

# Kapitel 3

## Methoden

### 3.1 Fragestellungen, Hypothesen und Ziele

Übergeordnetes Ziel dieser Studie ist die Erforschung der Phänomenologie von AR (SSV und SV) unter Gefangenen einer deutschen Gefangenenstichprobe. Es geht im Besonderen darum, sich einem im deutschen Sprachraum bisher kaum untersuchten Phänomen zu widmen. Eigene Erhebungen erscheinen unerlässlich, wenn den Bediensteten des Justizvollzuges im Hinblick auf den Umgang mit dem Problem praktische Hilfestellungen an die Hand gegeben werden sollen.

Die vorliegende Arbeit bearbeitet zwei große Fragestellungen. Dies sind einerseits Vergleiche innerhalb der Gefangenen mit AR nach verschiedenen Charakteristika (Frage 1) und andererseits der Vergleich der Gruppen der Gefangenen mit und ohne AR (Frage 2). Die Fragestellungen und Hypothesen werden im Folgenden vorgestellt.

### 3.1.1 Frage 1: Selbstschädigendes Verhalten und Suizidversuche als distinkte Phänomene

Lassen sich Gefangene mit autodestruktiven Reaktionen und dabei niedriger Suizidabsicht bzw. niedriger Letalität von solchen mit jeweils hohen Werten differenzieren und unterscheiden sie sich in ihrem klinischen Bild und der Art ihrer autodestruktiven Reaktion?

Wie bereits dargestellt gehen einige Untersuchungen davon aus, dass die Gruppe der Selbstschädiger mit niedriger Suizidabsicht bzw. niedriger Letalität eine homogene Untergruppe unter den Gefangenen mit AR bilden (siehe „deliberate self-harm syndrome“ in Abschnitt 2.6.2). Diese unterscheiden sich durch das gemeinsame psychiatrische Störungsbild und die Begleitumstände und Rahmenbedingungen ihrer AR von der eher heterogenen Gruppe der AR mit höherer Suizidabsicht und Letalität. Dementsprechend wird vermutet ( $H_1$ ),<sup>1</sup> dass bei Gefangenen mit niedrigerer Suizidabsicht bzw. Letalität im Gegensatz zu denen mit höheren Werten...

1. ...sich Unterschiede hinsichtlich der Begleitumstände und Rahmenbedingungen der AR ergeben, wie
  - (a) dem Zeitpunkt der AR (Wochentag, Tageszeit),
  - (b) den Methoden der AR,
  - (c) den Motiven der AR,
  - (d) der Impulsivität (gH: höhere bei weniger ernsthaften AR),
  - (e) dem begleitendem Schmerzerleben (gH: geringeres bei weniger ernsthaften AR),

---

<sup>1</sup>In Klammern wird angegeben, ob es sich um eine ungerichtete Hypothese (uH) oder gerichtete Hypothese (gH) handelt, im letzten Fall mit Angabe der Richtung.

- (f) der Anzahl bisheriger eigener AR (intra- und extramural) und
  - (g) der Anzahl bisheriger AR von Personen im Umfeld des Probanden (intra- und extramural).
2. ...sich Unterschiede hinsichtlich forensischer Charakteristika ergeben, wie bei
- (a) der Anstalt des Gefangenen,
  - (b) der Haftform (U-Haft vs. Strafhaft),
  - (c) der Häufigkeit von Gewaltdelikten (gH: weniger bei weniger ernsthaften AR),
  - (d) der Haftdauer bis zur AR,
  - (e) der Anzahl der Vorinhaftierungen,
  - (f) der Dauer der erwarteten Reststrafe,
  - (g) der Anzahl der Disziplinverstöße,
  - (h) der Anzahl der Unterbringungen in einem besonders gesicherten Haftraum (bgH) und
  - (i) der Häufigkeit der Bedrohung durch Mitgefangene.
3. ...sich Unterschiede hinsichtlich der Depressivität ergeben, wobei sie
- (a) niedrigere Depressionswerte (BDI; gH),
  - (b) geringere Hoffnungslosigkeit (BHS; gH) und
  - (c) weniger Suizidgedanken und -wünsche (BSS; gH) haben.
4. ...sich Unterschiede hinsichtlich der Charakterpathologie ergeben, wobei
- (a) sich die Gruppen hinsichtlich der Häufigkeit von Persönlichkeitsstörungen unterscheiden und
  - (b) ihre Werte der PCL-R höher ausfallen (inkl. beider Faktoren; gH).
5. ...eine andere Prävalenz
- (a) des Missbrauchs bzw.

- (b) der Abhängigkeit von psychotropen Substanzen aufweisen und
  - (c) es unterschiedlich viele psychotisch gestörte Gefangene in den Gruppen gibt.
6. ...sich Unterschiede hinsichtlich demografischer bzw. anamnestischer Variablen ergeben, wie
- (a) beim Alter,
  - (b) der Herkunft,
  - (c) den Deutschkenntnissen,
  - (d) etwaigen Beziehungsproblemen oder
  - (e) Verlusterlebnissen,
  - (f) der Obdachlosigkeit,
  - (g) früheren Kontakten zu Psychotherapeuten,
  - (h) früherer psychiatrischer / psychotherapeutischer Behandlung (vorwiegend stationär),
  - (i) sexuellem Missbrauch und
  - (j) körperlichen Misshandlungen.

Die ( $H_0$ ) geht von keinem signifikanten (uH) bzw. keinem signifikanten oder einem entgegengesetzten Unterschied (gH) zwischen Gefangenen mit niedriger und hoher Suizidabsicht bzw. Letalität der Verletzungen aus.

### 3.1.2 Anwendungsziel 1: Diagnostische Leitlinien

Falls die Ergebnisse die Existenz einer homogenen Untergruppe von Selbstschädigern nahe legen, kann dies zum besseren Verständnis des Phänomens beitragen und den Umgang der Justizvollzugsbediensteten und Anstaltsärzte im Hinblick auf bestmögliche Intervention erleichtern. Dieser Aspekt des Vorhabens ist insofern von gesellschaftspolitischer Relevanz, als er den Leidensdruck von betroffenen Gefangenen mindern und dem

Justizvollzug mehr Kompetenz vermitteln möchte.

Die vorliegende Studie hat zum Ziel, eventuell auftretende Unterschiede zwischen verschiedenen ernsthaften Formen von AR (erfasst über Suizidabsicht und Letalität) in Gefängnissen hinsichtlich Maßen der Depression, Charakterpathologie und einigen demografischen und gefängnispezifischen Variablen zu untersuchen. Das Ziel ist dabei eine Grundlage für spezifische klinische Einschätzungen, höhere Vorhersagegenauigkeit und eine adäquate Form des Umgangs bzw. der Behandlung zu schaffen. Auf die Entwicklung eines Leitfadens für die Praxis wird dabei bewusst verzichtet (siehe dazu Abschnitt 5.2.2).

### 3.1.3 Frage 2: Risikofaktoren für autodestruktive Reaktionen

Unterscheiden sich die Gefangenen mit autodestruktiven Reaktionen von einer Kontrollgruppe und lassen sich Risikofaktoren finden, die geeignet erscheinen die Risikopopulation zu identifizieren?

Die Angaben der Literatur zu Risikofaktoren sind uneinheitlich und widersprechen sich oft (siehe Tabellen in Abschnitt 2.8), außerdem existieren keine neueren Studien aus dem deutschen Sprachraum zu Risikofaktoren für AR unter Gefangenen. Gefangene mit AR (AR-Gruppe) und eine Kontrollgruppe ohne AR (K-Gruppe) werden bezüglich der soziodemografischen, forensischen und psychologisch-psychiatrischen Risikofaktoren internationaler Arbeiten untersucht. Es wird angenommen, dass sich die Gefangenen mit AR hinsichtlich dieser Variablen signifikant von der K-Gruppe unterscheiden, wobei...

1. ...sich Unterschiede hinsichtlich der Begleitumstände und Rahmenbedingungen der AR ergeben, wie
  - (a) der Anzahl bisheriger eigener AR (intra- und extramural / gH: mehr in der AR-Gruppe) und

- (b) der Anzahl bisheriger AR von Personen im Umfeld des Probanden (intra- und extramural / gH: mehr in der AR-Gruppe).
2. ...sich Unterschiede hinsichtlich forensischer Charakteristika ergeben, wie bei
- (a) der Häufigkeit von Gewaltdelikten (gH: mehr in der AR-Gruppe),
  - (b) der Anzahl der Vorinhaftierungen,
  - (c) der Dauer der erwarteten Reststrafe (gH: größer in der AR-Gruppe),
  - (d) der Anzahl der Disziplinverstöße (gH: mehr in der AR-Gruppe),
  - (e) der Anzahl der Unterbringungen in einem bgH (gH: mehr in der AR-Gruppe) und
  - (f) der Häufigkeit der Bedrohung durch Mitgefangene (gH: mehr in der AR-Gruppe).
3. ...sich Unterschiede hinsichtlich der Depressivität ergeben, wobei die AR-Gruppe
- (a) höhere Depressionswerte (BDI; gH),
  - (b) mehr Hoffnungslosigkeit (BHS; gH) und
  - (c) mehr Suizidgedanken und -wünsche (BSS; gH) zeigen und äußern.
4. ...sich Unterschiede hinsichtlich der Charakterpathologie ergeben, wobei
- (a) sich die Gruppen hinsichtlich der Häufigkeit von Persönlichkeitsstörungen unterscheiden und
  - (b) die Werte der PCL-R bei der AR-Gruppe höher ausfallen (inkl. beider Faktoren; gH).
5. ...sie eine andere Prävalenz
- (a) des Missbrauchs bzw.
  - (b) der Abhängigkeit von psychotropen Substanzen aufweisen und
  - (c) es unterschiedlich viele psychotisch gestörte Gefangene in den Gruppen gibt.

6. ...sich Unterschiede hinsichtlich demografischer bzw. anamnestischer Variablen ergeben, wie

- (a) beim Alter,
- (b) der Herkunft,
- (c) den Deutschkenntnissen,
- (d) etwaigen Beziehungsproblemen (gH: häufiger in der AR-Gruppe) oder
- (e) Verlusterlebnissen (gH: häufiger in der AR-Gruppe),
- (f) der Obdachlosigkeit (gH: häufiger in der AR-Gruppe),
- (g) früheren Kontakten zu Psychotherapeuten (gH: mehr in der AR-Gruppe),
- (h) früherer psychiatrischer / psychotherapeutischer Behandlung (vorwiegend stationär / gH: häufiger in der AR-Gruppe),
- (i) sexuellem Missbrauch (gH: häufiger in der AR-Gruppe) und
- (j) körperlichen Misshandlungen (gH: häufiger in der AR-Gruppe).

Die ( $H_0$ ) geht von keinem signifikanten (uH) bzw. keinem signifikanten oder einem entgegengesetzten Unterschied (gH) zwischen den Gefangenen der AR-Gruppe und der K-Gruppe aus.

Zeigen sich signifikante Unterschiede, so wird die prädiktive Validität des einzelnen Risikofaktors untersucht, wobei Zusammenhangshypothesen auf die Stärke der Korrelation zwischen dem Risikofaktor (Prädiktor) und der AR (Kriterium) geprüft werden.

### 3.1.4 Anwendungsziel 2: Risikoeinschätzung

Besonders valide Prädiktoren dienen der künftigen Risikoeinschätzung, mit deren Hilfe eine Erkennung und Risikoklassifikation ermöglicht werden soll.

### 3.1.5 Statistische Formulierung der Hypothesen

Ist von einer größeren Ausprägung des Merkmals in einer der beiden Gruppen auszugehen, lautet die statistische Umsetzung der Unterschiedshypothese:

$$\text{UH: } H_1: \mu_1 > \mu_2 \quad H_0: \mu_1 \leq \mu_2,$$

wobei dementsprechend von einem positivem Zusammenhang zwischen Prädiktor und Kriterium ausgegangen wird:

$$\text{ZH: } H_1: \varrho > 0 \quad H_0: \varrho \leq 0.$$

Lautet die statistische Umsetzung der Unterschiedshypothese:

$$\text{UH: } H_1: \mu_1 < \mu_2 \quad H_0: \mu_1 \geq \mu_2,$$

wird hier von einem negativen Zusammenhang zwischen Prädiktor und Kriterium ausgegangen:

$$\text{ZH: } H_1: \varrho < 0 \quad H_0: \varrho \geq 0.$$

In beiden Fällen gerichteter Hypothesen werden einseitige Signifikanzprüfungen durchgeführt.

Kann keine Richtung des Unterschieds angegeben werden, lautet die statistische Hypothese (ungerichtete Hypothese  $\Rightarrow$  zweiseitige Signifikanzprüfung):

$$\text{UH: } H_1: \mu_1 \neq \mu_2 \quad H_0: \mu_1 = \mu_2,$$

wobei bei einem signifikanten Unterschied anschließend auch keine Richtung im Zusammenhang unterstellt wird:

$$\text{ZH: } H_1: \varrho \neq 0 \quad H_0: \varrho = 0.$$



## 3.2 Operationale Definition der abhängigen Variable autodestruktives Verhalten

Hier wird nicht noch einmal auf die phänomenologische Definition von AR eingegangen, wie sie unter anderem im Kapitel 2.2 gegeben wurde. Vielmehr soll hier eine operationale Definition des Untersuchungsgegenstandes erklärt werden, d.h. welches Verhalten von Gefangenen betrachtet werden soll und wie es sich dabei von ähnlichen Verhaltensweisen abgrenzt.

Die vorliegende Arbeit hat sämtliche Verhaltensweisen von Gefangenen aus den Berliner Vollzugsanstalten zum Gegenstand, bei denen ein Gefangener mit der Absicht sich zu schädigen:

- ein Verhalten initiierte (zum Beispiel Schneiden, Wiedereröffnung alter Wunden, Sprung aus einer großen Höhe, Erhängen, Erdrosseln, Ersticken, Feuerlegen, Sichlegen oder Sprung vor einen sich bewegenden Gegenstand),
- eine Substanz zu sich nahm (injizierte, schluckte, inhalierte) oder
- ein Objekt in sich aufnahm (Schlucken, unter die Haut schieben, in natürliche Körperöffnungen einführen).

Eine genaue Aufführung von allen denkbaren Methoden und Arten von AR findet sich im ICD-10, Kapitel XX: äußere Ursachen für Morbidität und Mortalität – vorsätzliche Selbstbeschädigung (Weltgesundheitsorganisation, 1993, X60-X84). Diese Aufstellung wird auch für die Kodierung der verwendeten Methoden herangezogen und ist im Anhang E zu finden.

Ausgenommen sind dabei Verhaltensweisen bei denen das Ziel nicht in der Schädigung selbst lag, sondern die einem anderen Zweck dienten (zum Beispiel Tätowieren, Piercing und andere Formen die ausschließlich dem Körperschmuck dienen, sowie ausschließlich sexuell motiviertes Verhalten mit selbstschädigendem Charakter, siehe dazu

auch Conacher & Westwood, 1987). Ebenso findet das Verweigern der Nahrungsaufnahme keine Berücksichtigung, da sich kein exakter Zeitpunkt für die Tatbegehung angeben lässt und die Hintergründe für diese Verhaltensweise meist in einem „politischen“ (Yaroshevsky, 1975) oder allgemeinen Protestmotiv (Larkin, 1991) liegen und hier übergeordnete Kontextvariablen an Einfluss gewinnen, die diese Studie nicht mit berücksichtigen kann. Ein ungesunder oder riskanter Lebenswandel (siehe Boiko & Lester, 1995) wird hier ebenfalls nicht mit aufgenommen und untersucht (siehe auch Herpertz & Saß, 1994), da sich ebenfalls kein diskreter Zeitpunkt der Schädigung angeben ließe. Auch die heimliche Selbstbeschädigung<sup>2</sup> und artifizielle Störung werden nicht mit aufgenommen, für welche die Verleugnung insbesondere dem Arzt gegenüber charakteristisch ist (Herpertz & Saß, 1994).

Fälle von Überdosierungen von Substanzen, die unter das BtMG Fallen, wurden nur dann aufgenommen, wenn der Gefangenen davon ausging, dass es sich um eine potentiell letale Dosierung handelte und den Tod zumindest billigend in Kauf nahm. In solchen Fällen war also ein etwaiges Motiv hinter einer Überdosierung zu erfragen (siehe dazu O’Carroll et al., 1996).

In jedem Fall musste die Absicht einer AR in irgendeiner Form umgesetzt worden oder begonnen worden sein. Damit sollten Fälle reiner Drohungen (keinem Verhalten im eigentlichen Sinn, s.o.), zu denen es im Justizvollzug ebenfalls häufig kommt, von der Stichprobe ausgeschlossen werden.

### 3.3 Design der Studie

Die Untersuchung der Fragestellungen ist als quasiexperimentelles Gruppendesign konzipiert (Case-Control-Studie mit parallelisierter und randomisierter Zuweisung von Probanden in die Kontrollgruppe) und erhebt dabei die relevanten Variablen retrospektiv

---

<sup>2</sup>Beachte: Hierbei ist nicht die Rahmenbedingung einer heimlichen Tatbegehung, sondern vielmehr auch die spätere Absichtserklärung durch den Schädiger gemeint.

und querschnittlich (ein Messzeitpunkt).

### 3.4 Stichprobe

Diese Studie widmet sich ausschließlich den AR von männlichen Gefangenen, da AR von Frauen hinsichtlich der Begleitumstände und des klinischen Bildes der Patientinnen ein eigenständiges Phänomen zu sein scheinen (Hawton, Rodham, Evans & Weatherall, 2002b; Fulwiler et al., 1997; Jackson, 2000; Livingston, 1997; Maden et al., 1994; Meltzer et al., 2003). Zudem wurden trotz einer deutlich männlich dominierten Gefängnispopulation die meisten Arbeiten zum Thema an weiblichen Gefangenen durchgeführt (Fulwiler et al., 1997; Lloyd, 1990). Die vorliegende Arbeit versucht, die Lücke zu schließen, die durch den Mangel an Studien mit männlichen Probanden (Lloyd, 1990; Maden et al., 1994) und den fehlenden Arbeiten zum Thema aus dem deutschen Sprachraum während der letzten Jahrzehnte entstanden ist.

Die Probanden der K-Gruppe wurden parallelisiert zur AR-Gruppe erhoben, wobei die Kriterien für die Parallelisierung die Haftdauer bis zur Index-AR<sup>3</sup> und der Haftstatus<sup>4</sup> waren.

In den Tabellen D.1 bis D.5 des Anhangs sind demografische und forensische Charakteristika der Berliner Justizvollzugspopulation aufgeführt. Sie dienen der Information und der Beschreibung des Probandengutes.

In Tabelle D.6 ist die vorgesehene und die tatsächliche Belegung der JVA Moabit, Tegel und der JSA Berlin und die daraus resultierende prozentuale Auslastung aufgeführt.<sup>5</sup> Diese Zahlen werden wöchentlich einmal erhoben (so genannte „Mittwochsahlen“ der

---

<sup>3</sup>Im Falle der K-Gruppe also bis zum Interview

<sup>4</sup>Untersuchungshaft vs. Strafhaft

<sup>5</sup>Es zeigt sich in allen drei Anstalten eine Überbelegung.

Senatsverwaltung für Justiz), wobei die Zahlen vom 1. Februar 2006 gewählt wurden, da sie einen möglichst repräsentativen Überblick über die tatsächliche Belegungssituation vermitteln und nicht verzerrt sind, wie beispielsweise die drei Monate nach Oktober, in denen die jährliche Weihnachtsamnestie zu einem vorübergehenden Rückgang der Zahlen führt (siehe D.3).

### 3.4.1 Beschreibung der gesamten Stichprobe

Insgesamt wurden für die Studie 155 Gefangene gebeten, an einem Interview teilzunehmen, wobei 15 eine Teilnahme ablehnten (9.7% Ablehnungsquote). Aufgrund datenschutzrechtlicher Bestimmungen war es nicht möglich, Angaben über diese Gefangenen zu sammeln und auszuwerten. Als Gründe für die Verweigerung wurden dem Untersucher wiederholt die Motive genannt „keine Lust“ zu haben und nicht teilnehmen zu wollen, weil „man sich davon nichts verspreche“. Der Untersucher sah von weiteren Nachfragen ab. Manchmal stellte sich der Verdacht ein, dass die Gefangenen (sowohl der AR- wie auch der K-Gruppe) befürchteten, vom Psychologen, der sie untersuchte, gegen ihren Willen in die Psychiatrie eingewiesen zu werden (siehe auch Smith, 1984), obwohl der Untersucher ausdrücklich und mehrfach derartige Konsequenzen ausschloss. Von den  $N = 140$  Gefangenen, die bereit waren an der Studie teilzunehmen, entfielen jeweils  $n_1 = n_2 = 70$  auf die Gruppe der Gefangenen mit autodestruktiven Reaktionen (AR-Gruppe) und auf die Kontrollgruppe ohne AR (K-Gruppe). Das Durchschnittsalter aller Gefangenen lag bei 27.97 Jahren ( $SD = 9.223$ ; Streubreite = 14-56). In Tabelle 3.1 sind die Anstalten und der Haftstatus aller Probanden ersichtlich.

### 3.4.2 Gefangene mit autodestruktiven Reaktionen – AR-Gruppe

Es wurden insgesamt 70 Gefangene mit AR (Alter:  $M = 27.89$ ;  $SD = 9.152$ ; Streubreite = 14-56) untersucht. Die Gefangenen der AR-Gruppe waren bis zur Index-AR im Durchschnitt 360.06 Tage inhaftiert ( $SD = 901.971$ ; Streubreite = 1-5075). Aus Tabelle

**Tabelle 3.1:** Anzahl (und %) aller befragten Gefangener aus den jeweiligen Anstalten nach Haftstatus ( $N = 140$ )

Anstalt	Haftstatus		Gesamt
	U-Haft	Strafhaft	
JSA Berlin	27 (67.5%)	13 (32.5%)	40 (28.6%)
JVA Moabit	51 (71.8%)	20 (28.2%)	71 (50.7%)
JVA Tegel	0 (0.0%)	29 (100.0%)	29 (20.7%)
Gesamt	78 (55.7%)	62 (44.3%)	140 (100.0%)

**Tabelle 3.2:** Anzahl (und %) der Gefangenen mit AR aus den jeweiligen Anstalten nach Haftstatus ( $n_1 = 70$ )

Anstalt	Haftstatus		Gesamt
	U-Haft	Strafhaft	
JSA Berlin	14 (70.0%)	6 (30.0%)	20 (28.6%)
JVA Moabit	25 (71.4%)	10 (28.6%)	35 (50.0%)
JVA Tegel	0 (0.0%)	15 (100.0%)	15 (21.4%)
Gesamt	39 (55.7%)	31 (44.3%)	70 (100.0%)

3.2 werden die Anstalten in denen die Gefangenen einsaßen und der jeweilige Haftstatus ersichtlich. Sechs Gefangene, die AR gezeigt hatten, weigerten sich, an der Studie teilzunehmen (8.6% Ablehnungsquote). Die Datenaufnahme für diese Gruppe begann am 01.07.2004 und endete nach zwei Jahren, am 30.06.2006.

Um Ermessen zu können, inwieweit die Anzahl der erhobenen Fälle von AR den tatsächlich vorgefallenen entspricht, bietet sich als Möglichkeit der externen Validierung die Senatsverwaltung für Justiz an, die Meldungen über so genannte besondere Vorkommnisse erhält. Im Anhang (siehe B.1) sind die bei der Senatsverwaltung für Justiz im Jahr 2005 eingegangenen Meldungen hinsichtlich AR aus den Berliner Justizvollzugsanstalten aufgeführt und den Zahlen aus der eigenen Erhebung aus dem Jahr

2005 gegenübergestellt. Es wird ersichtlich dass die vorliegende Untersuchung in allen Bereichen<sup>6</sup> mehr Vorfälle entdeckte als bei der Senatsverwaltung gemeldet. Dies dürfte darauf zurückzuführen sein, dass die in dieser Arbeit aufgenommenen Fälle direkt bei den AGSt erfragt wurden. Bei Durchsicht der Meldungen verfestigte sich darüber hinaus der Eindruck, dass die Schwelle, ab wann eine AR gemeldet wird, nicht einheitlich gehandhabt wird. So sind dort einerseits Fälle aufgeführt, die offensichtlich von geringer Suizidabsicht und Letalität geprägt waren, wobei andere mit hoher Ernsthaftigkeit fehlen. Insofern scheint dieses Kriterium selbst von eingeschränkter Reliabilität und Validität zu sein.

### 3.4.3 Kontrollgruppe – K-Gruppe

Es wurden 70 Gefangene für die K-Gruppe untersucht und befragt. Das Durchschnittsalter der Probanden aus der Kontrollgruppe lag bei 28.06 Jahren ( $SD = 9.359$ ; Streubreite = 18-55). Damit unterscheidet sich das Alter der AR-Gruppe von dem der K-Gruppe nicht signifikant ( $t = .104$ ,  $df = 69$ ,  $p = .913$ , zweiseitige Prüfung). Die Gefangenen der K-Gruppe waren bis zum Interview 303.93 Tage in Haft ( $SD = 706.768$ ; Streubreite = 1-4745), wobei hier ebenfalls kein signifikanter Unterschied zum durchschnittlichen Wert der AR-Gruppe bestand<sup>7</sup> ( $t = .409$ ,  $df = 69$ ,  $p = .684$ , zweiseitige Prüfung). In Tabelle 3.3 sind die Anstalten der Gefangenen und ihr Haftstatus aufgeführt.<sup>8</sup>

Neun Gefangene, die zunächst per Zufallsauswahl für die Teilnahme vorgesehen waren, weigerten sich teilzunehmen (12.9% Ablehnungsquote). Aus den oben erwähnten Gründen konnten keine Angaben über diese Gefangenen gesammelt werden. Die genannten Motive für die Ablehnung unterschieden sich nicht von denen aus der AR-Gruppe. Bei einem Gefangenen wurde von der Durchführung eines Interviews abgesehen, weil er

---

<sup>6</sup>Sowohl hinsichtlich der Anstalten als auch hinsichtlich der Art des Vorfalls

<sup>7</sup>Beachte: Die Haftdauer ist das erste Kriterium nach dem die Parallelisierung erfolgte.

<sup>8</sup>Beachte: Der Haftstatus ist das zweite Kriterium nach dem die Parallelisierung erfolgte.

**Tabelle 3.3:** Anzahl (und %) der Gefangenen aus der K-Gruppe aus den jeweiligen Anstalten nach Haftstatus ( $n_2 = 70$ )

Anstalt	Haftstatus		Gesamt
	U-Haft	Strafhaft	
JSA Berlin	13 (65.0%)	7 (35.0%)	20 (28.6%)
JVA Moabit	26 (71.4%)	10 (28.6%)	36 (51.4%)
JVA Tegel	0 (0.0%)	14 (100.0%)	14 (20.0%)
Gesamt	39 (55.7%)	31 (44.3%)	70 (100.0%)

aufgrund einer starken Entzugssymptomatik nicht in der Lage gewesen wäre, einem Gespräch zu folgen. Aufgrund der angestrebten parallelisierten Erfassung konnte mit der Datenaufnahme erst relativ spät begonnen werden. Diese begann am 01.02.2006 und endete am 01.09.2006.

### 3.5 Erhebungsinstrumente

Da die Ergebnisse der hier verwendeten Skalen bei selbstschädigenden Gefangenen schon allein aufgrund unterschiedlicher Definitionen des Begriffes „autodestruktive Reaktion“ stark variieren dürften, werden Normwerte aus der Literatur nur für Gefangene ohne AR referiert. Dementsprechend werden hinsichtlich der beiden letzten Skalen (SIS und LSARS-II) keine Normwerte berichtet,<sup>9</sup> weil diese Skalen nur bei Gefangenen mit AR zum Einsatz kommen.

Bei der Diagnosestellung bedient sich die vorliegende Arbeit den Kriterien des DSM-IV (Saß, Wittchen & Zaudig, 2003), um die Ergebnisse mit denen internationaler Studien besser vergleichen zu können, die sich ebenfalls an diesem Manual orientiert haben (Fraedrich & Pfaefflin, 2000).

<sup>9</sup>Sofern diese überhaupt zu finden gewesen wären

### 3.5.1 Strukturiertes Klinisches Interview für DSM-IV / Achse-I / Sektion E (SKID-I / E)

Strukturiertes Klinisches Interview für DSM-IV, Achse-I, Sektion E – Störungen im Zusammenhang mit psychotropen Substanzen (SKID-I / E; Wittchen, Wunderlich, Gruschwitz & Zaudig, 1997). Das SKID-I dient der Erfassung und Diagnostik ausgewählter psychischer Syndrome und Störungen, wie sie im DSM-IV auf Achse I definiert werden. Um eine zu große Belastung des Probanden zu vermeiden, wird nur die Sektion E durchgeführt, die Missbrauch und Abhängigkeit von psychotropen Substanzen erfasst. Diese könnten sowohl auf die Lebensspanne als auch auf derzeitiges Vorliegen hin überprüft werden, wobei nur Störungen auf Lebenszeit kodiert werden. Der Fragebogen besteht aus 51 Items, die unnötig werden, wenn drei anfängliche Screeningfragen verneint werden. Ansonsten beträgt die Bearbeitungszeit ca. 45 Minuten.

*Normwerte, Cutoff & deutsche Adaption:* Da es sich um einen Fragebogen zur Erfassung psychischer Störungen handelt, liegen keine Angaben zu Normwerten vor. Angaben zur Prävalenz psychischer Störungen und damit auch zum Substanzgebrauch sind den Tabellen 2.5 und 2.6 zu entnehmen. Cutoffs werden durch das DSM-IV vorgegeben<sup>10</sup> und sind im SKID-I übernommen. Das SKID-I von Wittchen et al. (1997) ist eine deutsche Übersetzung des Originals aus den USA von First, Spitzer, Gibbon und Williams (1997 / SCID-I).

*Psychometrische Güte:* Mit Blick auf objektive Testdurchführung ist von Wichtigkeit, dass die Materialien in gedruckter Form vorliegen und die Fragen teilweise wörtlich vorgegeben sind. Beide Interviews (SKID-I und II) zeichnen sich jedoch durch die Flexibilität in der Durchführung aus. Der Interviewer sollte die Fragen an den Sprachgebrauch des Patienten anpassen und Zusatzfragen, die nicht im Manual enthalten sind, zum Verständnis einsetzen. Eine Phase der Einarbeitung und intensives Training mit dem Verfahren sind zur Gewährleistung von Objektivität und Reliabilität dringend angera-

---

<sup>10</sup>Durch die Anzahl an Symptomen, die für die Vergabe einer Diagnose vorhanden sein müssen



ten. Zur Beurteilerübereinstimmung der aktuellen SKID-Version liegen keine Ergebnisse vor. Hinsichtlich der Reliabilität weisen Wittchen et al. (1997, S. 16) darauf hin, dass das SKID-I für DSM-IV und seine Vorläuferversionen sowohl in amerikanischen als auch in deutschen Studien befriedigende psychometrische Ergebnisse aufwiesen. Im Manual wird lediglich auf die einschlägigen Studien verwiesen. Im Hinblick auf die Validität kann beim Verfahren von inhaltlich-logischer Gültigkeit ausgegangen werden.

*Besonderheiten der Anwendung im Justizvollzug:* Der Fragebogen wurde bereits häufig bei Studien unter Gefangenen angewendet (siehe Brink, 2005) und unter anderem von Peters et al. (2000) auch als Kriteriumsmaß zur Bestimmung der Effektivität der Diagnostik von Substanzgebrauch anderer Screeninginstrumente im Strafvollzug eingesetzt.

*Besonderheiten der Anwendung bei Jugendlichen und Heranwachsenden:* keine

### **3.5.2 Strukturiertes Klinisches Interview für DSM-IV / Achse-II (SKID-II)**

Strukturiertes Klinisches Interview für DSM-IV, Achse II – Persönlichkeitsstörungen (SKID-II; Fydrich, Renneberg, Schmitz & Wittchen, 1997). Das SKID-II ist ein Fragebogen zur Diagnostik der zehn Persönlichkeitsstörungen auf Achse II, sowie der zwei im Anhang des DSM-IV aufgeführten Persönlichkeitsstörungen (Negativistische und Depressive PS) und der Störung des Sozialverhaltens. Das SKID-II ist ein zweistufiges Verfahren, bestehend aus einem Fragebogen, dessen 131 Items die Kriterien des DSM-IV repräsentieren, und der als Screening für die Merkmale der zwölf erfassten PS dient. Im nachfolgenden strukturierten Interview brauchen nur diejenigen Fragen gestellt zu werden, für die im Fragebogen eine „ja“-Antwort angekreuzt wurde. Der Interviewer bewertet, ob einzelne Kriterien qualitativ und quantitativ der im DSM-IV vorgegebenen Form entsprechen. Durch dieses zweistufige Verfahren wird eine vergleichsweise geringe Durchführungszeit für das Interview von etwa 30 Minuten gewährleistet. Die Auswer-

tung kann hinsichtlich des Vorliegens einer PS nach DSM-IV (kategoriale Auswertung: Cutoffwert überschritten ja / nein), aber auch hinsichtlich eines dimensional Scores im Sinne entsprechender Persönlichkeitsmerkmale durchgeführt werden (siehe auch Blackburn & Coid, 1998). Letzteres ist auch in Anbetracht der Kritik an teilweise willkürlichen Cutoffs beispielsweise des DSM-IV (Ulrich & Marneros, 2004) von besonderem Interesse.

*Normwerte, Cutoff & deutsche Adaption:* Wie beim SKID-I werden die Cutoffs durch das DSM-IV vorgegeben und es gibt keine Normwerte. Prävalenzraten der Persönlichkeitsstörungen unter Gefangenen aus anderen Studien sind in den Tabellen 2.5 und 2.6 zu finden. Das SKID-II von Fydrich et al. (1997) ist die deutsche Bearbeitung des SCID-II von First, Gibbon, Spitzer, Williams und Benjamin (1997).

*Psychometrische Güte:* Die Analysen zur Retest- und Interrater-Reliabilität der amerikanischen Originalversion SCID-II weisen zusammenfassend akzeptable bis gute Reliabilitätskoeffizienten nach. In einer ersten Überprüfung der deutschsprachigen Version wurden von 435 durchgeführten Interviews 91 von einem unabhängigen zweiten Beurteiler nochmals nach den Kriterien des SKID-II eingeschätzt. Die prozentuale Übereinstimmung für die einzelnen PS liegen zwischen 89% und 100%, die Übereinstimmung bei der Diagnose mindestens einer PS beträgt 87%. Die Kappa-Koeffizienten liegen zwischen .55 (Histrionische PS) und .82 (Zwanghafte PS), konnten jedoch aufgrund zu niedriger Grundraten häufig nicht berechnet werden. Die grundratenunabhängigen Yule-Koeffizienten betragen zwischen  $Y = .72$  und  $Y = 1.00$ . Die Korrelationen der dimensional Scores einzelner PS und der Cluster weisen deutlich höhere Reliabilitätswerte ( $r = .69$  bis  $r = .95$ ) auf. Fraedrich und Pfaefflin (2000) berichten zusammenfassend ebenfalls von hoher Interrater-Reliabilität.

Hinsichtlich der Validität wurden Korrelationen zum „Fragebogen zu kognitiven Schemata“ (FKS) sowie Zusammenhänge und Übereinstimmungen zu den Diagnosen der behandelnden Therapeuten einer psychosomatischen Fachklinik berechnet. Die Zusammenhänge zwischen SKID-II und FKS erreichten nur für die Dependente und die Selbstunsichere PS höhere Werte. Auch bei den Übereinstimmungsmassen zwischen SKID-II

und Klinikerurteilen weichen die gestellten Diagnosen erheblich voneinander ab. Übereinstimmend erhielten nur 13 von 117 Patienten die Diagnose mindestens einer PS.

*Besonderheiten der Anwendung im Justizvollzug:* Der Fragebogen wurde bereits häufig bei Studien unter Gefangenen angewendet (siehe beispielsweise Coid, 2003; Gosden et al., 2003; Jenkins et al., 2005; Raymond, Coleman, Ohlerking, Christenson & Miner, 1999; Weizmann-Henelius, Sailas, Viemerö & Eronen, 2002). Das SKID-II eignet sich laut Fraedrich und Pfaefflin (2000) besonders für den Einsatz im Justizvollzug, da der Proband konkret gefragte Verhaltensweisen erläutern kann, ohne dabei zu viel von sich preiszugeben. Dies erscheint in einer Umgebung, die von gegenseitigem Misstrauen geprägt ist (wie dem Strafvollzug), besonders sinnvoll. Nach eigener Erfahrung ist bei der Durchführung besonders darauf zu achten, dass der Proband bei der Beantwortung der Fragen auch die Zeit vor seiner Haft mit in die Beurteilung einbezieht, wenn er beurteilt, ob er das jeweilige Kriterium erfüllt. Dies bezieht sich beispielsweise auf die Items „Flirtverhalten“ oder „Kleiderauswahl“ (Kriterium 2) der Histrionischen PS, Verhaltensweisen also, bei denen sich das Verhalten in Haft stark von dem außerhalb unterscheiden dürfte. Da das DSM-IV eine PS als ein andauerndes Muster von innerer Erfahrung und Verhalten definiert, das stark von den Erwartungen der Kultur der Person abweicht, stellen sich laut Rotter et al. (2002) besondere Probleme, wenn PS in Haft diagnostiziert werden sollen, da Gefängnisse ihre eigene Kultur und Normen der Anpassung hätten. Charakteristika wie Misstrauen, Feindseligkeit, sozialer Rückzug und Ich-Bezogenheit könnten adaptive und erwartete Muster von Verhalten in einer Kultur sein, in der auf sich selbst zu achten zum Überleben ebenso notwendig ist, wie anderen zu misstrauen. Die Anmerkungen von Rotter et al. (2002) diesbezüglich wurden bei der Diagnosestellung mit in Betracht gezogen.

*Besonderheiten der Anwendung bei Jugendlichen und Heranwachsenden:* Das DSM-IV fordert für die Vergabe einer PS bei Jugendlichen, dass die Persönlichkeitszüge mindestens ein Jahr bestehen müssen, bevor bei einer Person unter 18 Jahren die Diagnose einer Persönlichkeitsstörung gestellt werden darf (Saß et al., 2003). Da ein solcher Be-

obachtungszeitraum in der vorliegenden Studie nicht realisiert werden kann, sind die (kategorialen) Diagnosen als Verdachtsdiagnosen<sup>11</sup> zu verstehen. Darüber hinaus ist die Vergabe der Antisozialen PS an unter 18-Jährige nicht vorgesehen (siehe Dimond & Misch, 2002), wobei Eppright, Kashani, Robison und Reid (1993) jedoch Zweifel daran haben, ob diese Einschränkung angebracht ist. Da im vorliegenden Fall keine für den Probanden wichtigen Entscheidungen<sup>12</sup> getroffen werden, und das SKID-II auch im Sinne eines (dimensionalen) Persönlichkeitsfragebogens verwendet werden kann, stellt sich die Frage nach dem Alter der Probanden nicht. Im Übrigen gehen McManus, Brickman, Alessi und Grapentine (1984b) davon aus, dass PS von Jugendlichen über strukturierte Interviews für Erwachsene erfasst werden können.

### 3.5.3 Psychopathy Checklist-Revised (PCL-R)

Psychopathy CheckList – Revised (PCL-R; Hare, 2003). Die PCL-R misst den gemeinsamen Kern einer Persönlichkeitseigenschaft, die als Psychopathy definiert werden kann. Sie orientiert sich dabei am Konstrukt der Psychopathy von Cleckley (1976) und besteht aus einer 20 Items umfassenden Merkmalsliste, die sowohl mittels Interview (69 Fragekomplexe) wie auch durch Aktenstudium (Gefangenenpersonalakte und Gesundheitsakte; 34 Fragekomplexe) erhoben wird. Jedes der Items wird auf einer 3-Punkte Skala gewertet (0 = nicht vorhanden; 1 = dürfte vorhanden sein, jedoch reichen die Anhaltspunkte nicht für eine Wertung mit 2 aus; 2 = definitiv vorhanden). Die Gesamtergebnisse können somit von 0 bis 40 Punkte betragen. Je höher der Gesamtwert, desto mehr entspricht die Persönlichkeit des Probanden dem prototypischen Psychopath. Unter Punkt C.1 im Anhang ist der Erfassungsbogen der PCL-R beigelegt, damit der Leser, der nicht mit der PCL-R vertraut ist, sich einen Eindruck von den Kriterien und ihrer Zuordnung zu den beiden Faktoren machen kann. Hinsichtlich der zugrunde liegenden Struktur dieses Konstrukts wird in der Regel ein Zwei-Faktoren-Modell postuliert, das

---

<sup>11</sup>„Probable personality disorder“, siehe beispielsweise Gosden et al., 2003

<sup>12</sup>Wie beispielsweise bei einer Legalprognose

die so genannten „psychopathic traits“ (zwischenmenschliche und affektive Symptomatik) im Sinne traditioneller Persönlichkeitsmerkmale, sowie die „social deviance“ (sozial deviantes Verhalten und einen chronisch instabilen und antisozialen Lebensstil), in Form von antisozialem Verhalten abbildet. Die DSM-Diagnose einer Antisozialen PS steht dabei in Zusammenhang mit Faktor 2 der PCL-R, nicht aber mit Faktor 1 (Hare, 1991). Bezüglich der zugrunde liegenden Struktur der PCL-R zeigt sich jedoch, dass dieses – insbesondere im englischen Sprachraum häufig genannte (siehe beispielsweise Blackburn & Coid, 1998) – Zwei-Faktoren-Modell empirisch auch Widerspruch erfährt. Neuere Arbeiten, die mittels konfirmatorischer Faktorenanalysen strukturprüfend vorgehen, deuten hinsichtlich der Psychopathy auf eine Drei-Faktoren-Struktur hin (siehe Ullrich, Paelecke, Kahle & Marneros, 2003). Da jedoch die ursprüngliche Faktorenstruktur viel Bestätigung erfuhr (siehe Hare, 1991) und außerdem die Ergebnisse der vorliegenden Studie mit denen anderer vergleichbar sein sollen, folgt der Autor der Vorgehensweise von Verona, Hicks und Patrick (2005), die in ihren Studien zum gleichen Thema ebenfalls die Zwei-Faktoren-Lösung angewendet haben. Ein Großteil der Überlegungen zu dieser Arbeit, insbesondere zur Frage hinsichtlich Unterschieden zwischen Gefangenen mit SSV und SV, fußt auf den Arbeiten von Verona und Kollegen (Verona et al., 2001; Verona et al., 2005). Der Empfehlung von Fraedrich und Pfaefflin (2000) und Freese (1998), dass bei Verwendung der PCL-R nicht der deutsche Begriff der Psychopathie verwendet werden solle, sondern das englische Wort „psychopathy“, wird hier gefolgt (aus stilistischen Gründen wird dieses Substantiv „eingedeutscht“ und deshalb groß geschrieben). Dies geschieht, um deutlich zu machen, dass damit eine bestimmte amerikanische Tradition des Psychopathy-Konstruktes gemeint ist und nicht eine ältere, im deutschen Sprachraum übliche.

*Normwerte, Cutoff & deutsche Adaption:* Hinsichtlich der Prävalenz der Psychopathy zeichnen sich interkulturell deutliche Unterschiede ab (Cooke & Michie, 1999; Cooke, Michie, Hart & Clark, 2005a). Während Andersen, Sestoft, Lillebæk, Mortensen und Kramp (1999) Raten von 6% in Dänemark fanden, konnte Koehler (2004) 20.8% in einer deutschen Jugendstrafanstalt und Coid (2002) 73% in einer britischen Gefange-

nenpopulation feststellen. In europäischen Studien findet sich die Tendenz, den Cutoff der PCL-R (ursprünglich  $\geq 30$  Punkte, Hare, 1991) nach unten auf 25 (Moeller & Hell, 2003) bzw. 28 zu korrigieren (Cooke, Michie, Hart & Clark, 2005b), um die üblicherweise mit etwa 30% angegebenen Prävalenzraten zu erreichen. Aufgrund von Betrachtungen der Rohwertverteilungen der PCL-R finden sich inzwischen jedoch Überlegungen, Psychopathy als dimensionales Konstrukt zu erfassen, da bislang nicht eindeutig belegt ist, dass es sich um eine nosologische Entität handelt (Blackburn & Coid, 1998). Zudem erfolgen die Festlegungen der Cutoffs mit einer gewissen Willkür (siehe auch Ullrich & Marneros, 2004), selbst wenn diese Entscheidung auf Grundlage einer Prüfung von Sensitivität und Spezifität getroffen wird. Zum anderen ist anzunehmen, dass solchen Werten immer eine gewisse Stichprobenspezifität anhaftet (siehe Ullrich et al., 2003). Darüber hinaus erlaubt ein dimensionaler Ansatz der Erhebung von PS eine deutlich präzisere Vorhersage zukünftigen, beispielsweise kriminellen Verhaltens, als dies mit kategorialen Diagnosen möglich ist. Die vorliegende Arbeit untersucht neben der dimensionalen Auswertung die prädiktive Validität der Skala mit verschiedenen Cutoffs. Sie verwendet eine Übersetzung der PCL-R in Anlehnung an Loeffler und Welther (1999), wobei Freese (1998) die PCL-R im deutschen Sprachraum allgemein für anwendbar hält.

*Psychometrische Güte:* Verona et al. (2001) konnten in ihrer Studie eine hohe Interrater-Reliabilität zwischen  $.90 \leq r_{tt} \leq .95$  für Faktor 1 und 2 sowie den Gesamtscore erzielen. Auch die Werte hinsichtlich internaler Konsistenz liegen mit  $.83 \leq r_{tt} \leq .93$  hoch (Walters, Duncan & Geyer, 2003). Insgesamt werden die guten psychometrischen Eigenschaften der PCL-R hinsichtlich internaler Konsistenz (Alpha-Koeffizient von .87), Interrater-Reliabilität (Spearman-Brown Koeffizient von zwischen .91 und .93) und Test-Retest-Reliabilität ( $r_{tt} = .94$ ) (Hare, 1991) auch im interkulturellen Vergleich immer wieder bestätigt (Cooke & Michie, 1999).

Hinsichtlich kongruenter Validität berichtet Hare (1991) von einer Korrelation zwischen  $.80 \leq r \leq .90$  mit der Einschätzung der Psychopathy nach dem Konstrukt von Cleckley (1976), die von Klinikern vorgenommen wurde. Mit Blick auf die Diagnose einer Antisozialen PS nach DSM-III ergab sich eine Korrelation von  $r = .71$ . Hinsichtlich der

Vorhersagevalidität berichtet Hare (1991) von einer Reihe von Studien, die der PCL-R eine relativ hohe Rate an richtigen Vorhersagen bezüglich Rückfälligkeit und Haftverhalten bescheinigen.

Da nicht die selben Probanden von beiden Interviewern untersucht wurden, kann die Interraterreliabilität nicht über ein herkömmliches Maß berechnet werden (siehe auch Wirtz & Caspar, 2002). Die Übereinstimmung der Rater in ihren Urteilen hinsichtlich der PCL-R, deutet ein Vergleich der Mittelwerte ihrer Unterstichproben an (siehe Tabelle 3.4). Es wird ersichtlich, dass sich die Mittelwerte hinsichtlich Faktor 1, Faktor 2 und dem Gesamtergebnis der PCL-R Ratings nicht signifikant unterscheiden. Da sich auch diesbezüglich relevante Kontrollvariablen nicht in ihrem Mittelwert bzw. den Häufigkeiten zwischen den beiden Ratern unterschieden (Alter, Gewaltdelikt, Haftform), deutet sich an, dass es keine systematischen Verzerrungen bei der Urteilsbildung zwischen den Ratern gibt und somit eine akzeptable Übereinstimmung vorliegen sollte.

**Tabelle 3.4:** Durchschnittliche Ergebnisse der PCL-R ( $M$  mit  $SD$ ) der beiden Rater in ihren Subgruppen (unabhängiger  $t$ -Test; zweiseitige Signifikanzprüfung;  $n_2 = 70$ )

<i>Skala</i>	Lohner (n=43)	Häußermann (n=27)	<i>t</i>	<i>df</i>	<i>p</i>
PCL-R Faktor 1	4.30 (4.03)	4.15 (3.62)	.162	68	.872
PCL-R Faktor 2	9.21 (5.26)	8.26 (5.12)	.743	68	.460
PCL-R total	14.95 (8.48)	13.93 (9.16)	.478	68	.634

Aufgrund des eben erwähnten Problems, dass kein Objekt von beiden Ratern beurteilt wurde, wurde ansonsten von der Berechnung von Reliabilitätsindices oder Äquivalenten abgesehen. Die Berechnung der oben angegebenen Werte schien bei der PCL-R jedoch indiziert, da es sich hier überwiegend um ein „reines“ Ratinginstrument handelt. Bei allen anderen Instrumente beruhen die Ergebnisse größtenteils auf Selbstangaben oder objektiven Daten. Dementsprechend wurden hier die o.g. Berechnungen nicht an-

gestellt, da diese Ergebnisse (im Gegensatz zu denen der PCL-R) vom Rater unabhängig sein sollten.

*Besonderheiten der Anwendung im Justizvollzug:* Die PCL-R wurde für den Einsatz in Gefängnissen konzipiert, weshalb sich (im Gegensatz zu den anderen Tests) ein extramuraler Einsatz schwierig gestalten dürfte, da beispielsweise auf keine Akten zurückgegriffen werden kann. Die PCL-R hat sich als Vorhersageinstrument im Bereich der Legalprognose (siehe Gray et al., 2003; Stålenheim, 2001; Seagrave & Grisso, 2002; Ullrich et al., 2003) und hinsichtlich des Haftverhaltens (Coid, 2002; Coid et al., 2003a; Freese, 1998; Walters et al., 2003) bewährt.

*Besonderheiten der Anwendung bei Jugendlichen und Heranwachsenden:* Im Hinblick auf das Alter der Probanden und ihren PCL-R-Ergebnissen berichten Ullrich et al. (2003) und Freese (1998) zusammenfassend von Studien, die von einer Abnahme der Faktor-2-Ergebnisse mit zunehmendem Alter, jedoch einer relativen Stabilität der Faktor-1-Ergebnisse ausgehen. Seagrave und Grisso (2002) führen dies auch darauf zurück, dass viele Verhaltensweisen, die Bestandteil einer gesunden Entwicklung während der Adoleszenz seien, mit Eigenschaften eines Psychopath verwechselt werden könnten. Die daraus unvermeidlich resultierende Anzahl vieler Falsch-Positiver, lasse Zweifel an der Anwendbarkeit der PCL-R beispielsweise zu prognostischen Zwecken bei Jugendlichen aufkommen (auch Frick, 2002). Diese Zweifel beziehen sich aber vor allem auf die Anwendung im Rahmen klinischer Entscheidungen und nicht in der Forschung (Seagrave & Grisso, 2002). Obwohl auch eigene Instrumente zur Diagnostik unter Jugendlichen existieren (PCL-YV, PSD, CPS), wird auf ihre Anwendung in der vorliegenden Studie verzichtet, um die Vergleichbarkeit der Ergebnisse nicht zu gefährden, zumal Bedenken bestehen, ob diese Instrumente dasselbe wie die PCL-R messen (siehe Wiener, 2002). Außerdem kann die Studie die vielfach notwendigen Ratings von Lehrern und Erziehungsberechtigten nicht erbringen.



### 3.5.4 Beck Depressions Inventar (BDI)

Beck Depressions Inventar II (BDI; Beck, Steer & Brown, 1996) zur Messung der Depressivität. Das BDI ist ein 21 Items umfassendes Messinstrument, das auf der Selbstausskunft des Probanden beruht. Jedes Symptom wird auf einer 4 Punkte umfassenden Skala (0 bis 3 Punkte) eingestuft (siehe Anhang C.2). Mit den damit erfassten Symptomen werden sieben DSM-IV Kriteriensymptome vollständig, eines nur teilweise (Kriterium A8: verminderte Fähigkeit zu Konzentration und Entscheidung) und eines nicht abgedeckt (Kriterium A5: psychomotorische Unruhe). Im Normalfall kann der Fragebogen in 10-15 Minuten beantwortet werden. Mit fortschreitender Depressionstiefe können die Zeit und der Aufwand der Bearbeitung, zum Beispiel um einen Patienten zur Beantwortung zu motivieren, allerdings stark ansteigen. Wegen der dennoch relativ geringen Durchführungszeit, konnte auf die Verwendung des vereinfachten BDI (BDI-V, siehe Schmitt et al., 2003) verzichtet werden, was die Vergleichbarkeit der Ergebnisse der vorliegenden Studie mit denen anderer Arbeiten nicht gefährdet. Das BDI, als dimensionales Maß der Depressivität, wurde im Einklang mit Crichton (1997) einem kategorialen, wie etwa dem SKID-I, vorgezogen. Dadurch können Grade der Depressivität und nicht nur Depression in binärer Trennung (depressiv / nicht-depressiv) erfasst werden, was genauere und letztlich weiterreichende Analysen erlaubt.

*Normwerte, Cutoff & deutsche Adaption:* Normen, die zum Beispiel den signifikanten Zusammenhang von Geschlecht bzw. Alter und Depression berücksichtigen, existieren nicht. Zur besseren Orientierung und Einordnung von Einzelbefunden geben Hautzinger, Bailer, Worall und Keller (1994) Prozentrangwerte für depressive Patienten bei Behandlungsbeginn und Behandlungsende an. Ansonsten liegen Itemmittelwerte und Summenwerte verschiedener Stichproben (beispielsweise Berliner Psychatriepatienten mit  $M = 24.9$ ,  $SD = 8.6$ ) einschließlich gesunder Kontrollpersonen ( $M = 6.45$ ,  $SD = 5.20$ ) vor, die als Hilfsmittel für die diagnostische Urteilsbildung herangezogen werden können. Werte unter elf Punkten gelten nach Hautzinger et al. (1994) als unauffällig, Werte zwischen elf und 17 als mild oder mäßig depressiv, Werte von 18 und darüber

als klinisch relevant.<sup>13</sup> Kendall, Hollon, Beck, Hammen und Ingram (1987) wollen erst ab einem Wert von mehr als 20 Punkten von einer Depression sprechen, sofern dies durch ein strukturiertes klinisches Instrument abgesichert wurde (siehe auch Palmer & Connelly, 2005). Bei Werten darunter bevorzugen die Autoren den Begriff dysphorisch. Boothby und Durham (1999) schlagen, unter Hinweis auf eine Standardabweichung über dem Mittelwert ihrer Studie, als Cutoff bei Screeningmaßnahmen explizit für den Justizvollzug auch einen Wert von 20 oder mehr Punkten vor. Dieser Empfehlung schließt sich die vorliegende Arbeit an. Im Übrigen erwähnen auch sie Zweifel daran, ob das BDI tatsächlich Depression im Sinne einer nosologischen Entität oder eher generellen Distress misst.

Die Werte jugendlicher Gefangener (Sanislow et al., 2003:  $M = 16.4$ ,  $SD = 11.0$ ) sind mit denen von Erwachsenen in Haft vergleichbar (Eidhin et al., 2002:  $M = 16.31$ ,  $SD = 11.44$ ; Holden & Kroner, 2003:  $M = 12.46$ ,  $SD = 8.56$ ; Ivanoff et al., 1992:  $M = 15.21$ ,  $SD = 7.98$ ; Palmer & Connelly, 2005:  $M = 15.13$ ,  $SD = 13.99$ ; Ruchkin et al., 2003:  $M = 15.62$ ,  $SD = 10.21$ ) und dabei durchweg höher als die oben angegeben oder die von Beck, Steer, Beck und Newmann (1993) festgestellten Werte aus der gesunden Allgemeinbevölkerung ( $M = 11.89$ ,  $SD = 7.54$ ). Lediglich Dear et al. (2001b) fanden deutlich niedrigere Werte in ihrer Gefangenenstichprobe ( $M = 8.76$ ,  $SD = 8.33$ ), was unter Umständen auch auf den Zeitpunkt der Messung zurückgeführt werden kann. Smyth et al. (1994) oder auch Zamble und Porporino (1988) fanden eine signifikante Abnahme der BDI-Werte von  $T_1$  (kurz nach Inhaftierung) zu  $T_2$  (nach einem Jahr). Ebenso scheint es einen Zusammenhang zwischen den Haftbedingungen<sup>14</sup> und der Höhe des BDI Ergebnisses zu geben (Boothby & Durham, 1999). Letztgenannte Studie, die systematisch die Ergebnisse des BDI unter verschiedenen Bedingungen und Stichproben verglich, fand auch signifikant höhere Werte bei jugendlichen Gefangenen (Boothby & Durham, 1999). Das BDI liegt in der deutschen Adaption von Hautzinger et al. (1994) vor und wurde so angewendet.

---

<sup>13</sup>Die Werte weichen von den von Beck et al. (1996) genannten Werten ab.

<sup>14</sup>Zum Beispiel verschärfte Haft

*Psychometrische Güte:* Hinsichtlich der Objektivität kann bei schwer depressiven Personen das Problem auftauchen, diese zur Mitarbeit zu motivieren, wobei die Person des Testleiters an Einfluss gewinnt. Die Auswertung kann als objektiv gelten. Retest-Reliabilitäten unterschätzen die Zuverlässigkeit eher aufgrund der sich schnell ändernden Symptomatik bei Depressiven. Hautzinger et al. (1994) berichten dennoch Werte für die Originalfassung von  $r_{tt} = .60$  bis  $r_{tt} = .86$  für den Zeitraum von einer Woche. Die innere Konsistenz (Cronbachs Alpha) für psychiatrische Patienten liegt im Bereich von  $.76 \leq \alpha \leq .95$ , für nicht psychiatrische Stichproben von  $.73 \leq \alpha \leq .92$ . Bezüglich der Split-Half-Reliabilität ergeben sich bei Beck et al. (1996) mit  $r_{tt} = .86$  (korrigiert  $r_{tt} = .93$ ;  $N = 200$ ) zufriedenstellende Werte. Für die untersuchten deutschsprachigen Stichproben fanden Hautzinger et al. (1994) ein Alpha nach Cronbach von .88 über alle Probanden, einen Spearman-Brown-Koeffizienten von .84, einen Guttman-Koeffizienten von .84 und eine Split-Half-Reliabilität von .72. Logische und inhaltliche Validität sind insofern gegeben, als die beschriebenen Symptome der klinischen Praxis entstammen und in der Literatur beschrieben werden. Holden und Kroner (2003) berichten zusammenfassend von einer durchschnittlichen konvergenten Validität von  $r = .60$ , basierend auf der Korrelation zwischen dem BDI und relevanten klinischen Ratings für Stichproben außerhalb der Psychiatrie.

*Besonderheiten der Anwendung im Justizvollzug:* Das BDI wurde vielfach an Gefangenen angewendet (unter anderem Boothby & Durham, 1999; Cooper & Berwick, 2001; Ivanoff et al., 1996; Lovelace & Gannon, 1999; Rohde et al., 1997b). Boothby und Durham (1999) berichten als Ergebnis ihrer Itemanalyse von einem außergewöhnlich hohen Mittelwert von 2.25 ( $SD = 1.04$ ) für das Item F (Strafbedürfnis), wobei nahezu alle Gefangenen das Item positiv beantwortet hätten.<sup>15</sup> Der Itemname für Item F ist missverständlich, da bei diesem Item nicht nach dem *Wunsch*, sondern nach dem *Gefühl*, bestraft zu werden, gefragt wird, was – zumindest bei verurteilten Gefangenen – tatsächlich der Fall ist. Eine Faktorenanalyse erbrachte unter anderem einen Faktor

---

<sup>15</sup>Kein anderes Item erhielt einen Itemmittelwert von über .91.

– genannt Bestrafungsfaktor –, auf dem dieser und das Item A (Dysphorie) lagen. Dies lege nahe, dass die Items A und F etwas anderes als Depression bei Gefangenen messen würden, so Boothby und Durham (1999), und deshalb beim Depressionsscreening unter Gefangenen nicht hilfreich seien. Die vorliegende Arbeit wird die besprochenen Items jedoch weiterhin zum Gesamtergebnis hinzurechnen, um die Vergleichbarkeit mit anderen Studien zu gewährleisten.

*Besonderheiten der Anwendung bei Jugendlichen und Heranwachsenden:* Depression kann auch bei Jugendlichen reliabel und valide mittels standardisierter Verfahren erfasst werden (McManus et al., 1984a), wobei Steer und Beck (1988) das BDI auch bei Jugendlichen für geeignet erachten. Allerdings werden bei jugendlichen und heranwachsenden Probanden vielfach höhere Werte gefunden, was teilweise mit generellen Anpassungsschwierigkeiten erklärt werden kann (siehe Boothby & Durham, 1999; Steer & Beck, 1988), die regelmäßig mit dieser Entwicklungsphase einhergehen und nicht zwingend auf psychopathologische Auffälligkeiten schließen lassen. Aus diesem Grund schlagen Steer und Beck (1988) auch einen höheren Cutoff von 18<sup>16</sup> bei Heranwachsenden vor.

### 3.5.5 Beck Hoffnungslosigkeits Skala (BHS)

Beck Hoffnungslosigkeits Skala (BHS; Beck & Steer, 1993a) zur Messung der Hoffnungslosigkeit. Die BHS besteht aus 20 Items, die mit wahr oder falsch zu beantworten sind, und erhebt Pessimismus und negative Zukunftserwartungen. Das Fragebogenformular ist im Anhang unter Punkt C.3 zu finden. Die Bearbeitungszeit liegt bei ca. zehn Minuten.

*Normwerte, Cutoff & deutsche Adaption:* In einer Normierungsstudie der Allgemeinbevölkerung berichtet Greene (1981) von ungefähr einer Standardabweichung niedrigeren Werten ( $M = 4.45$ ,  $SD = 3.09$ ) als unter Patienten mit einem Suizidversuch ( $M$

---

<sup>16</sup>Im Gegensatz zur anders lautenden Empfehlung anderer Autoren von 16 Punkten

= 9.0,  $SD = 6.1$ ). Ähnlich fallen die Werte von Dear et al. (2001b) ( $M = 4.15$ ,  $SD = 3.34$ ), Eidhin et al. (2002) ( $M = 3.68$ ,  $SD = 2.69$ ), Holden und Kroner (2003) ( $M = 3.88$ ,  $SD = 4.24$ ) und Ivanoff et al. (1992) ( $M = 3.21$ ,  $SD = 3.00$ ) bei gefangenen Erwachsenen bzw. Jugendlichen aus (Biggam & Power, 1999a:  $M = 4.36$ ,  $SD = 2.73$ ). Ruchkin et al. (2003) dagegen fanden unter ihrer Stichprobe inhaftierter Jugendlicher leicht höhere Werte ( $M = 6.01$ ,  $SD = 3.69$ ), ebenso wie Palmer und Connelly (2005) bei erwachsenen Gefangenen ( $M = 6.29$ ,  $SD = 4.49$ ). Ein Grund für unterschiedliche Werte könnte auch dem Messzeitpunkt geschuldet sein, da Smyth et al. (1994) kurz nach Inhaftierung niedrigere Durchschnittswerte fanden ( $M = 5.4$ ,  $SD = 4.7$ ) als ein Jahr später ( $M = 4.4$ ,  $SD = 5.1$ ). Eidhin et al. (2002) und Gray et al. (2003) berufen sich auf die Arbeit von Beck et al. (1975b) und verwenden einen Cutoff von 9 oder mehr Punkten, weil diese Werte gute Prädiktoren für einen späteren Suizidversuch seien. Palmer und Connelly (2005) teilen die Ergebnisse (ebenfalls unter Berufung auf Beck & Steer, 1993a) in vier Bereiche ein, wobei 0-3 Punkte keine, 4-8 Punkte geringe, 9-14 Punkte moderate und Werte von 15 Punkten oder mehr, starke Hoffnungslosigkeit bedeuten. Diesen Cutoffwert von 15 Punkten verwendet die vorliegende Studie ebenfalls, was auch von den Ergebnissen von Biggam und Power (1999a) unterstützt wird. Die BHS liegt in der deutschen Adaption von Krampen (1994) vor und wird so verwendet.

*Psychometrische Güte:* Die Objektivität in Bezug auf die Durchführung und Auswertung des Verfahrens kann als weitgehend gesichert gelten. In den Untersuchungen von Krampen (1994) wurde für die Split-Half-Reliabilität (Spearman-Brown) ein Wert von  $r_{tt} = .87$  nachgewiesen ( $N = 212$ ). Die interne Konsistenz (Kuder-Richardson) wurde mit  $r = .86$  berechnet. Die in der Handanweisung angegebenen Koeffizienten für die Normierungsstichprobe und andere Stichproben liegen in einer ähnlichen Größenordnung. Beck, Weissman, Lester und Trexler (1974b) fanden eine hohe Interrater-Reliabilität von .93. Holden und Kroner (2003) berichten zusammenfassend von großer psychometrischer Stärke der Skala. So konnte ein Alpha-Koeffizient von .85 bzw. durch Ruchkin et al. (2003) von .79 festgestellt werden. Mit Blick auf die konvergente Validität wiesen Beck et al. (1974b) eine durchschnittliche Korrelation von  $r = .74$  zwischen der Ska-

la und klinischen Einschätzungen der Hoffnungslosigkeit nach. Insgesamt fanden Beck et al. (1974b) hervorragende Reliabilitäts- und Validitätsindizes für die BHS.

*Besonderheiten der Anwendung im Justizvollzug:* Die BHS wurde vielfach im Justizvollzug (Eidhin et al., 2002; Gray et al., 2003) und im Zusammenhang mit AR von Gefangenen (Biggam & Power, 1999a; Ivanoff et al., 1996) eingesetzt. Sie hat sich als guter Prädiktor für das Haftverhalten (Gray et al., 2003), die Anpassung des Gefangenen an die Haft (Smyth & Ivanoff, 1994) und Suizidalität unter Gefangenen (Bonner & Rich, 1990; Ivanoff & Jang, 1991) bewährt. Ein systematischer Unterschied bei der Beantwortung der Fragen nach der Zukunftsunsicherheit (Items 4, 7 und 13) ist zwischen Untersuchungsgefangenen und Strafgefangenen aufgrund des bevorstehenden Urteils zu erwarten.

*Besonderheiten der Anwendung bei Jugendlichen und Heranwachsenden:* Bei jugendlichen Gefangenen kam die BHS bereits vielfach zum Einsatz (Biggam & Power, 1999a; McGarvey, Kryzhanovskaya, Koopman, Waite & Canterbury, 1999; Ruchkin et al., 2003). Steer und Beck (1988) berichten zusammenfassend von guter prädiktiver Validität der Skala hinsichtlich Suizidalität bei Jugendlichen und ansonsten keinen Besonderheiten bei der Anwendung.

### 3.5.6 Beck Scale for Suicide Ideation (BSS)

Beck Scale for Suicide Ideation (BSS; Beck & Steer, 1993b), um die Stärke eines Suizidwunsches und den Grad suizidaler Gedanken zu messen. Die BSS ist ein Maß zur Erhebung der gegenwärtigen suizidalen Gedankenbildung bei Probanden mit und ohne AR in der Vergangenheit. Das Verfahren besteht aus 19 Items,<sup>17</sup> denen jeweils drei abgestufte Antwortmöglichkeiten zugeordnet sind. Während der Suizidwunsch (Items 1-8 und 11) den gegenwärtigen Wunsch zu sterben fokussiert, erfasst die Suizidvorbereitung

---

<sup>17</sup>Zuzüglich zweier Items, die Hintergrundinformationen erfassen

(Items 5, 10, 12-18) die aktiven Suizidpläne des Probanden (siehe Anhang C.4). Die Beantwortung der Items nimmt nicht mehr als fünf Minuten in Anspruch.

*Normwerte, Cutoff & deutsche Adaption:* Die BSS wurde als klinisches Forschungsinstrument konzipiert und hat diesen Status immer noch (Bonner & Rich, 1990). Dieser Status zeigt sich außerdem darin, dass große Uneinigkeit und Unsicherheit hinsichtlich eines möglichen Cutoffwertes, wie auch hinsichtlich des Namens der Skala besteht.<sup>18</sup> Durchschnittswerte werden aus der Allgemeinbevölkerung für Probanden ohne affektive Störung ( $N = 488$ ) mit  $M = 0.61$  ( $SD = 2.11$ ) und für Patienten mit affektiver Störung ( $N = 1360$ ) mit  $M = 2.70$  ( $SD = 4.79$ ) angegeben (Beck et al., 1993). Aus Haft berichten Palmer und Connelly (2005) Werte ( $M = 1.17$ ,  $SD = 4.46$ ), die mit denen aus der Allgemeinbevölkerung vergleichbar sind, wobei sie vom Haftzeitpunkt abhängig sind und mit zunehmender Dauer der Haft abnehmen (Smyth et al., 1994). Während Blaauw et al. im Jahr 2001 noch einen Cutoff von 4 oder mehr Punkten in der Summe bei den Fragen 1,2 und 4 vorschlagen, spricht eine andere Autorengruppe um Blaauw ein Jahr später von einem Cutoff von 6 aufsummierten Punkten bei denselben Fragen (Blaauw et al., 2002). Die Begründung bleibt in beiden Artikeln gleich und ist auf eine negativere Einstellung zum Leben und eine positivere zum Tode und suizidalem Verhalten in diesem Fall zurückzuführen. Esposito und Clum (2002) schlagen unter Berufung auf die Originalarbeit (Beck et al., 1979) einen Wert von 8 Punkten für das Gesamtergebnis vor. Dabei scheint es sich jedoch um einen Fehler von Esposito und Clum (2002) zu handeln, da in der Originalarbeit kein Cutoff für die BSS vorgeschlagen wird<sup>19</sup>. Palmer und Connelly (2005) verwenden keinen Cutoff und nehmen dabei Bezug auf das Manual der Skala von Beck und Steer (1993b). Dieser Maßgabe schließt sich die vorliegende Arbeit an. Die Skala wurde 1981 in deutscher Sprache von Beck, Rush, Shaw und Emery publiziert und wird so verwendet.

*Psychometrische Güte:* Aufgrund der Fragebogenform mit gebundener Beantwor-

---

<sup>18</sup>PSYNDEXplus zeigt vier „synonyme“ Namen an, darunter fälschlicher Weise den der SIS.

<sup>19</sup>Dagegen einer für die BHS von 8 Punkten

tung können Durchführung und Auswertung als objektiv gelten. Die interne Konsistenz wird mit einem Alpha-Koeffizienten von .89, die Interrater-Reliabilität mit  $r_{tt} = .83$  (Beck et al., 1979) bzw. mit  $r_{tt} = .91$  angegeben (Ivanoff & Jang, 1991). Eine ausführliche Diskussion zur psychometrischen Güte der Skala findet sich bei Beck et al. (1979) und Beck, Steer und Ranieri (1988). Blaauw et al. (2002) und Holden und Kroner (2003) berichten von angemessener internaler Konsistenz der Skala (Alpha-Koeffizient von .77 bzw.  $\geq .75$ ). Hinsichtlich der Konstruktvalidität des Verfahrens führen Range und Knott (1997) im Bezug auf konvergente Validität Untersuchungen an, die signifikante Korrelationen zur Selbstauskunft der Bildung suizidaler Gedanken ergaben. Wie bei den meisten Skalen, die Suizidalität vorhersagen wollen, fällt die prädiktive Validität im Vergleich zu den anderen Aspekten der Validität auch hier wegen der geringen Basisrate kleiner aus (Range & Knott, 1997).

*Besonderheiten der Anwendung im Justizvollzug:* Die Skala wurde bereits mehrfach zur Quantifizierung suizidaler Gedanken und Wünsche von psychiatrischen Patienten (beispielsweise Hawton, Houston, Haw, Townsend & Harriss, 2003) und im Strafvollzug eingesetzt (Eidhin et al., 2002; Ivanoff & Jang, 1991; Palmer & Connelly, 2005). Hinsichtlich der Items 12 und 13<sup>20</sup> ist, wie bereits erwähnt, insofern von einer Besonderheit auszugehen, als den Gefangenen nicht die Möglichkeiten zur Selbstbeschädigung zur Verfügung stehen, wie dies in Freiheit der Fall wäre. Dies sollte sich allerdings nicht bedeutsam auf die Anwendbarkeit der Skala oder Vergleichbarkeit der Ergebnisse auswirken.

*Besonderheiten der Anwendung bei Jugendlichen und Heranwachsenden:* In ihrer Arbeit sprechen Steer und Beck (1988) von keinen Besonderheiten bezüglich der Anwendung der BSS bei Jugendlichen. Sie gehen jedoch von geringfügig kleineren Werten bei Jugendlichen im Vergleich zu einer Stichprobe von depressiven Patienten aus.

---

<sup>20</sup>Merkmale eines erwogenen Suizidversuchs



### 3.5.7 Suicidal Intent Scale (SIS)

Suicidal Intent Scale (SIS; Beck, Schuyler & Herman, 1974a) zur Messung der Suizidabsicht. Die SIS ist eine semistrukturierte Ratingskala, die aus insgesamt 15 Items (inhaltlich getrennt in zwei Teile) besteht. Teil 1 (objektive Umstände; Items 1-8) deckt die objektiv erfassbaren Rahmenbedingungen der AR ab (Vorbereitungen, Art der Durchführung, Umstände, Versuche Entdeckung zu vermeiden etc.). Teil 2 (Selbstauskunft; Items 9-15) befasst sich mit introspektiven Daten (Erwartung, dass der Versuch zum Tode hätte führen können, Zweck des Versuches etc.). Alle Items sind mit 0, 1 oder 2 zu bewerten, wobei höhere Werte stärkere Absicht bedeuten. Zusätzlich werden in fünf weiteren Fragen interessante Informationen zur AR erfragt,<sup>21</sup> die aber nicht in die Wertung mit einfließen. Im Anhang ist der Erhebungsbogen bei C.5.1 enthalten. Die Bearbeitungszeit liegt bei ca. zehn Minuten. Während in der Originalversion der Skala (Beck et al., 1974a) oder auch den Anmerkungen von Dear (2003b) das Item 9 (Zweck des Suizidversuchs) zum ersten Teil gerechnet wurde, ist es in anderen Publikationen (Beck, Steer & Trexler, 1989b; Antretter, Dunkel, Seibl & Haring, 2002) Bestandteil des zweiten Teils (Selbstauskunft). Dies erscheint in Anbetracht des Charakters dieses Items, nämlich einer Angabe zum Motiv bzw. Sinn des Tuns, auch sinnvoll und wird so beibehalten (siehe auch Dear, 2003b). Ein Summenscore der Items 6 und 15 (max. 4 Punkte) erlaubt nach Suominen et al. (2000) eine Angabe der Impulsivität eines Suizidversuchs, wobei ein Wert von 0 einem impulsiven, Werte von 1 und 2 einem mittelgradig impulsiven und Werte von 3 und 4 einem nicht impulsiven Suizidversuch entsprechen. Während die BSS (s.o.) den aktuellen suizidalen Wunsch misst,<sup>22</sup> erfasst die SIS die Suizidabsicht zum Zeitpunkt einer AR.

*Normwerte, Cutoff & deutsche Adaption:* Dear (2000) nennen zusammenfassend drei mögliche Kategorisierungen der Werte: keine oder geringe Suizidabsicht (0 bis 3 Punkte), moderate Suizidabsicht (4 bis 10 Punkte) und starke Suizidabsicht (11 oder höher

---

<sup>21</sup>Anzahl früherer Versuche, Vorstellungen über den Tod etc.

<sup>22</sup>Also zum Zeitpunkt des Interviews

Punkte). Ebenso verwenden Blaauw et al. (2002) einen Cutoff von mehr als 10 Punkten. In diesem Fall sei zumindest bei zwei Dritteln der Fragen von moderater (je ein Punkt) bzw. bei einem Drittel der Fragen von starker Suizidabsicht (je 2 Punkte) auszugehen (insgesamt max. 30 Punkte). Diesen Cutoffwert von mehr als 10 Punkten verwendet die vorliegende Studie, um damit niedrige von hoher Suizidabsicht zu unterscheiden. Da keine deutsche Übersetzung vorliegt,<sup>23</sup> wurde der Fragebogen im Vorfeld der Untersuchung übersetzt (siehe C.5.2).

*Psychometrische Güte:* Beck et al. (1974a) berichten von hervorragender Interrater-Reliabilität ( $r_{tt} = .95$ ; für die Unterteile .74 und .90 siehe Range & Knott, 1997) und Goldston (2000) zusammenfassend von Intraklass Korrelationskoeffizienten zwischen  $.83 \leq ICC \leq 1.00$  und akzeptabler internaler Konsistenz (Spearman-Brown-Koeffizient von .82), wobei der Selbstauskunftsteil eine stärkere internale Konsistenz aufwies als der Teil der objektiven Umstände ( $\alpha = .85$  bzw. .60). Hinsichtlich der Reliabilität fanden Blaauw et al. (2002) bei der Anwendung der Skala an einer Gefangenenstichprobe einen Alpha-Koeffizienten von .83. Hinsichtlich der Kriteriumsvalidität, hier der kongruenten Validität, fanden Goldston (2000) deutliche, und Range und Knott (1997) moderate Zusammenhänge mit Werten der Depression, Hoffnungslosigkeit und der Bildung suizidaler Gedanken. Range und Knott (1997) berichten von guten Fähigkeiten hinsichtlich der Unterscheidung zwischen einfachem und wiederholtem SV bzw. vollendetem und versuchtem Suizid. Auch Dear (2003b) bestätigt zusammenfassend die befriedigenden Messeigenschaften der Skala.

*Besonderheiten der Anwendung im Justizvollzug:* Auch die SIS wurde von zahlreichen Studien zu AR im Justizvollzug angewendet (u. a. Blaauw et al., 2001b; Dear et al., 2000; Power et al., 1997), wobei jedoch einige Modifikationen sinnvoll erscheinen. Dear (2003b) schlägt für die Anwendung der SIS im australischen Strafvollzug einige kleinere Veränderungen bzw. Klärungen der Bewertungskriterien vor, die auch für den deutschen

---

<sup>23</sup>Es handelt sich um ein Instrument für die klinische Forschung (Goldston, 2000).

Vollzug Geltung haben. Diese beziehen sich vor allem auf die objektiven Umstände, wobei nur im Fall des Items 1 (Isolation) eine substantielle Neuformulierung der Kriterien für notwendig erachtet wurde. Da die Gefangenen die Isolation nicht wirklich beeinflussen könnten, verzerre die ursprüngliche Wertung die Realität des Justizvollzuges etwas. Eine volle Wertung von 2 Punkten komme nur in Frage, wenn sich der Gefangene absichtlich weiter als ohnehin gegeben isoliert habe. Diese Anmerkung und die anderen (eher geringfügigen) Empfehlungen werden in der vorliegenden Arbeit bei der Bearbeitung der Skala beachtet.

*Besonderheiten der Anwendung bei Jugendlichen und Heranwachsenden:* Es werden von Steer und Beck (1988) bezüglich der Anwendung bei Jugendlichen keine Angaben zur psychometrischen Güte der Skala gemacht. Im Übrigen ist bei der Anwendung der Skala bei Jugendlichen nicht von besonderen Bedingungen auszugehen und Range und Knott (1997) sehen trotz der ursprünglichen Konstruktion der Skala für Erwachsene keine Probleme, wenn diese (wie bereits vielfach geschehen) bei Heranwachsenden eingesetzt wird.

### **3.5.8 Lethality of Suicide Attempt Rating Scale-II (LSARS-II)**

Lethality of Suicide Attempt Rating Scale-II (LSARS-II; Smith, Conroy & Ehler, 1984 bzw. Berman, Sheperd & Silverman, 2003) zur Messung der medizinischen Gefährlichkeit der Handlungen. Da die Effektivität einer AR getrennt von der Suizidabsicht betrachtet werden sollte (Beck et al., 1974a; Leenaars et al., 1997) und folgerichtig auch nicht mittels der SIS erfasst werden kann, wird die Verwendung eines solchen Instrumentes notwendig. Die LSARS-II ist eine 11 Punkte umfassende Skala (0-10) zur Erfassung der Letalität einer AR. Mittels neun Beispielankern können AR wie Schneiden, Erhängen und andere Methoden hinsichtlich der Gefährlichkeit eingeschätzt werden. Dieser Merkmals- und Kriterienkatalog ist im Anhang unter Punkt C.6.1 enthalten. Eine umfangreiche Substanzliste mit Angabe der letalen Dosis abhängig vom Körperge-

wicht ist enthalten, um die Gefährlichkeit von Vergiftungen einschätzen zu können (siehe C.6.3; in der Übersetzung<sup>24</sup> siehe C.6.4). Dabei finden auch intervenierende Rahmenbedingungen Berücksichtigung, die das Ergebnis beeinflussen können, wie der Kontext und die Rahmenbedingungen der AR (siehe dazu Berman et al., 2003; Leenaars et al., 1997). Diese Aspekte betreffen dann das Konzept der Rettungsmöglichkeit und der Wahrscheinlichkeit mit der ein Akt (von außen) unterbrochen werden kann. Die LSARS-II kann von Nicht-Medizinern bearbeitet werden (Range & Knott, 1997; Smith et al., 1984). Die Skala stellt eine der wenigen verfügbaren Möglichkeiten dar, die Letalität der Verletzung von AR standardisiert zu erheben und dadurch einem Mangel bisheriger Studien zum Thema zu entgehen, die ihre Kriterien zu dieser Einschätzung nicht objektiviert haben (O'Carroll et al., 1996). Der Vorteil dieser Skala gegenüber Klinikereinschätzungen zur Gefährlichkeit liegt darin, dass die Urteile objektiv nachvollziehbar sind und ohne unquantifizierbare Urteile wie „niedrig“, „moderat“ und „schwer“ auskommen. Smith et al. (1984, S. 215) vergleichen andere Ansätze mit einem Physiker, der die Lichtstärke nicht in Lux angebe, sondern Wörter wie „dim“, „bright“, and „blinding“ verwendet. Die LSARS wird auch deshalb explizit vom National Institute of Mental Health (NIMH) zur Erhebung der Letalität eines SV empfohlen, sowie im Übrigen die SIS zur Erhebung der Suizidabsicht. Diese Empfehlung macht sich die Expertenkommission der IASR zu Eigen (Leenaars et al., 1997). Es ist von einer Bearbeitungszeit von unter fünf Minuten auszugehen.

*Normwerte, Cutoff & deutsche Adaption:* Zur LSARS-II sind keine Normwerte bekannt. Goldston (2000) sieht ihren Einsatzbereich in der klinischen Forschung. Beim Einsatz der Skala in der Forschung wird ein Cutoff für ernsthafte Handlungen von den Autoren bei 8 Punkten mit der folgender Begründung veranschlagt: „Because of the decreased level of ambivalence seen in the scale points of 8 and above, we will probably continue to use 8 as our research benchmark of 'seriousness'“ (Smith et al., 1984; S. 222). Dieser Cutoff erscheint bei einem maximalen Höchstwert von 10 Punkten sehr hoch

---

<sup>24</sup>Die Kennzeichnung / Gefahrstoffsymbole beziehen sich auf die EU-Grundrichtlinie zu Gefahrstoffen (RL 67 / 548 / EWG).

und seine theoretische Begründung ist nicht überzeugend. Smith et al. (1984) ignorieren damit die Ergebnisse selbst veranlasster Ratings von Klinikern (Psychiater, Psychologen, Sozialarbeiter, Pflegepersonal), welche den Cutoff bei  $> 3.5$  verortet sehen wollten. Wie auch von anderen Autoren, welche die LSARS angewendet haben (Daigle, Alarie & Lefebvre, 1999; Rohde et al., 1997a; Séguin, Lynch, Labelle & Gagnon, 2004), wird ein Wert von 5 oder mehr Punkten herangezogen, der in Anbetracht der Tatsache, dass hier zum ersten Mal der Tod wahrscheinlich ist,<sup>25</sup> sinnvoller erscheint. Berman et al. (2003), welche die Skala überarbeiteten, berichten von keinem festgelegten Cutoffwert für die Skala und unterschiedlichen Bestimmungsmethoden (inhaltlich vs. Mediansplitting) des Cutoff. Ein Cutoffwert von  $\geq 3.5$ , wie er beispielsweise von Handwerk, Larzelere, Friman und Mitchell (1998) angewendet wurde, ist weder inhaltlich noch durch eine statistische Berechnung gerechtfertigt. Eine Bestimmung des Cutoff im Nachhinein, um damit eine gewisse Gruppeneinteilung zu erreichen bzw. eine bestimmte Gruppengröße nicht zu unterschreiten, erscheint aus wissenschaftlicher Sicht nicht statthaft. Von der Festlegung des Cutoff durch ein Mediansplitting wird abgesehen, da durch die „künstliche Dichotomisierung (50% SSV und 50% SV)“ die tatsächliche Verteilung von SSV und SV verzerrt wird. Die vorliegende Arbeit wird deshalb einen in der Suizidforschung des Justizvollzugs international akzeptierten (siehe Range & Knott, 1997) und vorher festgelegten Wert von  $\geq 5$  heranziehen. Eine deutsche Adaption der englischsprachigen LSARS-II (C.6.1) lag noch nicht vor, sodass die Skala vom Autor der Arbeit zu übersetzen war (siehe C.6.2 und C.6.4).

*Psychometrische Güte:* Aufgrund der objektiv zu erhebenden Umstände und der festen Zuordnung der Verletzungen zu einem bestimmten Wert ist von guter Objektivität der LSARS-II auszugehen. Zusammenfassend berichten Berman et al. (2003) von guten Werten hinsichtlich Interrater-Reliabilität, wobei Smith et al. (1984) für letztere Werte zwischen  $.85 \leq r_{tt} \leq .99$  und Nasser und Overholser (1999)  $r_{tt} = .90$  ermitteln konnten. Handwerk et al. (1998) bestimmten die Übereinstimmung der Ratings verschiedener Kli-

---

<sup>25</sup> „Death is a fifty-fifty probability“ (Smith et al., 1984, S. 238)

niker und konnten in 93% der Fälle ein übereinstimmendes Rating ( $\pm 1$  Abweichung von der anderen Einschätzung) feststellen. Die Intraklass Korrelation ist ebenfalls sehr gut (ICC = .85; Nasser & Overholser, 1999). Bei einem Intervall von sechs Monaten lag die Test-Retest-Reliabilität bei  $r_{tt} = .72$  für den kontinuierlichen und bei  $\kappa = .76$  für den kategorialen Gebrauch der LSARS (Nasser & Overholser, 1999). Berman et al. (2003) und Goldston (2000) ziehen verschiedene andere Studien heran, um die kongruente Validität der Skala zu beschreiben und konnten hinsichtlich der Außenkriterien Depression, Hoffnungslosigkeit, Bildung suizidaler Gedanken und (dabei mit widersprüchlichen Ergebnissen) Suizidabsicht signifikante Korrelationen nachweisen.

*Besonderheiten der Anwendung im Justizvollzug:* Das LSARS kam bereits mehrfach in Studien über AR in Gefangenschaft zum Einsatz (beispielsweise Daigle et al., 1999; Farmer et al., 1996; Rohde et al., 1997a). Wie bereits beschrieben, ist die Zahl der Möglichkeiten zur Selbstschädigung im Strafvollzug eingeschränkt, sodass gewisse Methoden, wie sie in der Skala aufgeführt werden<sup>26</sup>, nicht oder weniger häufig vorkommen dürften. Da sich die Ratings der Skala jedoch nicht aufeinander beziehen, ist entgegen der Meinung von Daigle et al. (1999) nicht davon auszugehen, dass es Einschränkungen oder Verzerrungen des Grundgedankens dieses Ansatzes bei der Anwendung der Skala im Justizvollzug gibt. Die Formulierungen der Merkmalsliste der LSARS lassen Rückschlüsse auf ihren ursprünglichen Einsatzort und die Klientel zu (die SV von Patienten eines Krankenhauses), wobei die Formulierungen ohne Weiteres auf andere Situationen übertragen werden können, zumal es sich ebenfalls um eine „geschlossene Institution“ handelt und keinen privaten Rahmen.

*Besonderheiten der Anwendung bei Jugendlichen und Heranwachsenden:* keine

---

<sup>26</sup>Beispielsweise Sturz aus großer Höhe, Sturz vor ein sich bewegendes Objekt, Schusswaffengebrauch

### 3.5.9 Eigener Erhebungsbogen

Eigener Erhebungsbogen (Interview und nach Aktenlage) zur Erhebung von demografischen, forensischen und psychologisch-psychiatrischen Variablen. Diese unabhängige Variablen wurden mit folgenden Operationalisierungen erhoben:

- Haftform: Untersuchungshaft vs. Strafhaft (nach Aktenlage)?
- Geburtsland:<sup>27</sup> Deutschland vs. Ausland (nach Aktenlage)?
- Sprachkenntnisse: Spricht der Gefangene ausreichend deutsch, um sich über grundsätzliche Sachverhalte verständigen zu können (nach Einschätzung des Untersuchers)?
- Probleme in Beziehungen: Gab es in den letzten Wochen Probleme in zwischenmenschlichen Beziehungen, die sich nach Ansicht des Gefangenen auf seine aktuelle Befindlichkeit auswirken?
- Verlust: Hat der Gefangene in jüngster Zeit eine Bezugsperson durch Trennung, Tod etc. verloren (laut Selbstauskunft)?
- Obdachlosigkeit: Lebte der Gefangene vor der Inhaftierung auf der Straße (laut Selbstauskunft)? Allein der Aktenvermerk „ohne festen Wohnsitz“ bzw. eine dauerhafte Unterkunft bei Bekannten oder Verwandten oder beispielsweise in einer öffentlichen Einrichtung wurden nicht als Obdachlosigkeit gewertet.
- Bildung: Höchster Schulabschluss des Gefangenen (laut Selbstauskunft)?
- Unterbringung in einem besonders gesicherten Haftraum: Wie oft wurde der Gefangene während der aktuellen Inhaftierung in einem bgH, einem Kriseninterventionsraum oder einer Monitorzelle untergebracht (nach Aktenlage)?
- Vorfälle zur Meldung: Wie oft wurden Meldungen über „besondere Vorkommnisse“ über den Gefangenen geschrieben (nach Aktenlage)?

---

<sup>27</sup>Das Geburtsland wurde erhoben und nicht die tatsächliche Staatsangehörigkeit, da letztere weniger über den soziokulturellen Hintergrund des Pbn auszusagen vermag als seine Herkunft

- Bedrohung: Wurde oder wird der Gefangene in der aktuellen Haftsituation von Mitgefangenen bedroht oder belästigt (laut Selbstauskunft)?
- Vorinhaftierungen: Wie oft war der Gefangene zuvor inhaftiert (nach Aktenlage)?
- Delikt: Wegen welchen Deliktes / welcher Delikte wurde Anklage erhoben bzw. wurde der Gefangene verurteilt (nach Aktenlage)? Sämtliche Delikte wurden erhoben. Unter die Kategorie „Gewaltdelikt“ fallen folgende Delikte:
  1. §§ 176b-178 StGB: Sexueller Missbrauch von Kindern mit Todesfolge, Sexuelle Nötigung bzw. Vergewaltigung, Sexuelle Nötigung und Vergewaltigung mit Todesfolge
  2. §§ 211-222 StGB: Straftaten gegen das Leben (darunter vor allem Mord, Totschlag und minder schwerer Fall des Totschlags)
  3. §§ 223-231 StGB: Straftaten gegen die körperliche Unversehrtheit (darunter vor allem Körperverletzung und gefährliche Körperverletzung)
  4. § 234 StGB: Menschenraub
  5. § 234a StGB: Verschleppung
  6. § 239 StGB: Freiheitsberaubung
  7. § 239a StGB: Erpresserischer Menschenraub
  8. § 239b StGB: Geiselnahme
  9. § 249 StGB: Raub
  10. § 250 StGB: Schwerer Raub
  11. § 251 StGB: Raub mit Todesfolge
  12. § 252 StGB: Räuberischer Diebstahl
  13. §§ 306-323c StGB: Gemeingefährliche Straftaten<sup>28</sup>
  14. § 340 StGB: Körperverletzung im Amt

---

<sup>28</sup>Nur dann, wenn dabei Dritte zu Schaden kamen oder dies beabsichtigt war, wie im Fall des § 306c: Brandstiftung mit Todesfolge



- Straferwartung: Wie lange glaubt der Gefangene noch in Haft sein zu müssen (laut Selbstauskunft<sup>29</sup>)?
- Kontakte zu einem Psychologen, Psychiater oder Psychotherapeuten: Hatte der Gefangene vor der Inhaftierung jemals Kontakt zu diesen Berufsgruppen um sich beispielsweise beraten oder diagnostizieren zu lassen (laut Selbstauskunft und nach Aktenlage)?
- Psychiatrische bzw. psychotherapeutische Behandlung: Wurde der Gefangene jemals psychiatrisch oder psychotherapeutisch behandelt (laut Selbstauskunft und nach Aktenlage)?
- körperliche Misshandlungen: Wurde der Gefangene in seiner Kindheit körperlich misshandelt oder gequält (laut Selbstauskunft und nach Aktenlage)?
- sexueller Missbrauch: Wurde der Gefangene in seiner Kindheit sexuell missbraucht oder belästigt (laut Selbstauskunft und nach Aktenlage)?
- Analgesie: Spürte der Gefangene Schmerz während der aktuellen AR (laut Selbstauskunft)?

Im Bezug auf die AR wurde die Anzahl früherer AR (in Haft und in Freiheit, laut Selbstauskunft), der Zeitpunkt (Datum, Wochentag, Uhrzeit nach Aktenlage), das Motiv (laut Selbstauskunft), die Methode (laut Selbstauskunft und nach Aktenlage), Dauer und Grad der Planung (laut Selbstauskunft), Ankündigung gegenüber Dritten und AR im Umfeld der Person (in Freiheit und Gefangenschaft, laut Selbstauskunft) erfragt. Die Gesamtdauer beträgt ca. zehn Minuten. Das Formular ist im Anhang zu finden (C.7).

---

<sup>29</sup>Bei Untersuchungsgefangenen handelt es sich um die erwartete Strafe nach Einschätzung des Gefangenen, während es sich bei Strafgefangenen um die Reststrafe handelt, wobei hier bewusst das Wissen bzw. die Erwartung des Gefangenen und nicht Aktenlage erhoben wurde.

## 3.6 Prozedere

Die Daten wurden im Zeitraum von Juli 2004 bis Juli 2006 erhoben. Um neue Vorfälle von AR zu erfragen, kontaktierte der Untersucher jeden Morgen die Arztgeschäftsstellen (AGSt) der Häuser I-III der JVA Tegel, der JSA und die Stationen I-III des Krankenhauses der Berliner Vollzugsanstalten (KBVA). Fälle aus der JVA Moabit wurden entweder über das KBVA der JVA Moabit erfragt oder die Hausbüros der Teilanstalten informierten den Untersucher mit einem Anruf ihrerseits. Die jeweiligen Ansprechpartner, Anstalts- und Teilanstaltsleiter und das Ärztekollegium des KBVA wurden im Vorfeld der Untersuchung in Informationsgesprächen und der Teilnahme an Ärztedienstbesprechungen sowie über die Studie und die beabsichtigte Zusammenarbeit informiert.

Die Daten wurden in den JVA und der JSA erhoben. Dazu wurden, soweit dies die Sicherheitsbestimmungen im individuellen Fall erlaubten, separate Räume abseits des Vollzugsaalltags aufgesucht, um eine Atmosphäre der Ruhe und Vertraulichkeit zu gewährleisten. In einigen Fällen ließ dies der Sicherheitsstatus des Gefangenen nicht zu, sodass die Interviews in diesen Fällen in den bgH erfolgen mussten, in welche die Gefangenen als Folge ihrer AR verbracht worden waren. Zunächst wurde der Gefangene über den Hintergrund der Untersuchung und die absolute Freiwilligkeit seiner Teilnahme informiert.<sup>30</sup> Bei dieser Gelegenheit wurde der Gefangene darauf hingewiesen, dass ihm weder durch Ablehnung noch durch seine Zustimmung zur Teilnahme oder irgendwie geartete Antworten Vor- oder Nachteile entstehen würden. Außerdem wurde hervorgehoben, dass es weder der Zweck der Studie sei noch im Einflussbereich des Untersuchers läge, die individuelle Situation<sup>31</sup> des Gefangenen zu verändern. In diesem Zusammenhang bekräftigte der Untersucher seine Rolle als von der Anstalt, Gerichten und der Polizei unabhängige Person und verwies auf seine Schweigepflicht in Zusammenhang mit § 203 Abs. 2 StGB. Im Falle der Zustimmung hatte der Gefangene eine „Erlaubnis zur Einsichtnahme in die Gefangenenpersonalakte“ und „vorübergehende Entbindung von der ärztlichen Schwei-

---

<sup>30</sup>Der Gefangene erhielt dazu ein ausführliches Informationsblatt (siehe Anlage F).

<sup>31</sup>Beispielsweise Verlegung in einen anderen Haftraum, Kontaktaufnahme zu Verwandten, Medikation des Gefangenen

gepflicht“ (siehe Anlage F) zu unterzeichnen. In der Folge wurden im Rahmen des Interviews dem Gefangenen die Fragebögen vorgelegt (SKID-II, BDI, BHS) bzw. die Interviewfragen gestellt (SKID-I, PCL-R, SIS, BSS). Bei Probanden mit Leseschwäche wurden die Fragen vorgelesen und der Fragebogen gemäß der Antworten des Probanden vom Untersucher ausgefüllt (vgl. Morgan & Hawton, 2004).

Im Anschluss an das Interview wurden zur weiteren Datengewinnung<sup>32</sup> und zur Validierung der Interviewergebnisse die Gefangenenpersonalakte (GPA) und die Gesundheitsakte (GA) analysiert und die Einverständniserklärungen beigelegt. Nach Rücksprache mit dem behandelnden Arzt wurde ein Rating der Letalität der Verletzungen auf der LSARS-II vorgenommen. Die gewonnenen Daten wurden auf einem dafür angefertigten Auswertungsbogen zusammengefasst (siehe Anlage G) und später zur Verrechnung in SPSS eingegeben.

Auf das vielfach anzutreffende Phänomen der Nachfragen der JVA-Bediensteten, des Krankenhauspersonals und der Anstaltsärzte beispielsweise nach dem Zustand des Gefangenen oder der Einschätzung der Ernsthaftigkeit der AR durch den Untersucher, antwortete dieser stets ausweichend und sehr allgemein. Dies war geboten, um einerseits nicht die Schweigepflicht zu verletzen und andererseits auch ein positives Verhältnis zur Untersuchung nicht zu gefährden, wobei stets ersterem Ziel unbedingte Priorität gewährt wurde.

Alle Interviews der AR-Gruppe wurden vom Autor der vorliegenden Arbeit durchgeführt. Zur Arbeitsentlastung wurde ein Teil der KG (29 Probanden) von einer Diplomandin erhoben. Um die Durchführungs- und Auswertungs- und Interpretationsobjektivität zu gewährleisten, wurde die Studentin im Hauptstudium (Psychologie) im Rahmen von Interviewübungen und ersten gemeinsamen Datenerhebungen eingearbeitet und in der Folge ständig vom Autor supervidiert.

---

<sup>32</sup>Unter anderem nötig für die PCL-R und Angaben beispielsweise über das Haftverhalten und die Umstände der AR etc.

## 3.7 Statistische Auswertung

Zur Prüfung, ob sich die zwei Stichproben bezüglich eines Alternativmerkmals signifikant unterscheiden, wurde der Vierfelder- $\chi^2$ -Test angewendet (Prüfung der UH). Aufgrund der teilweise kleinen Stichprobengrößen in den Untergruppen wurde der exakte Fisher-Yates-Test angewendet. Mit diesem wurden Unterschiede der Gruppen bei kategorialen Variablen, deren erwartete Häufigkeiten kleiner als 5 waren, auf ihre statistische Signifikanz hin untersucht (siehe Bortz & Lienert, 2003).

Um die Verteilung der Vorfälle von AR über die Zeit (Wochentage, Uhrzeiten) hinsichtlich ihrer Gleichverteilung bzw. überzufälliger Häufungen zu überprüfen, wurde Pearsons  $\chi^2$ -Test angewendet.

Wegen der guten bis sehr guten Reliabilitätsindices der verwendeten Fragebögen wurde zur Überprüfung von Unterschiedshypothesen bei dimensionalen Daten (Vergleich zwischen AR-Gruppen und K-Gruppe) der  $t$ -Test für abhängige Stichproben herangezogen. Bei einigen der verwendeten Maße liegt zwar eher Ordinal- als Intervallskalenniveau vor (siehe zu dieser Thematik Bortz & Lienert, 2003), doch reagiert dieser Test gegenüber Verletzung der Voraussetzungen (beispielsweise keine Normalverteilung der Messwerte) relativ robust hinsichtlich möglicher Verzerrungen (Bortz, 1999).

Bei Vergleichen dimensionaler Daten innerhalb der AR-Gruppe, getrennt nach Ernsthaftigkeit, fand der  $t$ -Test für unabhängige Stichproben Anwendung.

Korrelationen zwischen den Tests wurden über Spearmans *Rho* ermittelt, weil die verwendeten Skalen auf Ordinalniveau konstruiert wurden (Prüfung der ZH).

Zur Feststellung der Vorhersagekraft der gefundenen Risikofaktoren wurden bei signifikanten Ergebnissen im Rahmen der abhängigen  $t$ -Tests Receiver-Operating-Characteristics-Analysen (ROC-Analysen) durchgeführt (s.u.). Wenn sich signifikante Unterschiede zwischen der AR- und der K-Gruppe hinsichtlich kategorialer Variablen ergaben,

wurde der Kontingenzkoeffizient berechnet. Ziel war es, über dieses Zusammenhangsmaß einen Indikator der prädiktiven Validität des entsprechenden Risikofaktors zu gewinnen.

Die ROC-Analyse stellt sich speziell bei der Entwicklung von Prognoseinstrumenten als Erfolg versprechend dar. Die mittels der ROC generierte Kurve ist eine grafische Gegenüberstellung der Rate richtig positiv klassifizierter Fälle und der falsch positiv klassifizierten Fälle eines Maßes oder diagnostischen Instrumentes bei bestimmten Schwellenwerten (Cutoffs). Durch die ROC-Analyse soll ein optimaler Schwellenwert für Prädiktoren und die entsprechende Sensitivität und Spezifität bestimmt werden. Die ROC-Analyse erlaubt unter anderem einen Vergleich der verschiedenen möglichen Cutoffs des Prädiktors und gibt eine Übersicht über die Genauigkeit der Vorhersage, die sich durch unterschiedliche Schwellenwerte ergibt. Sie kann an Datensätzen eingesetzt werden, wie sie bei dieser Untersuchung vorliegen, nämlich einem kontinuierlich skalierten Prädiktor (beispielsweise Ergebnis des BDI, der BHS, BSS oder auch der PCL-R) und einem nominal skalierten Kriterium (AR). Die ROC ist dabei weniger abhängig von kleinen Basisraten des Kriteriums<sup>33</sup> als dies bei Korrelationsmaßen der Fall ist, die aus herkömmlichen 2x2 Kontingenztafeln (mit Falsch-Positiven und Falsch-Negativen) gewonnen werden. Die Fläche unter der abgebildeten Kurve (area under curve; AUC) kann darüber hinaus als genereller Indikator der Genauigkeit und der praktischen Nützlichkeit interpretiert werden. Diese Fläche reicht von 0 (perfekte negative Vorhersage) über 0.5 (Diagonale = Zufallsvorhersage gleich einem Münzwurf) bis 1 (perfekte Vorhersage). Dabei gibt dieser Wert an einer bestimmten Stelle den Wert für die Wahrscheinlichkeit an, mit der ein – im vorliegenden Fall AR-Gefangener – wirklich eine AR zeigt. So bedeutet beispielsweise ein Wert von 0.75 eine 75%ige Wahrscheinlichkeit, dass sich der Gefangene mit positivem Score auch tatsächlich selbst geschädigt hat (retrospektive Analyse). Dabei gilt, dass AUCs zwischen 0.600 und 0.700 moderate Klassifikationsgenauigkeit widerspiegeln.

---

<sup>33</sup>Auf das Problem der geringen Basisrate bei der Vorhersage seltener Zufallsereignisse wurde bereits in Abschnitt 2.7 eingegangen.

Mit Hilfe des Segmentierungs- und Klassifikationsprogramms AnswerTree wurde anhand der Entwicklungsstichprobe ein baumartig strukturiertes Klassifikationssystem erstellt. Diese Analyseverfahren der Daten soll Populationssegmente ermitteln, definiert durch bestimmte Ausprägungskombinationen der beteiligten Prädiktorvariablen, die sich bezüglich eines vorgegebenen Kriteriums (AR) maximal unterscheiden (Baltes-Götz, 2004). Ziel war die Identifizierung von Risikokombinationen, um auf diese Weise eine Entscheidungshilfe für die Risikoeinschätzung individueller Gefangener in Zukunft zu ermöglichen.

Die Wahl der von AnswerTree angebotenen Algorithmen (CHAID<sup>34</sup>, Exhaustive-CHAID<sup>35</sup>, C&RT<sup>36</sup>, QUEST<sup>37</sup>) fiel auf Exhaustive-CHAID. Der klassische CHAID-Algorithmus stoppt bei der assoziations-steigernden Prädiktorvorbehandlung, wenn keine Kategorienpaare mehr aufgrund insignifikanter Kriteriumsunterschiede zu vereinen sind. Der Exhaustive-CHAID-Algorithmus hat dagegen den Vorteil, alle von einem Prädiktor ermöglichten Zerlegungen auf eine möglichst signifikante Kriteriums-Assoziation hin zu untersuchen. Darüber hinaus „beseitigt die exhaustive Variante einige Ungereimtheiten in der vom klassischen Verfahren eingesetzten Bonferroni-Adjustierung“ (Baltes-Götz, 2004, S. 29). Dieser Algorithmus war auch dem C&RT- bzw. dem QUEST-Verfahren vorzuziehen, da Letztgenannte nur binäre Prädiktorenzerlegungen erlauben und ihr Vorteil, bei vielen metrischen Prädiktoren überlegen zu sein (siehe Baltes-Götz, 2004), angesichts der verwendeten, vorwiegend ordinalskalierten Maße nicht relevant erschien.

CHAID ist eine nicht-parametrische Analyseverfahren, die auf sich wiederholenden Partitionierungsalgorithmen beruht. Sie kann angewendet werden, um gleichzeitig zwei oder mehrere Cutoffwerte für mehr als eine Prädiktorvariable zu errechnen, ohne dass diese Prädiktoren in linearem Verhältnis zur Zielvariable (Kriterium) stehen müssen. Diese

---

<sup>34</sup>Chi-squared Automatic Interaction Detection (Kass, 1980)

<sup>35</sup>Biggs, De Ville und Suen (1991)

<sup>36</sup>Classification and Regression Trees (Breimann, Friedman, Olshen & Stone, 1984)

<sup>37</sup>Quick, Unbiased, Efficient, Statistical Tree (Loh & Shih, 1997)

Art der Analyse macht dabei keine Annahmen über die zugrunde liegende Verteilung, minimiert den Effekt von Ausreißern, erfasst kategoriale und ordinale Daten<sup>38</sup> und ist dabei robust. Dieser Algorithmus fand in einer Reihe von psychologischen Forschungsfeldern Anwendung (siehe beispielsweise Dahle, 2006 oder auch O'Connell et al., 2006).

Da das AnswerTree-Verfahren die Prädiktoren einer sequentiellen Überprüfung unterzieht und die Auswahl des jeweiligen Prädiktors nur von der momentanen Position im Entscheidungsbaum abhängig ist, werden unter Umständen leistungsstarke Prädiktoren nicht berücksichtigt. Um auch deren Beitrag zur Validität der Gesamtvorhersage prüfen zu können, wurden die signifikanten Variablen einer multivariaten Prüfung unterzogen. Das Mittel der Wahl war im hier vorliegenden Fall (nominal skaliertes, dichotomes Kriterium) eine logistische Regression. Sie kam im Rahmen der Prüfung der prädiktiven Validität der Risikofaktoren (2. Fragestellung) zum Einsatz.

In Anbetracht der Vielzahl der nacheinander durchgeführten Hypothesenprüfungen konnte die übliche Irrtumswahrscheinlichkeit von 5% nicht angelegt werden. Eine Bonferroni-Korrektur des Signifikanzniveaus gilt als sehr konservativ (Bortz & Lienert, 2003) und erscheint (wie beispielsweise auch die Holm-Korrektur) mit Blick auf die teilweise recht kleine Stichprobengröße der Subgruppen und dem explorativen Charakter der Studie nicht angebracht. Es galt dementsprechend einen Kompromiss zu finden, bei dem einerseits eine Kumulierung des alpha-Fehlers vermieden bzw. durch Adjustierung ausgeglichen würde und der andererseits auch dem explorativen Charakter der Studie gerecht wird. Es sollte also nicht so niedrig gewählt werden, dass jede Signifikanz von vorneherein verhindert würde.

Deshalb wurde nach Diskussion mit Kollegen aus einer Gruppe internationaler Forscher zur Suizidalität in Gefangenschaft (Lohner, 2005, Juli) wie bei der Veröffentlichung erster Ergebnisse dieses Projektes (Lohner & Konrad, 2006), ein Signifikanzniveau von 1% angelegt. Ergebnisse mit einer Irrtumswahrscheinlichkeit zwischen 1% und 5% wer-

---

<sup>38</sup>Metrische Daten werden auf Ordinalskalenniveau rekodiert.

den als statistischer Trend bezeichnet. Die Irrtumswahrscheinlichkeit wurde meist durch zweiseitige Testung bestimmt, da die insgesamt unklare Ergebnislage aus anderen Studien meist keine Formulierung gerichteter Hypothesen erlaubt. In den Fällen, in denen gerichtete Hypothesen formuliert wurden, wurde dementsprechend eine einseitige Testung durchgeführt.

Für die CHAID-Analyse<sup>39</sup> wurde eine Irrtumswahrscheinlichkeit von 5% gewählt, da im Rahmen dieses Verfahrens ohnehin nur Prädiktoren Aufnahme fanden, die sich zuvor auf dem eben beschriebenen niedrigeren Signifikanzniveau bewährt hatten. Der  $p$ -Wert wurde bei Analysen mit rekodierten Prädiktoren einer Bonferroni-Adjustierung unterzogen, um die künstliche Deflationierung durch die optimale Zusammenlegung von Kategorien auszugleichen (siehe Baltés-Götz, 2004).

Die Daten wurden mittels des Statistikprogramms „SPSS für Windows“ Version 12.0 G statistisch ausgewertet. Zur Segmentierungsanalyse wurde das auf SPSS basierende Programm „SPSS AnswerTree“ Version 3.0 verwendet.

## 3.8 Rahmenbedingungen der Studie

### 3.8.1 Ethische Aspekte

Beim Umgang mit AR stehen Ärzte und Psychologen vor einer Reihe von ethischen Fragestellungen. Dies trifft insbesondere auf den Justizvollzug zu. So werden immer wieder der Einsatz von Isolation und die Rolle von Psychiatern und Psychologen als Erfüllungsgehilfen eines reibungslosen Ablaufs des Gefängnisalltags kritisch diskutiert (Corcos & Lewin, 2001). Dies trifft auch und besonders auf die Erforschung von AR zu. Mishara und Weisstub (2005) diskutieren in diesem Zusammenhang eine Reihe von speziellen ethischen und rechtlichen Problemstellungen. Dabei ist neben den bereits

---

<sup>39</sup>Obwohl es sich um den Exhaustive-CHAID-Algorithmus und nicht den CHAID-Algorithmus handelt, wird der Einfachheit halber und wegen der großen Ähnlichkeit der Verfahren im Folgenden auf die Zusatzbezeichnung verzichtet.



erwähnten ethisch fragwürdigen prospektiven Untersuchungen mit „non-treatment“-Kontrollgruppen (Lloyd, 1990) unter anderem auch die Fragestellung von besonderem Interesse, wie im Hinblick auf die Schweigepflicht zu verfahren ist, wenn der Proband dem Untersucher die Absicht eines (weiteren) SV mitteilt und dies (im Falle von Gefangenen) der Anstaltsleitung nicht bekannt ist. Die vorliegende Arbeit schließt sich diesbezüglich einer – um in der Terminologie von Mishara und Weisstub zu bleiben – moralistischen Sichtweise an, die von der Erhaltung des Lebens und deshalb Verhinderung eines potentiellen SV als höchstem Gut ausgeht und deshalb in einem solchen Fall die Schweigepflicht brechen würde.

Mishara und Weisstub (2005) diskutieren darüber hinaus die Freiwilligkeit der Teilnahme von Gefangenen an Untersuchungen über AR angesichts ungleicher Machtverhältnisse und externen Drucks durch Bedienstete. Deshalb wurde, wie bereits erwähnt, besonders die individuelle Folgenlosigkeit der Untersuchung gegenüber den Gefangenen betont. Im Übrigen verweisen Fazel und Lubbe (2005) diesbezüglich darauf, dass Gefangene grundsätzlich zumindest aus psychiatrischer Sicht durchaus einwilligungs- und zustimmungsfähig seien.

Bei der schwierigen Frage nach der Bewertung von falschen Vorhersagen<sup>40</sup> neigt der Autor dieser Arbeit dazu, sich der relativistischen Sichtweise nach Mishara und Weisstub (2005) anzuschließen, die angesichts knapper Ressourcen von einem ausgewogenen Verhältnis von größtmöglicher Sensitivität bei größtmöglicher Spezifität ausgeht. Eine endgültige Bewertung dieser Fragestellung ist allerdings weder möglich noch scheint der Versuch im Rahmen dieser Arbeit geboten.

---

<sup>40</sup>Mehr Falsch-Positive bedeuten fehlgeleitete Hilfsanstrengungen, mehr Falsch-Negative dagegen unentdeckte Risikogefangene.

### 3.8.2 Rechtliche Aspekte

Alle notwendigen Genehmigungen wurden von der Senatsverwaltung für Justiz eingeholt<sup>41</sup> und das Vorgehen wurde mit den Anstaltsleitungen abgestimmt.

Bevor Interviews mit Untersuchungsgefangenen durchgeführt werden konnten, war in der Regel die Erteilung eines Sprechscheins durch den / die Vorsitzende(n) Richter(in) bzw. die Staatsanwaltschaft von Nöten. Um die richterliche Unabhängigkeit nicht zu untergraben, konnte der Präsident des Landgerichts keine pauschale Erlaubnis erteilen. Er unterstützte und begrüßte jedoch das Vorhaben,<sup>42</sup> wodurch die Erteilung der Sprechscheine im Folgenden beschleunigt werden sollte.

Im Hinblick auf den bereits erwähnten (hypothetischen) Fall, dass ein Gefangener gegenüber dem Untersucher eine Suizidabsicht äußerte, die der Anstalt bis dato noch nicht bekannt wäre, entschied sich der Untersucher grundsätzlich und bereits im Vorfeld für folgendes Vorgehen:

Zunächst sollte versucht werden, den Gefangenen dazu zu bewegen die Hilfsangebote seitens der Anstalt anzunehmen, d. h. es würde ihm nahe gelegt, er möge sich diesbezüglich an einen Anstaltsarzt wenden.

Würde dies der Gefangene ausschlagen oder dieser Vorschlag nicht glaubhaft verfolgt, so würde der Untersucher versuchen, das Einverständnis des Gefangenen zu einer Entbindung des Untersuchers von seiner Schweigepflicht gegenüber dem Anstaltsarzt zu bewegen.

Sollte auch dies nicht zu erreichen sein, so würde sich der Untersucher nach einer Abwägung der in § 203 Abs. 2 StGB<sup>43</sup> bzw. § 323c StGB<sup>44</sup> beschriebenen Rechtsgü-

---

<sup>41</sup>Az. III AF 2-4557 E-V / 12.04

<sup>42</sup>Mitgeteilt mit Schreiben vom 18. August 2004.

<sup>43</sup>Verletzung von Privatgeheimnissen

<sup>44</sup>Unterlassene Hilfeleistung

ter, wie bereits im letzten Abschnitt beschrieben, für die Sicherheit des Gefangenen entscheiden und die Schweigepflicht brechen.

Glücklicherweise konnten in den wenigen Fällen, in denen sich der Untersucher vor dieses Problem gestellt sah, die Gefangenen dazu bewegt werden, sich nicht nur dem Interviewer, sondern auch der Ärzteschaft anzuvertrauen, sodass der dritte Schritt nicht nötig wurde.

### 3.8.3 Datenschutzrechtliche Prüfung

Die Datenerhebung erfolgte in kodierter Form mit getrennten Erfassungsbögen und Codeschlüsseln. Die Erfassungsbögen enthielten keine Namen, Geburtsdaten oder Aktennummern. Die relevanten Informationen wurden in den Stahlcontainern des Gutachtenarchivs des Instituts für Forensische Psychiatrie der Charité – Universitätsmedizin Berlin gesichert aufbewahrt und nach Abschluss des Projektes im Reißwolf vernichtet, wobei dies namentlich und mit Datum versehen protokolliert wurde. Die Verfahrensweise wurde vom Landesdatenschutzbeauftragten des Landes Berlin geprüft<sup>45</sup>.

---

<sup>45</sup>Vorgangsnr. 5612.75