

**Chronische Rückfalldelinquenz im Individuellen Menschlichen  
Entwicklungsverlauf**

**Die Berliner CRIME – Studie**

**Ausgewählte Ergebnisse  
aus dem Endbericht für die  
Deutsche Forschungsgemeinschaft  
(DA 522/1-1, DA 522/1-2, DA 522/1-4)**

**PD Dr. phil. Klaus-Peter Dahle  
und cand. psych. Thorsten Kahnt**

**2006**

## **Inhalt**

1 Methodische Vorbemerkung .....	1
2 Stichprobenbeschreibung und Rückfallraten.....	1
3 Korrelate ausgewählter Einzelmerkmale mit Rückfälligkeit .....	2
4 Verteilungs-, Item- und Skalencharakteristika von LSI-R, HCR-20 und PCL-R sowie der idiographischen Einschätzungen .....	3
5 Beurteilerübereinstimmung .....	4
6 Vorhersagegenauigkeit der aktuarischen Prognosen.....	4
7 Vorhersagegenauigkeit der klinisch-idiographischen Prognosen .....	6
8 Zusammenfassende Bewertung der bislang erreichten Ergebnisse.....	7

## 1 Methodische Vorbemerkung

Die im Rahmen der CRIME-Studie erstellten klinisch-idiographischen Prognosen folgten methodisch dem Strukturmodell klinischer Kriminalprognosen (Dahle, 1997, 2000) und wurden explizit in jedem Einzelfall nach der Einschätzung der Basisrückfallrate und unter Berücksichtigung der Ergebnisse einiger Prognoseinstrumente – dem *Level of Service Inventory - Revised* (LSI-R; Andrews & Bonta, 1995), dem *HCR-20 Schema* (Webster, Douglas, Eaves & Hart, 1997) und der *Psychopathy Checklist – Revised* (PCL-R, Hare, 1991) – vorgenommen. Sie basieren insoweit auf aktuarischen Einschätzungen der zu erwartenden Rückfallwahrscheinlichkeiten und sollten diese bei entsprechendem Anlass durch den Einbezug etwaiger individueller Besonderheiten ergänzen bzw. modifizieren. Für Auswertungszwecke waren die Prognostiker gehalten, getrennte Aussagen zur Wahrscheinlichkeit allgemeiner strafrechtlicher Rückfälle und zur Wahrscheinlichkeit speziell gewalttätiger Straftaten zu treffen und diese jeweils auf einer 5-stufigen Ratingskala (sehr günstig – eher günstig – unklar/mittleres Risiko – eher ungünstig – sehr ungünstig) einzuschätzen. Diese Ratings dienten als Grundlage für die Validitätsanalyse.

Aus der Fülle in Frage kommender Rückfallkriterien für die Analyse der Vorhersagevalidität der Prognosen wurde für den Bericht eine Auswahl nach Katamnesezeit und Schweregrad differenzierter Kriterien getroffen. Die meisten Analysen differenzieren einerseits kurz- (2 Jahre), mittel- (5 Jahre) und langfristige (Gesamtkatamnese:  $M=19,61$  Jahre;  $SD=1,74$ ) Beobachtungszeiträume und andererseits zwischen allgemeinen Rückfällen (jede Verurteilung wegen einer nach Haftentlassung begangenen Tat), erneutem (vollzogenen) Freiheitsentzug, Gewaltdelikten und gravierenden Gewaltdelikten (Gewalttaten mit Freiheitsstrafe  $> 2$  Jahre). Für einzelne Auswertungen wurde zudem ein Rückfallschwereindex mit folgenden Stufen gebildet: kein Rückfall – geringfügiger Rückfall ohne erneute Haft (nur Geld- oder erfolgreiche Bewährungsstrafen) – moderater Rückfall mit Freiheitsstrafe(n) bis zu 2 Jahren (ggf. auch mehrere) – erheblicher Rückfall mit Freiheitsstrafen (auch) über 2 Jahren – gravierende Gewalttat(en) (Gewalttaten mit FS  $> 2$  Jahre). Grundlage waren in jedem Fall entsprechende Einträge im BZR aufgrund von Handlungen, die nach Entlassung aus der Indexhaft erfolgten.

Soweit für einzelne Personen aus unterschiedlichen Gründen (Verstorben, Abgeschoben, andauernde Schuldunfähigkeit) keine BZR-Auszüge verfügbar waren, wurden diese ausgeschlossen, so dass sich die Auswertungen insgesamt auf  $N = 307$  Pbn (77,3 % der Ursprungsstichprobe) beziehen.

## 2 Stichprobenbeschreibung und Rückfallraten

Zum Zeitpunkt der Erstuntersuchung waren die Pbn im Durchschnitt 30,52 Jahre alt ( $SD=5,38$ ). Knapp die Hälfte trat damals ihre Freiheitsstrafe aufgrund eines Diebstahlsdelikts an, gefolgt von Gewaltdelikten (19%), Betrugsdelikten (13%) und

Unterhaltspflichtverletzungen (10%; jeweils Hauptdelikt). Die für das Indexdelikt angeordnete Freiheitsstrafe betrug im Mittel 18,93 Monate ( $SD = 18,10$ ). Die Pbn wurden zwischen 1976 und 1990 aus der Indexhaft entlassen, die Katamnesezeit nach der Entlassung bis zur Einsichtnahme in das Bundeszentralregister (BZR) im Jahr 1997 beträgt im Durchschnitt 19,6 Jahre ( $SD = 1,7$ ).

Für insgesamt rd. 72% der Pbn fanden sich bis zu 20 erneute Einträge im Strafregister ( $M=4,68$ ;  $SD=4,35$ ), 64% verbüßten während der Katamnesezeit erneute Haftstrafen. Insgesamt wurden 863 erneute Freiheitsstrafen mit einer Gesamtdauer von rund 783 Jahren ausgesprochen, von denen geschätzte 676 Jahre vollstreckt wurden. Bei den einzelnen Deliktgruppen variierten die allgemeinen Rückfallquoten zwischen 54% (Unterhaltspflichtverletzer) und 80% (Betrüger). Tabelle 1 gibt eine Übersicht über die Rückfallquoten der Gesamtgruppe und von relevanten Subgruppen im Hinblick auf unterschiedliche Rückfallereignisse.

Tabelle 1 Basisrückfallraten der Gesamtgruppe und relevanter Teilgruppen für die Gesamtkatamnese

	jede neue Verurteilung	erneute Strafhaft	Gewaltdelikt	gravierendes Gewaltdelikt*
<b>alle Probanden</b>				
2 Jahre nach Haftentlassung	47 %	38 %	7 %	4 %
5 Jahre nach Haftentlassung	63 %	57 %	13 %	6 %
Gesamtkatamnese	72 %	64 %	23 %	9 %
<b>nach Deliktgruppen (Gesamtkatamnese)</b>				
Eigentumsdelikte	75 %	70 %	22 %	11 %
Betrugsdelikte	80 %	64 %	18 %	5 %
Gewaltdelikte	68 %	59 %	36 %	14 %
Unterhaltspflichtverletzung	54 %	39 %	6 %	0 %
Sonstige Delikte	78 %	72 %	17 %	6 %
<b>nach Beginn der strafrechtlichen Karriere (Gesamtkatamnese)</b>				
Ersthäft als Jugendlicher	86 %	81 %	31 %	18 %
Ersthäft als Heranwachsender	75 %	67 %	33 %	9 %
Ersthäft als Erwachsener	61 %	50 %	12 %	4 %

\* Gewaltdelikt mit Freiheitsstrafe > 2 Jahre

### 3 Korrelate ausgewählter Einzelmerkmale mit Rückfälligkeit

Zunächst wurde eine Reihe von Einzelmerkmalen auf ihre Zusammenhänge mit dem Rückfallverhalten hin untersucht. Erwartungsgemäß zeigten von den untersuchten statischen Merkmalen insbesondere Variablen zur strafrechtlichen Vorschichte relativ hohe und auch stabile Zusammenhänge mit Rückfälligkeit. So imponiert das Alter bei Ersteintrag im BZR mit einem, über die Beobachtungszeiträume relativ stabilen Zusammenhang von  $-.33$  bis  $-.38$

( $p < 0.01$ ) mit dem Rückfallschwereindex. Für die EEG-Befunde (Teilgruppe mit  $n=102$ ) ergaben sich indessen kaum nennenswerte Korrelationen; nur ein vergleichsweise unspezifischer Parameter (unspezifische Abweichung vom Norm-EEG) zeigte mäßige Zusammenhänge mit der allgemeinen Rückfälligkeit der Pbn ( $r=.23-.24$ ), bedeutsame Zusammenhänge mit der Rückfallschwere ergaben sich indessen nicht.

Auch für die dynamischen Variablen – insbesondere die Qualität der sozialen Einbettung bei Haftentlassung, die Arbeitsperspektiven und das Ausmaß kriminogener Einstellungen – ergaben sich einige recht beachtliche Zusammenhänge, deren Größenordnung mit Zunahme der Beobachtungsdauer jedoch tendenziell abnahm. Da eine Teilgruppe der Pbn während der Indexhaft in der Sozialtherapeutischen Anstalt der JVA Tegel behandelt wurde ( $n=26$ ), wurden etwaige Behandlungseffekte gesondert untersucht. Da die behandelte Gruppe tendenziell gravierendere Anlasstaten aufwies und entsprechend längere Haftstrafen verbüßte als die Gesamtgruppe, wurden für die Vergleichsuntersuchung per Paarzuordnung aus der unbehandelten Gruppe diejenigen Pbn ausgewählt, die dem jeweiligen Partner der Behandlungsgruppe hinsichtlich Anlasstat, Dauer der Haftstrafe, Alter und Vorstrafenzahl am ähnlichsten war ( $n=26$ ). Es zeigten sich im Vergleich zu dieser Kontrollgruppe teilweise recht beachtliche Behandlungseffekte von längerfristig bis zu .33.

#### **4 Verteilungs-, Item- und Skalencharakteristika von LSI-R, HCR-20 und PCL-R sowie der idiographischen Einschätzungen**

Beim Summenscore des LSI-R ergab sich für die Gesamtgruppe ein Mittelwert von  $M=24,65$  ( $Mdn=25$ ) und eine Standardabweichung von  $SD=7,35$ . Dies entspricht recht genau den von den Testautoren für ihre Normstichprobe genannten Verhältnissen (Andrews & Bonta, 1995). Die Verteilung zeigte eine leicht rechtsschiefe Charakteristik (K-S  $z=1,40$ ;  $p < .05$ ). Die mittlere Itemschwierigkeit lag bei  $0,46$  ( $SD=0,27$ ). Die interne Konsistenz (Cronbachs Alpha) betrug für das gesamte Verfahren  $\alpha=.84$ , sie lag im (ungewichteten) Mittel der Subskalen bei  $\alpha=.46$ , was an einzelnen Einbrüchen bei Skalen mit nur wenigen Items lag.

Der Summenscore des HCR-20(+3) lag im Mittel bei  $M=16,52$  ( $Mdn=16$ ;  $SD=6,31$ ) und war weitgehend normalverteilt (K-S  $z=0,99$ ,  $p=.28$ ). Die mittlere Itemausprägung betrug  $.75$  ( $SD=0,21$ ), wobei die Items Werte zwischen 0 und 2 annehmen können. Die interne Konsistenz betrug für den Summenscore ebenfalls  $\alpha=.84$  und für die Subskalen  $.69$ ,  $.49$  und  $.80$ <sup>1</sup>.

Mit einem mittleren Summenscore von  $M=12,03$  ( $Mdn=12$ ;  $SD=4,70$ ) ergab sich auch für die PCL-R eine vergleichsweise geringe mittlere Ausprägung (wobei ähnliche Gründe, wie sie für die HCR-20 genannt wurden, eine Rolle spielen dürften; vgl. FN 1). Die

---

<sup>1</sup> Hinsichtlich der, im Vergleich zu anderen Untersuchungen relativ gering erscheinenden mittleren Ausprägung des HCR-Scores ist zu beachten, dass es sich bei der CRIME-Stichprobe vor der Hand nicht um psychisch gestörte Personen handelt (weswegen insb. die psychopathologischen Items eine sehr geringe Auftretenswahrscheinlichkeit hatten) und es sich nicht um eine Selektion gewalttätiger Straftäter, sondern um eine Querschnittsstichprobe aus dem Strafvollzug handelt.

Verteilung wies dabei eine leicht linksschiefe Charakteristik auf (K-S  $z=1,57$ ,  $p<.05$ ). Die mittlere Itemausprägung lag bei 0,30 ( $SD=0,22$ ), wobei auch im PCL Werte zwischen 0 und 2 auftreten können. Die interne Konsistenz der Gesamtskala lag bei  $\alpha=.71$ , für die beiden Faktoren bei  $\alpha=.59$  bzw.  $\alpha=.71$ .

Im Gegensatz zu den normal- bzw. nur leicht schiefen Verteilungen der zusammenfassenden Summenscores der aktuarischen Instrumente wiesen die idiographischen Prognoseeinschätzungen eine erhebliche rechts- (allgemeine Rückfallerwartung) bzw. linksschiefe (Gewaltprognose) Verteilungsform auf. Dabei ist zu bedenken, dass diese Verteilungsformen eher den Basisraten der entsprechenden Rückfallereignisse entsprechen. Eine weitere Konsequenz der nicht-normalverteilten Verhältnisse ist es, dass für größere Personengruppen vergleichsweise klare Prognosen getroffen wurden, wohingegen bei den aktuarischen Systemen überproportional viele Pbn mittlere Ausprägungsgrade nahe dem Gesamtmittelwert aufwiesen. Dies lässt erwarten, dass mit den aktuarischen Systemen eindeutige Prognosen nur für vergleichsweise kleine Personengruppen mit entsprechend extremen Verhältnissen möglich sind.

Die verschiedenen Prognoseverfahren korrelierten erwartungsgemäß deutlich miteinander. Die Interkorrelationen variierten dabei zwischen  $r_{\min}=.45$  (idiographische Gewaltprognose und LSI-R) und  $r_{\max}=.80$  (LSI-R und HCR-20), im Mittel lag sie bei  $r_{\text{mean}}=.61$  ( $SD=0,12$ ).

## 5 Beurteilerübereinstimmung

Um Hinweise auf die Beurteilerübereinstimmungen zu erhalten, wurde eine Teilgruppe von  $n=30$  Pbn unabhängig voneinander durch zwei Auswerter mit den verschiedenen Verfahren eingeschätzt. Für den LSI-Gesamtscore und die idiographische Gewaltprognose ergaben entsprechende Überprüfungen leichte, aber statistisch signifikante Niveaueffekte zwischen den Beurteilern (d. h. ein Prognostiker wertete die Prognosen systematisch etwas skeptischer), wobei die Mittelwertdifferenzen für den LSI-R mit  $\Delta = 1,2$  ( $t_{(df=29)}=2,36$ ,  $p<.05$ ) und für die Gewaltprognose  $\Delta = 0,37$  ( $t_{(df=29)}=2,48$ ,  $p<.05$ ) jedoch eher gering waren.

Insgesamt wurden bei den aktuarischen Systemen sehr gute Übereinstimmungen erreicht (ICC .91 bis .94). Diese waren bei den klinisch-idiographischen Einschätzungen erwartungsgemäß etwas niedriger, sie liegen jedoch noch auf einem auch für individualdiagnostische Zwecke hinreichendem Niveau (ICC bei .77 bis .79).

## 6 Vorhersagegenauigkeit der aktuarischen Prognosen

Zur Beurteilung der Vorhersageleistung der aktuarischen Instrumente wurden für die eingangs definierten Katamnesezeiträume und Rückfallereignisse zunächst die jeweiligen Korrelationen mit dem Gesamtscore berechnet und für die dichotomen Rückfallkriterien die Flächen unter den ROC-Kurven (area under curve; AUC) mit den zugehörigen

Konfidenzintervallen (95%) bestimmt. Sämtliche Validitätsmaße erreichten signifikante, zumeist hochsignifikante Größenordnungen, die Korrelationen reichten dabei bis zu  $r_{\max}=.43$  (LSI-R mit Rückfallindex für 2- und 5-Jahreskatamnese), die Flächenmaße bis zu  $AUC_{\max}=.77$  (HCR-20 mit gravierenden Gewaltdelikten 5-Jahreskatamnese). Die meisten Validitätskoeffizienten ergaben vergleichbare Größenordnungen, wie sie auch aus internationalen Studien berichtet wurden.

Eine differenzierte Betrachtung der Werte für die verschiedenen Rückfallereignisse und Katamnesezeiten deutet zunächst auf folgende Besonderheiten: (1.) Das LSI zeigt seine Stärken eher bei der Vorhersage kurzfristiger nicht-gewalttätiger Rückfälle, während für die (längerfristige) Vorhersage gewalttätiger Straftaten eher die anderen beiden Instrumente indiziert scheinen. (2.) Die Korrelationen mit den zusammenfassenden Rückfallindizes sind insgesamt höher als mit den dichotomen Rückfallkriterien, was auf lineare Zusammenhänge der Scores sowohl mit der Rückfallwahrscheinlichkeit als auch mit der Rückfallschwere deutet. (3.) Die Vorhersage nicht-gewalttätiger Rückfälle scheint über kürzere Beobachtungszeiträume (2 Jahre) valider als für längere Zeiträume, während die Gewaltprognose erst bei längeren Beobachtungszeiträumen (mindestens 5 Jahre) eine nennenswerte Validität erreicht. (4.) Bei den Gewaltprognosen erscheinen einige Validitätsscores teilweise etwas inhomogen in dem Sinne, dass eher schwache Korrelationen mit höher erscheinenden AUC-Maßen einhergehen. Dies deutet darauf hin, dass der Informationsgewinn durch die betreffenden Instrumente nicht gleichmäßig über den Messbereich der jeweiligen Skala verteilt ist, sondern in Teilabschnitten Schwerpunkte aufweist – dies wird auch durch die Sichtung der ROC-Kurven bestätigt.

Trotz einiger leichter spezifischer Stärken und Schwächen, unterschieden sich die Gütewerte der Instrumente für die meisten Rückfallereignisse und Katamnesezeiträume nicht signifikant voneinander. Sie lagen insgesamt auf weitgehend befriedigendem bis gutem Niveau, ähnlich wie sie international berichtet werden. Nähere Analysen deuteten jedoch auch auf einige Begrenzungen hin. Bedingt durch die weitgehend symmetrische Verteilungscharakteristik erzielten zunächst vergleichsweise viele Pbn Werte nahe des Gesamtmittelwertes (Mittelfeldproblematik) und insoweit unspezifische Prognosen. Für relativ viele Pbn war der erzielbare prognostische Informationsgewinn insoweit gering. Ferner fluktuierte die Prognosegüte in Abhängigkeit von einigen kriminologischen, demographischen und psychopathologischen Merkmalen der Probanden, es ließen sich also gut einschätzbare („predictables“) und weniger gut einschätzbare („unpredictables“) Pbn identifizieren. Last but not least schienen die untersuchten Instrumente den vorhandenen Bestand empirisch gesicherter Kenntnisse über Einflussfaktoren auf die Rückfälligkeit von Rechtsbrechern noch nicht ganz optimal zu nutzen, da ein einfaches Vorhersagemodell, das einige wenige Aspekte (Alter, Behandlung, typische Hoch- oder Niedrigrisikokonfiguration) zusätzlich einbezieht, bereits systematisch zuverlässigere Vorhersagen erzielte und die o. g. Schwächen zu reduzieren vermochte.

## 7 Vorhersagegenauigkeit der klinisch-idiographischen Prognosen

Auch für die klinisch-idiographischen Prognosen wurden zur Beurteilung der Vorhersagegüte zunächst die Validitätskoeffizienten, wie im Vorabschnitt beschrieben, berechnet. Für alle geprüften Rückfallkriterien und Beobachtungszeiträume ergaben sich dabei hochsignifikante Werte in Größenordnungen bis zu  $r_{\max}=.59$  bzw.  $AUC_{\max}=.85$ . Wie bei den aktuarischen Instrumenten erzielten die Gewaltprognosen insgesamt geringere Werte, die jedoch mit zunehmender Beobachtungsdauer besser wurden. Hervorzuheben ist, dass sämtliche Koeffizienten höher waren, als die entsprechenden Werte der aktuarischen Instrumente, die Leistungsverbesserung war für die meisten Rückfallereignisse und Katamnesezeiträume statistisch abzusichern (vgl. Tabelle 2).

Tabelle 2 Vergleiche<sup>a</sup> der prognostischen Leistungen (AUC-Maße) der statistischen und klinischen Prognosen im Hinblick auf verschiedene Rückfallereignisse und Vorhersagezeiträume

	jede neue Verurteilung	erneute Strafhaft	Gewaltdelikt	gravierendes Gewaltdelikt <sup>b</sup>
<b>2 Jahre:</b>				
klinisch vs. statistisch allg.	$\chi^2(df=1)=11,26$ $p<.001$	$\chi^2(df=1)=9,16$ $p<.01$	$\chi^2(df=1)=4,59$ $p<.05$	siehe <sup>b</sup>
klinisch vs. LSI – R	$z = 4,13$ $p<.001$	$z = 3,27$ $p<.001$	$z = 2,55$ $p<.01$	siehe <sup>b</sup>
klinisch vs. HCR-20	$z = 3,98$ $p<.001$	$z = 3,73$ $p<.001$	$z = 1,43$ n.s.	siehe <sup>b</sup>
klinisch vs. PCL – R	$z = 4,41$ $p<.001$	$z = 4,36$ $p<.001$	$z = 2,99$ $p<.01$	siehe <sup>b</sup>
<b>5 Jahre:</b>				
klinisch vs. statistisch allg.	$\chi^2(df=1)=13,99$ $p<.001$	$\chi^2(df=1)=13,32$ $p<.001$	$\chi^2(df=1)=1,67$ n.s.	$\chi^2(df=1)=1,47$ n.s.
klinisch vs. LSI – R	$z = 5,75$ $p<.001$	$z = 4,89$ $p<.001$	$z = 2,21$ $p<.05$	$z = 1,39$ n.s.
klinisch vs. HCR-20	$z = 4,08$ $p<.001$	$z = 4,74$ $p<.001$	$z = 1,23$ n.s.	$z = 1,03$ n.s.
klinisch vs. PCL – R	$z = 3,51$ $p<.001$	$z = 4,21$ $p<.001$	$z = 1,59$ n.s.	$z = 1,48$ n.s.
<b>Gesamtkatamnese:</b>				
klinisch vs. statistisch allg.	$\chi^2(df=1)=3,41$ n.s.	$\chi^2(df=1)=10,82$ $p<.01$	$\chi^2(df=1)=8,20$ $p<.01$	$\chi^2(df=1)=3,90$ $p<.05$
klinisch vs. LSI – R	$z = 4,11$ $p<.001$	$z = 5,34$ $p<.001$	$z = 4,04$ $p<.001$	$z = 2,87$ $p<.01$
klinisch vs. HCR-20	$z = 2,36$ $p<.05$	$z = 4,24$ $p<.001$	$z = 2,96$ $p<.01$	$z = 1,75$ ( $p<.10$ )
klinisch vs. PCL – R	$z = 1,57$ n.s.	$z = 3,19$ $p<.001$	$z = 3,11$ $p<.001$	$z = 2,13$ $p<.05$

- <sup>a</sup> Alle Tests wurden zweiseitig durchgeführt. Die einzelnen Paarvergleiche wurden in der Tabelle nicht alphaadjustiert.
- <sup>b</sup> Gewaltdelikt mit Freiheitsstrafe über 2 Jahren; für 2-Jahres-Beobachtungszeitraum war die Basisrate zu gering für sinnvolle statistische Auswertungen

Der Nachweis besserer Leistungen klinisch-idiographischer gegenüber nomothetisch-aktuarischer Prognosen stellt ein vergleichsweise seltenes Ergebnis von Validierungsstudien dar. Zu bedenken ist dabei, dass es sich bei den klinisch-idiographischen Prognosen in der hiesigen Studie (1.) um informierte, d. h. in expliziter Kenntnis der statistischen Zusammenhänge erstellter Individualprognosen handelt und diese Beurteilungen (2.) einer ausformulierten methodischen Vorgehensweise folgten. Es handelte sich insoweit nicht um intuitive Beurteilungen von Experten im Sinne einer „unguided clinical prediction“, wie sie in sehr vielen vorliegenden Vergleichsstudien zugrunde gelegt wurden.

## **8 Zusammenfassende Bewertung der bislang erreichten Ergebnisse**

Mit der CRIME-Studie liegen erstmals Untersuchungsergebnisse zur Anwendbarkeit und zur prognostischen Güte für einige der international meistdiskutierten Prognoseinstrumente an einer unselektierten deutschen Gruppe ehemaliger Strafgefangener vor. Die Befunde sprechen grundsätzlich dafür, dass die Instrumente mit leichten Adaptationen auch bei hiesigen Strafgefangenen reliabel anwendbar sind und vergleichbare Vorhersageleistungen erzielen, wie sie auch aus anderen Ländern berichtet werden. Eine eindeutige und übergreifende Überlegenheit eines bestimmten Verfahrens zeigte sich nicht, wenngleich sich in den Ergebnissen einige leichte spezifische Stärken und Schwächen andeuteten (die jedoch zumeist nicht statistisch abzusichern waren). Trotz hoher Interkorrelationen konnte durch eine Integration der Befunde der Instrumente eine, gegenüber den einzelnen Instrumenten, leicht verbesserte Vorhersageleistung erzielt werden. Auch dies lässt vermuten, dass die nicht gemeinsamen Varianzanteile der Verfahren durchaus spezifische Bedeutungen haben könnten.

Genauere Analysen zeigten indessen auch einige Grenzen der aktuarischen, gewöhnlich auf Summenbildungen von Risikofaktoren basierenden Instrumente. Die wichtigste Einschränkung liegt in der Neigung dieser Scores, bei größeren Fallzahlen eine Gauß'sche Verteilungsform (bzw. hiervon nur unwesentlich abweichende Form) anzunehmen. Eigenart bedingt eine hohe Grundwahrscheinlichkeit, dass Personen durch die Instrumente vorzugsweise in mittlere Risikobereiche klassifiziert werden, wo die Prognose nicht nennenswert von den zugehörigen Basisraten abweicht und der Informationsgewinn entsprechend niedrig ist. Klare Prognosen mit eindeutig erhöhter oder reduzierter Risikoerwartung sind demgegenüber – jedenfalls bei Anwendung an unselektierten Strafgefangenenpopulationen – eher selten.

Mit der CRIME-Studie liegen weiterhin erstmals Befunde zur Anwendbarkeit und Vorhersagegüte einer dezidierten klinisch-idiographischen Prognosemethodik vor, die den derzeitigen rechtlichen Anforderungen an individuelle Kriminalprognosen im Strafrecht eher entspricht als es ein rein aktuarischer Ansatz vermag (vgl. KG – Berlin, Beschluss vom 11.12.1998 – 5 Ws 672/98 sowie LG Nürnberg, Beschluss vom 22.8.2001 – Ws 942/01). Die mit dieser Methodik erstellten Rückfall- und Gewaltprognosen erwiesen sich sowohl gegenüber den einzelnen aktuarischen Instrumenten als auch gegenüber diversen Verknüpfungen als durchgängig treffsicherer. Insbesondere wiesen diese Prognosen nicht die o.g. Mittelfeldproblematik der aktuarischen Instrumente auf, so dass für deutlich größere Personengruppen eindeutigere prognostische Aussagen möglich waren. Bei der Würdigung des, in der bisherigen Literatur eher ungewöhnlichen Befundes einer gegenüber statistischen Verfahren überlegenen klinisch-idiographischen Beurteilung ist indessen zu bedenken, dass die hiesigen individuellen Beurteilungen explizit in Kenntnis der durch die Instrumente erzielbaren Informationen erfolgten und hierauf aufbauten und zudem einer dezidierten Methodik folgten. In diesem Sinne besteht, nach den bisherigen Ergebnissen der Studie, die zuverlässigste Strategie bei der individuellen Prognosebeurteilung in einem zweistufigen Vorgehen, das sowohl empirisch belegte Zusammenhänge, wie sie die aktuarischen Instrumente in systematischer Form zusammenstellt, berücksichtigt als auch eine hierauf aufbauende individuelle Beurteilung vornimmt, um etwaigen Besonderheiten Rechnung tragen zu können.

Fragt man abschließend nach wichtigen Forschungsdesideraten, so scheinen derzeit zwei Aspekte vordringlich. Zunächst scheint für einige Teilgruppen der Probanden weiterer Forschungsbedarf zur Entwicklung angemessener Prognosemethoden zu bestehen, da die Beurteilung dieser speziellen Gruppen mit dem derzeit verfügbaren Methodenarsenal noch wenig befriedigende Ergebnisse erbrachte. Im Hinblick auf die aktuellen Diskussionen um Verschärfungen der Maßregel der Sicherungsverwahrung erscheinen vor allem aber weitere Validierungsstudien an Zielgruppen mit besonders gravierender Gewaltkriminalität geboten – die hier vorgelegte Arbeit betraf eine Querschnittsklientel aus dem Strafvollzug und keine Selektion von Hochrisikoprobanden. Zwar erscheinen die hier untersuchten und z.T. entwickelten Methoden grundsätzlich auch für Personen mit schweren Gewalttaten anwendbar, im Gegenteil besteht aus theoretischen und empirischen Erwägungen Anlass zur Erwartung, dass valide Prognosen bei dieser Zielgruppe mit schweren Anlasstaten vielleicht sogar einfacher und treffsicherer sein könnten als bei Personen mit mittelschwerer Delinquenz (vgl. Dahle, 2005). Hinreichende empirische Befunde hierzu und zur Frage der Größenordnung und Verteilung von Fehlbeurteilungen stehen indessen noch aus.

Andrews, D.A. & Bonta, J. (1995). *LSI-R: The Level of Service Inventory-Revised*. Toronto: Multi-Health Systems.

Dahle, K.-P. (1997). Kriminalprognosen Im Strafrecht – Psychologische Aspekte individueller Verhaltensvorhersagen. In M. Steller & R. Volbert (Hg.), *Psychologie im Strafverfahren* (S. 118-139). Bern: Huber.

- Dahle, K.-P. (2000). Psychologische Begutachtung zur Kriminalprognose. In H.-L. Kröber & M. Steller (Hrsg.), *Psychologische Diagnostik im Strafverfahren: Indikation, Methodik und Qualitätsstandards* (S. 77-111). Darmstadt: Steinkopff-Verlag. (neue und überarbeitete Auflage 2005)
- Hare, R.D. (1991). *The Hare Psychopathy Checklist – Revised: Manual*. Toronto, CA: Multi-Health Systems.
- Webster, C. D., Douglas, K. S., Eaves, D. & Hart, S. D. (1997). *HCR-20: Assessing risk of violence* (Version 2). Vancouver: Mental Health Law & Policy Institute, Simon Fraser University.

Eine vollständige Beschreibung der Studie findet sich in:

- Dahle, K.-P. (2005) *Psychologische Kriminalprognose. Wege zu einer integrativen Methodik für die Beurteilung der Rückfallwahrscheinlichkeit bei Strafgefangenen*. Herbolzheim: Centaurus Verlag.