eventuell erhöhten.

Für die anhand der Nu-DESC gemessenen Inzidenzen eines postoperativen Delirs gibt es bisher bis auf eine im speziellen Setting des Aufwachraums durchgeführte⁸³ noch keine Studien, so dass die Einordnung der 28 von der Nu-DESC als delirant eingestuft Patienten in einen Gesamtkontext aktueller Literatur leider noch nicht vorgenommen werden kann.

5.2 Risikofaktoren

Von den in dieser Studie erhobenen Parametern kristallisierte sich lediglich die Narkosedauer als eindeutiger Risikofaktor heraus. Obwohl in anderen Studien wiederholt sowohl das Alter als auch das Geschlecht^{9, 26} als Risikofaktoren genannt werden, konnte hier kein signifikanter Unterschied zwischen den Gruppen gefunden werden. Dies kann zum einen daran liegen, dass durch die Einschlusskriterien (u.a. ASA-PS Klassifikation ab III) das Alter der untersuchten Personen bereits recht homogen und

vergleichsweise hoch war, andererseits war eventuell die Studienpopulation zu klein, um gewisse kleinere Unterschiede statistisch signifikant werden zu lassen.

Die Dauer der Operation war unter den deliranten Patienten nicht eindeutig verlängert, wohl aber die der Narkose. Dieses Phänomen wurde bereits in anderen Studien beschrieben^{22, 84, 85}. Dies könnte ein Hinweis auf die unter 1.4.2 diskutierte mögliche präzipitierende Funktion des Narkosetraumas sein.

5.3 Outcome

Patienten, die postoperativ ein Delir entwickelten, hatten nach der Operation eine signifikant längere Liegedauer insgesamt sowie auf der Intensivstation. Dieser Aspekt eines im Ganzen schlechteren Outcomes wird in der Literatur wiederholt beschrieben⁸⁶⁻⁸⁸. Interessant hierbei ist, dass es unerheblich war, welcher Test zur Klassifizierung eines Delirs zugrunde gelegt wurde. Sowohl die CAM, die Nu-DESC als auch die DSM-IV Kriterien wiesen für oben genannte Parameter signifikante Differenzen auf, wobei die Unterschiede zwischen den Gruppen bei der Nu-DESC am deutlichsten ausfielen ($p=0.001$). Es ließe sich somit argumentieren, dass alle angewandten Testmethoden in ihrer Anwendung einen Vorteil für den Patienten darstellen würden, da sie alle die potentiell kranken Patienten, die ein schlechteres Outcome bezüglich der Liegedauer entwickeln, identifizieren können.

5.4 Testgüte

Im Folgenden werden die Ergebnisse von 4.3 bis 4.4.15 jeweils für die einzelnen untersuchten Testmethoden zusammengefasst und bewertet.

5.4.1 Nu-DESC

In dieser Studie erwies sich die Nu-DESC mit einer mittleren Sensitivität von fast 98% als der sensitivere Test verglichen mit der CAM. Sie zeigte auch eine hohe Spezifität von 92%. Diese Ergebnisse stimmen mit anderen Arbeiten, die diese Aspekte untersuchten, überein bzw. übertrafen sie sogar bezüglich der ermittelten Spezifität, die bisher mit Werten zwischen 79-87% beschrieben wurde^{50, 83, 89}. Sowohl Sensitivität als auch Spezifität ließen sich im Verlauf der beobachteten sechs postoperativen Tage wiederholt auf diesem hohen Niveau reproduzieren, was zeigt, dass die ermittelten hohen Durchschnittswerte nicht durch Ausreißer erzeugt wurden (siehe Abbildung 5+6).

Diskussion

In der ROC-Analyse bewies die Nu-DESC darüber hinaus mit einer AUC von 0,99 eine ausgezeichnete Trennschärfe gegenüber den DSM-IV Kriterien und bestätigte den in der Literatur angegebenen Grenzwert von > 1 ⁵⁰. Außerdem war die Nu-DESC die am wenigsten zeitaufwendige Methode mit einer Durchführungszeit von im Schnitt nur 76 Sekunden und zeigte eine hohe Interrater-Reliabilität (kappa 0,83).

Allerdings wies die Nu-DESC eine relativ hohe Rate falsch positiver Messergebnisse auf (7,7%), was wiederum einen negativen Einfluss auf den positiv prädiktiven Wert hatte - dieser betrug nur 56,9%. Ein solches Resultat könnte mit der Annahme begründet werden, dass die Nu-DESC auch Patienten detektiert, die sich in einem Prodromalstadium des Delirs befinden. Dieses erfüllt formal nicht alle Kriterien der DSM-IV, kann aber in ein voll ausgebildetes Delir übergehen^{9, 90}. In der vorliegenden Arbeit wurden falsch positive Ergebnisse der Nu-DESC vorwiegend durch die Erfassung der Symptome „psychomotorische Retardierung“ unterschiedlichen Ausmaßes sowie „Desorientierung“ verursacht. Es kann kritisch angemerkt werden, dass eben diese Merkmale nach einer Operation, insbesondere am ersten postoperativen Tag, häufiger auftreten können, ohne von besonderem Krankheitswert zu sein. In der vorliegenden Studie entwickelten allerdings 21,4% der mit der Nu-DESC falsch positiv Getesteten im Verlauf von ein oder zwei Tagen ein Delir nach DSM-IV Kriterien. Bei weiteren 25% der falsch Positiven schlug die Nu-DESC über ein bis zwei Tage nach einem durchgemachten Delir nach DSM-IV weiter aus. Dies unterstützt die These, dass die Nu-DESC Prodromalphasen und abklingende Stadien eines Deliriums detektiert. Es wurde in verschiedenen Studien bereits nachgewiesen, dass Patienten mit einem subsyndromalen Delir ein ähnlich schlechtes Outcome haben wie Patienten mit dem Vollbild der Erkrankung^{7, 45}. Auch in der hier präsentierten Analyse hatten Nu-DESC positive Patienten, also auch solche mit falsch positiven Testergebnissen in Prodromalphasen und abklingenden Stadien, eine längere postoperative Liegedauer im Krankenhaus und auf der Intensivstation. Somit ist zu diskutieren, ob es nicht sogar für den Patienten von größerem Nutzen wäre, einen Test anzuwenden, der eben diese Vorstufe eines Delirs mit erfasst. Denn scheinbar teilt diese das schlechte Outcome der manifesten Delirien, so dass betroffene Patienten von einer Intervention potentiell ebenso profitieren könnten.

Darüber hinaus macht gerade die Berücksichtigung der Aspekte psychomotorischer Retardierung und fehlender Orientierung die Nu-DESC besonders sensitiv in der Detektion des hypoaktiven Delirs⁵⁰, welches unter den postoperativen Delirien bis zu

85% der Fälle ausmacht^{45, 87}. Angesichts seiner schlechteren Prognose im Vergleich zu anderen Formen des Delirs^{91, 92} sollte gerade die Erkennung des hypoaktiven Delirs durch die Anwendung eines geeigneten Tests forciert werden.

5.4.2 CAM

Die CAM erwies sich verglichen mit der Nu-DESC als weniger sensitiv. Die in unserer Studie ermittelte Sensitivität von knapp 75% liegt zudem auch etwas unter den in der Literatur angegebenen Werten - in Untersuchungen auf geriatrischen Stationen und Notaufnahmen ergaben sich Zahlen zwischen 76 und 86%^{53, 54, 89}. Dies spiegelt sich auch in der hohen Rate der falsch negativen Ergebnisse von 26% wider.

Allerdings scheint die Sensitivität der CAM zum Teil auch sehr abhängig vom untersuchenden Personal zu sein: Bei Rolfson et al. sank sie auf 13%, wenn die CAM von Krankenschwestern statt von Ärzten angewendet wurde⁹³.

Insbesondere im Zeitverlauf über die sechs beobachteten postoperativen Tage schwankte die Sensitivität stark und lieferte somit keine reproduzierbaren Werte (siehe Abbildung 5).

Die Ursache für die niedrige Sensitivität der CAM lag in der vorliegenden Studie vor allem daran, dass oftmals das Kriterium „Unaufmerksamkeit“ bei der CAM von den Patienten nicht erfüllt wurde. Es ist ein viel diskutiertes Problem, ob dieses Symptom zwingend erfüllt sein muss, um ein Delir zu diagnostizieren. Beispielsweise ist noch nicht genau definiert, welches Ausmaß einer Aufmerksamkeitsstörung für ein Delir typisch ist und wie milde Einschränkungen im klinischen Alltag erkannt werden können^{11, 94}. Daher wurde in dieser Studie die Auslegung der DSM-IV Kriterien nach Cole et al. verwendet⁹⁵. Seine Arbeitsgruppe zeigte, dass die Interpretation des Kriteriums A („Bewusstseinsstörung mit reduzierter Fähigkeit, die Aufmerksamkeit zu richten, aufrecht zu erhalten oder zu verlagern“) einen wesentlichen Einfluss auf Sensitivität und Spezifität der DSM-IV Kriterien hatte. Forderte man die Erfüllung beider Aspekte, lagen Sensitivität und Spezifität bei nur 47,5% und 90%; definierte man Kriterium A nur als Bewusstseinsstörung allein, war das Ergebnis dasselbe. Sah man dagegen das Kriterium als erfüllt an, wenn entweder eine Bewusstseinsstörung oder eine Aufmerksamkeitsstörung vorlag, stieg die Sensitivität auf 100% an, die Spezifität lag bei 71%. Da ein Goldstandard mit einer Sensitivität unter 50% nicht akzeptabel wäre, richtete man sich in der vorliegenden Untersuchung nach den Empfehlungen dieser Studie und entschied sich für letztere Interpretation bezüglich der DSM-IV Kriterien.

Diskussion

So kam es, dass Patienten, die dem Verlauf des Interviews folgen konnten, jedoch dennoch eine milde Bewusstseinsstörung zeigten, als delirant entsprechend der DSM-IV Kriterien eingestuft wurden, die Anforderungen der CAM jedoch nicht voll erfüllten, da diese explizit beide Punkte als zwingend notwendig ansieht. Konkret kam diese Diskrepanz bei 12 von 501 Messungen vor.

Ein anderer möglicher Erklärungsansatz für diese Unterschiede könnte darin gesehen werden, dass gerade milde Einschränkungen der Aufmerksamkeit durch die verschiedenen Untersucher von Goldstandard und CAM unterschiedlich eingeschätzt wurden. Die DSM-IV Kriterien können bei Zweifeln um weitere Tests zur Feststellung der Aufmerksamkeit, wie beispielsweise den Digit-Span-Test, ergänzt werden, während dies für die Anwendung der CAM nicht vorgesehen ist. Somit haben die Untersucher mit dem Goldstandard unter Umständen mildere Aufmerksamkeitsstörungen eher detektiert als die Untersucher mit der CAM.

Die Spezifität der CAM dagegen lag durchgehend bei 100% und bestätigt damit die Ergebnisse von Leung et al. und Monette et al.^{54, 89}. Durch die restriktive Form der CAM wurden keine Patienten diagnostiziert, die nach DSM-IV kein Delir gehabt hätten. Somit ließe sich die Relevanz der CAM als Rule-In-Test diskutieren, mit dem man in anderen Tests gewonnene, eventuell noch fragliche Ergebnisse verifizieren kann.

Die AUC der CAM lag mit 0,85 zwar deutlich unter der der Nu-DESC, bewies aber immer noch eine gute Trennschärfe gegenüber den DSM-IV Kriterien und lag höher als in einer anderen Studie, in der die AUC nur 0,71 betrug⁸³. Die Ermittlung eines Cut-Off erübrigte sich aufgrund der binären Struktur des Tests.

Die CAM war in der Anwendung die eindeutig aufwändigere Methode mit einem durchschnittlichen Zeitbedarf von ca. 5,3 Minuten pro Durchgang. Es dürfte damit zur Diskussion gestellt werden, ob sich dieser Test zum Screening im Alltag einer peripheren Station eignet oder nicht doch auf Methoden zurückgegriffen werden sollte, die sich schneller durchführen lassen. Nichtsdestoweniger wies die CAM eine exzellente Interrater-Reliabilität von 100% auf, was wiederum an ihrer restriktiven Struktur liegen könnte, die wenig Raum für persönliche Interpretationen der einzelnen Prüfer lässt und im Gegensatz zur Nu-DESC auch keine Abstufung in Schweregrade erfordert.

5.5 Methodenkritik

Im Folgenden soll die interne und externe Validität der vorliegenden Studie gemäß den Forderungen der STARD ⁹⁶ und den entsprechenden Systematiken ^{97, 98} diskutiert werden.

5.5.1 Bias

Unter einem Bias wird die Verzerrung des Versuchsergebnisses durch falsche Untersuchungsmethoden verstanden. Dabei gibt es verschiedene Formen.

5.5.1.1 Selection Bias

Unterformen der *Selection Bias*, der *Assembly* und der *Sampling Bias* im Sinne einer verzerrten Auswahl der Studienteilnehmer durch fehlenden Zugang zum System oder zu restriktive Vorauswahl, sind in dieser Studie eher unwahrscheinlich. Da konsekutiv alle in der einzigen präoperativen anästhesiologischen Ambulanz erfassten Patienten auf ihre Eignung zur Versuchsteilnahme hin überprüft wurden und bis auf eine in den Ausschlusskriterien formulierte Fachrichtung alle großen chirurgischen Disziplinen vertreten waren, kann man davon ausgehen, dass das rekrutierte Patientenkollektiv einem repräsentativen Durchschnitt des Patientenguts auf der peripheren Station entspricht. Es kann angemerkt werden, dass allgemeinchirurgische Patienten überdurchschnittlich häufig vertreten waren. Eine Verzerrung der Messergebnisse ist hierdurch allerdings nicht zu erwarten, da bisher nicht bekannt ist, dass diese Patienten ein erhöhtes Risiko für die Entwicklung eines Delirs im Vergleich zu Patienten anderer Fachrichtungen hätten. Die einzige wirklich restriktive Vorauswahl war die Beschränkung auf ASA-PS-III Patienten, die natürlich nur einen Ausschnitt der Patienten der chirurgischen Stationen darstellen. Diese Einschränkung musste aus Gründen der Logistik in Kauf genommen werden, um eine genügend hohe Fallzahl an Delirien überhaupt wahrscheinlich zu machen.

5.5.1.2 Information Bias

Eine wichtige Form des *Information Bias* ist der *Observer Bias* und hierbei insbesondere der *Expectation Bias*, also die Beeinflussung des Gesamturteils des Untersuchers durch das Wissen um einzelne diagnoserelevante Symptome des Patienten. Dem versucht die Strukturierung der verwendeten Testmethoden entgegenzuwirken, die das Gesamtkonstrukt „Delir“ bewusst in einzelne durch Punkte zu bewertende Dimensionen

zergliedert und damit die Gesamtbeurteilung des Tests von der klinischen Beurteilung entkoppelt und objektiviert. Das Erkennen der einzelnen Symptome ist hier also ein gewollter Prozess, deren Addierung zu einer Diagnose allerdings erst in einem zweiten Schritt erfolgt. Inwiefern ein erster Gesamteindruck des Patienten die Beurteilung der einzelnen Items durch den Untersucher beeinflusst, ist in diesem Rahmen leider nicht abzuschätzen, weder in der Stärke noch in der Richtung des eventuell entstehenden Bias. Dies ist allerdings keine Schwäche speziell unseres Studiendesigns, sondern ein generelles Problem klinischer Tests.

Einem *Misclassification Bias* im Sinne einer Verzerrung der Ergebnisse durch das Messinstrument wurde versucht durch die Schulung aller beteiligten Untersucher in der Anwendung der Methoden entgegenzuwirken. Darüber hinaus wurde die Interrater-Reliabilität überprüft, die sowohl für die CAM als auch für die Nu-DESC „gut“ bis „sehr gut“ war.

Ein *Analysis Bias* entsteht typischerweise in randomisierten Studien, wenn Patienten einzelner Studienarme nicht mit den gleichen Methoden untersucht oder behandelt werden. Da die vorliegende Studie nicht randomisiert wurde, wurden alle Patienten mit allen Testmethoden gescreent und der *Analysis Bias* insbesondere in Form eines *work-up Bias* (Beeinflussung der Referenz-Durchführung durch die Testergebnisse) somit vermieden⁹⁹.

5.5.2 Confounder

Confounder sind Störgrößen, die mit dem beobachteten Faktor assoziiert sind, jedoch keinen nachweisbaren kausalen Zusammenhang zeigen. Da diese Studie nicht randomisiert wurde, können vielfältige Confounder aufgetreten sein. Die wichtigsten wurden versucht durch die Ein- und Ausschlusskriterien zu minimieren. So wurden Patienten, bei denen beispielsweise eine Demenz, stattgehabte Infarkte, zahlreiche neurologische Erkrankungen sowie Substanzdefekte nach intrakraniellen Eingriffen Symptome eines Delirs imitieren und somit das Ergebnis hätten verzerren können, ausgeschlossen. Aus demselben Grunde wurden auch Patienten mit Alkohol- und Drogenabusus, insbesondere im Entzug, ausgeschlossen. Inwiefern nicht erfasste, eventuell dem Individuum zuzuordnende Confounder das Ergebnis beeinflusst haben, ist mit diesem Studiendesign nicht abzuschätzen.

5.5.3 Zufall

In der vorliegenden Arbeit wurde wie in medizinischen Studien meist üblich eine Irrtumswahrscheinlichkeit von $\alpha < 0.05$ angenommen, das heißt bei den als signifikant postulierten Ergebnissen liegt die Wahrscheinlichkeit eines Fehlers 1. Art bei weniger als 5%. Um diese Aussage zu präzisieren, wurde ein Konfidenzintervall mit angegeben, dass mit einer Wahrscheinlichkeit von 95% die wahren Ergebnisse mit einschließt.

5.5.4 Fallzahl

Es ist kritisch anzumerken, dass diese Studie mit 88 Patienten eine relativ geringe Fallzahl aufweist. Diese erklärt sich daher, dass die Rekrutierung der Patienten eingebettet in den Rahmen einer übergeordneten Studie stattfand, für die zwei Blutentnahmen notwendig waren, denen viele potentiell geeignete Patienten nicht zustimmten. Wurde die Blutentnahme unter der Operation verpasst, wurden die Patienten aus dem weiteren Screening ausgeschlossen. Dies war bei 25 Patienten der Fall. Die hier gewonnenen Ergebnisse sollten somit in einer weiteren Studie validiert werden, die diese Schwäche nicht aufweist.

5.5.5 Reproduzierbarkeit

In der vorliegenden Arbeit wurden Patienten wiederholt und von verschiedenen Testern gescreent. Somit könnte die Frage nach dem Einfluss wiederholter Testungen auf das Messergebnis gestellt werden. Die Interrater-Reliabilität wurde überprüft und erwies sich für alle angewendeten Testmethoden als hoch (siehe 4.4.15). Die Test-Retest-Reliabilität konnte mit dem vorliegenden Design nicht untersucht werden, da zum einen durch die fluktuierende Natur des Delirs keine Konstanz der zu messenden Eigenschaft vorausgesetzt werden konnte und zudem Erinnerungseffekte der Patienten das Messergebnis hätten beeinflussen können. Allerdings ist gerade die Verlaufsbeobachtung bei dem Versuch der Validierung eines Delir-Screeninginstrumentes wichtig, da das Delir als dynamischer Prozess mit fluktuierender Symptomatik in einer einmaligen Messung verzerrte, eventuell falsch negative Ergebnisse liefern kann. Aufgrund ihrer intensiven Schulung waren die Mitarbeiter weitestgehend in der Lage, eine echte Fluktuation der Symptome zu erkennen, so dass Schwankungen der Testergebnisse eher durch diesen krankheitstypischen Verlauf als durch eine geringe Test-Retest-Reliabilität begründet wären.

5.5.6 Qualität des Goldstandards

Die Kriterien des DSM-IV gelten derzeit zwar als Goldstandard in der Diagnose eines Delirs, stellen dabei aber eine eher unspezifische Zusammenfassung verschiedener Symptome eines vielschichtigen Syndroms dar. Die Kriterien des DSM-IV Katalogs haben sich aus dem DSM-III und DSM-III-R entwickelt und werden aktuell parallel zum Katalog der ICD-10 angewendet. In einer Vergleichsstudie dieser Diagnoseschemata konnte gezeigt werden, dass sich die Sensitivitäten und Spezifitäten aller vier Systeme signifikant voneinander unterscheiden ⁴⁷. Die DSM-IV Kriterien schließen dabei das größte Patientenkollektiv ein, sind zum Teil jedoch weniger spezifisch, je nachdem, welcher Referenzstandard zugrunde gelegt wird (Abbildung 8).

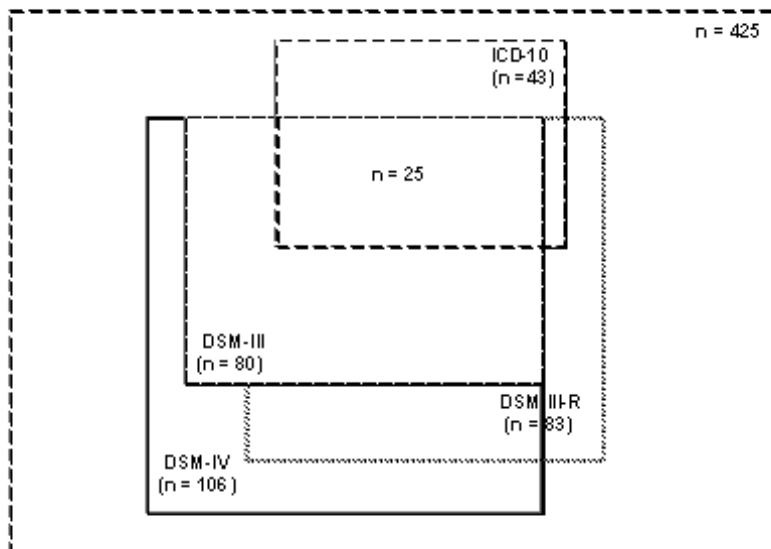


Abbildung 8: Übereinstimmung verschiedener Diagnosesysteme, modifiziert nach Laurila et al. 2004⁴⁷

Den absoluten, in der Praxis oft jedoch nicht realisierbaren Goldstandard bildet auch heute noch der Expertenkonsens ^{3, 100-102}, der auch der Diagnose durch einen Psychiater überlegen zu sein scheint ¹⁰⁰.

Zur Test-Retest-Reliabilität der DSM-IV Kriterien wurden in der Literatur keine Studien gefunden.

Es lässt sich somit zusammenfassen, dass der gewählte Goldstandard dem Stand der wissenschaftlichen Diskussion in der aktuellen Literatur entspricht, jedoch noch weitere

Überprüfungen seiner Validität und Reliabilität bestehen müsste, um auch die hier gewonnenen Ergebnisse zu bestätigen.

5.5.7 Praxisrelevanz der gefundenen Ergebnisse

Unabhängig von der Validität von Nu-DESC und CAM ließe sich allgemein fragen, ob es dem Patienten überhaupt nützlich ist, ein Delir zu diagnostizieren. Wie bereits in der Einleitung erwähnt, wird ein Delir in bis zu 50% der Fälle nicht erkannt^{20, 75} und führt entsprechend zu einem schlechteren Outcome³⁴⁻⁴¹. Die rechtzeitige Detektion eines postoperativen Delirs wäre für den Patienten daher also sowohl für die Verringerung der Mortalität als auch für die Zuführung zu einer Intervention zur Vermeidung von Folgen und Komplikationen sinnvoll – besonders, wenn man neueste Studienergebnisse berücksichtigt, die belegen, dass das Outcome umso schlechter ist, je länger ein Delir besteht⁴².

Darüber hinaus ließen sich perspektivisch für den Patienten wertvolle, akut eher wissenschaftliche Fragestellungen wie z.B.

- eine Erforschung des Beitrags des chirurgisch-anästhesiologischen Traumas zur Entwicklung eines Delirs
- die Verbesserung der Qualitätssicherung der klinischen Prozesse
- die Erhöhung der Effektivität von Präventionsprogrammen

klären.

Insbesondere die Nu-DESC stellt eine Testmethode dar, die sich im Vergleich zu den wesentlich aufwendigeren DSM-IV Kriterien leicht in die klinische Routine einbetten ließe. Da sie als reine Beobachtungsmethode entwickelt wurde, könnte sie bei Hochrisikopatienten ohne allzu großen Aufwand in der „kleinen Krankenakte“ mitgeführt und täglich vom Pflegepersonal ausgefüllt werden. Die Nu-DESC zeigte auch die wichtigste Forderung an ein medizinisches Screening-Verfahren, nämlich eine geringe Rate falsch negativer Ergebnisse von nur 2.3%. Allerdings bewies sie einen Positiv Prädiktiven Wert (PPW) von nur knapp 57% und eine geringe Post-Test-Wahrscheinlichkeit von nur 52% und wäre damit allein angewendet streng genommen ökonomisch und medizinisch eher unbefriedigend. Angesichts der negativen Folgen und der derzeit geringen Feststellungsrate eines Delirs wäre es aber dennoch wünschenswert, einen besonders sensitiven Test anzuwenden. Es wäre daher sinnvoll, die Nu-DESC tatsächlich als breitbasiges Screening-Instrument in die bestehenden *Standard Operating Procedures* (SOPs) zu integrieren und durch einen *Rule-in-Test* zu ergänzen. Hier würde sich die

Diskussion

CAM anbieten, die in dieser Studie keinen einzigen Fall falsch positiv eingestuft hat und somit in ihrer Spezifität mit der der DSM-IV Kriterien vergleichbar war. Sie kann auch von geschultem nichtpsychiatrischen Personal angewendet werden und würde als etwas zeitaufwendigere Methode in der Abfolge nach der Nu-DESC deren Ergebnisse validieren können. Somit wäre auch eine sinnvolle Staffelung vom weniger zeit-, personal- und kostenintensiven Verfahren zum nächst höheren gewährleistet.

Ein weiteres Argument, das für die bevorzugte Anwendung der Nu-DESC spricht, ist die Beobachtung, dass im klinischen Alltag vor allem das hyperaktive Delir erkannt und behandelt wird - vielleicht auch, weil die psychotischen Symptome den Ablauf von Behandlung und Pflege stören. Ein hypoaktives Delir bleibt dagegen eher unerkant ¹². Daher ist zu überlegen, ob ein neu hinzugezogenes Screening-Instrument nicht besonders zur Erfassung dieser Form des Delirs geeignet sein sollte, was die Nu-DESC ist.

Außerdem lässt sich zur weiteren Begründung der Forderung einer Implementierung eines verlässlichen Delirium-Screening-Instruments in die SOPs anbringen, dass ein in den täglichen Ablauf eingebettetes Delir-Screening für den Patienten im Verhältnis zum potentiellen Nutzen keine zusätzliche Belastung oder gar einen Schaden (im Sinne von Schmerz, Nebenwirkung oder unzumutbarer Durchführung) birgt.

Eine endgültige Aussage hierzu müsste jedoch durch weitere Studien, die randomisiert verschiedene Diagnose-Schemata hinsichtlich einer Kosten-Nutzen-Analyse vergleichen sollten, untermauert werden.

6 Zusammenfassung

6.1 Hintergrund

Die Entwicklung eines postoperativen Delirs hat erhebliche negative Folgen sowohl für das Outcome des einzelnen Patienten als auch für den Ressourcenverbrauch im Allgemeinen. Trotz seiner Häufigkeit werden die existierenden Screening-Instrumente im klinischen Alltag bisher selten angewandt, allerdings fehlten auch geeignete Studien, die die verschiedenen Testmethoden hinsichtlich ihrer Anwendbarkeit und Validität verglichen hätten. Ziel dieser Studie war es daher, einen validen und praktikablen Delirium-Screening-Score für die periphere chirurgische Station zu finden.

6.2 Methoden

In dieser prospektiven Observationsstudie wurden 88 Patienten der peripheren chirurgischen Station an sechs aufeinanderfolgenden postoperativen Tagen auf die Entwicklung eines Delirs hin gescreent. Dabei wurden die Nursing Delirium Screening Scale (Nu-DESC) und die Confusion Assessment Method (CAM) von geschulten wissenschaftlichen Mitarbeitern angewendet. Als Goldstandard wurden die DSM-IV Kriterien genutzt.

6.3 Ergebnisse

Bei 17 Patienten (19%) wurde an insgesamt 39 Tagen ein Delir nach DSM-IV Kriterien festgestellt. Mit der CAM wurden 15 Patienten an 28 Tagen, mit der Nu-DESC 28 Patienten an 65 Tagen positiv getestet. Die Sensitivität des CAM lag bei insgesamt knapp 75%, die der Nu-DESC bei 98%. Die Spezifitäten betrugen 100% für die CAM und 92% für die Nu-DESC. Die Falsch-Positiv-Raten ergaben 0,0% bei der CAM und 7,7% bei der Nu-DESC. Die Falsch-Negativ-Rate der CAM betrug dagegen 26,1%, die der Nu-DESC nur 2,3%.

Die Interrater-Reliabilität war für beide Tests „gut“ bis „sehr gut“, die Nu-DESC zeigte außerdem die schnellere Anwendbarkeit mit einer Durchführungszeit von durchschnittlich 76 Sekunden.

Im Outcome wiesen delirante Patienten eine signifikant längere postoperative Liegezeit auf der Intensivstation sowie im Krankenhaus auf.

6.4 Schlussfolgerung

Sowohl CAM als auch Nu-DESC zeigten eine gute Spezifität, wobei die Sensitivität der Nu-DESC signifikant höher war. Im Zusammenhang mit ihrer niedrigen Falsch-Negativ-Rate scheint die Nu-DESC somit am ehesten geeignet für ein Screening auf der peripheren Station, eventuell ergänzt durch einen Rule-in-Test, z.B. die CAM.

7 Literaturverzeichnis

- 1 Francis J. Delirium in older patients. *J Am Geriatr Soc* 1992; 40: 829-38
- 2 Tucker GJ. The diagnosis of delirium and DSM-IV. *Dement Geriatr Cogn Disord* 1999; 10: 359-63
- 3 Liptzin B. What criteria should be used for the diagnosis of delirium? *Dement Geriatr Cogn Disord* 1999; 10: 364-7
- 4 O'Keeffe ST, Ni Chonchubhair A. Postoperative delirium in the elderly. *Br J Anaesth* 1994; 73: 673-87
- 5 Practice guideline for the treatment of patients with delirium. American Psychiatric Association. *Am J Psychiatry* 1999; 156: 1-20
- 6 Lipowski ZJ. Delirium updated. *Compr Psychiatry* 1980; 21: 190-6
- 7 Cole M, McCusker J, Dendukuri N, Han L. The prognostic significance of subsyndromal delirium in elderly medical inpatients. *J Am Geriatr Soc* 2003; 51: 754-60
- 8 Levkoff SE, Besdine RW, Wetle T. Acute confusional states (delirium) in the hospitalized elderly. *Annu Rev Gerontol Geriatr* 1986; 6: 1-26
- 9 Elie M, Cole MG, Primeau FJ, Bellavance F. Delirium risk factors in elderly hospitalized patients. *J Gen Intern Med* 1998; 13: 204-12
- 10 Fick DM, Agostini JV, Inouye SK. Delirium superimposed on dementia: a systematic review. *J Am Geriatr Soc* 2002; 50: 1723-32
- 11 Armstrong SC, Cozza KL, Watanabe KS. The misdiagnosis of delirium. *Psychosomatics* 1997; 38: 433-9
- 12 Inouye SK. The dilemma of delirium: clinical and research controversies regarding diagnosis and evaluation of delirium in hospitalized elderly medical patients. *Am J Med* 1994; 97: 278-88
- 13 Olin K, Eriksson-Jonhagen M, Jansson A, Herrington MK, Kristiansson M, Permert J. Postoperative delirium in elderly patients after major abdominal surgery. *Br J Surg* 2005; 92: 1559-64
- 14 Bucht G, Gustafson Y, Sandberg O. Epidemiology of delirium. *Dement Geriatr Cogn Disord* 1999; 10: 315-8
- 15 Cole MG. Delirium in elderly patients. *Am J Geriatr Psychiatry* 2004; 12: 7-21
- 16 Alagiakrishnan K, Wiens CA. An approach to drug induced delirium in the elderly. *Postgrad Med J* 2004; 80: 388-93

Literaturverzeichnis

- 17 Hshieh TT, Fong TG, Marcantonio ER, Inouye SK. Cholinergic deficiency hypothesis in delirium: a synthesis of current evidence. *J Gerontol A Biol Sci Med Sci* 2008; 63: 764-72
- 18 Flacker JM, Lipsitz LA. Neural mechanisms of delirium: current hypotheses and evolving concepts. *J Gerontol A Biol Sci Med Sci* 1999; 54: B239-46
- 19 Hala M. Pathophysiology of postoperative delirium: systemic inflammation as a response to surgical trauma causes diffuse microcirculatory impairment. *Med Hypotheses* 2007; 68: 194-6
- 20 Young J, Inouye SK. Delirium in older people. *Bmj* 2007; 334: 842-6
- 21 White S, Calver BL, Newsway V, Wade R, Patel S, Bayer A, O'Mahony MS. Enzymes of drug metabolism during delirium. *Age Ageing* 2005; 34: 603-8
- 22 Inouye SK, Charpentier PA. Precipitating factors for delirium in hospitalized elderly persons. Predictive model and interrelationship with baseline vulnerability. *Jama* 1996; 275: 852-7
- 23 O'Keeffe ST, Lavan JN. Predicting delirium in elderly patients: development and validation of a risk-stratification model. *Age Ageing* 1996; 25: 317-21
- 24 Inouye SK, Zhang Y, Jones RN, Kiely DK, Yang F, Marcantonio ER. Risk factors for delirium at discharge: development and validation of a predictive model. *Arch Intern Med* 2007; 167: 1406-13
- 25 Kalisvaart KJ, Vreeswijk R, de Jonghe JF, van der Ploeg T, van Gool WA, Eikelenboom P. Risk factors and prediction of postoperative delirium in elderly hip-surgery patients: implementation and validation of a medical risk factor model. *J Am Geriatr Soc* 2006; 54: 817-22
- 26 Dasgupta M, Dumbrell AC. Preoperative risk assessment for delirium after noncardiac surgery: a systematic review. *J Am Geriatr Soc* 2006; 54: 1578-89
- 27 Litaker D, Locala J, Franco K, Bronson DL, Tannous Z. Preoperative risk factors for postoperative delirium. *Gen Hosp Psychiatry* 2001; 23: 84-9
- 28 Marcantonio ER, Goldman L, Mangione CM, Ludwig LE, Muraca B, Haslauer CM, Donaldson MC, Whittlemore AD, Sugarbaker DJ, Poss R, et al. A clinical prediction rule for delirium after elective noncardiac surgery. *Jama* 1994; 271: 134-9
- 29 Bryson GL, Wyand A. Evidence-based clinical update: general anesthesia and the risk of delirium and postoperative cognitive dysfunction. *Can J Anaesth* 2006; 53: 669-77
- 30 Marcantonio ER, Goldman L, Orav EJ, Cook EF, Lee TH. The association of intraoperative factors with the development of postoperative delirium. *Am J Med* 1998; 105: 380-4

Literaturverzeichnis

- 31 Inouye SK. Predisposing and precipitating factors for delirium in hospitalized older patients. *Dement Geriatr Cogn Disord* 1999; 10: 393-400
- 32 Gallagher R. Nicotine withdrawal as an etiologic factor in delirium. *J Pain Symptom Manage* 1998; 16: 76-7
- 33 Pandharipande P, Shintani A, Peterson J, Pun BT, Wilkinson GR, Dittus RS, Bernard GR, Ely EW. Lorazepam is an independent risk factor for transitioning to delirium in intensive care unit patients. *Anesthesiology* 2006; 104: 21-6
- 34 Zakriya K, Sieber FE, Christmas C, Wenz JF, Sr., Franckowiak S. Brief postoperative delirium in hip fracture patients affects functional outcome at three months. *Anesth Analg* 2004; 98: 1798-802, table of contents
- 35 Inouye SK, Rushing JT, Foreman MD, Palmer RM, Pompei P. Does delirium contribute to poor hospital outcomes? A three-site epidemiologic study. *J Gen Intern Med* 1998; 13: 234-42
- 36 Edlund A, Lundstrom M, Lundstrom G, Hedqvist B, Gustafson Y. Clinical profile of delirium in patients treated for femoral neck fractures. *Dement Geriatr Cogn Disord* 1999; 10: 325-9
- 37 McCusker J, Cole M, Abrahamowicz M, Primeau F, Belzile E. Delirium predicts 12-month mortality. *Arch Intern Med* 2002; 162: 457-63
- 38 Leslie DL, Zhang Y, Holford TR, Bogardus ST, Leo-Summers LS, Inouye SK. Premature death associated with delirium at 1-year follow-up. *Arch Intern Med* 2005; 165: 1657-62
- 39 Curyto KJ, Johnson J, TenHave T, Mossey J, Knott K, Katz IR. Survival of hospitalized elderly patients with delirium: a prospective study. *Am J Geriatr Psychiatry* 2001; 9: 141-7
- 40 Bickel H, Gradingner R, Kochs E, Forstl H. High risk of cognitive and functional decline after postoperative delirium. A three-year prospective study. *Dement Geriatr Cogn Disord* 2008; 26: 26-31
- 41 McAvay GJ, Van Ness PH, Bogardus ST, Jr., Zhang Y, Leslie DL, Leo-Summers LS, Inouye SK. Older adults discharged from the hospital with delirium: 1-year outcomes. *J Am Geriatr Soc* 2006; 54: 1245-50
- 42 Pisani MA, Kong SY, Kasl SV, Murphy TE, Araujo KL, Van Ness PH. Days of Delirium are Associated with 1-year Mortality in an Older Intensive Care Unit Population. *Am J Respir Crit Care Med* 2009
- 43 Wacker P, Nunes PV, Cabrita H, Forlenza OV. Post-operative delirium is associated with poor cognitive outcome and dementia. *Dement Geriatr Cogn Disord* 2006; 21: 221-7
- 44 Rudolph JL, Marcantonio ER, Culley DJ, Silverstein JH, Rasmussen LS, Crosby GJ, Inouye SK. Delirium is associated with early postoperative cognitive dysfunction. *Anaesthesia* 2008; 63: 941-7

Literaturverzeichnis

- 45 Marcantonio E, Ta T, Duthie E, Resnick NM. Delirium severity and psychomotor types: their relationship with outcomes after hip fracture repair. *J Am Geriatr Soc* 2002; 50: 850-7
- 46 American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 4th ed. Washington, D.C. 1994:124-133
- 47 Laurila JV, Pitkala KH, Strandberg TE, Tilvis RS. Delirium among patients with and without dementia: does the diagnosis according to the DSM-IV differ from the previous classifications? *Int J Geriatr Psychiatry* 2004; 19: 271-7
- 48 Laurila JV, Pitkala KH, Strandberg TE, Tilvis RS. Impact of different diagnostic criteria on prognosis of delirium: a prospective study. *Dement Geriatr Cogn Disord* 2004; 18: 240-4
- 49 Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegel AP, Horwitz RI. Clarifying confusion: the confusion assessment method. A new method for detection of delirium. *Ann Intern Med* 1990; 113: 941-8
- 50 Gaudreau JD, Gagnon P, Harel F, Tremblay A, Roy MA. Fast, systematic, and continuous delirium assessment in hospitalized patients: the nursing delirium screening scale. *J Pain Symptom Manage* 2005; 29: 368-75
- 51 Otter H, Martin J, Basell K, von Heymann C, Hein OV, Bollert P, Jansch P, Behnisch I, Wernecke KD, Konertz W, Loening S, Blohmer JU, Spies C. Validity and reliability of the DDS for severity of delirium in the ICU. *Neurocrit Care* 2005; 2: 150-8
- 52 Adamis D, Treloar A, MacDonald AJ, Martin FC. Concurrent validity of two instruments (the Confusion Assessment Method and the Delirium Rating Scale) in the detection of delirium among older medical inpatients. *Age Ageing* 2005; 34: 72-5
- 53 Laurila JV, Pitkala KH, Strandberg TE, Tilvis RS. Confusion assessment method in the diagnostics of delirium among aged hospital patients: would it serve better in screening than as a diagnostic instrument? *Int J Geriatr Psychiatry* 2002; 17: 1112-9
- 54 Monette J, Galbaud du Fort G, Fung SH, Massoud F, Moride Y, Arsenault L, Afilalo M. Evaluation of the Confusion Assessment Method (CAM) as a screening tool for delirium in the emergency room. *Gen Hosp Psychiatry* 2001; 23: 20-5
- 55 Lundstrom M, Edlund A, Karlsson S, Brannstrom B, Bucht G, Gustafson Y. A multifactorial intervention program reduces the duration of delirium, length of hospitalization, and mortality in delirious patients. *J Am Geriatr Soc* 2005; 53: 622-8
- 56 Milisen K, Foreman MD, Abraham IL, De Geest S, Godderis J, Vandermeulen E, Fischler B, Delooz HH, Spiessens B, Broos PL. A nurse-led interdisciplinary intervention program for delirium in elderly hip-fracture patients. *J Am Geriatr Soc* 2001; 49: 523-32

Literaturverzeichnis

- 57 Johnson J. Identifying and recognizing delirium. *Dement Geriatr Cogn Disord* 1999; 10: 353-8
- 58 Naughton BJ, Saltzman S, Ramadan F, Chadha N, Priore R, Mylotte JM. A multifactorial intervention to reduce prevalence of delirium and shorten hospital length of stay. *J Am Geriatr Soc* 2005; 53: 18-23
- 59 Cole MG, McCusker J, Bellavance F, Primeau FJ, Bailey RF, Bonnycastle MJ, Laplante J. Systematic detection and multidisciplinary care of delirium in older medical inpatients: a randomized trial. *Cmaj* 2002; 167: 753-9
- 60 Pitkala KH, Laurila JV, Strandberg TE, Tilvis RS. Multicomponent geriatric intervention for elderly inpatients with delirium: a randomized, controlled trial. *J Gerontol A Biol Sci Med Sci* 2006; 61: 176-81
- 61 Flaherty JH, Tariq SH, Raghavan S, Bakshi S, Moinuddin A, Morley JE. A model for managing delirious older inpatients. *J Am Geriatr Soc* 2003; 51: 1031-5
- 62 O'Keeffe ST. Down with bedrails? *Lancet* 2004; 363: 343-4
- 63 Lonergan E, Britton AM, Luxenberg J, Wyller T. Antipsychotics for delirium. *Cochrane Database Syst Rev* 2007: CD005594
- 64 Overshott R, Karim S, Burns A. Cholinesterase inhibitors for delirium. *Cochrane Database Syst Rev* 2008: CD005317
- 65 Jackson KC, Lipman, A.G. Drug therapie for delirium in terminally ill patients. *The Cochrane Library* 2007: 1-11
- 66 Inouye SK, Bogardus ST, Jr., Charpentier PA, Leo-Summers L, Acampora D, Holford TR, Cooney LM, Jr. A multicomponent intervention to prevent delirium in hospitalized older patients. *N Engl J Med* 1999; 340: 669-76
- 67 Siddiqi N, Stockdale R, Britton AM, Holmes J. Interventions for preventing delirium in hospitalised patients. *Cochrane Database Syst Rev* 2007: CD005563
- 68 Lundstrom M, Edlund A, Lundstrom G, Gustafson Y. Reorganization of nursing and medical care to reduce the incidence of postoperative delirium and improve rehabilitation outcome in elderly patients treated for femoral neck fractures. *Scand J Caring Sci* 1999; 13: 193-200
- 69 Young LJ, George J. Do guidelines improve the process and outcomes of care in delirium? *Age Ageing* 2003; 32: 525-8
- 70 Inouye SK, Bogardus ST, Jr., Williams CS, Leo-Summers L, Agostini JV. The role of adherence on the effectiveness of nonpharmacologic interventions: evidence from the delirium prevention trial. *Arch Intern Med* 2003; 163: 958-64
- 71 Inouye SK, Bogardus ST, Jr., Baker DI, Leo-Summers L, Cooney LM, Jr. The Hospital Elder Life Program: a model of care to prevent cognitive and functional decline in older hospitalized patients. *Hospital Elder Life Program. J Am Geriatr Soc* 2000; 48: 1697-706

Literaturverzeichnis

- 72 Inouye SK, Baker DI, Fugal P, Bradley EH. Dissemination of the hospital elder life program: implementation, adaptation, and successes. *J Am Geriatr Soc* 2006; 54: 1492-9
- 73 McCusker J, Cole M, Abrahamowicz M, Han L, Podoba JE, Ramman-Haddad L. Environmental risk factors for delirium in hospitalized older people. *J Am Geriatr Soc* 2001; 49: 1327-34
- 74 Caplan GA, Coconis J, Board N, Sayers A, Woods J. Does home treatment affect delirium? A randomised controlled trial of rehabilitation of elderly and care at home or usual treatment (The REACH-OUT trial). *Age Ageing* 2006; 35: 53-60
- 75 Rockwood K, Cosway S, Stolee P, Kydd D, Carver D, Jarrett P, O'Brien B. Increasing the recognition of delirium in elderly patients. *J Am Geriatr Soc* 1994; 42: 252-6
- 76 Ely EW, Gautam S, Margolin R, Francis J, May L, Speroff T, Truman B, Dittus R, Bernard R, Inouye SK. The impact of delirium in the intensive care unit on hospital length of stay. *Intensive Care Med* 2001; 27: 1892-900
- 77 Wild DG, A.; Martin, M.; Eremenco, S.; McElroy, S.; Verjee-Lorenz, A.; Erikson, P. . Principles of Good Practice for the Translation and Cultural Adaption Process for Patient-Reported Outcomes (PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural Adaption. *Value Health* 2005; 8: 94-104
- 78 Lutz A, Radtke,F.M., Franck,M., Seeling,M., Gaudreau,J.D., Kleinwachter,R., Kork,F., Zieb,A., Heymann,A., Spies,C.D. [The Nursing Delirium Screening Scale (NU-DESC)]. *Anesthesiol Intensivmed Notfallmed Schmerzther* 2008; 43: 98-102
- 79 Lee EW, Dubin N. Estimation and sample size considerations for clustered binary responses. *Stat Med* 1994; 13: 1241-52
- 80 Koster S, Hensens AG, Oosterveld FG, Wijma A, van der Palen J. The delirium observation screening scale recognizes delirium early after cardiac surgery. *Eur J Cardiovasc Nurs* 2009
- 81 Morimoto Y, Yoshimura M, Utada K, Setoyama K, Matsumoto M, Sakabe T. Prediction of postoperative delirium after abdominal surgery in the elderly. *J Anesth* 2009; 23: 51-6
- 82 Koebrugge B, Koek HL, van Wensen RJ, Dautzenberg PL, Bosscha K. Delirium after abdominal surgery at a surgical ward with a high standard of delirium care: incidence, risk factors and outcomes. *Dig Surg* 2009; 26: 63-8
- 83 Radtke FM, Franck M, Schneider M, Luetz A, Seeling M, Heinz A, Wernecke KD, Spies CD. Comparison of three scores to screen for delirium in the recovery room. *Br J Anaesth* 2008; 101: 338-43
- 84 Redelmeier DA, Thiruchelvam D, Daneman N. Delirium after elective surgery among elderly patients taking statins. *Cmaj* 2008; 179: 645-52

Literaturverzeichnis

- 85 Yildizeli B, Ozyurtkan MO, Batirel HF, Kuscu K, Bekiroglu N, Yuksel M. Factors associated with postoperative delirium after thoracic surgery. *Ann Thorac Surg* 2005; 79: 1004-9
- 86 Minden SL, Carbone LA, Barsky A, Borus JF, Fife A, Fricchione GL, Orav EJ. Predictors and outcomes of delirium. *Gen Hosp Psychiatry* 2005; 27: 209-14
- 87 Klugkist M, Sedemund-Adib B, Schmidtke C, Schmucker P, Sievers HH, Huppe M. [Confusion Assessment Method for the Intensive Care Unit (CAM-ICU): diagnosis of postoperative delirium in cardiac surgery]. *Anaesthesist* 2008; 57: 464-74
- 88 Inouye SK. Delirium in older persons. *N Engl J Med* 2006; 354: 1157-65
- 89 Leung JM, Leung VW, Leung CM, Pan PC. Clinical utility and validation of two instruments (the Confusion Assessment Method Algorithm and the Chinese version of Nursing Delirium Screening Scale) to detect delirium in geriatric inpatients. *Gen Hosp Psychiatry* 2008; 30: 171-6
- 90 de Jonghe JF, Kalisvaart KJ, Dijkstra M, van Dis H, Vreeswijk R, Kat MG, Eikelenboom P, van der Ploeg T, van Gool WA. Early symptoms in the prodromal phase of delirium: a prospective cohort study in elderly patients undergoing hip surgery. *Am J Geriatr Psychiatry* 2007; 15: 112-21
- 91 O'Keeffe ST. Clinical subtypes of delirium in the elderly. *Dement Geriatr Cogn Disord* 1999; 10: 380-5
- 92 Kiely DK, Jones RN, Bergmann MA, Marcantonio ER. Association between psychomotor activity delirium subtypes and mortality among newly admitted post-acute facility patients. *J Gerontol A Biol Sci Med Sci* 2007; 62: 174-9
- 93 Rolfson DB, McElhaney JE, Jhangri GS, Rockwood K. Validity of the confusion assessment method in detecting postoperative delirium in the elderly. *Int Psychogeriatr* 1999; 11: 431-8
- 94 Meagher DJ, Maclullich AM, Laurila JV. Defining delirium for the International Classification of Diseases, 11th Revision. *J Psychosom Res* 2008; 65: 207-14
- 95 Cole MG, Dendukuri N, McCusker J, Han L. An empirical study of different diagnostic criteria for delirium among elderly medical inpatients. *J Neuropsychiatry Clin Neurosci* 2003; 15: 200-7
- 96 Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HC. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Bmj* 2003; 326: 41-4
- 97 Sackett DL. Bias in analytic research. *J Chronic Dis* 1979; 32: 51-63
- 98 Greenhalgh T. How to read a paper. Papers that report diagnostic or screening tests. *Bmj* 1997; 315: 540-3

Literaturverzeichnis

- 99 Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. *Jama* 1995; 274: 645-51
- 100 Zou Y, Cole MG, Primeau FJ, McCusker J, Bellavance F, Laplante J. Detection and diagnosis of delirium in the elderly: psychiatrist diagnosis, confusion assessment method, or consensus diagnosis? *Int Psychogeriatr* 1998; 10: 303-8
- 101 Liptzin B, Levkoff SE, Cleary PD, Pilgrim DM, Reilly CH, Albert M, Wetle TT. An empirical study of diagnostic criteria for delirium. *Am J Psychiatry* 1991; 148: 454-7
- 102 Potter J, George J. The prevention, diagnosis and management of delirium in older people: concise guidelines. *Clin Med* 2006; 6: 303-8

8 Publikationsliste

Radtke,F.M., Franck,M., Schust,S., Boehme,L., Pascher,A., Bail,H., Seeling,M., Luetz,A., Wernecke,K., Heinz,A., Spies,C.D. A comparison of three scores to screen for delirium on surgical ward. World Journal of Surgery in press 2010

9 Abbildungsverzeichnis

	Seite
Abbildung 1: Kriterien zur Diagnose nach DSM-IV	16
Abbildung 2: Nu-DESC nach Gaudreau et al. 2005 und Lütz et al. 2008	17
Abbildung 3: Die CAM nach Inouye et al. 1990	18
Abbildung 4: Flussdiagramm des Studienverlaufs gemäß STARD (Bossuyt et al. 2003)	20
Abbildung 5: Sensitivitäten über sechs postoperative Tage	32
Abbildung 6: Spezifitäten über sechs postoperative Tage	33
Abbildung 7: ROC-Kurve für CAM und Nu-DESC, Goldstandard = DSM-IV Kriterien.....	37
Abbildung 8: Übereinstimmung verschiedener Diagnosesysteme, modifiziert nach Laurila et al. 2004.....	47

10 Tabellenverzeichnis

	Seite
Tabelle 1: präoperative Basisdaten	21
Tabelle 2: Basisdaten zu Operation und Narkose	21
Tabelle 3: postoperative Daten.....	22
Tabelle 4: Vergleich Basisdaten Patienten mit und ohne Delir nach DSM-IV	22
Tabelle 5: Delirium-Score und Outcome, Median, 25- und 75%-Perzentile in Klammern, p-Werte für Mann-Whitney-U-Test	23
Tabelle 6: Anzahl der Patienten CAM versus DSM-IV Kriterien	24
Tabelle 7: Anzahl der Patienten Nu-DESC versus DSM-IV Kriterien.....	24
Tabelle 8: mittlere Dauer des Delirs nach DSM-IV, CAM und Nu-DESC in Tagen, Median, 25%- und 75%-Perzentile in Klammern	24
Tabelle 9: Anzahl der Patiententage CAM versus DSM-IV Kriterien	25
Tabelle 10: Anzahl der Patiententage Nu-DESC versus DSM-IV Kriterien.....	25
Tabelle 11: DSM-IV – Erfüllungsgrad der Einzelkriterien Bewusstseinsstörung, Veränderung kognitiver Funktionen und akute Entwicklung und Fluktuation in Anzahl der Patienten, Prozentangaben in Klammern. 0 = nicht erfüllt, 1 = erfüllt.....	26
Tabelle 12: An mindestens einem Tag maximal erreichte Anzahl der Faktoren im DSM-IV, Anzahl der Patienten (% in Klammern)	26
Tabelle 13: DSM-IV – Erfüllungsgrad der Einzelkriterien Bewusstseinsstörung, Veränderung kognitiver Funktionen und akute Entwicklung und Fluktuation in Anzahl der Patiententage, Prozentangaben in Klammern. 0 = nicht erfüllt, 1 = erfüllt	27
Tabelle 14: An mindestens einem Tag maximal erfüllte Anzahl der Faktoren im DSM-IV, Anzahl der Patiententage (% in Klammern)	27
Tabelle 15: CAM – Erfüllungsgrad der Einzelkriterien akuter Beginn, Unaufmerksamkeit, unorganisiertes Denken und veränderte Bewusstseinslage in Anzahl der Patienten, Prozentangaben in Klammern. 0 = nicht erfüllt, 1 = erfüllt	28
Tabelle 16: CAM – Erfüllungsgrad der Einzelkriterien akuter Beginn,	

Tabellenverzeichnis

Unaufmerksamkeit, unorganisiertes Denken und veränderte Bewusstseinslage in Anzahl der Patiententage, Prozentangaben in Klammern. 0 = nicht erfüllt, 1 = erfüllt	28
Tabelle 17: Nu-DESC – Erfüllungsgrad der Einzelkriterien Desorientierung, unangemessenes Verhalten, unangemessene Kommunikation, Illusion/Halluzination und psychomotorische Retardierung in Anzahl der Patienten, Prozentangaben in Klammern. 0 = nicht erfüllt, 1 = leichte Ausprägung, 2 = schwere Ausprägung	29
Tabelle 18: An mindestens einem Tag maximal erreichte Punktzahl im Nu-DESC, Anzahl der Patienten (% in Klammern)	29
Tabelle 19: Nu-DESC – Erfüllungsgrad der Einzelkriterien Desorientierung, unangemessenes Verhalten, unangemessene Kommunikation, Illusion/Halluzination und psychomotorische Retardierung in Anzahl der Patiententage, Prozentangaben in Klammern. 0 = nicht erfüllt, 1 = leichte Ausprägung, 2 = schwere Ausprägung	30
Tabelle 20: Erreichte Punktzahl im Nu-DESC, Anzahl der Patiententage (% in Klammern)	30
Tabelle 21: Berechnung der Hauptmerkmale von Screening-Tests; a, b, c und d entsprechen den jeweiligen Zellen in der Kontingenztafel mit a – DSM-IV positiv und Test positiv, b – DSM-IV negativ und Test positiv, c – DSM-IV positiv und Test negativ, d – DSM-IV negativ und Test negativ (siehe Tab. 9+10)	31
Tabelle 22: Anzahl der Patiententage mit Delir (DSM-IV) zeitlich vor oder nach einem falsch positiven Testergebnis bei Anwendung der Nu-DESC	33
Tabelle 23: Testgütekriterien von CAM und Nu-DESC im Vergleich	36
Tabelle 24: Area under the Curve für CAM und Nu-DESC im Vergleich, p-Wert für Unterschied zur Indifferenzfläche von 0.5.....	37
Tabelle 25: Interrater-Reliabilität für CAM und Nu-DESC im Vergleich, p-Wert für McNemar	38
Tabelle 26: Durchschnittlicher Zeitaufwand zur Durchführung von CAM und Nu-DESC, p-Wert für t-Test	38

11 Danksagung

Zunächst möchte ich allen Patienten danken, die sich - zum Teil erst nach ausführlichster Aufklärungsarbeit unsererseits - bereit erklärt haben, an dieser Studie teilzunehmen.

Des Weiteren danke ich Fr. Prof. Dr. med. Claudia Spies für die Überlassung des Themas, das Einreichen der Genehmigungen zur Ermöglichung der Arbeit, die Betreuung der Arbeit, die Diskussion in den Doktorandenseminaren sowie die Korrektur meiner Arbeit.

Herrn Dr. med. Radtke sowie Herrn Dr. med. Martin Franck danke ich für die unmittelbare hervorragende Unterstützung bei der Umsetzung dieser Arbeit und die Unterstützung beim Schreiben der Promotionsarbeit.

In diesem Zusammenhang möchte ich auch Lina Böhme für die Zusammenarbeit beim Einschließen und Untersuchen der Patienten danken.

Herrn Prof. Dr. rer. nat. Klaus-Dieter Wernecke (Geschäftsführer der SOSTANA GmbH und ehem. Direktor des Instituts für Medizinische Biometrie des Universitätsklinikums Charite) danke ich für die Beratung in allen biometrischen und statistischen Fragen.

Mein besonderer Dank gilt meiner Familie, die immer Verständnis für zeitliche Engpässe meinerseits hatte und trotz viel wichtigerer Ereignisse in unserem Leben auf ihre angenehme, unaufgeregte Art regen Anteil am Fortgang meiner Arbeit nahm.

Ganz besonders möchte ich auch meinem Freund Benjamin Bleicken für seinen grenzenlosen Optimismus, seine Geduld und die vielen kleinen Aufmerksamkeiten danken. Du hattest so einige Tränen zu trocknen, was du irgendwie immer hingekriegt hast. Wir saßen schließlich beide im selben Boot...

Und zu guter Letzt: Mama! Du weißt, warum du hier stehst und was ich dir mit diesen Zeilen sagen will. Ohne dich wäre diese Arbeit so nicht entstanden. Und weil alles andere sehr schwer in Worte zu fassen ist: Danke!

12 Lebenslauf

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

13 Erklärung an Eides Statt

„Ich, Sabine Schust, erkläre, dass ich die vorgelegte Dissertationsschrift mit dem Thema: „Das postoperative Delirium. Vergleich der Nursing Delirium Screening Scale und der Confusion Assessment Method im Screening auf der peripheren chirurgischen Station“ selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt, ohne die (unzulässige) Hilfe Dritter verfasst und auch in Teilen keine Kopien anderer Arbeiten dargestellt habe.“

Berlin, den 25.11.2009

(Sabine Schust)