

Aus der Universitätsklinik und Poliklinik für Psychiatrie,
Psychotherapie und Psychosomatik
der Martin-Luther-Universität Halle-Wittenberg
(Direktor: Prof. Dr. med. Dr. h. c. mult. Andreas Marneros)

**Kriminalprognostische Validität der
Psychopathy-Checklist: Screening Version.
Eine kritische Untersuchung.**

Dissertation

Zur Erlangung des akademischen Grades
Doktor rerum medicarum (Dr. rer. medic.) für das Fachgebiet
Medizinische Psychologie

vorgelegt der Medizinischen Fakultät
der Martin-Luther-Universität Halle-Wittenberg

Renate Reichel

Gutachter:

- 1 Prof. Dr. med. Dr. h. c. mult. Andreas Marneros
- 2 Prof. Dr. phil. Max Steller

Verteidigt am 21.10.2009

Kurzreferat zur Arbeit: *Kriminalprognostische Validität der PCL:SV*

Fragestellung: Das primäre Ziel der vorliegenden Arbeit war es, die prognostische Validität der Psychopathy-Checklist: Screening Version (PCL:SV) zur Vorhersage von sowohl genereller als auch gewalttätiger Rückfalldelinquenz an einer heterogenen deutschen Straftäterpopulation zu ermitteln und mit den Befunden der internationalen Literatur zu vergleichen. Zusätzlich sollten weitere, in der einschlägigen Literatur diskutierte Risikofaktoren erfasst und ebenfalls hinsichtlich ihres Zusammenhangs mit kriminellem Rückfall analysiert werden.

Methode: Eingeschlossen in die Auswertungen wurden 244 Straftäter aus deutschen Haftanstalten. Die prognostische Validität der PCL:SV sowie ihrer Subskalen wurde, getrennt für generelle und gewalttätige Rückfalldelinquenz, mit Hilfe von ROC-Analysen (und aus Vergleichsgründen mit weiteren statistisch äquivalenten Verfahren) ermittelt und statistisch evaluiert. Schließlich wurde die Wirkungsweise der PCL:SV für konkrete (basisratenabhängige) Entscheidungsregeln eingehend untersucht.

Befunde: Die Ergebnisse der vorliegenden Untersuchung stehen im Einklang mit einem Großteil der Studien aus dem angloamerikanischen Sprachraum und der Schweiz. Abgesehen vom PCL:SV-Faktor-1-Score bei der Vorhersage gewalttätiger Rückfalldelinquenz weichen die untersuchten AUC-Werte für den PCL:SV-Summenscore sowie für seine Subskalen zur Vorhersage von sowohl genereller als auch gewalttätiger Rückfalldelinquenz signifikant von 0.5 ab und liegen alle in einem Bereich von 0.61 bis 0.70 (analoge Resultate liefern die verwendeten äquivalenten Verfahren).

Schlussfolgerungen: 1) Die Analyse von konkreten Entscheidungsregeln macht in der untersuchten Straftätergruppe deutlich, dass trotz vorliegender *statistischer Bedeutsamkeit*, die *praktische Bedeutsamkeit* der untersuchten AUC-Werte nicht befriedigen kann, da durchweg der Zugewinn an Trefferwahrscheinlichkeit im Vergleich zu zugehörigen Zufallsverfahren gering ist. 2) Die verschiedentlich hervorgehobene Fähigkeit des PCL:SV-Summenscores, eine hoch rückfallgefährdete Risikogruppe zu erfassen, erweist sich bei formaler Betrachtung als eine Eigenschaft, die *jedes* brauchbare Prognoseinstrument notwendigerweise besitzen muss. 3) Wesentlich einfacher feststellbare Merkmale, nämlich *Alter bei erster Verurteilung* und *Anzahl der Vorstrafen* schneiden hinsichtlich ihrer prognostischen Güte gleich gut oder sogar leicht besser als der *PCL:SV-Summenscore* ab. 4) Ein mit den Variablen *PCL:SV-Summenscore*, *Alter bei erster Verurteilung* und *Anzahl der Vorstrafen* konstruiertes entscheidungsbaumbasiertes Modell zur Vorhersage eines kriminellen Rückfalls führt zu einer niedrigeren Fehlerrate als vergleichbare lineare Vorhersagemodelle.

Inhaltsverzeichnis

1	Theoretischer Hintergrund	1
1.1	Einleitung	1
1.2	Von der Kriminalprognose zur Risikoeinschätzung	3
1.2.1	<i>Geschichte der Prognoseforschung – ein kurzer historischer Überblick</i>	<i>3</i>
1.2.2	<i>Prognostische Vorgehensweisen.....</i>	<i>3</i>
1.2.3	<i>Prognoseinstrumente – Nutzen und Grenzen.....</i>	<i>5</i>
1.3	Die Psychopathy Checklist-Revised (PCL-R, Hare 1991).....	6
1.3.1	<i>Validierungsstudien zur PCL-R</i>	<i>7</i>
1.4	Psychopathy-Checklist: Screening Version (PCL:SV, Hart et al. 1995)	8
1.4.1	<i>Validierungsstudien zur PCL:SV</i>	<i>9</i>
1.5	Einige methodologische Bemerkungen.....	10
1.5.1	<i>Entscheidungsstafeln und ihre Charakterisierungen.....</i>	<i>10</i>
1.5.2	<i>Ein statistisches Prognosemodell.....</i>	<i>16</i>
1.5.3	<i>Einige Bemerkungen zur Basisrate</i>	<i>21</i>
2	Fragestellung	24
3	Material und Methoden.....	25
3.1	Terminologisches	25
3.2	Das Bundeszentralregister.....	26
3.3	Stichprobenbeschreibung	27
3.4	Erhebungsinstrumente.....	28
3.5	Methoden und Hypothesen	29
4	Ergebnisdarstellung.....	31
4.1	Deskriptive Analysen.....	31
4.1.1	<i>Charakterisierung der Gesamtstichprobe.....</i>	<i>31</i>
4.1.2	<i>Charakterisierung der Rückfalltäter und Nichtrückfalltäter</i>	<i>31</i>
4.1.3	<i>Deliktverteilungen der Index- und Rückfalldelikte</i>	<i>31</i>
4.2	Statistische Analysen zu den ausgewählten Einzelmerkmalen	33
4.2.1	<i>Relevante Merkmale der strafrechtlichen Vorgeschichte</i>	<i>34</i>
4.2.2	<i>Relevante Merkmale der Soziobiographie</i>	<i>35</i>
4.2.3	<i>Psychopathologische Auffälligkeiten zum Zeitpunkt des Indexdeliktes</i>	<i>36</i>
4.2.4	<i>Zusammenhänge der ausgewählten Einzelmerkmale mit kriminellem Rückfall</i>	<i>38</i>
4.3	Zentrale Analysen: Die PCL:SV	40
4.3.1	<i>Deskriptive Analysen zur PCL:SV</i>	<i>40</i>
4.3.2	<i>Statistische Analysen zur PCL:SV.....</i>	<i>41</i>

4.3.3	<i>Korrelationen der PCL:SV mit generellem und gewalttätigem Rückfall</i>	42
4.3.4	<i>Prädiktive Validität der PCL:SV</i>	42
4.3.5	<i>Prognoserelevante Kennwerte der PCL:SV-Gesamt</i>	44
4.3.6	<i>Prognoserelevante Kennwerte für die Subskalen der PCL:SV</i>	48
4.3.7	<i>Weiterführende Datenanalysen</i>	49
5	Diskussion	51
5.1	Prädiktive Validität der PCL:SV.....	51
5.2	PCL:SV-Gesamt und krimineller Rückfall	53
5.3	PCL:SV-Gesamt, <i>Psychopathy</i> und die Identifikation einer Hochrisikogruppe.....	55
5.4	PCL:SV-SD bzw. PCL:SV-PP und krimineller Rückfall	57
5.5	Soziobiographische und psychopathologische Merkmale und krimineller Rückfall.....	57
5.6	Merkmale der strafrechtlichen Vorgeschichte und krimineller Rückfall.....	58
5.7	Komplexere Entscheidungsregeln.....	61
5.7.1	<i>Lineare Kombinationen der Prädiktoren</i>	61
5.7.2	<i>Nicht lineare Kombinationen der Prädiktoren</i>	62
5.8	Limitierungen.....	65
6	Zusammenfassung	67
7	Literaturverzeichnis	70
8	Thesen	77
9	Anhang	
	Tabellarischer Lebenslauf	
	Selbstständigkeitserklärung	
	Erklärung über frühere Promotionsversuche	
	Danksagung	

Verzeichnis der verwendeten Abkürzungen

<i>AUC</i>	area under curve (Fläche unter der ROC-Kurve)
<i>d</i>	Effektgröße
<i>FN</i>	Falsch-Negative
<i>FP</i>	Falsch-Positive
μ	Erwartungswert
<i>M</i>	Stichprobenmittelwert (Schätzer für Erwartungswert μ)
<i>N</i>	Gesamtstichprobengröße
<i>NPP</i>	Negative-Predictive-Power
<i>PCL:SV</i>	Psychopathy-Checklist: Screening Version
<i>PCL:SV-Gesamt</i>	PCL:SV-Summenscore
<i>PCL:SV-PP</i>	PCL:SV-Faktor-1-Score
<i>PCL:SV-SD</i>	PCL:SV-Faktor-2-Score
<i>PPP</i>	Positive-Predictive-Power
<i>ROC</i>	Receiver Operating Characteristic
<i>R₀</i>	Teilpopulation der Nicht-Rückfälligen
<i>R₁</i>	Teilpopulation der Rückfälligen
<i>r₀</i>	Prognose: "gehört zu R ₀ "
<i>r₁</i>	Prognose: "gehört zu R ₁ "
<i>RN</i>	Richtig-Negative
<i>RP</i>	Richtig-Positive
<i>S</i>	Grundgesamtheit der Straftäter
σ	Standardabweichung
<i>SD</i>	Schätzer für Standardabweichung σ
β	Regressionskoeffizient der einfachen Logistischen Regression
<i>B</i>	Schätzer des Regressionskoeffizienten β
π	Basisrate
τ	Selektionsrate
ρ_{XR}	punktbiseriale Korrelation
<i>r_{XR}</i>	Schätzer der Korrelation ρ_{XR}

1 Theoretischer Hintergrund

1.1 Einleitung

Die Frage nach den Ursachen und Bedingungen für kriminelle Rückfälle ist nicht nur von wissenschaftlichem, sondern auch von kriminalpolitischem Interesse. In der Praxis ist man daran interessiert, über gültige Kriterien und effektive Instrumente zu verfügen, mit deren Hilfe sich rückfallgefährdete Personen von jenen trennen lassen, von denen nach Haftentlassung bzw. während einer Vollzugslockerung keine Komplikationen im Sinne einer erneuten Straftat zu erwarten sind. Die Identifikation von Risikofaktoren für kriminelles Verhalten im Allgemeinen sowie für kriminelle Rückfälle im Besonderen stellt somit ein wichtiges Forschungsanliegen in den forensischen Wissenschaften dar.

Trotz einer mittlerweile vorliegenden Vielzahl von Theorien und einer Fülle von Untersuchungen ist der heutige Wissensstand im Hinblick auf die Entstehungsbedingungen der Kriminalität und ihre vielfältigen Erscheinungsformen nicht eben befriedigend. Je nach theoretischer Ausrichtung werden soziokulturelle, sozialökonomische oder psychologische Aspekte für das Auftreten kriminellen Verhaltens vorrangig verantwortlich gemacht. Bis zum heutigen Tag scheint es keine Kriminalitätstheorie zu geben, die geeignet wäre, den gesamten Bereich strafrechtlicher Verfehlungen in einer hinlänglichen Breite konzeptuell zu erfassen.

Ein ähnliches Bild zeigt sich, wenn man sich die Forschungsergebnisse auf dem Gebiet der Kriminalprognose anschaut. Insgesamt haben sich zwar eine Menge von Variablen (Prädiktoren) herausgeschält, von denen man annimmt, dass sie mit kriminellem Verhalten assoziiert sind, jedoch konnte die theoretische Frage, warum gewisse Variablen prognostisch bedeutsam sind und andere nicht, bis in die jüngste Zeit nicht überzeugend geklärt werden. Zudem wird die geringe Reliabilität vieler der in der Literatur als relevant erachteten Prädiktorvariablen beklagt. So bewegen sich Korrelationen für Einzelprädiktoren gewöhnlich in Bereichen von .06 (z. B. sozioökonomischer Status) bis .18 (z. B. strafrechtliche Vorgeschichte oder antisoziale Persönlichkeit). Bei der Kombination verschiedener Variablen zu sogenannten Risikoskalen steigen die Korrelationen zwar erwartungsgemäß an, erreichen aber auch dann kaum Werte, die .30 übersteigen (siehe dazu z. B. Gendreau et al. 1996).

Als eine Weiterentwicklung in der Prognoseforschung der letzten Jahre kann die ausdrückliche Trennung der Risikofaktoren hinsichtlich ihrer unterschiedlichen Qualität in statische und dynamische Risikofaktoren betrachtet werden (Andrews & Bonta 1998, 2003). Geht es bei den statischen Faktoren um Kriterien wie die strafrechtliche Vorgeschichte oder soziobiographische Merkmale, beziehen sich dynamische Faktoren auf potenziell veränderbare

Merkmale wie beispielsweise Einstellungen oder soziale Kompetenzen. Mit dieser Trennung wurde die Prognose auf eine insgesamt breitere und differenziertere konzeptuelle Grundlage gestellt, da der Blickwinkel unter anderem um situative Faktoren und soziale Aspekte erweitert wurde. Insbesondere wurden dadurch Perspektiven für Verlaufsbetrachtungen eröffnet und der Fokus auch auf differenzielle Veränderungen in der prognostischen Bedeutsamkeit einzelner Merkmale infolge therapeutischer Interventionen gerichtet (Dahle et al. 2007).

Eine Zusammenstellung bedeutsamer Prädiktoren, die sich über zahlreiche Studien hinweg als stabile Einflussfaktoren erwiesen haben, liefern in besonderer Weise Metaanalysen. Sie liegen inzwischen sowohl für jugendliche und erwachsene Straftäter als auch für Straftäter mit psychischen Störungen sowie für Sexualstraftäter vor. So fanden beispielsweise Gendreau et al. (1996), die im Rahmen ihrer Metaanalyse 131 Studien auswerteten, dass für erwachsene Straftäter Merkmale wie die strafrechtliche Vorgeschichte, antisoziales Verhalten, Alter, Geschlecht und ethnische Zugehörigkeit sowie Verhaltensmuster und Einstellungen, die bei einer Person inhaltlich mit strafrechtlichem Verhalten verknüpft sind, eine höhere prädiktive Gültigkeit zur Vorhersage krimineller Rückfälle haben als intellektuelle Fähigkeiten, sozioökonomischer Status oder Merkmale der Herkunftsfamilie (siehe auch Bonta et al. 1998, Cottle et al. 2001, Hanson & Bussière 1998). Zudem sprechen die Ergebnisse einiger Untersuchungen dafür, dass die jeweiligen Merkmale bzw. Merkmalsbereiche in unterschiedlichen forensischen Teilpopulationen (z. B. Gewaltstraftäter oder Sexualstraftäter) unterschiedliche prognostische Relevanz haben können und dementsprechend unterschiedlich zu gewichten sind (z. B. Kröber 2006). Darüber hinaus weist die Datenlage auf spezielle, nur in bestimmten Straftäterpopulationen gültige, Prognosefaktoren hin (z. B. Dahle 2006, Dahle et al. 2008, Egg 1999, 2001, Hanson & Bussière 1998, Stadtland & Nedopil 2005).

Zu den bis heute wohl robustesten Merkmalen zur Vorhersage eines kriminellen Rückfalls werden die sogenannten „*big four*“ gezählt (Andrews & Bonta 1998, 2003). Im Einzelnen sind das 1) die Vorgeschichte antisozialen und delinquenten Verhaltens, 2) Merkmale einer antisozialen (psychopathischen) Persönlichkeit, 3) antisoziale Kognitionen und Einstellungen und 4) ein antisoziales Umfeld (siehe dazu auch Dahle 2005, Nedopil 2005 und Rasch 1999).

Wie sich der derzeitige Forschungsstand im Hinblick auf die Ermittlung relevanter Risikofaktoren und die Verbesserung kriminalprognostischer Methodik entwickelt hat und welche Bemühungen in den letzten Jahren unternommen wurden, um die Prognosegenauigkeit zu erhöhen, soll in den folgenden Abschnitten näher erläutert werden.

1.2 Von der Kriminalprognose zur Risikoeinschätzung

1.2.1 *Geschichte der Prognoseforschung – ein kurzer historischer Überblick*

Die Geschichte der Kriminalprognose ist relativ jung. Empirische Forschungen auf diesem Gebiet sind seit den 20er Jahren des letzten Jahrhunderts zu verzeichnen (ausführlich dazu z. B. Dahle 2005, 2006 und Nedopil 2005). Zu dieser Zeit begannen die Bemühungen zahlreicher Wissenschaftler, Merkmalsammlungen und Prognosetafeln zu erstellen, in denen Kriterien aufgelistet wurden, die für die Vorhersage kriminellen Verhaltens von Bedeutung sind. Ein wesentliches Prinzip bei der Erstellung derartiger Sammlungen bestand darin, bedeutsame Unterschiede im Lebenslauf und der Persönlichkeit zwischen Rückfälligen und Nichtrückfälligen zu ermitteln. Interessanterweise zählten Merkmale, die bis heute nichts an ihrer prognostischen Relevanz eingebüßt haben, wie beispielsweise die kriminelle Vorgeschichte des Täters oder eine antisoziale Persönlichkeit, schon damals zu den wesentlichen Faktoren (Burgess 1928, 1968).

Der anfängliche Optimismus, mit derartigen Instrumentarien hinreichend valide Prognosen erstellen zu können, wich jedoch schon bald realistischeren Einschätzungen der komplexen Sachlage, die ihren Ausdruck in einigen kritischen Analysen zu den konzeptuellen Unzulänglichkeiten prognostischer Entscheidungen fanden (Monahan 1978, 1981, 1984, 1996, Rasch 1999, Steadman 1973, Steadman & Coccozza 1974, Steadman et al. 1994). Kritisiert wurden besonders die hohen Raten sogenannter Falsch-Positiver, also Personen, deren weiterer Lebenslauf gezeigt hatte, dass man sie fälschlicherweise als gefährlich eingeschätzt hatte. Auch eine in jüngster Zeit in Deutschland durchgeführte Untersuchung zu Häufigkeit und Gründen falscher Kriminalprognosen bei psychisch kranken Gewaltverbrechern weist mit ihren Ergebnissen in diese Richtung (Rusche 2004).

Die in den 80er und 90er Jahren des letzten Jahrhunderts verbreitete Skepsis und Kritik an den prognostischen Methoden und Vorgehensweisen (z. B. Monahan 1981, Steadman et al. 1994) führte in der Folgezeit zu intensiven Anstrengungen, die Methodik von Prognosen zu verfeinern und sich mit den zugrundeliegenden statistischen Modellen genauer als bisher auseinander zu setzen (Cohen 1992, Dahle 2005, Endres 2000, Groß 2004, Kühl & Schumann 1989, Nedopil 1998, 2005, Prenky & Burgess 2000, Urbaniok et al. 2007c, Volckart 2002).

1.2.2 *Prognostische Vorgehensweisen*

In der Literatur werden unterschiedliche Traditionen des prognostischen Vorgehens diskutiert. Bei der eher intuitiv orientierten Vorgehensweise wird eine Prognose – so die Vorstellung –

aufgrund von langjähriger beruflicher Erfahrung und Fachwissen abgeben, ohne dass dabei theoretische Fundierungen oder empirische Befunde maßgeblich in die prognostische Urteilsbildung einfließen müssen. Das eher klinisch orientierte Vorgehen ist gekennzeichnet durch eine umfassende Anamneseerhebung der Krankheits- und Delinquenzgeschichte der zu begutachtenden Person. Hier steht eine sorgfältige Erfassung des Einzelfalls im Vordergrund. Auf der Grundlage der so gewonnenen Informationen wird dann versucht, zukünftiges Verhalten zu prognostizieren (Nedopil 2005). Intuitives und klinisches Vorgehen wurden in der Vergangenheit zunehmend kritisch hinterfragt (Monahan 1981, Rasch 1984, Steadman 1973, Stone 1985). Als hauptsächliche Kritikpunkte wurden mangelnde Objektivität und geringe Transparenz genannt (Bonta 1996, Monahan 1981, Webster et al. 1994). In den letzten Jahren ist daher vermehrt die statistische (aktuarische) Methode in den Fokus des wissenschaftlichen – aber auch praktischen – Interesses gerückt. Im Rahmen dieser Methodik bedient man sich solcher Merkmale, die im statistischen Sinn bedeutsam mit Rückfälligkeit korrelieren. Im besten Fall stammen derartige Merkmale aus multivariaten Analysen prospektiver Längsschnittstudien, wobei davon ausgegangen wird, dass diese Merkmale geeignet sind, für vergleichbare Personengruppen auch zukünftige Rückfälle vorherzusagen. Die rein statistische Prognose beruht auf einer vollständig regelgeleiteten Einschätzung der Person (Meehl 1996). Die Treffsicherheit von Risikoeinschätzungen ließ sich, wie verschiedene Studien belegen, durch den Gebrauch statistischer Methoden wesentlich verbessern (z. B. Meehl 1996, Monahan 1981, 1996). Sie gilt nach Ansicht einiger Autoren deshalb der klinischen (und intuitiven) Prognose als überlegen (Dawes et al. 1989, Grove & Meehl 1996, Monahan 1981). So fanden beispielsweise Bonta et al. (1998) für die Vorhersage eines gewalttätigen Rückfalls bei psychisch kranken Rechtsbrechern, dass klinische Einschätzungen im Vergleich zu statistischen Vorhersagemethoden eine geringere Korrelation mit Rückfall aufwiesen ($r = .09$ für klinische Einschätzungen und $r = .30$ für statistische Einschätzungen). Allerdings muss an dieser Stelle kritisch bemerkt werden, dass es bis heute keine einzige empirische Studie zu geben scheint, die bei derselben Stichprobe statistische Vorhersagen und Vorhersagen aufgrund klinischer Urteilsbildung einander gegenübergestellt und in ihrer Prognosegenauigkeit miteinander verglichen hat (Endres 2002, siehe auch Leygraf 2004 und Leygraf & Nowara 1992).

Da mit Hilfe der klinischen – im Vergleich zur statistischen – Methode den jeweiligen Besonderheiten des Einzelfalls Rechnung getragen werden kann, wird heute meist eine Kombination aus klinischer und statistischer Vorgehensweise bei der Prognoseerstellung bevorzugt (Douglas et al. 2003, Lewis & Webster 2004).

1.2.3 Prognoseinstrumente – Nutzen und Grenzen

Um die Validität und Reliabilität von Prognosen zu erhöhen und damit in der Folge den Anteil der Falsch-Positiven und Falsch-Negativen zu verringern, wurden umfangreiche Bemühungen zur weiteren Verbesserung der Methodik der Prognosestellung unternommen (z. B. Dahle 2005, Dünkel & Geng 2003, Kröber 2006, Nedopil 2005). So wurden neben bereits beschriebenen klassischen prognostischen Vorgehensweisen in wachsendem Maße auch systematisch konstruierte Instrumente als geeignete Hilfsmittel herangezogen und für prognostische Zwecke nutzbar gemacht.

Vorteile, die bei standardisierten Prognoseinstrumenten auf der Hand liegen, sind deren Transparenz, Objektivität und Ökonomie. Darüber hinaus lassen sich testtheoretische Gütekriterien wie Reliabilität und Validität einerseits sowie – nach Festlegung eines *cut-off* Wertes – Sensitivität, Spezifität, *positive* und *negative predictive power* (*PPP* und *NPP*) andererseits ermitteln.

Viele der momentan gebräuchlichsten Instrumente wurden zur Vorhersage ganz bestimmter Rückfallereignisse oder für spezielle Risikogruppen entwickelt (Hanson & Thornton 1999, Harris et al. 1993, Quinsey et al. 1995, Webster et al. 1997). Einige, wie etwa das Historical-Clinical-Risk-Management-20-Item-Schema (HCR-20, Webster et al. 1997) sind inzwischen auch an den deutschen Sprachraum adaptiert worden (Müller-Isberner et al. 1998). Daneben finden sich Instrumente speziell für psychisch gestörte Rechtsbrecher oder jugendliche Straftäter (zusammenfassend siehe dazu beispielsweise Dahle et al. 2007).

Die meisten der bisher entwickelten Prognoseinstrumente zeichnen sich durch ein Übergewicht an statischen und eine Vernachlässigung von dynamischen Variablen aus, deren Bedeutsamkeit in der kriminellen Prognostik sich in den letzten Jahren allerdings zunehmend herauskristallisiert hat (Andrews & Bonta 1998, 2003).

Im Laufe der 90er Jahre wurde deshalb vor allem in Nordamerika eine neue Generation von Instrumenten entwickelt, die neben statischen auch dynamische und damit potenziell veränderbare Faktoren in die Prognosebeurteilung einbezogen. Stellvertretend und beispielhaft sei hier das Level-of-Service-Inventory-Revised (LSI-R) von Andrews & Bonta (1995) genannt.

Der überwiegende Teil der bislang entwickelten Verfahren wird in Bezug auf ihre prognostische Validität in vielen Studien als hinreichend gut eingeschätzt (Andrews & Bonta 1995, Claix & Pham 2004, Hanson & Thornton 1999, Harris et al. 2004, Harris et al. 1993, Webster et al. 1997). Eine Vielzahl der Validierungsstudien stammt dabei aus dem angloamerikanischen und skandinavischen Sprachraum.

Allerdings gibt es auch andere Stimmen, die weniger optimistisch klingen. So weisen Dahle (2004), Nedopil und Stadtland (2005) sowie Rusche (2004) in ihren Untersuchungen darauf hin, dass viele Personen mit auffälligen Werten in den entsprechenden Prognoseinstrumenten nicht rückfällig wurden. Die Rückfallrate der sogenannten „High-Scorer“ lag hierbei im Mittel nie über 50 %, was – wie die Autoren diskutieren – darauf hindeutet, dass die Rate der Falsch-Positiven in der Gruppe der „High-Scorer“ sehr hoch ist (siehe auch Freedmann 2001 sowie Ross & Pfäfflin 2005). Auch der Anteil der Falsch-Negativen liegt nicht selten in vergleichbaren Größenordnungen, wobei kritisch anzumerken ist, dass in der Literatur nicht immer ganz klar wird, was die betrachtete Bezugsgröße ist (Dietiker et al. 2007). Auch wenn die gegenwärtige Forschungslage auf dem Gebiet der Kriminalprognose als nicht übermäßig optimistisch eingeschätzt werden kann, besteht doch bei den meisten Wissenschaftlern ein solider Konsens dahingehend, dass sich die Genauigkeit der Vorhersagen durch den Einsatz von (statistischen) Prognoseinstrumenten nachweislich verbessern lässt (Endres 2000, Webster 1997).

1.3 Die Psychopathy Checklist-Revised (PCL-R, Hare 1991)

Ein Instrument, das in der Kriminalprognose in den letzten Jahren immer mehr an Bedeutung gewonnen hat, ist zweifelsohne die PCL-R (Hare 1991). Obwohl sie ursprünglich nicht als ein Prognoseinstrument konzipiert wurde, sondern zur Erfassung des Persönlichkeitskonstruktes *Psychopathy*, hat sie sich besonders im englischen Sprachraum zu einem häufig eingesetzten und anerkannten Prognoseinstrument entwickelt (z. B. Hare et al. 2000, Laurell & Daderman 2005, Salekin et al. 1996, Shine & Hobson 1997).

Nach Hare (1991) wird *Psychopathy* als ein zweifaktorielles Persönlichkeitskonstrukt aufgefasst, das sich aus den Faktoren *psychopathic traits* (Faktor 1) und *social deviance* (Faktor 2) zusammensetzen soll. Ersterer soll Kernmerkmale der psychopathischen Persönlichkeit, letzterer Merkmale antisozialen Verhaltens beschreiben. Die Kriterien des ersten Faktors basieren dabei auf der klinischen Beurteilung und Einschätzung des Interviewers. Die Kriterien des zweiten Faktors spiegeln im Großen und Ganzen sozial deviante Verhaltensweisen wider. Die PCL-R beinhaltet sowohl statische als auch dynamische Risikofaktoren.

Neben dem ursprünglichen Zwei-Faktoren-Modell (Hare 1991) legen neuere Untersuchungen auch andere Faktorenlösungen nahe. So werden auch Drei- (Cooke & Michie 2001) und Vier-Faktoren-Modelle (Hare 2003, Neuman et al. 2007) der *Psychopathy* vorgeschlagen. Im Rahmen der vorliegenden Untersuchung wurde dem klassischen Zwei-Faktoren-Modell nach Hare

der Vorzug gegeben, unter anderem deshalb, weil sich ein Großteil der empirischen Arbeiten ebenfalls auf das ursprüngliche Zwei-Faktoren-Modell bezieht.

Obwohl nicht alle *Psychopaths* zwingend in Konflikt mit dem Rechtssystem geraten (Babiak 1995, Hare 1998a), scheinen die dem Konzept zugrundeliegenden Merkmale mit einem hohen Risiko für kriminelles bzw. aggressives und gewalttätiges Verhalten assoziiert zu sein (Hare 1998ab, 1999, Hart & Hare 1997).

So ist es nicht verwunderlich, dass *Psychopathy*, im Sinne Hares (1991), als ein wesentlicher Prädiktor kriminellen Handelns, insbesondere von rezidivierender Gewaltkriminalität betrachtet wird (Grann et al. 1999, Hare 1999, 2001, Hare et al. 2000, Harris et al. 1991, Harris et al. 1993, Hart et al. 1994, Hemphill et al. 1998a, Rice & Harris 1995, Salekin et al. 1996, Serin 1991, 1996, Serin & Amos 1995, Ullrich et al. 2003). Beispielsweise fanden Hemphill et al. (1998a) in ihrer Metaanalyse, dass *Psychopaths* im Vergleich zu Straftätern ohne diese Diagnose ein Jahr nach Entlassung aus dem Gefängnis dreimal häufiger wiederverurteilt und viermal häufiger mit einem Gewaltdelikt rückfällig wurden. Die Korrelation zwischen PCL-R und Gewalttätigkeit soll darüber hinaus als zeitlich stabil gelten und weitgehend unabhängig von der Länge des Beobachtungszeitraumes sein (Hare et al. 2000, Quinsey et al. 1995). Im Besonderen sollen die Merkmale des Faktors 2 *social deviance* relativ hoch mit Rückfälligkeit korrelieren (Grann et al. 1999, Hemphill et al. 1998b, Leistico et al. 2008, Salekin et al. 1996, Walters 2003). Zu gegenteiligen Ergebnissen kommen Heilbrun et al. (1998), Serin (1996) sowie Wilson (2003). Sie bescheinigen den Merkmalen des Faktors 1 eine höhere prädiktive Valenz, speziell in Bezug auf die Vorhersage gewalttätiger Rückfälle, als dem Faktor 2. Des Weiteren findet sich ein enger Zusammenhang sowohl mit Bewährungsversagen als auch mit schlechtem Behandlungserfolg (Hemphill et al. 1998a, Ogloff et al. 1990, Rice & Harris 1992, Serin 1996).

Der PCL-R werden gute psychometrische Eigenschaften bescheinigt, die über Kulturen hinweg stabil sein sollen (Hare 1985, Shine & Hobson 1997). So fand man in entsprechenden Untersuchungen eine adäquate interne Konsistenz (Alpha-Koeffizient .87) eine hohe Interrater-Reliabilität ($r = .91$) sowie eine hohe Test-Retest-Reliabilität ($r = .94$) (Cooke & Michie 1997, Hare et al. 1990, Harris et al. 1994).

1.3.1 Validierungsstudien zur PCL-R

Eine Vielzahl von Validierungsstudien konnte inzwischen zwar eine stabile, aber doch eher moderat hohe prognostische Validität der PCL-R in forensischen Stichproben belegen (z. B.

Hare 1998a, 1999, Hemphill et al. 1998a, Salekin et al. 1996). Die meisten Untersuchungen stammen dabei aus den USA und Kanada. Aber auch Studien aus England, Schweden, Dänemark, Norwegen, Belgien, Spanien und Portugal konnten eine akzeptable prädiktive Validität der PCL-R aufzeigen (zusammenfassend siehe dazu Hare et al. 2000). Seit jüngster Zeit liegen auch Validierungsstudien aus Deutschland vor (z. B. Dahle 2006, Nygaard 2005, Stadtland et al. 2005).

Als Validitätsmaß wird neben Korrelationen in neueren Untersuchungen auch die sogenannte AUC erwähnt.¹ So berichten beispielsweise Grann et al. (1999), die bei gewalttätigen Straftätern mit Persönlichkeitsstörungen den Zusammenhang zwischen PCL-R-Summscore und einschlägigem Rückfall untersucht haben, eine geschätzte AUC von .72. Zu ähnlichen Ergebnissen kommen auch Studien aus dem deutschen Sprachraum, die, je nach Vorhersageereignis (z. B. generelle bzw. gewalttätige Rückfalldelinquenz), geschätzte AUCs von .61 bis .72 berichten (Dahle 2006, Stadtland et al. 2005, Urbaniok et al. 2007b).² Darüber hinaus liefern verschiedene Metaanalysen mittlere Korrelationen von .28 (Gendreau et al. 1996) bis .29 (Hemphill et al. 1998a) zwischen PCL-R und unterschiedlichen Rückfallereignissen.

1.4 Psychopathy-Checklist: Screening Version (PCL:SV, Hart et al. 1995)

Die PCL:SV ist eine gekürzte und vereinfachte Version der PCL-R und dient der Erfassung von *Psychopathy* sowohl bei forensischen als auch bei nicht kriminellen und allgemeinpsychiatrischen Patienten. Sie besteht aus 12 Items, wobei je sechs Items Faktor 1 (*psychopathic traits*) bzw. Faktor 2 (*social deviance*) erfassen sollen (siehe Tab. 1). Die Probanden werden – wie bei der PCL-R – mit Hilfe dreifach abgestufter Items (0 = Merkmal nicht vorhanden, 1 = fraglich, 2 = Merkmal sicher vorhanden) eingeschätzt. Der Summscore variiert also zwischen 0 und 24. Zur Diagnose einer *Psychopathy* schlagen die Autoren auf der Grundlage ihrer Untersuchungen einen *cut-off* Wert von 18 Punkten vor. Probanden mit Summscores von größer oder gleich 18 werden in krimineller Hinsicht zur sogenannten Hochrisikogruppe gezählt. Bei ihnen wird eine erhöhte Rückfallgefahr angenommen (Andrews & Bonta 1998, 2003, Dahle 2005, Douglas et al. 1997, Nedopil 2005, Urbaniok et al. 2007a).

¹ Das Akronym AUC steht für *area under curve*. Dieser Wert gibt die Fläche unter einer ROC-Kurve (Receiver Operating Characteristic-Kurve) an und liegt in der Regel zwischen 0.5 und 1 (vgl. dazu die Ausführungen in Abschnitt 1.5). Die Bewertung der Höhe der AUC-Werte erfolgt in der Literatur häufig mit Hinweis auf eine Klassifizierung, die auf Cohen (1988) zurückgeht. Danach werden AUC-Werte von mindestens .72 als „hoch“, Werte im Intervall von .64 bis .72 als „moderat“ bezeichnet (vgl. aber dazu Kapitel 5).

² Für generelle und gewalttätige Rückfalldelinquenz sei im Folgenden – in Anlehnung an die Sprechweise einschlägiger Publikationen – die Abkürzung „genereller“ bzw. „gewalttätiger Rückfall“ verwendet.

Tabelle 1: Die 12 Items der PCL:SV (Hart et al. 1995, deutsch von Freese 1999).

Faktor 1 „ <i>psychopathic traits</i> “	Faktor 2 „ <i>social deviance</i> “
1 oberflächlich	7 impulsiv
2 grandios, großspurig	8 schlechte Verhaltenssteuerung
3 betrügerisch-manipulativ	9 fehlende Lebensziele
4 Fehlen von Reue, Gewissensbissen	10 verantwortungslos
5 Fehlen von Empathie	11 antisoziales Verhalten in der Adoleszenz
6 übernimmt keine Verantwortung	12 antisoziales Verhalten im Erwachsenenalter

Im Vergleich zur PCL-R besteht ein wesentlicher Vorteil der PCL:SV offensichtlich in ihrem geringeren Zeitaufwand bei der Durchführung, da sie, neben der geringeren Itemanzahl, nicht zwingend eine aufwendige Einbindung aktenbasierter Informationen erfordert. Gleichwohl werden der PCL:SV vergleichbare psychometrische Eigenschaften wie der PCL-R bescheinigt (z. B. Cooke et al. 1999, Hart et al. 1995). Darüber hinaus korrelieren in Abhängigkeit der betrachteten Populationen PCL-R und PCL:SV zwischen .51 und .85 miteinander (Hart et al. 1995). Die PCL:SV wurde nicht nur für männliche, sondern auch für weibliche Straftäter validiert (Hart et al. 1995).

Wenn auch die Struktur der *Psychopathy*, wie bereits weiter oben erwähnt wurde, als kulturunabhängig gelten soll (Cooke et al. 2001), zeichnen sich hinsichtlich der Prävalenz deutliche kulturelle Unterschiede ab (Cooke 1998). Die Prävalenzraten in kriminellen Populationen variieren von 3 % bis 30 %, wobei sie in forensisch-psychiatrischen Populationen vergleichsweise geringer ausfallen als in anderen Straftäterpopulationen. Insgesamt erreichen die Prävalenzraten in nordamerikanischen Populationen, verglichen mit europäischen, durchschnittlich höhere Werte (Hart et al. 1995). Beispielsweise berichtet Cooke (1998) von etwa 3 % schottischer Strafgefangener mit dieser Diagnose im Vergleich zu 23 % nordamerikanischer Straftäter. Bei einer in Deutschland an inhaftierten Straftätern durchgeführten Untersuchung fiel der Anteil an diagnostizierten *Psychopaths* mit 7 % – im Vergleich zu den Befunden aus Nordamerika – ebenfalls vergleichsweise gering aus (Ullrich et al. 2003).

1.4.1 Validierungsstudien zur PCL:SV

Die meisten der bisherigen Studien, die *Psychopathy* im Zusammenhang mit gewalttätigem Verhalten oder kriminellen Rückfällen untersucht haben, haben diese Persönlichkeitsstörung mit Hilfe der PCL-R erfasst (Serin & Brown 2000). Die Frage, ob sich auch die PCL:SV zur Vorhersage von kriminellem Verhalten eignet, wurde erst in den letzten Jahren eingehender erforscht. Die überwiegende Zahl der Studien stammt – wie schon bei der PCL-R – aus dem

englischsprachigen Raum und konzentriert sich dabei hauptsächlich auf die Vorhersage gewalttätigen Verhaltens. So bescheinigen beispielsweise Hill et al. (1996) der PCL:SV eine angemessene prädiktive Validität bei der Vorhersage institutioneller Gewalt. Douglas et al. (1999) berichten, dass der PCL:SV-Gesamt mit gewalttätigem kriminellen Verhalten assoziiert ist (geschätzte AUC = .79). Auch die Ergebnisse neuerer Studien legen nahe, dass die PCL:SV ein angemessener Prädiktor für die Vorhersage gewalttätigen Verhaltens ist (Skeem & Mulvey 2001). Die Autoren fanden unter Benutzung von Daten aus der MacArthur Risk Assessment Study (Monahan et al. 2001) für die Vorhersage gewalttätigen Verhaltens eine geschätzte AUC von .73.

Im Hinblick auf die Frage, ob sich die PCL:SV als Prognoseinstrument zur Vorhersage sowohl genereller als auch gewalttätiger krimineller Rückfälle bei deutschen Straftätern eignet, ist die Forschungslage noch rar. So gibt es anscheinend bislang nur eine Studie aus der Schweiz, in der die Autoren eine geschätzte AUC von .69 hinsichtlich des PCL:SV Summenscores und generellem Rückfall berichten (Urbaniok et al. 2007a). Ebenfalls aus der Schweiz stammt eine zweite Untersuchung, die allerdings die prädiktive Validität der PCL:SV ausschließlich in Bezug auf die Vorhersage sexueller Rückfälle untersucht hat, was die Generalisierbarkeit der Befunde deutlich einschränkt (Dietiker et al. 2007).

1.5 Einige methodologische Bemerkungen

Bereits im Abschnitt 1.2 wurde darauf hingewiesen, dass Prognoseinstrumente durch Kennwerte wie *area under curve (AUC)*, *Sensitivität*, *Spezifität*, *positive predictive power (PPP)*, *negative predictive power (NPP)* sowie durch die *Fehler- und Trefferraten* charakterisiert werden können. Die genannten Kennwerte und Beziehungen zu anderen Statistiken sollen in diesem Abschnitt wegen ihrer zentralen Bedeutung in der aktuarischen Rückfallprognose nun eingeführt und näher erläutert werden, nicht zuletzt auch deswegen, weil sie in der Literatur nicht immer in der wünschenswerten Deutlichkeit diskutiert werden.

1.5.1 Entscheidungstabellen und ihre Charakterisierungen

Prognostische Entscheidungen lassen sich übersichtlich mit Hilfe von Vier-Felder-Tabellen (Kontingenztabellen) darstellen, die hier als *Entscheidungstabellen* bezeichnet werden. Die Begriffe, die in diesem Abschnitt diskutiert werden, stellen Methoden zur Verfügung, mit denen Rückfallprognosen evaluiert werden können. Hierbei spielt es keine Rolle, auf welche Weise die Entscheidungen/Prognosen zustande gekommen sind, ob sich der Prognostiker in seinem

Vorgehen einer eher idiographischen oder nomothetischen Methodik verpflichtet fühlt. Wichtig in diesem Abschnitt ist lediglich, dass mit Hilfe eines *Entscheidungsverfahrens*, das zur Bewertung ansteht, für eine größere Menge von Personen Prognosen getroffen worden sind, die in einer Entscheidungstafel übersichtlich zusammengefasst werden können. Eine solche (empirische) Entscheidungstafel wird dann aufgefasst als eine Schätzung jener (theoretischen) Tafel, die man erhielte, würde man für alle Personen aus der betrachteten Grundgesamtheit eine Rückfallprognose vornehmen.

Zur Vorhersage krimineller Rückfälle betrachtet man – auf theoretischer Ebene – dabei die Grundgesamtheit aller Straftäter, zerlegt in die beiden Teilpopulationen der Nicht-Rückfälligen und der Rückfälligen.³ Die Grundgesamtheit der Straftäter soll im Folgenden mit S , die Teilpopulation der Nicht-Rückfälligen mit R_0 und die der Rückfälligen mit R_1 bezeichnet werden. Für die Prognosen „gehört zu R_0 “ und „gehört zu R_1 “ sollen die Abkürzungen r_0 und r_1 verwendet werden.⁴

Will man nun für eine Person aus der Grundgesamtheit aller Straftäter mit Hilfe eines Entscheidungsverfahrens vorhersagen, ob sie rückfällig wird oder nicht, können vier verschiedene Ereignisse eintreten. Eine Person kann

1. richtigerweise als zu R_0 gehörend (richtig negativ: RN)
2. fälschlicherweise als zu R_0 gehörend (falsch negativ: FN)
3. richtigerweise als zu R_1 gehörend (richtig positiv: RP)
4. fälschlicherweise als zu R_1 gehörend (falsch positiv: FP)

prognostiziert werden.

Die Entscheidungstafel in Tab. 2 zeigt die Wahrscheinlichkeiten der Prognoseereignisse, d. h. die Anteile der Ereignisse RN , FN , RP und FP an allen getroffenen Entscheidungen. Die Summe in der zweiten Spalte gibt dabei den Anteil der Rückfälligen an der Gesamtgruppe S an und wird auch als *Basisrate* (sie soll im Folgenden mit π abgekürzt werden) bezeichnet. Die Summe der Wahrscheinlichkeiten in der ersten Spalte, also der Anteil der Nicht-Rückfälligen an der Gesamtgruppe S , ergibt folglich $1 - \pi$. Summiert man die Wahrscheinlichkeiten in der zweiten Zeile, erhält man die *Selektionsrate* (sie soll im Folgenden mit τ be-

³ Es sei an dieser Stelle angemerkt, dass „Rückfälligkeit“ empirisch nur feststellbar für jene Straftäter ist, die – als ungefährlich eingestuft – in die Freiheit entlassen worden sind. Der Anteil der Rückfälligen (bzw. Nicht-Rückfälligen) an den Nicht-Entlassenen ist offensichtlich eine fiktive Größe.

⁴ Im Rahmen dieses Abschnitts ist es aus Gründen der Übersichtlichkeit in den Tabellen und Formeln notwendig, einige Kürzel einzuführen. Diese werden an entsprechender Stelle im Text bzw. im Verzeichnis der verwendeten Abkürzungen am Anfang der Arbeit erläutert.

zeichnet werden); die Summe der ersten Zeile lässt sich dann als $1 - \tau$ schreiben. Die Summe der Wahrscheinlichkeiten aller vier Zellen ergibt Eins.

Tabelle 2: Entscheidungstafel mit Wahrscheinlichkeiten der Ereignisse RN , FN , RP und FP ; τ steht für die Selektions- und π für die Basisrate. Weitere Erläuterungen im Text.

		Realität		
		R_0	R_1	
Prognose	r_0	$P(RN)$	$P(FN)$	$1 - \tau$
	r_1	$P(FP)$	$P(RP)$	τ
		$1 - \pi$	π	

Fehler- und Trefferquote. Ein gutes Entscheidungsverfahren sollte sich nun durch möglichst wenige Fehlprognosen, bzw. durch möglichst viele Treffer auszeichnen. Die Fehlerquote (bzw. Trefferquote) ist daher eine wichtige Größe zur Evaluation eines solchen Verfahrens. Fehlprognosen treten genau dann ein, wenn das Ereignis FN oder FP eintritt. Die Wahrscheinlichkeit von Fehlprognosen insgesamt ist somit $P(\text{Fehler}) = P(FN) + P(FP)$. Ganz analog ergibt sich für die Trefferwahrscheinlichkeit $P(\text{Treffer}) = P(RN) + P(RP)$, denn ein Treffer tritt genau dann ein, wenn das Ereignis RN oder RP eintritt. In manchen Fällen wird man bei der Bewertung eines Entscheidungsverfahrens nicht unbedingt die Wahrscheinlichkeit des Gesamtfehlers, sondern beispielsweise eine gewichtete Summe der beiden Fehlerwahrscheinlichkeiten $P(FN)$ und $P(FP)$ vorziehen, da die „Kosten“ der beiden Fehlprognosetypen FN und FP als verschieden hoch angesehen werden. Wie groß die Wahrscheinlichkeit des gesamten Prognosefehlers höchstens sein sollte, damit man von einem brauchbaren Entscheidungsverfahren sprechen kann, wird im Abschnitt φ^2 -Koeffizient und Korrelation erörtert.

Spezifität und Sensitivität. Betrachtet man nun nicht die Grundgesamtheit aller Straftäter, sondern lediglich die Teilgruppe der Nicht-Rückfälligen (R_0), dann bezeichnet man dort den Anteil der *Richtig-Negativen*, also die (bedingte) Wahrscheinlichkeit $P(RN) / [P(RN) + P(FP)]$, als die *Spezifität*; der Anteil der *Falsch-Positiven* in dieser Teilgruppe bestimmt sich dann durch $P(FP) / [P(RN) + P(FP)]$ und ist gerade identisch der Größe $1 - \text{Spezifität}$. Für die *Spezifität* schreibt man auch $P(r_0/R_0)$, für $1 - \text{Spezifität}$ $P(r_1/R_0)$. Man erhält diese beiden bedingten Wahrscheinlichkeiten, indem man die Wahrscheinlichkeit in der ersten Spalte der Tab. 2 durch $1 - \pi = P(RN) + P(FP)$ dividiert.

Entsprechend bestimmt man die *Sensitivität* eines Entscheidungsverfahrens. Darunter versteht man in der Teilgruppe der Rückfälligen (R_1) den Anteil der *Richtig-Positiven* $P(RP) / [P(RP) + P(FN)]$. Der Anteil der *Falsch-Negativen* in dieser Teilgruppe ist dann gleich $1 - \text{Sensitivität}$ ($= P(FN) / [P(RP) + P(FN)]$). Diese beiden bedingten Wahrscheinlichkeiten (die auch als $P(r_1/R_1)$ und $P(r_0/R_1)$ geschrieben werden) erhält man, indem man die Wahrscheinlichkeiten der zweiten Spalte der Tab. 2 durch die Basisrate $\pi = P(RP) + P(FN)$ dividiert.

Besitzt ein Entscheidungsverfahren für die Population S eine hohe *Sensitivität*, so garantiert es, dass von den Rückfälligen ein hoher Anteil auch tatsächlich die Prognose „wird rückfällig“ (r_1) erhält; besitzt es eine hohe *Spezifität*, so heißt dies, dass von den Nicht-Rückfälligen ein hoher Anteil als nicht rückfällig klassifiziert wird, also die Prognose „wird nicht rückfällig“ (r_0) erhält (vgl. linke Tafel in Tab. 3).

Tabelle 3: Verteilungen der Prognosen r_0 und r_1 in den beiden Teilpopulationen R_0 und R_1 getrennt (linke Tafel) und Verteilung von Nicht-Rückfälligkeit und Rückfälligkeit (R_0 und R_1) in beiden Prognosegruppen (r_0 und r_1) getrennt (rechte Tafel).

		Realität		Realität	
		R_0	R_1	R_0	R_1
Prognose	r_0	Spez.	1 - Sens.	r_0	NPP 1-NPP
	r_1	1- Spez.	Sens.	r_1	1-PPP PPP

Negative Predictive Power, Positive Predictive Power. Zur Charakterisierung des Begriffs der *negative predictive power* (*NPP*) betrachtet man alle Straftäter aus S , die durch das Entscheidungsverfahren als „gehört zu R_0 “ (r_0) eingestuft werden. Der Anteil der Nicht-Rückfälligen in dieser Subgruppe (gleichbedeutend: der Anteil der Richtig-Negativen) wird als *negative predictive power* (*NPP*) bezeichnet, man schreibt dafür auch $P(R_0/r_0)$; $1 - NPP$ gibt dann den Anteil der Rückfälligen (bzw. den Anteil der Falsch-Negativen) in dieser Teilgruppe an. Dieser Anteil wird auch als $P(R_1/r_0)$ notiert. Man erhält diese (bedingten) Wahrscheinlichkeiten, indem man die erste Zeile von Tab. 2 durch die Wahrscheinlichkeit $1 - \tau$ dividiert.

Betrachtet man schließlich jene Straftäter aus S , die die Prognose „gehört zu R_1 “ (r_1) erhalten, dann bezeichnet die *positive predictive power* (*PPP*) den Anteil der Rückfälligen (gleichbedeutend: den Anteil der Richtig-Positiven). Dieser Anteil wird auch als $P(R_1/r_1)$ notiert; $1 - PPP$ beschreibt dann den Anteil der Nicht-Rückfälligen (bzw. den Anteil der Falsch-

Positiven) in dieser Teilgruppe, man schreibt dafür $P(R_0/r_1)$. Diese Werte erhält man, indem man die zweite Zeile in Tab. 2 durch die Selektionsrate τ dividiert (vgl. auch die rechte Tafel in Tab. 3).

Hat ein Entscheidungsverfahren eine hohe *PPP*, so wird unter den als „wird rückfällig“ Eingestuften der Anteil der (tatsächlich) Rückfälligen hoch sein; besitzt es eine hohe *NPP*, so ist unter den als „wird nicht rückfällig“ Eingestuften der Anteil der Nicht-Rückfälligen hoch. Zusammenfassend lässt sich am Ende dieses Abschnittes festhalten, dass den vier möglichen Ereignissen *RN*, *FN*, *RP* und *FP* jeweils *drei* unterschiedliche Wahrscheinlichkeiten zugeordnet werden (jeweils eine unbedingte und zwei bedingte). Beispielsweise lassen sich dem Ereignis *RP* die Wahrscheinlichkeiten $P(RP)$, die *Sensitivität* ($= P(r_1/R_1)$) und die *PPP* ($= P(R_1/r_1)$) zuordnen.

ϕ^2 -Koeffizient und Korrelation. Ein weiteres wichtiges Maß zur Erfassung der Güte eines Entscheidungsverfahrens ist der ϕ^2 -Koeffizient der Entscheidungstafel aus Tab. 2, der sich mit Hilfe der Formel χ^2/N schätzen lässt (N = Stichprobengröße). Der ϕ^2 -Koeffizient beschreibt den Zusammenhang zwischen den beiden binären Variablen *Prognose* (*P*) mit den Ausprägungen r_0 und r_1 und *Realität* (*R*) mit den Ausprägungen R_0 und R_1 . Er kann bei Vier-Felder-Tafeln Werte zwischen Null und Eins annehmen.

Ein brauchbares Entscheidungsverfahren sollte in der betrachteten Grundgesamtheit einen möglichst hohen ϕ^2 -Koeffizienten besitzen. Besteht zwischen beiden Variablen ein vollständiger (deterministischer) Zusammenhang, ist dieser Koeffizient Eins. In diesem Fall würde ein Straftäter aus *S* genau dann die Prognose „gehört zu R_1 “ erhalten, wenn er tatsächlich rückfällig wird. Die Fehlerquote wäre dann Null und die Trefferquote Eins.⁵ Der andere Extremfall ($\phi^2 = 0$) tritt genau dann ein, wenn die beiden Variablen unabhängig voneinander sind. Dies ist genau dann der Fall, wenn der Anteil der als rückfällig Prognostizierten sowohl unter den Nicht-Rückfälligen als auch unter den Rückfälligen identisch der Selektionsrate τ ist, mit anderen Worten, wenn *Sensitivität* = *Selektionsrate* = $1 - \textit{Spezifität}$. Unabhängigkeit liegt also dann vor, wenn die Verteilung der Prognosen r_0 und r_1 in den beiden Gruppen R_0 und R_1 dieselbe ist. Ein Entscheidungsverfahren, dessen ϕ^2 -Koeffizient Null ist, ist statistisch nicht von einem Zufallsverfahren zu unterscheiden. Das wäre ein Verfahren, das Straftäter –

⁵ Diese Aussage ist nicht ganz korrekt. Der ϕ^2 -Koeffizient ist auch für den Fall Eins, wenn nicht die Fehlerquote, sondern die Trefferquote verschwindet. Dies wäre aber nur ein Hinweis darauf, dass das Entscheidungsverfahren eine „verdrehte“ Kodierung besitzt.

egal ob aus R_0 oder R_1 – den Kategorien r_0 bzw. r_1 zufällig so zuordnet, dass die Selektionsrate gerade τ ist. Die obigen Überlegungen implizieren, dass sich im Fall der Unabhängigkeit der Gesamtprognosefehler über die Formel $P(\text{Fehler}) = (1 - \pi) \tau + \pi (1 - \tau)$ bestimmen lässt. Hieraus folgt, dass die Fehlerwahrscheinlichkeit beim Vorliegen von Unabhängigkeit stets zwischen $\min(\pi, 1 - \pi)$ und $\max(\pi, 1 - \pi)$ liegen muss. Überdies ergibt sich bei Unabhängigkeit, dass die *PPP* gleich π und die *NPP* gleich $1 - \pi$ ist.

Wie hat man in der obigen Gleichung für $P(\text{Fehler})$ die Selektionsrate zu wählen, damit die Fehlerrate möglichst klein wird? Zur Beantwortung dieser Frage sind drei Fälle zu unterscheiden. Ist die Basisrate π gleich 0.5, so bleibt die Fehlerwahrscheinlichkeit bei Änderung der Selektionsrate konstant ($= \pi$); ist die Basisrate größer als 0.5, so nimmt mit wachsender Selektionsrate die Fehlerwahrscheinlichkeit linear ab und nimmt im Grenzfall $\tau = 1$ den Wert $1 - \pi$ an; ist schließlich die Basisrate kleiner als 0.5, so fällt die Fehlerrate mit fallender Selektionsrate und nimmt im Grenzfall $\tau = 0$ den Wert π an. Mit der Selektionsrate $\tau = 0$ ist die Regel „prognostiziere stets R_0 “ und mit $\tau = 1$ die Regel „prognostiziere stets R_1 “ verknüpft. Diese beiden Regeln liefern also gerade – je nachdem, ob die Basisrate kleiner oder größer 0.5 ist – die *Zufallsregel* mit der geringsten Fehlerrate. Daher sollte die Fehlerrate eines Entscheidungsverfahrens, insbesondere wenn die beiden Fehlertypen nicht als unterschiedlich schwerwiegend angesehen werden, bei einer Basisrate größer als 0.5 kleiner als $1 - \pi$ und bei einer Basisrate kleiner als 0.5 kleiner als π sein, denn ein vernünftiges Entscheidungsverfahren sollte doch wohl eine Fehlerrate besitzen, die kleiner ist als die der zugehörigen Zufallsregel.

Da für Vier-Felder-Tafeln der φ^2 -Koeffizient identisch dem zugehörigen quadrierten Korrelationskoeffizienten zwischen den eingangs erwähnten (binären) Variablen *Prognose* und *Realität* ist (er soll als ρ^2_{PR} bezeichnet werden) lässt sich der φ^2 -Koeffizient also auch über die Korrelation der beiden Variablen P und R bestimmen. Für eine angemessene Bewertung der Höhe des Korrelationskoeffizienten (bzw. des φ -Koeffizienten) sollte man wissen, dass seine maximal erreichbare Höhe vom Verhältnis der Basisrate zur Selektionsrate abhängt. Wenn beispielsweise die Korrelation ρ_{PR} positiv und die Basisrate π größer als die Selektionsrate τ ist, so kann die Korrelation (der φ -Koeffizient) nicht größer als

$$Q = \sqrt{\frac{1 - \pi}{\pi} \cdot \frac{\tau}{1 - \tau}}$$

werden. Die Korrelation ρ_{PR} liegt also stets unter einer oberen Schranke Q . Sie kann demnach überhaupt nur dann Eins werden, wenn die Basisrate identisch der Selektionsrate ist. Ist die

Selektionsrate größer als die Basisrate, so erhält man die zugehörige obere Schranke, wenn man in der obigen Formel die Rollen von Basis- und Selektionsrate vertauscht.

Es sei noch erwähnt, dass sich der φ^2 -Koeffizient (und damit auch der quadrierte Korrelationskoeffizient) direkt durch die Größen *Spezifität*, *Sensitivität*, *NPP* und *PPP* bestimmen lässt, da die folgende Beziehung gilt:

$$\varphi^2 = (1 - \text{Sens.} - \text{Spez.})(1 - \text{PPP} - \text{NPP}).$$

Wie die Formel zeigt, ist der φ^2 -Koeffizient eine von der Basisrate abhängige Größe, da ja *NPP* und *PPP* von der Basisrate abhängen.

1.5.2 Ein statistisches Prognosemodell

Die bisherigen Ausführungen sollen nun erweitert und für Entscheidungssituationen angepasst werden, in denen Entscheidungen auf der Grundlage eines Prognoseinstrumentes X (etwa der PCL:SV-Gesamt) in einer Population S von Straftätern vorgenommen werden. Hierbei soll davon ausgegangen werden, dass sich Rückfällige (R_1) und Nicht-Rückfällige (R_0) „im Mittel“ hinsichtlich ihrer Testwerte voneinander unterscheiden, dass insbesondere der Mittelwert der Rückfälligen größer ist als der der Nicht-Rückfälligen.⁶ Weiterhin soll angenommen werden, dass X in den beiden Teilpopulationen R_0 und R_1 jeweils normalverteilt ist und dass die Standardabweichungen von X in den beiden Teilpopulationen identisch sind (sie werden im Folgenden mit σ bezeichnet). Der Mittelwert der Variablen X in R_0 soll mit μ_0 und der in R_1 mit μ_1 bezeichnet werden (vgl. Abb. 1).⁷ Dieses (einfache) Modell wird im Weiteren als Standardmodell bezeichnet.

Soll nun auf der Grundlage eines vorliegenden x -Wertes für eine Person aus S die Frage beantwortet werden, ob diese der Gruppe R_0 oder R_1 angehört, so steht man vor dem Problem, dass in beiden Teilpopulationen – theoretisch – alle x -Werte auftreten können, wenn auch mit unterschiedlichen Wahrscheinlichkeiten. Die Frage nach der Gruppenzugehörigkeit kann also nicht mit Sicherheit beantwortet, sondern nur mit wahrscheinlichkeitstheoretischem Vorbehalt formuliert werden. Schaut man sich die Graphik in Abb. 1 an, so wird darüber hinaus deutlich, dass die Entscheidungssituation umso schwieriger und auch fehleranfälliger wird, je

⁶ Wenn die Mittelwerte verschieden sind, lässt sich durch eine geeignete Transformation der Variablen X stets diese Situation herstellen.

⁷ Aus Gründen der einfacheren Darstellbarkeit soll hier lediglich das Standardmodell diskutiert werden. Dieses Modell hat seine Wurzeln in dem sogenannten Signal-Entdeckungs-Modell, das in den 40er und 50er Jahren des vergangenen Jahrhunderts im Bereich der Nachrichtentechnik entstand und das einen starken Einfluss auf die Entwicklung entscheidungstheoretischer und psychophysikalischer Modelle vor allem in der Psychologie hatte (vgl. Wickens 2002).

mehr sich die beiden Verteilungen überlappen, das heißt, je kleiner der Abstand $\mu_1 - \mu_0$ oder je größer die Standardabweichung σ ist.

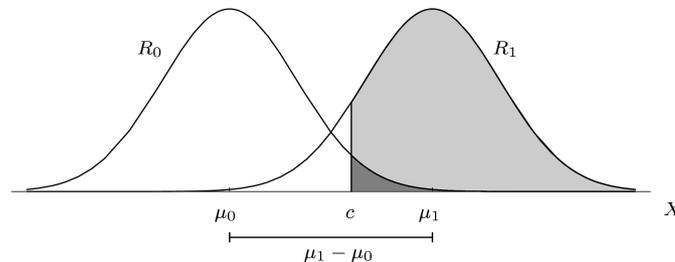


Abbildung 1: Verteilung des Prognoseinstrumentes X in den beiden Grundgesamtheiten R_0 und R_1 (weitere Erläuterungen im Text).

Da die Verteilung der x -Werte der Rückfälligen im Vergleich zu der der Nicht-Rückfälligen weiter rechts liegt, ist es naheliegend, wie folgt vorzugehen: Man bestimmt einen Wert c , den sogenannten *cut-off* Wert, so, dass Personen mit x -Werten größer oder gleich c der Gruppe R_1 (d. h. Prognose „ r_1 “) und solche mit Werten kleiner als c der Gruppe R_0 (d. h. Prognose „ r_0 “) zugeordnet werden. Die Auswahl eines *cut-off* Wertes führt also zu einer *Entscheidungsregel* („prognostiziere *wird rückfällig* gdw. $x \geq c$ “), deren Wirkungsweise in einer Grundgesamtheit wieder durch eine zugehörige Entscheidungstafel dargestellt werden kann.

Es sollen nun die wichtigsten Kenngrößen diskutiert werden, die bei der Evaluation eines Prognoseinstrumentes insgesamt (d , AUC , ρ_{XR}) bzw. bei der Evaluation von Entscheidungsregeln (Spezifität, Sensitivität, NPP , PPP , Fehler- und Trefferquote, φ -Koeffizient) eine Rolle spielen.⁸

Die wichtigen Kenngrößen im vorliegenden Modell. Trifft man – beispielsweise unter Verwendung der eben beschriebenen einfachen Entscheidungsregel – eine prognostische Entscheidung, können einem, wie bereits im Abschnitt 1.5.1 diskutiert wurde, zwei Arten von Fehlern unterlaufen: Man kann fälschlicherweise eine Person der Gruppe R_1 (Rückfällige) zuordnen, obwohl sie zur Gruppe R_0 (Nicht-Rückfällige) gehört, oder man kann eine Person fälschlicherweise der Gruppe R_0 zuordnen, obgleich sie der Gruppe R_1 angehört. Die Wahrscheinlichkeit des Fehlers erster Art ist der Anteil von Personen in R_0 (Population der Nicht-

⁸ Mit ρ_{XR} wird die Korrelation zwischen einer metrischen Prädiktorvariablen (z. B. PCL:SV-Gesamt) und einer binären Kriteriumsvariablen (hier: „Rückfall“ ja/nein) bezeichnet. Korrelationen, bei der eine Variable metrisch, die andere binär ist, nennt man auch punktbiseriale Korrelationen.

Rückfälligen), deren x -Wert größer oder gleich c (gewählter *cut-off* Wert) ist. Dieser Anteil wird in Abb. 1 durch die dunkelgrau hervorgehobene Fläche unter der R_0 -Kurve rechts von c repräsentiert und ist identisch $P(r_1/R_0) = 1 - \text{Spezifität}$ (Falsch-Positiv-Rate in der Teilpopulation R_0). Bei Personen aus dieser Gruppe tritt das prognostizierte Verhalten *nicht* ein, obwohl es vorhergesagt wird.

Die Wahrscheinlichkeit des Fehlers zweiter Art ist der Anteil von Personen in Gruppe R_1 (Population der Rückfälligen), die einen x -Wert besitzen, der kleiner als c ist. Dieser Anteil wird durch die weiße Fläche links von c unter der R_1 -Kurve dargestellt und ist identisch $P(r_0/R_1) = 1 - \text{Sensitivität}$ (Falsch-Negativ-Rate in der Teilpopulation R_1). Bei Personen aus dieser Gruppe tritt das prognostizierte Verhalten ein, obwohl es *nicht* vorhergesagt wird.

Der Anteil von Personen in R_0 , deren x -Wert kleiner als c ist (repräsentiert durch die weiße Fläche unter der R_0 -Kurve links vom *cut-off* Wert c), ist die Spezifität und der Anteil der Personen in R_1 , deren x -Wert größer oder gleich c ist (repräsentiert durch die hellgrau hervorgehobene Fläche), die Sensitivität.

Ein gewählter *cut-off* Wert determiniert also die Spezifität und Sensitivität der dazugehörigen Entscheidungsregel und damit eine Tafel wie in Tab. 3 (linke Seite). Die Werte der ersten Spalte können als komplementäre Flächen unter der R_0 -Kurve und die der zweiten als komplementäre Flächen unter der R_1 -Kurve gelesen werden.

Zusammenfassend kann man Folgendes festhalten: Jeder mögliche *cut-off* Wert determiniert eine Entscheidungsregel mit dazugehörigen Werten für Sensitivität und Spezifität. Zusammen mit der Basisrate lassen sich dann alle weiteren zu diesem *cut-off* Wert gehörenden Kenngrößen leicht bestimmen. Dies unterstreicht noch einmal, dass im Gegensatz zu Spezifität und Sensitivität die Größen NPP und PPP , $P(\text{Fehler})$ und $P(\text{Treffer})$ sowie der ϕ -Koeffizient nicht nur vom *cut-off* Wert, sondern auch von der Basisrate abhängen.

Dass NPP und PPP von der Basisrate abhängig sind, macht sie allerdings nicht zu weniger nützlichen Kenngrößen. Sie sind vielmehr in einer konkreten Anwendungssituation wichtige und unverzichtbare Kennwerte zur Beurteilung der Leistung einer gewählten Entscheidungsregel. Denn bei einer zu treffenden Prognose ist es wichtiger zu wissen, mit welcher Wahrscheinlichkeit eine Person tatsächlich zur Gruppe der Rückfälligen (R_1) gehört, wenn man sie als zu dieser Gruppe gehörend prognostiziert (PPP), als zu wissen, mit welcher Wahrscheinlichkeit man eine Person zur Gruppe R_1 gehörend prognostiziert, wenn sie zur Gruppe R_1 gehört (*Sensitivität*). Die praktische Verwendbarkeit der Kennwerte NPP und PPP wird aller-

dings dadurch eingeschränkt, dass Basisraten häufig nicht vorliegen und sich nicht immer ohne Weiteres ermitteln lassen.

ROC-Kurve und AUC. Die Wirkungsweise eines Prognoseinstrumentes X lässt sich für ein gegebenes Paar von Teilpopulationen R_0 und R_1 (mit Populationsmittelwerten μ_0 und μ_1 und Standardabweichung σ) dadurch charakterisieren, dass man für *jeden* möglichen *cut-off* Wert c die zugehörige Spezifität und Sensitivität angibt (oder wahlweise einen der jeweiligen komplementären Anteile). Eine solche Beschreibung liefert die sogenannte *Receiver-Operating-Characteristic-Kurve* (ROC-Kurve), die für jeden möglichen c -Wert die zugehörigen Größen $1 - \text{Spezifität}$ und Sensitivität als Punkt in einem Koordinatensystem darstellt.

Die linke Seite der Abb. 2 zeigt verschiedene ROC-Kurven. Es soll nun diejenige Kurve etwas genauer betrachtet werden, auf der einige Punkte markiert sind (sie repräsentiere die für die in Abb. 1 dargestellte Situation zugehörige ROC-Kurve).

Mit jedem Punkt auf der ROC-Kurve ist genau ein Wertepaar ($1 - \text{Spez.}$, Sens.) verbunden, das wiederum mit genau einem *cut-off* Wert c auf der X -Achse in Abb. 1 verknüpft ist. Umgekehrt gehört zu jedem *cut-off* Wert c genau ein solches Wertepaar, das – durch einen Punkt repräsentiert – auf der zugehörigen ROC-Kurve liegt. Dabei durchlaufen die zugehörigen Punkte auf der ROC-Kurve, wenn die c -Werte auf der X -Achse von ganz links kommend nach rechts laufen, die Kurve von rechts oben nach links unten. Der spezifische Verlauf der ROC-Kurven liegt begründet in der Veränderung von Spezifität und Sensitivität bei Verschiebung des *cut-off* Wertes: Verschiebt man den *cut-off* Wert c immer weiter nach rechts, so sinkt die Sensitivität und nähert sich (asymptotisch) dem Wert Null; verschiebt man c hingegen immer weiter nach links, so vergrößert sie sich und nähert sich (asymptotisch) der Eins. Mit der Spezifität ist es genau umgekehrt: Mit wachsendem c geht der Wert der Spezifität gegen Eins, mit fallendem c gegen Null.

Der Verlauf der ROC-Kurve macht auch die gegenseitige Abhängigkeit der Wahrscheinlichkeiten der beiden Fehlertypen deutlich: Verschiebt man den c -Wert nach rechts, um die Wahrscheinlichkeit des Fehlers erster Art ($= 1 - \text{Spezifität}$) zu verringern, so vergrößert sich die Wahrscheinlichkeit des Fehlers zweiter Art ($= 1 - \text{Sensitivität}$); umgekehrt verringert sich zwar die Wahrscheinlichkeit des Fehlers zweiter Art bei Verschiebung des *cut-off* Wertes nach links, gleichzeitig steigt dann aber die Wahrscheinlichkeit des Fehlers erster Art.

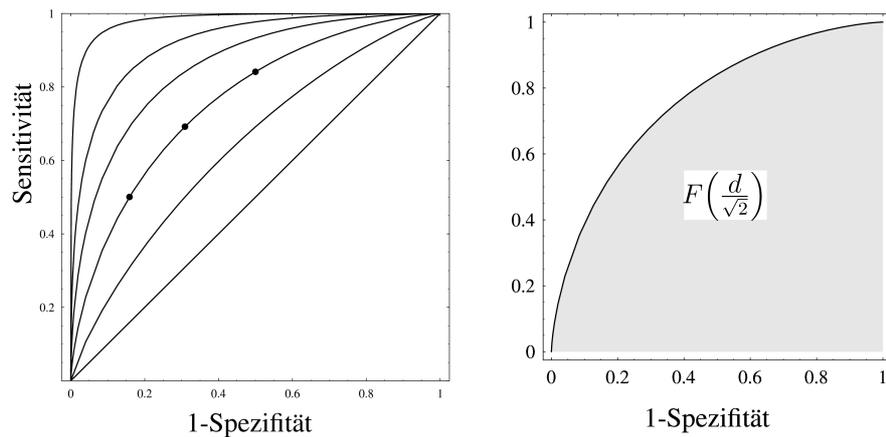


Abbildung 2: Linke Seite: Verschiedene ROC-Kurven. Rechte Seite: Fläche unter einer ROC-Kurve (AUC). Weitere Erläuterungen im Text.

Ändert man die in Abb. 1 dargestellte Situation dahingehend ab, dass man den Abstand der Populationsmittelwerte $\mu_1 - \mu_0$ vergrößert oder die Standardabweichung σ verringert, so erhält man eine neue ROC-Kurve, und zwar eine mit einer stärkeren Konvexität. Die Konvexität ist abhängig von dem Maß d , welches sich wie folgt bestimmen lässt:

$$d := \frac{\mu_1 - \mu_0}{\sigma}.$$

Mit jedem d ist genau eine ROC-Kurve verknüpft; dabei nimmt die Konvexität der ROC-Kurve mit wachsendem d zu und mit sinkendem d ab. Wenn d sich der Null nähert, dann gehen die zugehörigen ROC-Kurven in die Diagonale in Abb. 2 über.⁹ Für die Prognostik ist es sicherlich wünschenswert, wenn die zugehörige ROC-Kurve eine ausgeprägte Konvexität besitzt, denn dann lassen sich *cut-off* Werte wählen, die sich dadurch auszeichnen, dass hohe Spezifitäten mit hohen Sensitivitäten einhergehen. Die Konvexitätsstärke einer ROC-Kurve kann auch durch die Fläche unter ihr ausgedrückt werden. Diese Fläche wird als *area under curve* (AUC) bezeichnet (vgl. rechte Seite von Abb. 2). Sie wird als Index für die Validität eines Prognoseinstrumentes herangezogen. Da die Konvexität von der Größe d abhängt, hängt auch die AUC von d ab. Der genaue Zusammenhang wird durch die nachfolgende Gleichung gegeben. Für die Fläche unter einer ROC-Kurve gilt

$$\text{AUC} = F\left(\frac{1}{\sqrt{2}} \cdot d\right),$$

⁹ In diesem Fall sind die Verteilungen der Variablen X in beiden Teilpopulationen R_0 und R_1 identisch, d. h., es liegt die Situation vor, bei der Sensitivität = τ = 1 - Spezifität ist (Unabhängigkeit von Prognose und Realität). Vgl. die Ausführungen im Abschnitt *ϕ^2 -Koeffizient und Korrelation*.

wobei F die kumulative Verteilungsfunktion der Standardnormalverteilung bezeichnet (vgl. Wickens 2002). Die AUC hat eine interessante probabilistische Interpretation: Greift man sich aus den beiden Teilpopulationen R_0 und R_1 (unabhängig voneinander) jeweils eine Person heraus, dann gibt die AUC die Wahrscheinlichkeit dafür an, dass der x -Wert der aus R_0 kommenden Person kleiner ist als jener der aus R_1 kommenden Person.

Die Größe d ist in der Statistik als Effektstärkemaß bekannt (im Zusammenhang mit dem t -Test für unabhängige Stichproben). Von Cohen (1988) stammt eine populäre Ad-hoc-Klassifikation dieser (und anderer) Effektgrößen, die auf das AUC-Maß übertragen wurde und für die Einschätzung der Güte eines Prognoseinstrumentes immer wieder herangezogen wird. Danach werden AUC-Werte größer oder gleich .72 als „hoch“ und AUC-Werte größer oder gleich .64 als „moderat“ klassifiziert.

Als ein Vorteil für die Verwendung dieses Gütemaßes wird häufig dessen Unabhängigkeit von der Basisrate angeführt (Mossman 1994, Rice & Harris 2005). Denn das AUC-Maß ist nur von der Größe d abhängig. Es beschreibt sozusagen die Wirkungsweise des Prognoseinstrumentes X (bei vorgegebenem d) für alle möglichen *cut-off* Werte, aber unabhängig von der jeweils vorliegenden Basisrate.

Am Ende dieses Abschnitts sei noch darauf hingewiesen, dass sich das Quadrat der punktbiserialen Korrelation zwischen X und R über die Formel

$$\rho_{XR}^2 = \frac{d^2}{d^2 + [\pi(1 - \pi)]^{-1}}$$

bestimmen lässt. Diese Formel macht deutlich, dass die Korrelation ρ_{XR}^2 – wie die AUC – von d abhängt und sich deshalb ebenfalls als eine globale Charakterisierung der prognostischen Güte von X auffassen lässt, aber offenbar als eine, die, im Gegensatz zur AUC, *basisratenabhängig* ist.

1.5.3 Einige Bemerkungen zur Basisrate

Die ROC-Kurve und die AUC beschreiben, wie in den vorangegangenen Ausführungen gezeigt wurde, die Güte eines Prognoseinstrumentes global, d. h., sie beschreiben Spezifität und Sensitivität der mit allen möglichen *cut-off* Werten verbundenen prognostischen Entscheidungen. Dadurch, dass diese Beschreibung auf den Spezifitäten und Sensitivitäten beruht, sind ROC-Kurve und AUC unabhängig von der Basisrate: Eine Veränderung der Basisrate in der betrachteten Population (ohne dass sich dabei wesentliche Charakteristika der Rückfälligen

und Nicht-Rückfälligen ändern) würde nicht zu einer Veränderung der globalen Beschreibung des Prognoseinstrumentes führen.¹⁰

So brauchbar eine basisraten-unabhängige globale Beschreibung eines Prognoseinstrumentes auch ist, so wichtig ist es dennoch, zur Einschätzung der Nützlichkeit einer konkreten (d. h. durch einen bestimmten *cut-off* Wert gegebenen) Entscheidungsregel basisraten-abhängige Kennwerte zu untersuchen. In diesem Abschnitt soll deshalb gezeigt werden, dass sich ein Prognoseinstrument (wie beispielsweise die PCL:SV-Gesamt), angewendet in verschiedenen Grundgesamtheiten, in denen *verschiedene* Basisraten, aber *dieselben* ROC-Kurven vorliegen, in seiner prognostischen Leistung deutlich unterscheidet (eine solche Situation wird beispielsweise in Rice und Harris (1995) besprochen).

Dieser Sachverhalt soll am Beispiel der Trefferquote diskutiert werden. Es wird eine Entscheidungsregel E_c (Prognose: „wird rückfällig, falls $x \geq c$) betrachtet und die zugehörige Trefferquote, die sich ja durch die Formel $(1 - \pi) \cdot \text{Spez}(c) + \pi \cdot \text{Sens}(c)$ bestimmt (die vom *cut-off* Wert abhängigen Größen sind mit c indiziert). Mit Hilfe der Überlegungen des letzten Abschnitts kann man erkennen, dass für jede gegebene Basisrate π die Trefferquote bei sinkendem *cut-off* Wert c gegen π und bei wachsendem *cut-off* Wert c gegen $1 - \pi$ läuft. Es sollen nun zwei Fälle unterschieden werden: $\pi > 0.5$ und $\pi < 0.5$.

Falls die Basisrate größer als 0.5 ist, dann sollte für jeden *cut-off* Wert c die Trefferquote des Verfahrens E_c mindestens so hoch wie die Basisrate sein, da man ja bereits durch das „starre Verfahren“, bei dem man *stets* die Entscheidung „gehört zu R_I “ trifft, eine Trefferquote von π erreicht (dieses Verfahren sei mit E_π bezeichnet). Den Gewinn an Trefferwahrscheinlichkeit, den man durch die Verwendung des Verfahrens E_c gegenüber dem Verfahren E_π erreicht, erhält man also, indem von der Trefferquote des Verfahrens E_c die Basisrate π abgezogen wird. Für diesen Gewinn gilt nun, dass er mit sinkendem c gegen Null und mit wachsendem c gegen $1 - 2\pi$ läuft. Abb. 3 (linke Seite) zeigt für die Basisraten $\pi = 0.5$, $\pi = 0.7$ und $\pi = 0.9$ die zugehörigen Kurven (dieser von c und π abhängige Gewinn wird mit $G_\pi(c)$ bezeichnet).¹¹ Der größte Gewinn – gegenüber dem zugehörigen „starreren Verfahren“ – ist bei einer Basisrate von $\pi = 0.5$ zu erwarten. Ferner nimmt der maximale Zugewinn mit wachsender Basisrate ab.

¹⁰ ROC-Kurve und AUC sind zwar basisraten-unabhängig, aber nicht populations-unabhängig. Ein und dasselbe Prognoseinstrument X kann selbstverständlich in verschiedenen Straftätergrundgesamtheiten unterschiedliche ROC-Kurven und damit auch unterschiedliche AUCs besitzen.

¹¹ Die Kurven sind für den Fall $d = 1$ berechnet. Auf der x-Achse stehen nicht die c -Werte selbst, sondern deren Standardisierungen.

Analoges gilt für den Fall $\pi < 0.5$. Jetzt sollte für jeden c Wert die Trefferquote des Verfahrens $1 - \pi$ betragen, da nun das „starre Verfahren“ (es sei mit $E_{1-\pi}$ abgekürzt), bei dem stets die Entscheidung „gehört zu R_0 “ gefällt wird, schon eine Trefferquote von $1 - \pi$ besitzt. Der Gewinn an Trefferwahrscheinlichkeit gegenüber dem Verfahren $E_{1-\pi}$ hat die Eigenschaft, dass er mit sinkendem c gegen $2\pi - 1$ und mit wachsendem c gegen Null läuft. Die Kurven in Abb. 3 (rechte Seite) zeigen wieder, dass der größte Gewinn bei einer Basisrate von $\pi = 0.5$ zu erwarten ist und dass der maximale Zugewinn nun mit schwindender Basisrate abnimmt (hier wird der von c und π abhängige Gewinn mit $G_{1-\pi}(c)$ bezeichnet).

Zusammenfassend lässt sich im Hinblick auf die Trefferquote für eine Entscheidungsregel festhalten: Bei gleichbleibender AUC ist bei Verwendung einer Entscheidungsregel E_c der maximale Zugewinn an Trefferwahrscheinlichkeit gegenüber den „starrten Verfahren“ E_π und $E_{1-\pi}$ am höchsten für Basisraten von $\pi = 0.5$; er nimmt mit der Entfernung der Basisraten von $\pi = 0.5$ ab und nähert sich der Null an. Dies demonstriert deutlich, wie wichtig neben der Berücksichtigung des AUC-Maßes die Analyse von basisratenabhängigen Kenngrößen (etwa der Trefferquote) ist. Diese können, wie die Ausführungen gezeigt haben, je nach Größe der Basisrate, sehr unterschiedlich ausfallen.

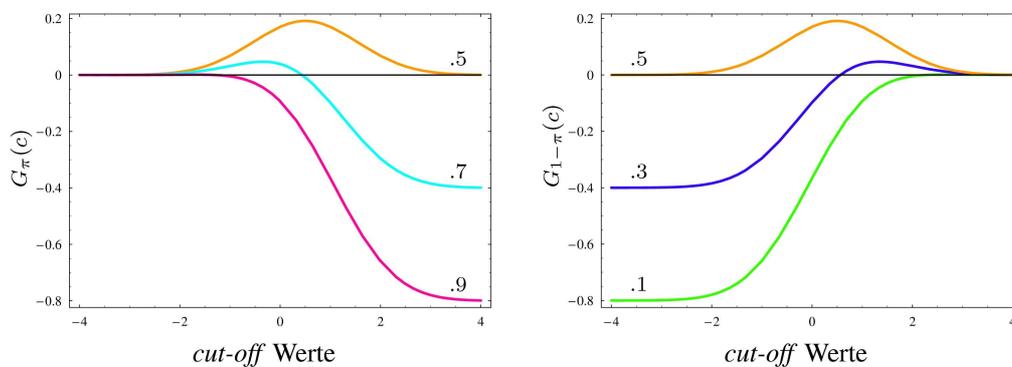


Abbildung 3: Linke Seite: $G_\pi(c)$ in Abhängigkeit vom *cut-off* Wert (standardisiert) für die Basisraten 0.5, 0.7 und 0.9. Rechte Seite: $G_{1-\pi}(c)$ in Abhängigkeit vom *cut-off* Wert (standardisiert) für die Basisraten 0.1, 0.3 und 0.5.

2 Fragestellung

Das Konzept der *Psychopathy* spielt unter den Persönlichkeitsmerkmalen, die mit kriminellem Verhalten assoziiert sein sollen, eine bedeutende Rolle. In einschlägigen Untersuchungen konnte *Psychopathy*, operationalisiert mit Hilfe der PCL-R (Hare 1991) bzw. PCL:SV (Hart et al. 1995), als Risikofaktor für kriminelle, insbesondere gewalttätige Rückfälle identifiziert werden (Andrews & Bonta 1998, 2003, Gendreau et al. 1996, Serin 1991, 1996).

Obwohl die PCL-R ursprünglich nicht als Prognoseinstrument für kriminelle Rückfälle konzipiert wurde, wurde vielfach eine moderate prognostische Validität im Hinblick auf die Vorhersage sowohl genereller als auch gewalttätiger Rückfälle berichtet. Inwieweit auch die PCL:SV als ein valides Instrument zur Kriminalprognose gelten kann, wurde im deutschen Sprachraum bislang kaum untersucht.

Ziel dieser Studie ist deshalb die Evaluierung der prädiktiven Validität der PCL:SV für kriminelle Rückfälle erstmals an einer heterogenen deutschen Straftäterpopulation aus dem Regelvollzug. Die prognostische Güte der PCL:SV-Gesamt und ihrer Subskalen wird dabei im Hinblick auf die Vorhersage sowohl genereller als auch gewalttätiger Rückfälle analysiert.

Zur Anwendung kommt die deutschsprachige Version der PCL:SV (Freese 1999). Auch wird die Wirkungsweise dieses Verfahrens für konkrete Entscheidungsregeln, die durch die Festlegung spezifischer *cut-off* Werte bestimmt werden, im Einzelnen untersucht. Darüber hinaus werden die prädiktiven Eigenschaften der PCL:SV im Vergleich zu prädiktiven Eigenschaften weiterer prognoserelevanter Merkmale diskutiert.

3 Material und Methoden

Vor den weiteren Ausführungen sollen einige Bezeichnungen festgelegt werden, deren Verwendung in der kriminologischen Literatur nicht immer ganz einheitlich zu sein scheint. Zudem soll eine kurze Beschreibung der wesentlichen Inhalte des Bundeszentralregisters (BZR) erfolgen, da dieses Register eine wichtige Dateninformationsquelle der vorliegenden Studie darstellt. An dieser Stelle soll auch erläutert werden, warum auf eine sogenannte „*Time-at-risk*“ Auswertung verzichtet wurde.

3.1 Terminologisches

Als *krimineller Rückfall* wird jeder erneute Eintrag in das Bundeszentralregister (BZR) nach dem *Indexdelikt* gewertet, wobei unter einem *generellen Rückfall* jeder Rückfall (inklusive Rückfall mit einem *Gewaltdelikt*) nach Haftentlassung verstanden wird. Als *Indexdelikt(e)* werden die Delikte bezeichnet, wegen derer die Probanden zum Zeitpunkt der ersten Datenerhebung in Haft sitzen. Diese Haft wird hier im Folgenden als *Indexhaft* bezeichnet. Zu den *Gewaltdelikten* werden (versuchter) Mord, (versuchter) Totschlag, (versuchte) Körperverletzung, Raub und Erpressung gezählt. Folglich werden als *gewalttätige Rückfälle* ausschließlich solche Rückfälle gewertet, die eines der vorab beschriebenen Gewaltdelikte aufweisen. Es sei darauf hingewiesen, dass sexuell motivierte Gewalt hier nicht unter die Gewaltdelikte subsumiert wird, da in der Fachliteratur wiederholt gezeigt werden konnte, dass bei der Vorhersage sexuell motivierter Gewaltkriminalität andere Faktoren eine Rolle spielen als bei der Vorhersage von Gewaltkriminalität im Allgemeinen (Dahle et al. 2008, Hanson & Bussière 1998, Hanson et al. 1995).

Die *Anzahl der Vorstrafen* umfasst alle Eintragungen im BZR bis zum *Indexdelikt* und die *Anzahl aller rechtskräftiger Verurteilungen* die Summe aller BZR-Eintragungen der jeweiligen Probanden bis zum Ende des *Beobachtungszeitraumes*. Die Variable *Anzahl aller rechtskräftiger Verurteilungen* setzt sich somit zusammen aus der *Anzahl der Vorstrafen* bis zum *Indexdelikt* und allen nachfolgenden rechtskräftigen Verurteilungen (inklusive des *Indexdeliktes*) bis zum Ende des *Beobachtungszeitraumes*. Das *Alter bei erster Verurteilung* wird definiert durch den ersten Eintrag im BZR. Täter ohne Vorstrafen werden als *Monotäter* bezeichnet. Bei diesen Tätern stellt somit das *Indexdelikt* die erste, im BZR registrierte, Straftat dar. Als *Polytäter* werden entsprechend alle Täter klassifiziert, die vor ihrem *Indexdelikt* mindestens einen weiteren Eintrag im BZR aufweisen.

Als *Beobachtungszeitraum* (im Sinne einer „*Time-at-risk*“) wird der Zeitraum von der Entlassung aus der *Indexhaft* bis zu einer erneuten Straftat (hier: erneuter Eintrag in das BZR) bzw. bei Straffreiheit der Zeitraum bis zum Abschluss der Untersuchung (hier: Anforderung der BZR-Auszüge) betrachtet. Dieser Zeitraum wird auf der Grundlage der vorhandenen Daten aus dem BZR ermittelt.

3.2 Das Bundeszentralregister

In das Bundeszentralregister (BZR) werden „rechtskräftige Entscheidungen der Strafgerichte sowie bestimmte Entscheidungen der Vormundschaftsgerichte und von Verwaltungsbehörden sowie – nach einer rechtsvergleichenden Begutachtung – ausländische strafrechtliche Verurteilungen gegen Deutsche oder gegen in Deutschland wohnende ausländische Personen eingetragen. Zudem können Suchvermerke im Register niedergelegt werden“ (Bundesamt für Justiz, <http://www.bundeszentralregister.de>). Die entsprechenden Einträge erfolgen grundsätzlich nach einer rechtskräftigen Verurteilung. Dies hat zur Folge, dass der eigentliche Bewährungszeitraum für eine untersuchte Person geringer sein kann, als er sich im BZR darstellt. So hat ein (erneut) strafrechtlich in Erscheinung getretener Rechtsbrecher keinen bzw. noch keinen Eintrag im BZR, wenn er hinsichtlich seiner begangenen Tat ausschließlich eine richterliche Verwarnung erhalten hat oder aber wenn er in U-Haft sitzt, d. h., wenn noch keine rechtskräftige Verurteilung erfolgt ist.

Informationen über laufende bzw. noch nicht eröffnete Verfahren findet man ebenfalls nicht im BZR. Darüber hinaus werden Eintragungen über rechtskräftige Verurteilungen nicht auf Dauer im BZR gespeichert. Somit kann es vorkommen, dass ein angeforderter BZR-Auszug leer ist. Der Vermerk „keine Eintragung“ kann dabei u. a. bedeuten, dass die sogenannte Tilgungsfrist erreicht ist (Löschung nach fünf Jahren), dass der entsprechende Straftäter verstorben ist (Löschung drei Jahre nach Todestag), dass der Proband über 90 Jahre alt ist oder dass ein Verfahren wegen Schuldunfähigkeit eingestellt wurde (Löschung nach zehn Jahren).

Auch ist dem BZR nicht in jedem Fall die genaue Vollzugsdauer zu entnehmen. So deckt sich beispielsweise die Angabe „Datum der (letzten) Tat“ nicht in jedem Fall mit dem Datum der Inhaftierung. Das korrekte Datum der Inhaftierung muss aber als Grundlage dienen, um die tatsächlich in Haft verbrachte Zeit exakt berechnen zu können, die wiederum für die Errechnung eines korrekten Entlassungsdatums notwendig ist. Auch der Eintrag über die Rechtskräftigkeit des Urteils ist nicht immer gleichbedeutend mit dem Tag des Antretens der Haftstrafe. Nicht in allen Fällen findet man darüber hinaus genaue Angaben zur „Erledigung der Straf-

vollstreckung“. In Fällen erfolgreicher Bewährungsaussetzung von Reststrafen ist dem BZR die genaue Vollzugsdauer ebenfalls nicht zu entnehmen.

Um an diese, für eine korrekte „*Time-at-risk*“ Bestimmung nötigen Informationen zu gelangen, hätten auch die jeweiligen Straftaten der hier untersuchten Probanden vorliegen müssen. Da dies nicht der Fall war, wurde in der vorliegenden Untersuchung auf eine „*Time-at-risk*“ Auswertung verzichtet, da man bei einer solchen Auswertung allein auf der Grundlage von BZR-Auszügen mit erheblichen Verzerrungseffekten hätte rechnen müssen.

Für die vorliegende Untersuchung waren die folgenden Einträge von Relevanz: bisherige Verurteilungen, Urteilsdatum, Datum der Rechtskraft, Deliktbezeichnungen und Strafmaß. Probanden mit Einträgen wegen Bewährungsversagen ohne Delikt wurde als nicht rückfällig gewertet.

3.3 Stichprobenbeschreibung

Ausgangspunkt der Studie war eine Stichprobe von 414 Straftätern, die von Amts- und Landesgerichten Sachsen-Anhalts aufgrund verschiedener Delikte rechtskräftig zu Haftstrafen verurteilt worden waren. Die Datenerhebung erstreckte sich auf zwei Phasen. Die erste Datenerhebung fand in der Zeit von August 1999 bis Februar 2002 statt. Eingeschlossen wurden Probanden, die zum Zeitpunkt der Datenerhebung das 21. Lebensjahr vollendet hatten und deren Muttersprache Deutsch war. Insgesamt wurden ein breites Spektrum an Variablen aus den Bereichen Psychopathologie, Persönlichkeit (inklusive PCL:SV) und Soziobiographie sowie die kriminelle Vorgeschichte der Probanden erfasst. Die Bandbreite der Delikte reichte von Bagatelldelikten wie Fahren ohne Fahrerlaubnis bis hin zu schweren Gewaltdelikten und Mord.

Im Januar 2006 erfolgte die zweite Datenerhebung, in deren Rahmen die Bundeszentralregisterauszüge (BZR-Auszüge) angefordert wurden. Für dieses Vorhaben war aus datenschutzrechtlichen Gründen eine erneute Einholung einer Einwilligungserklärung aller Probanden hinsichtlich der Zustimmung ihrer BZR-Auswertungen notwendig, was eine Verzögerung der Datenauswertung bis März 2007 zur Folge hatte. Bis zum 31. März 2007, der gleichzeitig als Stichtag für das Ende des Beobachtungszeitraumes galt, lagen von 280 Probanden, die ihre Einwilligung zur Auswertung ihrer BZR-Auszüge gegeben hatten, die entsprechenden Auszüge vor.¹² Von diesen 280 Probanden wiesen 4 Probanden keinen Eintrag im BZR-Auszug auf.

¹² Probanden, deren BZR-Auszüge nicht angefordert werden konnten, hatten entweder ihre Einwilligung verweigert oder konnten von uns fünf Jahre nach Abschluss der ersten Datenerhebung nicht mehr erreicht werden.

Darüber hinaus war bei 11 Probanden keine korrekte Zuordnung der Daten zu den BZR-Informationen möglich. Somit reduzierte sich die Stichprobe auf 265 Straftäter. Von diesen 265 Straftätern waren seit ihrer Indexhaft bis zum Ende des Beobachtungszeitraumes 21 Probanden durchgängig in Haft. Da diese Probanden nur begrenzt rückfällig werden konnten, mussten sie von den weiteren Analysen ausgeschlossen werden. Somit bestand die Untersuchungsstichprobe aus 244 Straftätern. Der mittlere Beobachtungszeitraum lag – unter Berücksichtigung der in Abschnitt 3.2 genannten Einschränkungen bei der Berechnung – bei 5.2 Jahren. Für jeden rückfälligen Probanden wurde der Beobachtungszeitraum mit dem ersten neuen Eintrag in das BZR nach Entlassung aus der Indexhaft als beendet gewertet. Für alle nicht rückfälligen Probanden endete der Beobachtungszeitraum am 31.03.2007.

3.4 Erhebungsinstrumente

Zur Erfassung von *Psychopathy* nach Hare (1991) kam die Psychopathy-Checklist: Screening Version (PCL:SV, Hart et al. 1995, deutschsprachige Version von Freese 1999) zur Anwendung (ausführliche Beschreibung siehe Abschnitt 1.4). Die Kodierung der PCL:SV-Daten basiert ausschließlich auf Experteneinschätzungen, die während des direkten Kontaktes mit dem Straftäter vorgenommen wurden.

Psychische Störungen nach ICD-10 (WHO 1993) wurden mit Hilfe zweier strukturierter klinischer Interviews erfasst: der „*Schedules for Clinical Assessment in Neuropsychiatry*“ (SCAN, WHO 1995) und der „*International Personality Disorder Examination*“ (IPDE, WHO 1996). Während es sich beim SCAN um ein Instrument zur Diagnostik der grundlegenden Psychopathologie der wichtigsten psychiatrischen Störungen des Erwachsenenalters – mit Ausnahme der Persönlichkeitsstörungen – handelt, erfasst die IPDE differenziert den Bereich der Persönlichkeitsstörungen.

Relevante soziobiographische Daten wurden mit einem in der Klinik für Psychiatrie, Psychotherapie und Psychosomatik der Martin-Luther-Universität Halle-Wittenberg entwickelten soziobiographischen Interview erfasst. Im Rahmen der vorliegenden Studie wurden daraus Angaben zur Herkunftsfamilie sowie zur familiären und beruflichen Situation des Probanden zum Zeitpunkt des Indexdeliktes ausgewertet.

Die Erfassung der Index- und Rückfalldelikte, des Alters bei erster Verurteilung und der Anzahl der Vorstrafen erfolgte auf der Grundlage der BZR-Auszüge.

3.5 Methoden und Hypothesen

Die meisten der neueren Studien zu den hier untersuchten Fragestellungen werten ihre Daten mit Hilfe von ROC-Analysen oder mit dem Modell der Logistischen Regression aus. Neben diesen Verfahren werden in der Literatur auch die Analyse von Mittelwertsdifferenzen (bzw. χ^2 -Analysen) sowie die Analyse (punktbiserialer) Korrelationen herangezogen. Zuweilen stehen diese Verfahren auch gleichzeitig nebeneinander. Dass diese, auf den ersten Blick ganz unterschiedlich erscheinenden Analysemethoden in enger Beziehung zueinander stehen, zeigen die nachfolgenden Überlegungen, die den weiteren Ausführungen (statistische Hypothesen) vorangestellt werden sollen.

Unter der Voraussetzung, dass für ein gegebenes Prognoseinstrument das im Abschnitt 1.5.2 diskutierte Standardmodell gilt (vgl. Abb. 1), sind die AUC (Fläche unter der ROC-Kurve) und β (Regressionskoeffizient der logistischen Regression) auf theoretischer Ebene eng miteinander verknüpft. Denn es kann gezeigt werden, dass die Effektgröße d identisch $\beta \cdot \sigma$ ist.¹³ Da die AUC im Standardmodell ebenfalls eine direkte Funktion der Effektgröße d ist (vgl. im Abschnitt 1.5.2 den Unterpunkt *ROC-Kurve und AUC*), sind beide Parameter, AUC und β -Koeffizient, Funktionen von d . Aus diesen Beziehungen ergibt sich die für die statistischen Hypothesen interessante Verknüpfung äquivalenter Aussagen:

$$\text{AUC} > 0.5 \Leftrightarrow d > 0 \Leftrightarrow \beta > 0 \quad (\Leftrightarrow \exp(\beta) > 1).$$

Diese Verknüpfung von Aussagen zeigt, dass die beiden Hypothesen „AUC > 0.5“ und „ $\beta > 0$ “ genau dann gelten, wenn die Hypothese, dass die Erwartungswertdifferenz $\mu_1 - \mu_0$ größer Null ist, zutrifft. Aber auch die (punktbiserialer) Korrelation ρ_{XR} der binären Variablen R (Rückfälliger vs. Nicht-Rückfälliger) mit der untersuchten Prognosevariablen X ist eine Funktion der Effektgröße d (vgl. im Abschnitt 1.5.2 den Unterpunkt *ROC-Kurve und AUC*). Daher kann der Verknüpfung äquivalenter Aussagen noch die Hypothese „ $\rho_{XR} > 0$ “ hinzugefügt werden. Aufgrund dieser Überlegungen ist zu erwarten, dass auch die zugehörigen statistischen Verfahren zu konkordanten Ergebnissen führen: Entweder es liefern alle statistischen Tests der vier Hypothesen ein signifikantes Ergebnis oder keiner.

Zusammenfassend kann also festgehalten werden, dass die in der Literatur zur Untersuchung der prädiktiven Validität betrachteten statistischen Hypothesen formal äquivalent sind und dass auch die statistischen Tests dieser Hypothesen zu konkordanten Ergebnissen führen.

¹³ Auf diese Beziehung hat mich Prof. Dieter Heyer, Universität Halle, aufmerksam gemacht.

Aus Gründen der Vergleichbarkeit der Ergebnisse mit anderen Studien sollen hier – trotz der gerade beschriebenen Redundanz – nicht nur die Ergebnisse der ROC-Analysen, sondern auch die der Logistischen Regression berichtet werden. Darüber hinaus werden auch statistische Tests zu den Mittelwertsdifferenzen und zu den Korrelationsanalysen aufgeführt werden.

Für die ROC-Analysen haben alle zu untersuchenden Hypothesen die Form

$$H_0: AUC = 0.5 \text{ vs. } H_1: AUC > 0.5,$$

und im Rahmen der Logistischen Regressionen haben sie die Form

$$H_0: \exp(\beta) = 1 \text{ vs. } H_1: \exp(\beta) > 1.$$

Die erste Klasse von H_1 -Hypothesen ($AUC > 0.5$) besagt, dass die betrachteten Verteilungen (PCL:SV-Gesamt, PCL:SV-PP und PCL:SV-SD) sich in den Gruppen „Nicht-Rückfällige“, „Rückfällige“ unterscheiden (das ist äquivalent dazu, dass $d > 0$ ist). Die zweite Klasse von H_1 -Hypothesen besagt für die betrachteten Variablen (PCL:SV-Gesamt, PCL:SV-PP und PCL:SV-SD), dass mit wachsendem Variablen-Wert die Rückfallwahrscheinlichkeit steigt.

In der einschlägigen Literatur wird verschiedentlich darauf hingewiesen, dass sich die prädikative Validität der PCL:SV bei der Vorhersage genereller Rückfälle von der Validität bei der Vorhersage gewalttätiger Rückfälle unterscheidet. Um dies zu untersuchen, wurden getrennte Analysen für generelle und gewalttätige Rückfälle durchgeführt.¹⁴ Hierbei wurde für die statistischen Analysen als (binäre) abhängige Variable einmal die Variable „genereller Rückfall vs. kein Rückfall“, einmal die Variable „gewalttätiger Rückfall vs. kein gewalttätiger Rückfall“ gewählt (die Kategorie „kein gewalttätiger Rückfall“ umfasst dabei sowohl Nicht-Rückfällige als auch Rückfällige mit einem nicht gewalttätigen Delikt). Zur Evaluation der Validität der PCL:SV wurden neben den zentralen Analysen noch weitere Merkmale in ihrem Zusammenhang mit Rückfälligkeit untersucht. In diesem Zusammenhang wurden die prognostischen Eigenschaften der PCL:SV mit den prognostischen Eigenschaften der Variablen „Alter bei erster Verurteilung“ und „Anzahl der Vorstrafen“ verglichen und in einem weiteren Schritt die Frage nach alternativen Konstruktionsmöglichkeiten von Entscheidungsregeln diskutiert.

¹⁴ Die Vorhersage sexueller Rückfälle konnte aufgrund der geringen Fallzahlen nicht gesondert analysiert werden.

4 Ergebnisdarstellung

4.1 Deskriptive Analysen

4.1.1 Charakterisierung der Gesamtstichprobe

Die Untersuchungsstichprobe bestand aus 244 Straftätern (davon waren 34 Frauen (13.9 %) und 210 (86.1 %) Männer)¹⁵. Zum Zeitpunkt des Indexdeliktes betrug das mittlere Alter der Probanden 31.65 Jahre (SD = 8.02; Range 20-56). Das mittlere Alter zum Zeitpunkt ihrer ersten Verurteilung lag bei 24.35 Jahren (SD = 7.72; Range 15-56). Von den 244 Straftätern wiesen 36 Probanden (14.8 %) keine Vorstrafen auf (Monotäter). Entsprechend waren 208 (85.2 %) Probanden Täter, die mit mindestens einem Eintrag in das BZR vorbelastet waren (Polytäter).

4.1.2 Charakterisierung der Rückfalltäter und Nichtrückfalltäter

Insgesamt wurden von den 244 Straftätern im Beobachtungszeitraum 167 (68.4 %) im oben definierten Sinn rückfällig, entsprechend wurden 77 (31.6 %) nicht rückfällig. Von den 167 rückfälligen Probanden wurden 38 (22.8 %) mit einem Gewaltdelikt und 129 (77.2 %) mit einem nicht gewalttätigen Delikt rückfällig. Von den 36 Monotätern (Ersttätern) wurden 12 Personen (7.2 %) rückfällig, davon eine Person mit einem Gewaltdelikt. Von den Tätern mit mindestens einer Vorstrafe (Polytäter) waren es 155 (74.5 %), davon 37 mit einem Gewaltdelikt.

4.1.3 Deliktverteilungen der Index- und Rückfalldelikte

Zu den am häufigsten vorkommenden Indexdelikten zählten, wie Tab. 4 zeigt, Diebstahls- und Unterschlagungsdelikte (34.0 %), gefolgt von Körperverletzung (25.4 %) und Delikten gegen die Straßenverkehrsordnung (21.3 %). Interessanterweise zählten dieselben Delikte auch zu den am häufigsten vorkommenden Rückfalldelikten, wobei hier Diebstahls- und Unterschlagungsdelikte mit 21.7 % den Hauptanteil ausmachten, gefolgt von Körperverletzung (13.1 %) und Delikten gegen die Straßenverkehrsordnung (11.9 %). Wie aus Tab. 4 weiterhin ersichtlich ist, kommen Straftaten gegen das Leben als Rückfalldelikte nicht vor.

¹⁵ Die Resultate vieler Studien sprechen für das Vorhandensein von Geschlechtseffekten für die hier untersuchten Merkmale. Allerdings wäre bei einer geschlechtsspezifischen Auswertung aufgrund des geringen Anteils von Frauen die Schätzung der entsprechenden Parameter außerordentlich instabil ausgefallen. Da andererseits ein Ausschluss der Frauen zu keinen nennenswerten Veränderungen in den Ergebnissen geführt hätte, wurden sie in der Untersuchungsstichprobe belassen.

Am Rande sei hier noch bemerkt, dass von den 12 Probanden, die als Indexdelikt eine Straftat gegen das Leben begangen hatten, immerhin neun Personen zu den Monotätern zählen. Zu den Häufigkeitsverteilungen der Index- und Rückfalldelikte siehe Tab. 4.

Tabelle 4: Häufigkeiten Indexdelikte und Rückfalldelikte nach StGB (Auswahl)¹⁶

Delikte nach StGB	Indexdelikte		Rückfalldelikte	
	n	%	n	%
Straftaten gegen das Leben (Totschlag, Mord)	12	4.9	0	0
Körperverletzung	62	25.4	32	13.1
Raub und Erpressung	39	16.0	12	4.9
Straftaten gegen die persönliche Freiheit	17	7.0	5	2.0
Straftaten gegen die sexuelle Selbstbestimmung	18	7.4	2	0.8
Gemeingefährliche Straftaten	33	13.5	18	7.4
Diebstahl/ Unterschlagung	83	34.0	53	21.7
Urkundenfälschung	14	5.7	7	2.9
Betrug/Untreue	36	14.8	20	8.2
Sachbeschädigung	8	3.3	7	2.9
Straftaten gegen die öffentliche Ordnung	10	4.1	5	2.0
Straftaten gegen die Straßenverkehrsordnung	52	21.3	29	11.9
Verstoß gegen das BtMG	21	8.6	14	5.7

Aufgrund der Erfassung von Tatmehrheit sind Mehrfachnennungen möglich.

Schaut man sich die Rückfallraten in Abhängigkeit des jeweiligen Indexdeliktes an, so stellt man fest, dass Straftäter, die sich wegen Diebstahls- und Unterschlagungsdelikten, wegen Betruges und Untreue sowie wegen Straftaten gegen die öffentliche Ordnung in Haft befanden, am häufigsten rückfällig wurden, gefolgt von Straftätern mit Indexdelikten gegen die Straßenverkehrsordnung und Straftaten gegen die persönliche Freiheit (siehe Tab. 5). Am seltensten rückfällig nach Entlassung aus der Indexhaft wurden Straftäter, die eine Straftat gegen das Leben begangen hatten. Bei der Interpretation dieser Prozentzahl muss allerdings beachtet werden, dass jene 12 Probanden, die wegen einer Straftat gegen das Leben in Indexhaft saßen, im Vergleich zu den anderen Straftätern im Mittel den kürzesten Beobachtungszeitraum aufwiesen und dass sich zudem unter den 21 Probanden, die aus den Analysen aufgrund ihrer sich über den gesamten Beobachtungszeitraum erstreckenden Haftstrafen ausgeschlossen werden mussten, immerhin 14 Straftäter mit Straftaten gegen das Leben befanden.

¹⁶ Kriterien für die Auswahl der in Tab. 4 und 5 aufgeführten Delikte waren zum einen deren gesellschaftliche Bedeutsamkeit und zum anderen die Häufigkeit ihres Auftretens.

Tabelle 5: Rückfallraten in Abhängigkeit vom Indexdelikt (Auswahl)

Indexdelikte	Kein Rückfall		Rückfall	
	n	%	n	%
Straftaten gegen das Leben (Totschlag, Mord) (n = 12)	11	91.7	1	8.3
Körperverletzung (n = 62)	20	32.3	42	67.7
Raub und Erpressung (n = 39)	13	33.3	26	66.7
Straftaten gegen die persönliche Freiheit (n = 17)	4	23.5	13	76.5
Straftaten gegen die sexuelle Selbstbestimmung (n = 18)	10	55.6	8	44.4
Gemeingefährliche Straftaten (n = 33)	9	27.3	24	72.7
Diebstahl/ Unterschlagung (n = 83)	16	19.3	67	80.7
Urkundenfälschung (n = 14)	4	28.6	10	71.4
Betrug/Untreue (n = 36)	7	19.4	29	80.6
Sachbeschädigung (n = 8)	3	37.5	5	62.5
Straftaten gegen die öffentliche Ordnung (n = 10)	2	20.0	8	80.0
Straftaten gegen die Straßenverkehrsordnung (n = 52)	12	23.1	40	76.9
Verstoß gegen das BtMG (n = 21)	10	47.6	11	52.4

Aufgrund der Erfassung von Tatmehrheit sind Mehrfachnennungen möglich.

4.2 Statistische Analysen zu den ausgewählten Einzelmerkmalen

Bevor die Ergebnisse der zentralen Analysen dargestellt werden, soll vorab eine vergleichende Gegenüberstellung der Rückfalltäter und Nichtrückfalltäter hinsichtlich ausgewählter Einzelmerkmale erfolgen. Die Auswahl der Merkmale orientiert sich an einschlägigen Untersuchungen, wobei dem aktuellen Stand der Rückfallforschung entsprechend sowohl statische als auch dynamische Merkmale berücksichtigt wurden. Um den Rahmen der Studie nicht zu sprengen, wurde die Auswahl auf solche Variablen beschränkt, die bei der Rückfallprognose als besonders relevant angesehen werden.

Um Aufschluss darüber zu erlangen, ob die betrachteten Merkmale bei Tätern mit gewalttätigen Rückfällen möglicherweise eine andere Relevanz als bei Tätern mit generellen Rückfällen haben, wurden die Daten für diese beiden Gruppen getrennt ausgewertet. Es sei an dieser Stelle noch einmal darauf hingewiesen, dass Straftäter mit einem gewalttätigen Rückfall eine Teilgruppe aller Rückfalltäter darstellen und dass die Vergleichsgruppe „kein gewalttätiger Rückfall“ sowohl alle Nichtrückfälligen als auch jene Täter umfasst, die mit einem nicht gewalttätigen Delikt rückfällig geworden sind.¹⁷ Die jeweiligen Gruppenvergleiche wurden – je nach Datentyp – mit zweiseitigen t-Tests bzw. χ^2 -Tests durchgeführt.

¹⁷ Es dürfte klar sein, dass die Straftäter mit einem gewalttätigen Rückfall zu vergleichen sind mit der Menge aller Straftäter, da zum Zeitpunkt der Prognose nicht bekannt ist, ob der Proband überhaupt rückfällig wird oder nicht.

4.2.1 Relevante Merkmale der strafrechtlichen Vorgeschichte

Zunächst wurden die Merkmale der strafrechtlichen Vorgeschichte analysiert. Wie aufgrund empirischer Befunde zu erwarten war, waren Rückfalltäter (genereller Rückfall) im Vergleich zu Nichtrückfalltätern zum Zeitpunkt ihrer ersten Verurteilung im Mittel jünger ($p < .000$). Auch wiesen Rückfalltäter mehr Vorstrafen auf ($p < .000$) und wurden bis zum Ende des Beobachtungszeitraumes insgesamt häufiger rechtskräftig verurteilt ($p < .000$). Tab. 6 fasst die Ergebnisse zusammen.

Tabelle 6: Strafrechtliche Vorgeschichte: genereller Rückfall vs. kein Rückfall

N = 244	Genereller Rückfall nach Indexdelikt (n = 167)		Kein Rückfall nach Indexdelikt (n = 77)		d	t-Wert (df = 242)
	M (SD)	Range	M (SD)	Range		
Alter bei erster Verurteilung	22.94 (6.65)	15-50	27.40 (8.96)	16-56	.599	-4.34***
Anzahl der Vorstrafen	7.28 (5.50)	0-25	4.58 (5.45)	0-28	.489	3.56***
Anzahl aller Verurteilungen bis Ende Beobachtungszeitraum	10.71 (6.20)	2-33	5.58 (5.45)	1-29	.858	6.23***

T-Test für unabhängige Stichproben, zweiseitige Testung; * $p < .05$; ** $p < .01$; *** $p < .001$, |d| bezeichnet den Absolutbetrag der geschätzten d-Werte.

Beim Vergleich der Mittelwerte der Gruppen „gewalttätiger Rückfall“ vs. „kein gewalttätiger Rückfall“ zeigte sich hinsichtlich dieser Merkmale ein etwas anderes Bild. So unterschieden sich diese beiden Gruppen nur bezüglich des Merkmals *Alter bei erster Verurteilung* ($p = .006$) statistisch bedeutsam voneinander, wobei die Täter, die mit einem Gewaltdelikt rückfällig wurden, im Mittel jünger waren. (zusammenfassend siehe Tab. 7).

Tabelle 7: Strafrechtliche Vorgeschichte: gewalttätiger Rückfall vs. kein gewalttätiger Rückfall

N = 244	Gewalttätiger Rückfall nach Indexdelikt (n = 38)		Kein gewalttätiger Rückfall nach Indexdelikt (n = 206)		d	t-Wert (df = 242)
	M (SD)	Range	M (SD)	Range		
Alter bei erster Verurteilung	21.21 (4.96)	16-34	24.93 (8.01)	15-56	.488	-2.762**
Anzahl der Vorstrafen	6.71 (5.19)	0-18	6.37 (5.70)	0-28	.062	.339
Anzahl aller Verurteilungen bis Ende Beobachtungszeitraum	9.61 (5.80)	3-22	9.00 (6.54)	1-33	.094	.533

T-Test für unabhängige Stichproben, zweiseitige Testung; * $p < .05$; ** $p < .01$; *** $p < .001$, |d| bezeichnet den Absolutbetrag der geschätzten d-Werte.

4.2.2 Relevante Merkmale der Soziobiographie

Es folgen nun die Merkmale aus der Soziobiographie. Auch hier sollen zunächst die Rückfalltäter (genereller Rückfall) mit den Nichtrückfalltätern verglichen werden. Anschließend folgt, den vorangegangenen Analysen entsprechend, ein Vergleich zwischen den Gruppen „gewalttätiger Rückfall“ vs. „kein gewalttätiger Rückfall“. Betrachtet man die Daten aus Tab. 8 zunächst auf deskriptiver Ebene, so stellt man fest, dass Rückfalltäter im Vergleich zu Nichtrückfalltätern häufiger aus sogenannten Broken-Home-Situationen kamen und zudem häufiger Gewalt in ihrer Herkunftsfamilie erlebt hatten.¹⁸ Auch waren die Eltern der Rückfalltäter häufiger selbst straffällig gewesen. Etwa 92 % der Nichtrückfalltäter hatten einen Haupt- oder Realschulabschluss bzw. das Abitur. Bei den Rückfalltätern waren es etwa 89 %. Eine Intelligenzminderung wurde insgesamt gesehen bei 14 Straftätern diagnostiziert, wobei 5 % der Rückfalltäter und etwa 7 % der Nichtrückfalligen diese Diagnose erhalten hatten.

Tabelle 8: Soziobiographie: genereller Rückfall vs. kein Rückfall

N = 244	Genereller Rückfall nach Indexdelikt (n = 167)		Kein Rückfall nach Indexdelikt (n = 77)		χ^2
	n	%	n	%	
Broken-Home	120	71.9	45	58.4	4.332*
Gewalt in der Herkunftsfamilie	67	40.1	23	29.9	2.378
Straffälligkeit Eltern	22	13.2	7	9.1	.839
Schulabschluss (Haupt-, Realschule oder Abitur) ^a	146	88.5	71	92.2	.786
Berufsausbildung (abgeschlossen)	109	56.3	56	72.7	1.339
Intelligenzminderung ^b	9	5.4	5	6.6	0.136

χ^2 -Test, df = 1, *p < .05; **p < .01; ***p < .001, ^a bei zwei Personen keine Angaben, ^b bei einer Person keine Angaben.

Im statistischen Sinne unterschieden sich die beiden betrachteten Gruppen allerdings nur hinsichtlich der Variablen *Broken-Home* (p = .037). Für alle anderen Variablen konnten keine statistisch bedeutsamen Unterschiede zwischen den beiden Tätergruppen gefunden werden (*Gewalt in der Herkunftsfamilie*, p = .123; *Straffälligkeit der Eltern*, p = .360; *Schulabschluss*, p = .423; *Berufsausbildung*, p = .247 und *Intelligenzminderung*, p = .712).

¹⁸ Unter einer Broken-Home-Situation wird die Zerstörung der primären familiären Strukturen vor dem 15. Lebensjahr der betroffenen Person verstanden.

Tabelle 9: Soziobiographie: gewalttätiger Rückfall vs. kein gewalttätiger Rückfall

N = 244	Gewalttätiger Rückfall nach Indexdelikt (n = 38)		Kein gewalttätiger Rückfall nach Indexdelikt (n = 206)		χ^2
	n	%	n	%	
Broken-Home	29	76.3	136	66.0	1.553
Gewalt in der Herkunftsfamilie	16	42.1	74	35.9	0.527
Straffälligkeit Eltern	7	18.4	22	10.7	1.836
Schulbildung (Haupt-, Realschule oder Abitur) ^a	32	84.2	185	90.7	1.450
Berufsausbildung (abgeschlossen)	21	55.3	144	69.9	3.140
Intelligenzminderung ^b	2	5.3	12	5.9	0.021

χ^2 -Test, df = 1, *p < .05; **p < .01; ***p < .001, ^a bei zwei Personen keine Angaben, ^b bei einer Person keine Angaben.

Ähnlich, wie schon bei den Merkmalen zur strafrechtlichen Vorgeschichte, zeigte sich auch im Bereich der Soziobiographie beim Vergleich der Gruppen „gewalttätiger Rückfall“ vs. „kein gewalttätiger Rückfall“ ein anderes Bild als bei den Analysen zum generellen Rückfall. So unterschieden sich die in Tab. 9 untersuchten Gruppen hinsichtlich keiner der aufgeführten Merkmale statistisch bedeutsam voneinander (*Broken-Home*, p = .213; *Gewalt in der Herkunftsfamilie*, p = .468; *Straffälligkeit der Eltern*, p = .175; *Schulbildung*, p = .229; *Berufsausbildung*, p = .229 und *Intelligenzminderung*, p = .886).

4.2.3 Psychopathologische Auffälligkeiten zum Zeitpunkt des Indexdeliktes

Es folgen nun die Angaben zu relevanten Merkmalen der Psychopathologie zum Zeitpunkt des Indexdeliktes. Wie die Daten in Tab. 10 zeigen, konnten statistisch signifikante Unterschiede zwischen Rückfalltätern und Nichtrückfalltätern nur bzgl. zweier Diagnosegruppen gefunden werden: *psychische Störung(en) durch psychotrope Substanzen* (p = .037) sowie *psychische Störung(en) inklusive Persönlichkeitsstörung(en)* (p = .028). In allen anderen der in Tab. 10 aufgeführten Störungsbildern war der Unterschied zwischen Rückfalltätern und Nichtrückfalltätern statistisch nicht signifikant (*psychische Störung(en)*, p = .074; *psychische und Verhaltensstörung(en) durch Alkohol*, p = .060; *psychische Störung(en) durch psychotrope Substanzen außer Alkohol*, p = .809; *Persönlichkeitsstörung(en)*, p = .104).

Tabelle 10: Psychopathologische Auffälligkeiten: genereller Rückfall vs. kein Rückfall

N = 244	Genereller Rückfall nach Indexdelikt (n = 167)		Kein Rückfall nach Indexdelikt (n = 77)		χ^2
	n	%	n	%	
Psychische Störung(en) (F00-F50)	117	70.1	45	58.4	3.188
Psychische Störung(en) durch psychotrope Substanzen inklusive Alkohol (F10-F19)	104	62.3	37	48.1	4.371*
Psychische und Verhaltensstörung(en) durch Alkohol (F10.x)	80	47.9	27	35.1	3.528
Psychische Störung(en) durch psychotrope Substanzen außer Alkohol (F11-F19)	37	22.2	16	20.8	.059
Persönlichkeitsstörung(en) (F60.x)	44	26.3	13	16.9	2.637
Psychische Störung(en) inklusive Persönlichkeitsstörung(en)	125	74.9	47	61.0	4.833*

χ^2 -Test, df = 1, *p < .05; **p < .01; ***p < .001; Aufgrund von Erfassung von Komorbiditäten sind Mehrfachnennungen möglich.

Vergleicht man die Gruppen „gewalttätiger“ vs. „kein gewalttätiger“ Rückfall (siehe Tab. 11), so zeigt sich hier lediglich ein signifikanter Gruppenunterschied hinsichtlich des Merkmals *Persönlichkeitsstörung(en)* (p = .003).

Tabelle 11: Psychopathologische Auffälligkeiten: gewalttätiger Rückfall vs. kein gewalttätiger Rückfall

N = 244	Gewalttätiger Rückfall nach Indexdelikt (n = 38)		Kein gewalttätiger Rück- fall nach Indexdelikt (n = 206)		χ^2
	n	%	n	%	
Psychische Störung(en) (F00-F50)	27	71.1	135	65.5	0.438
Psychische Störung(en) durch psychotrope Substanzen inklusive Alkohol (F10-F19)	26	68.4	115	55.8	2.087
Psychische und Verhaltensstörung(en) durch Alkohol (F10.x)	22	57.9	85	41.3	3.605
Psychische Störung(en) durch psychotrope Substanzen außer Alkohol (F11-F19)	8	21.1	45	21.8	0.012
Persönlichkeitsstörung(en) (F60.x)	16	42.1	41	19.9	8.833**
Psychische Störung(en) inklusive Persönlichkeitsstörung(en)	31	81.6	141	68.4	2.660

χ^2 -Test, df = 1, *p < .05; **p < .01; ***p < .001; Aufgrund von Erfassung von Komorbiditäten sind Mehrfachnennungen möglich.

Hinsichtlich des Merkmals *psychische Störung(en) durch Alkohol* ist eine Tendenz erkennbar (p = .058). In allen anderen der in Tab. 11 aufgeführten Merkmale war der Unterschied zwischen Tätern mit einem gewalttätigen Rückfall und Tätern mit keinem gewalttätigen Rückfall statistisch nicht signifikant (*psychische Störung(en)*, p = .508; *psychische Störung(en) durch*

psychotrope Substanzen, $p = .149$; *psychische Störung(en) durch psychotrope Substanzen außer Alkohol* $p = .913$; *psychische Störung(en) inklusive Persönlichkeitsstörung(en)* $p = .103$).

4.2.4 Zusammenhänge der ausgewählten Einzelmerkmale mit kriminellem Rückfall

Einige Studien diskutieren dort, wo hier t- bzw. χ^2 -Werte berücksichtigt wurden, Korrelationen zwischen den binären Variablen „genereller Rückfall/ kein Rückfall“ bzw. „gewalttätiger Rückfall/ kein gewalttätiger Rückfall“ einerseits und relevanten Merkmalen zur Rückfallprognose andererseits (z. B. Dahle 2004, Gendreau et al. 1996, Hanson & Bussière 1998, Wilson 2003). Um direkte Vergleiche mit den Ergebnissen dieser und anderer Studien zu ermöglichen, sollen, wie bereits eingangs erwähnt wurde, auch die Resultate der Korrelationsanalysen aufgeführt werden. Man beachte aber, dass die Tab. 12 und 13 dieselben Informationen nur in anderer Form wiedergeben (siehe dazu auch noch einmal die Anmerkungen weiter oben). Aus formalen Gründen ist daher zu erwarten, dass in den beiden nachfolgenden Tabellen die Korrelationen mit jenen Merkmalen signifikant werden, deren entsprechende statistische Analysen auch in den vorangegangenen Auswertungen statistisch bedeutsam geworden sind. Es kann daher auf eine wiederholte Diskussion der Ergebnisse verzichtet werden.

Tabelle 12: Korrelationen zwischen ausgewählten Einzelmerkmalen und genereller Rückfälligkeit

N = 244	Genereller Rückfall nach Indexdelikt	
	Korrelation	p-Wert
Alter bei erster Verurteilung	-.269	< .01**
Anzahl der Vorstrafen	.223	< .01**
Anzahl aller rechtskräftiger Verurteilungen bis Ende Beobachtungszeitraum	.372	< .01**
Broken-Home	.133	.038*
Gewalt in der Herkunftsfamilie	.099	.124
Straffälligkeit Eltern	.059	.362
Schulbildung (Haupt-, Realschule oder Abitur) ^a	-.057	.378
Berufsausbildung (abgeschlossen)	.074	.249
Intelligenzminderung ^b	-.024	.714
Psychische Störungen (F00-F50) (Diagnose mindestens einer psychischen Störung, inklusive organische psychische Störungen)	.144	.075
Psychische Störung(en) durch psychotrope Substanzen, inklusive Alkohol (F10-F19)	.134*	.037*
Psychische und Verhaltensstörung(en) durch Alkohol (F10.x)	.120	.061
Psychische Störung(en) durch psychotrope Substanzen außer Alkohol (F11-F19)	.016	.809
Persönlichkeitsstörung(en) (F60.x) (Diagnose mindestens einer Persönlichkeitsstörung)	.104	.105
Psychische Störung(en) inklusive Persönlichkeitsstörung(en)	.141	.028*

*p < .05; **p < .01; ***p < .001; zweiseitige Testung, ^a bei zwei Personen keine Angaben, ^b bei einer Person keine Angaben.

Tabelle 13: Korrelationen zwischen ausgewählten Einzelmerkmalen und gewalttätiger Rückfälligkeit

N = 244	Gewalttätiger Rückfall nach Indexdelikt	
	Korrelation	p-Wert
Alter bei erster Verurteilung	-.175	.006**
Anzahl der Vorstrafen	.022	.735
Anzahl aller rechtskräftiger Verurteilungen bis Ende Beobachtungszeitraum	.034	.595
Broken-Home	.080	.214
Gewalt in der Herkunftsfamilie	.046	.470
Straffälligkeit Eltern	.087	.177
Schulbildung (Haupt-, Realschule oder Abitur) ^a	-.113	.077
Berufsausbildung (abgeschlossen)	-.077	.230
Intelligenzminderung ^b	-.009	.886
Psychische Störungen (F00-F50) (Diagnose mindestens einer psychischen Störung, inklusive organische psychische Störungen)	.042	.510
Psychische Störung(en) durch psychotrope Substanzen, inklusive Alkohol (F10-F19)	.092	.150
Psychische und Verhaltensstörung(en) durch Alkohol (F10.x)	.122	.058
Psychische Störung(en) durch psychotrope Substanzen außer Alkohol (F11-F19)	-.007	.914
Persönlichkeitsstörung(en) (F60.x) (Diagnose mindestens einer Persönlichkeitsstörung)	.190	.003**
Psychische Störung(en) inklusive Persönlichkeitsstörung(en)	.104	.104

*p < .05; **p < .01; ***p < .001; zweiseitige Testung, ^a bei zwei Personen keine Angaben, ^b bei einer Person keine Angaben.

4.3 Zentrale Analysen: Die PCL:SV

4.3.1 Deskriptive Analysen zur PCL:SV

Für den PC:SV-Gesamt ergab sich bei den rückfälligen Straftätern (genereller Rückfall) ein Mittelwert von 9.89 (SD = 5.42; Range 0-23) und bei den Nichtrückfälligen einer von 6.94 (SD = 4.69; Range 0-18). Straftäter mit einem gewalttätigen Rückfall hatten im Mittel einen PC:SV-Gesamt von 11.29 (SD = 4.80; Range 0-21).

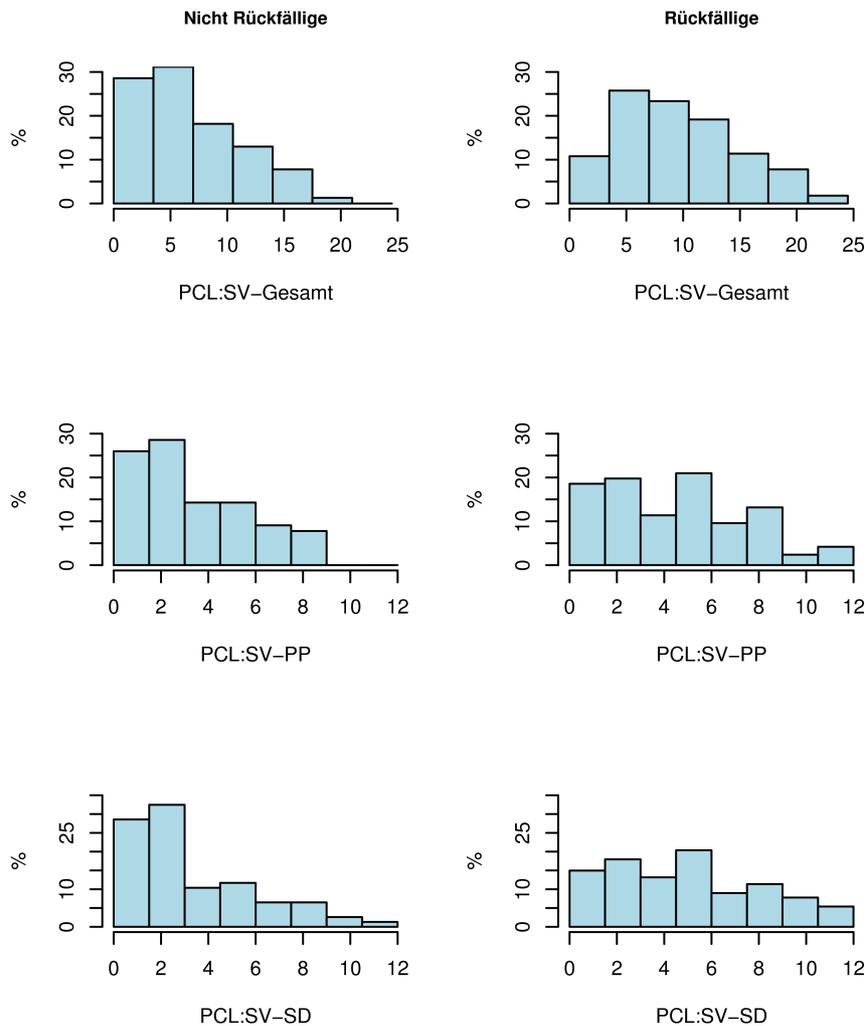


Abbildung 4: Verteilung der PCL:SV Summenscores, Faktor 1- und Faktor 2-Scores bei Rückfalltätern und Nichtrückfalltätern im Hinblick auf generelle Rückfälle.

Der Mittelwert des PC:SV-Gesamt war für alle Straftäter 8.96 (SD = 5.37) mit einem Minimum von 0 und einem Maximum von 23 Punkten (vgl. auch die Mittelwerte in den Tab. 14 und 15). Abb. 4 zeigt die Verteilungen der einzelnen PCL:SV-Scores (PCL:SV-Gesamt, PCL:SV-SD und PCL:SV-PP) bei den Rückfalltätern und den Nicht-Rückfalltätern im Hinblick auf generelle Rückfälle.

Die interne Konsistenz (*cronbachs alpha* als Schätzer für die Reliabilität) beträgt für die PCL:SV-Gesamt $\alpha = 0.834$, für die PCL:SV-PP $\alpha = 0.757$ und für die PCL:SV-SD $\alpha = 0.777$.

4.3.2 Statistische Analysen zur PCL:SV

Es soll nun geprüft werden, ob sich die empirischen Mittelwerte der jeweiligen PCL:SV-Scores bei rückfälligen (genereller und gewalttätiger Rückfall) und nicht-rückfälligen Straftätern statistisch signifikant voneinander unterscheiden.

Tabelle 14: PCL:SV-Scores und genereller Rückfall

N = 244	Genereller Rückfall nach Indexdelikt (n = 167)		Kein Rückfall nach Indexdelikt (n = 77)		d	t-Wert (df = 242)
	M	SD	M	SD		
Scores:						
PCL:SV-Gesamt	9.89	5.42	6.94	4.69	.568	4.124***
PCL:SV-PP	4.69	3.07	3.51	2.56	.405	2.939**
PCL:SV-SD	5.20	3.18	3.43	2.85	.576	4.184***

T-Test für unabhängige Stichproben, zweiseitige Testung; * $p < .05$; ** $p < .01$; *** $p < .001$, |d| bezeichnet den Absolutbetrag der geschätzten d-Werte.

Wie aus Tab. 14 ersichtlich ist, wiesen die Probanden mit einem generellen Rückfall sowohl hinsichtlich der PCL:SV-Gesamt als auch hinsichtlich der PCL:SV-PP und PCL:SV-SD im Mittel höhere Werte auf als die nicht-rückfälligen Probanden. Bei der Überprüfung der Mittelwertsunterschiede auf statistische Signifikanz zeigte sich für alle drei Scores ein signifikantes Ergebnis (PCL:SV-Gesamt, $p < .000$; PCL:SV-PP, $p = .004$; PCL:SV-SD, $p < .000$).

Tabelle 15: PCL:SV-Scores und gewalttätiger Rückfall

N = 244	Gewalttätiger Rückfall nach Indexdelikt (n = 167)		Kein gewalttätiger Rück- fall nach Indexdelikt (n = 77)		d	t-Wert (df = 242)
	M	SD	M	SD		
Scores:						
PCL:SV-Gesamt	11.29	4.80	8.53	5.37	.522	2.955**
PCL:SV-PP	4.84	2.81	4.22	2.99	.211	1.192
PCL:SV-SD	6.45	2.81	4.31	3.14	.691	3.913***

T-Test für unabhängige Stichproben, zweiseitige Testung; * $p < .05$; ** $p < .01$; *** $p < .001$, |d| bezeichnet den Absolutbetrag der geschätzten d-Werte.

Beim Vergleich der Gruppen „gewalttätiger Rückfall“ vs. „kein gewalttätiger Rückfall“ ergaben die statistischen Analysen signifikante Unterschiede hinsichtlich der PCL:SV-Gesamt ($p = .003$) und der PCL:SV-SD ($p < .000$). Die Mittelwertsunterschiede der PCL:SV-PP wurden nicht signifikant ($p = .234$) (siehe dazu Tab. 15).

4.3.3 Korrelationen der PCL:SV mit generellem und gewalttätigem Rückfall

Wie schon bei den Auswertungen zu relevanten Einzelmerkmalen sollen auch für die PCL:SV aus Vergleichbarkeitsgründen mit anderen Studien die (punktserialen) Korrelationen der einzelnen PCL:SV-Scores mit Rückfälligkeit berichtet werden (siehe dazu Tab. 16).

Tabelle 16: Korrelationen der PCL:SV mit generellem und gewalttätigem Rückfall

N = 244	Genereller Rückfall		Gewalttätiger Rückfall	
	r	p-Wert	r	p-Wert
Scores:				
PCL:SV-Gesamt	.256***	.000	.187**	.003
PCL:SV-PP	.186**	.004	.076	.234
PCL:SV-SD	.260***	.000	.244***	.000

*p < .05; **p < .01; ***p < .001; n (genereller Rückfall) = 167, n (gewalttätiger Rückfall) = 38.

Es zeigten sich erwartungsgemäß zu den soeben berichteten Resultaten analoge Ergebnisse: Die Korrelationen aller drei PCL:SV-Scores mit generellem Rückfall wurden signifikant, jene mit gewalttätigem Rückfall nur im Hinblick auf die PCL:SV-Gesamt und PCL:SV-SD.

4.3.4 Prädiktive Validität der PCL:SV

Es folgen nun die Ergebnisse der ROC-Analysen sowie der Logistischen Regressionen.

Tabelle 17: ROC-Analysen: genereller und gewalttätiger Rückfall

N = 244	Genereller Rückfall		Gewalttätiger Rückfall	
	AUC	95%KI	AUC	95%KI
Scores:				
PCL:SV-Gesamt	.660***	0.586-0.733	.660***	0.576-0.745
PCL:SV-PP	.609**	0.535-0.682	.569	0.500-0.662
PCL:SV-SD	.666***	0.592-0.739	.700***	0.618-0.782

*p < .05 ; **p < .01; ***p < .001; n (genereller Rückfall) = 167, n (gewalttätiger Rückfall) = 38.

Wie Tab. 17 zu entnehmen ist, konnten bei den Analysen zum generellen Rückfall für die PCL:SV-Gesamt (geschätzte AUC = .660), PCL:SV-PP (geschätzte AUC = .609) und PCL:SV-SD (geschätzte AUC = .666) die zugehörigen H_0 -Hypothesen (H_0 : AUC = 0.5/ H_1 : AUC > 0.5) zurückgewiesen werden. Für den gewalttätigen Rückfall gilt dies nur für die PCL:SV-Gesamt (geschätzte AUC = .660) und PCL:SV-SD (geschätzte AUC = .700). Für die PCL:SV-PP konnte die zugehörige H_0 -Hypothese nicht verworfen werden (geschätzte AUC = .569). In Abb. 5 sind die dazugehörigen empirischen ROC-Kurven dargestellt (durch Striche

verbundene Punkte; die durchgezogenen Kurven stellen Schätzungen der ROC-Kurven unter der Annahme dar, dass das den ROC-Analysen zugrunde liegenden Standardmodell gilt, vgl. dazu Abschnitt 1.5).

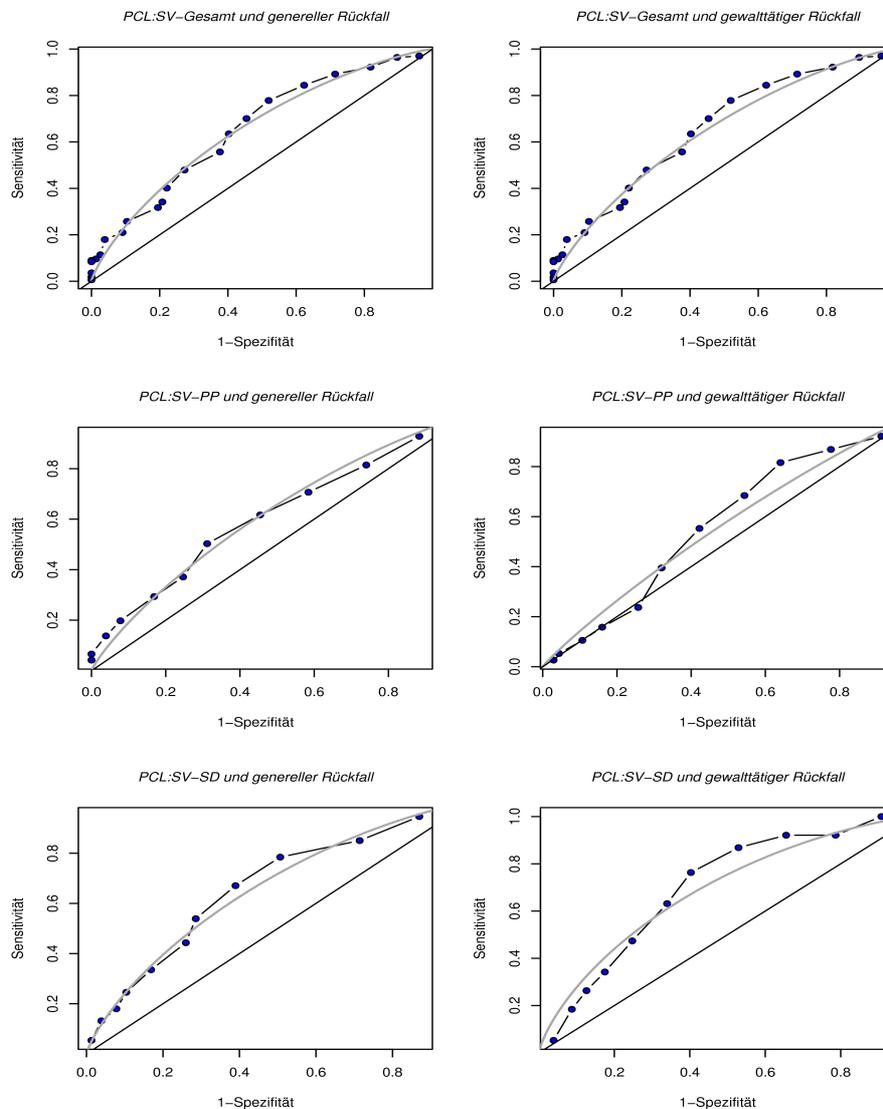


Abbildung 5: ROC-Kurven für generelle und gewalttätige Rückfälle für PCL:SV-Gesamt, PCL:SV-PP und PCL:SV-SD.

Bei den Logistischen Regressionen zeigten sich erwartungsgemäß zu den in Tab. 17 berichteten analoge Ergebnisse (siehe Tab. 18). So konnten für den generellen Rückfall sämtliche H_0 -Hypothesen ($H_0: \exp(\beta) = 1$ vs. $H_1: \exp(\beta) > 1$) zurückgewiesen werden, wobei für die PCL:SV-Gesamt ein $\exp(\beta)$ von 1.122 ($R^2 = 0.068$), für die PCL:SV-PP von 1.155 ($R^2 =$

0.035) und für die PCL:SV-SD von 1.217 ($R^2 = 0.069$) errechnet wurde.¹⁹ Zur Vorhersage eines gewalttätigen Rückfalls zeigte sich ein statistisch bedeutsamer Einfluss der PCL:SV-Gesamt ($\exp(B) = 1.097$; $R^2 = 0.033$) und PCL:SV-SD ($\exp(B) = 1.232$; $R^2 = 0.056$). Für die PCL:SV-PP hingegen erreichte das Ergebnis, wie schon in der ROC-Analyse, keine statistische Signifikanz ($\exp(B) = 1.072$; $R^2 = 0.006$).

Beide Analysemethoden zeigen somit, dass für die Vorhersage eines generellen Rückfalls die PCL:SV-Gesamt, die PCL:SV-PP sowie die PCL:SV-SD und für die Vorhersage eines gewalttätigen Rückfalls die PCL:SV-Gesamt und PCL:SV-SD im statistischen Sinne bedeutsam sind. Allerdings fallen sämtliche R^2 -Werte der Logistischen Regressionen sehr klein aus, die Effekte fallen also recht gering aus.

Tabelle 18: Logistische Regressionen: jeweils für generellen und gewalttätigen Rückfall

N = 244	Genereller Rückfall		Gewalttätiger Rückfall	
	exp (B)	95%KI	exp (B)	95%KI
Scores:				
PCL:SV-Summenscore	1.122***	1.059-1.189	1.097**	1.029-1.170
PCL:SV-Faktor 1: psychopathic traits	1.155**	1.046-1.275	1.072	0.956-1.203
PCL:SV-Faktor 2: social deviance	1.217***	1.103-1.342	1.232***	1.102-1.377

* $p < .05$; ** $p < .01$; *** $p < .001$; n (genereller Rückfall) = 167, n (gewalttätiger Rückfall) = 38.

4.3.5 Prognoserelevante Kennwerte der PCL:SV-Gesamt

Es folgt nun die Darstellung der prognoserelevanten Kennwerte im Hinblick auf jeden möglichen *cut-off* Wert der PCL:SV-Gesamt, sowohl für die Vorhersage genereller als auch für die Vorhersage gewalttätiger Rückfälle. Neben Spezifität, Sensitivität sowie *PPP*, *NPP* und Gesamtfehler zeigen die Tab. 19 und 20 auch die zugehörige Basis- und Selektionsrate sowie die Korrelation bzw. den ϕ -Koeffizienten (der ϕ -Koeffizient der durch den *cut-off* Wert gegebenen Vier-Felder-Tafel ist identisch der Korrelation der durch den *cut-off* Wert bestimmten beiden binären Variablen).

¹⁹ Pseudo- R^2 nach Cox & Snell.

Tabelle 19: Prognoserelevante Kennwerte im Hinblick auf jeden möglichen *cut-off* Wert c der PCL:SV-Gesamt zur Vorhersage genereller Rückfälle (d. h. für alle Entscheidungsregeln der Form: Prognose eines generellen Rückfalls *cut-off* Wert $\geq c$).

PCL:SV-Gesamt	Spezifität in %	Sensitivität in %	PPP in %	NPP in %	% Fehler insgesamt	Basisrate in %	Selektions- rate in %	φ/r
0	0	100	68.4	n.d.	31.6	68.4	100	n.d.
1	3.9	97.0	68.6	37.5	32.4	68.4	96.7	0.02
2	10.4	96.4	70.0	57.1	30.7	68.4	94.3	0.14
3	18.2	92.2	70.9	51.8	31.1	68.4	88.9	0.15
4	28.6	89.2	73.0	55.0	30.0	68.4	83.6	0.22
5	37.7	84.4	74.6	52.7	30.3	68.4	77.5	0.25
6	48.1	77.8	76.5	50.0	31.6	68.4	69.7	0.26
7	54.5	70.1	76.9	45.7	34.8	68.4	62.3	0.24
8	59.7	63.5	77.4	43.0	37.7	68.4	56.1	0.22
9	62.3	55.7	76.2	39.3	42.2	68.4	50.0	0.17
10	72.7	47.9	79.2	39.2	44.3	68.4	41.4	0.19
11	77.9	40.1	79.8	37.5	48.0	68.4	34.4	0.18
12	79.2	34.1	78.1	35.7	51.6	68.4	30.0	0.14
13	80.5	31.7	77.9	35.2	52.9	68.4	27.9	0.13
14	89.6	25.7	84.3	35.8	54.1	68.4	21.0	0.18
15	90.9	21.0	83.3	34.7	57.0	68.4	17.2	0.15
16	96.1	18.0	91.0	35.1	57.4	68.4	13.5	0.19
17	97.4	11.4	90.4	33.6	61.5	68.4	8.6	0.15
18	98.7	9.6	94.1	33.5	62.3	68.4	7.0	0.15
19	100	9.0	100	33.6	62.3	68.4	6.1	0.17
20	100	8.4	100	33.5	62.7	68.4	5.7	0.17
21	100	3.6	100	32.4	66.0	68.4	2.5	0.11
22	100	1.8	100	32.0	67.2	68.4	1.2	0.08
23	100	0.6	100	31.7	68.0	68.4	0.4	0.04
24	100	0	n.d.	31.6	68.4	68.4	0	n.d.

Anmerkung: n.d.: wegen Division durch Null-Wert nicht definiert.

Abb. 6 liefert eine graphische Veranschaulichung der Kennwerte aus Tab. 19. Auf der x-Achse sind die *cut-off* Werte der PCL:SV-Gesamt abgetragen und auf der y-Achse stehen für alle Kurven – mit Ausnahme der „ φ -Kurve“ – die relativen Häufigkeiten. Für die „ φ -Kurve“ entsprechen die Angaben auf der y-Achse der Höhe des φ -Koeffizienten bzw. Korrelationskoeffizienten.

Gemäß den theoretischen Erörterungen in Abschnitt 1.5 ergibt sich: Durchläuft c alle Werte von 0 bis 24, wächst die Spezifität, von Null kommend bis auf Eins an, während die Sensitivität, von Eins startend, abfällt und sich der Null annähert. Die Fehlerrate beginnt mit dem Wert 1 – Basisrate (.32) und nähert sich mit wachsendem c schließlich der Basisrate an (.68). Die *PPP* startet bei der Basisrate und steigt dann bis zur Eins an, während die *NPP* bis auf 1 – Basisrate abfällt. Die Korrelationen (Validitäten) sind bei *cut-off* Werten zwischen 5 und 6 am höchsten, bei vergleichsweise geringer Fehlerrate.

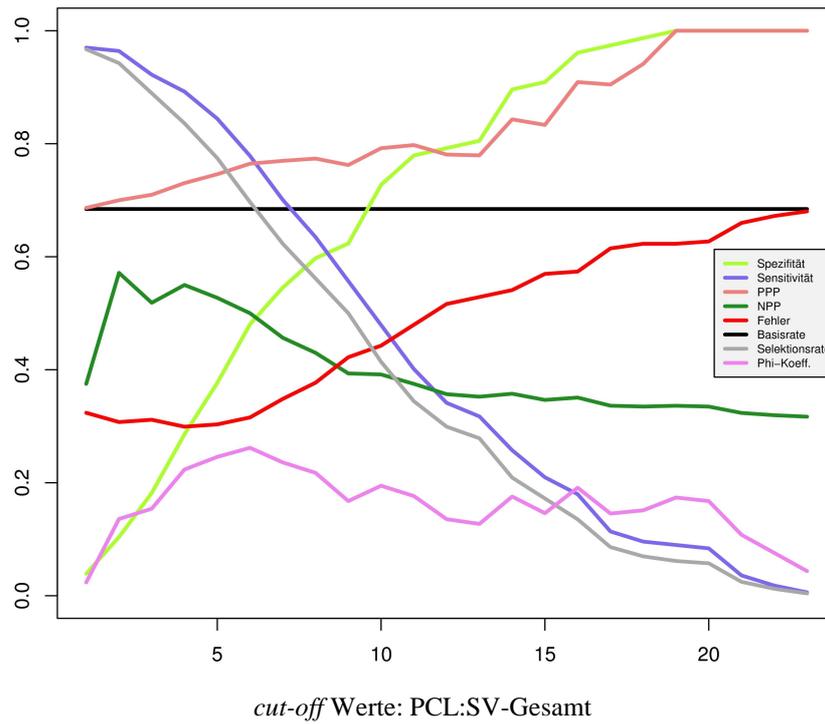


Abbildung 6: Kennwertverlauf bei variierendem *cut-off* Wert für generellen Rückfall (Graphische Veranschaulichung der Kennwerte aus Tab. 19)

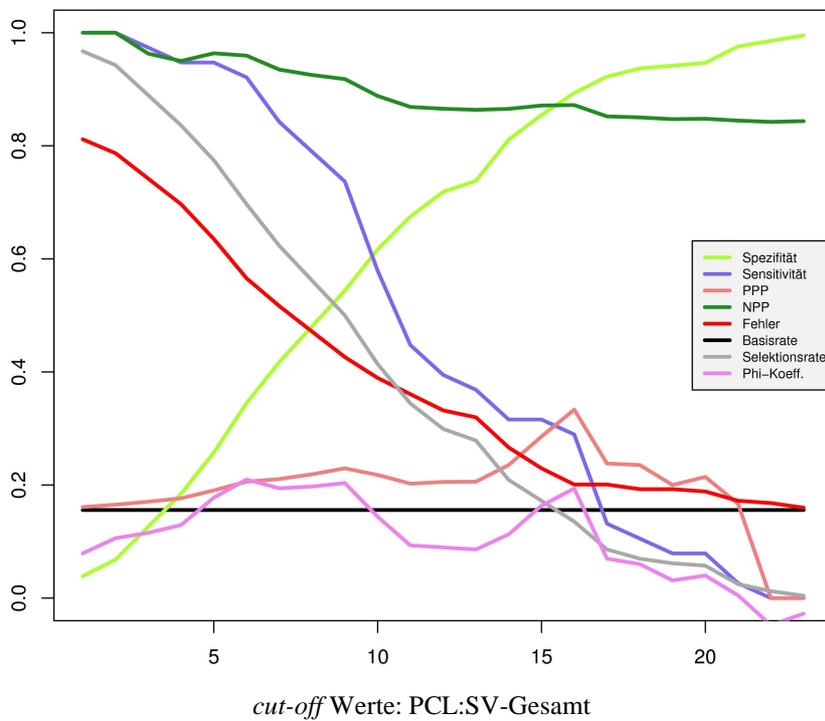


Abbildung 7: Kennwertverlauf bei variierendem *cut-off* Wert für gewalttätigen Rückfall (Graphische Veranschaulichung der Kennwerte aus Tab. 20)

Tabelle 20: Prognoserelevante Kennwerte im Hinblick auf jeden möglichen *cut-off* Wert der PCL:SV-Gesamt zur Vorhersage gewalttätiger Rückfälle (d. h. für alle Entscheidungsregeln der Form: Prognose eines gewalttätigen Rückfalls *cut-off* Wert $\geq c$).

PCL:SV-Gesamt	Spezifität in %	Sensitivität in %	PPP in %	NPP in %	% Fehler insgesamt	Basisrate in %	Selektionsrate in %	ϕ/r
0	0	100	15.6	n.d.	84.4	15.6	100	n.d.
1	3.9	100	16.1	100	81.1	15.6	96.7	0.08
2	6.8	100	16.5	100	78.7	15.6	94.3	0.12
3	12.6	97.4	17.1	96.3	74.2	15.6	88.9	0.12
4	18.4	94.7	17.6	95.0	69.7	15.6	83.6	0.13
5	25.7	94.7	19.0	96.4	63.5	15.6	77.5	0.18
6	34.5	92.1	20.6	95.9	56.6	15.6	69.7	0.21
7	41.7	84.2	21.1	93.5	51.6	15.6	62.3	0.19
8	48.1	78.9	21.9	92.5	47.1	15.6	56.1	0.20
9	54.4	73.7	23.0	91.8	42.6	15.6	50.0	0.20
10	61.7	57.9	21.8	88.8	38.9	15.6	41.4	0.14
11	67.5	44.7	20.2	86.9	36.1	15.6	34.4	0.09
12	71.8	39.5	20.5	86.5	33.2	15.6	30.0	0.09
13	73.8	36.8	20.6	86.4	32.0	15.6	27.9	0.09
14	81.1	31.6	23.5	86.5	26.6	15.6	20.9	0.11
15	85.4	31.6	28.6	87.1	23.0	15.6	17.2	0.16
16	89.3	28.9	33.3	87.2	20.1	15.6	13.5	0.19
17	92.2	13.2	23.8	85.2	20.1	15.6	8.6	0.07
18	93.7	10.2	23.5	85.0	19.3	15.6	7.0	0.06
19	94.2	7.9	20.0	84.7	19.3	15.6	6.1	0.03
20	94.7	7.9	21.4	84.8	18.9	15.6	5.7	0.04
21	97.6	2.6	16.7	84.5	17.2	15.6	2.5	0.00
22	98.5	0	0	84.2	16.8	15.6	1.2	-0.05
23	99.5	0	0	84.4	16.0	15.6	0.4	-0.03
24	100	0	n.d.	84.4	15.6	15.6	0	n.d.

Anmerkung: n.d.: wegen Division durch Null-Wert nicht definiert.

Erwartungsgemäß stehen auch bei der Prognose eines gewalttätigen Rückfalls Spezifität und Sensitivität in einem inversen Verhältnis zueinander und der Gesamtfehler nähert sich mit zunehmendem *cut-off* Wert der Basisrate an (siehe Tab. 20). Allerdings ist bei der Interpretation seiner Höhe zu beachten, dass die Basisrate nun wesentlich kleiner als bei der Prognose genereller Rückfälle ist. Durchläuft der *cut-off* Wert c die Werte von 0 bis 24, so beginnt die *NPP* bei 1, fällt ab und endet bei $1 - \text{Basisrate}$ (hier: 84.4%). Aufgrund der geringeren Basisrate erreicht die *NPP* hier höhere Werte als die *PPP* (bei der Prognose eines generellen Rückfalls war es genau umgekehrt). In Abb. 7 wurden die Verläufe der einzelnen Kennwerte ebenfalls graphisch veranschaulicht. Wie schon in Abb. 6 sind auf der x-Achse die einzelnen *cut-off* Werte und auf der y-Achse die relativen Häufigkeiten abgetragen (wieder ausgenommen: „ ϕ -Kurve“).

4.3.6 Prognoserelevante Kennwerte für die Subskalen der PCL:SV

Aus Gründen der Vollständigkeit der Darstellung sollen hier auch die prognoserelevanten Kennwerte im Hinblick auf jeden möglichen *cut-off* Wert der beiden Subskalen der PCL:SV berichtet werden. Zunächst folgen in den Tab. 21 und Tab. 22 die Kennwerte für die PCL:SV-PP sowohl für generelle als auch für gewalttätige Rückfälle. Im Anschluss daran werden in den Tab. 23 und Tab. 24 die entsprechenden Werte für die PCL:SV-SD zusammenfassend dargestellt.

Tabelle 21: Prognoserelevante Kennwerte im Hinblick auf jeden möglichen *cut-off* Wert der PCL:SV-PP zur Vorhersage genereller Rückfälle (d. h. für alle Entscheidungsregeln der Form: Prognose eines generellen Rückfalls *cut-off* Wert $\geq c$).

PCL:SV-PP	Spezifität in %	Sensitivität in %	PPP in %	NPP in %	% Fehler insgesamt	Basisrate in %	Selektions- rate in %	ϕ/r
0	0	100	68.4	n.d.	31.6	68.4	100	n.d.
1	11.7	92.8	69.5	42.9	32.8	68.4	91.4	0.07
2	25.9	81.4	70.5	39.2	36.1	68.4	79.1	0.08
3	41.6	70.7	72.4	39.5	38.5	68.4	66.8	0.12
4	54.5	61.7	74.6	39.6	40.6	68.4	56.6	0.15
5	68.8	50.3	77.8	38.9	43.9	68.4	44.3	0.17
6	75.3	37.1	76.5	35.6	50.8	68.4	33.2	0.12
7	83.1	29.3	79.0	35.2	53.7	68.4	25.4	0.13
8	92.2	19.8	84.6	34.6	57.4	68.4	15.9	0.15
9	96.1	13.8	88.5	33.9	60.2	68.4	10.7	0.14
10	100	6.6	100	33.0	63.9	68.4	4.5	0.14
11	100	4.2	100	32.5	65.6	68.4	2.9	0.11
12	100	1.2	100	31.8	67.6	68.4	0.8	0.06

Tabelle 22: Prognoserelevante Kennwerte im Hinblick auf jeden möglichen *cut-off* Wert der PCL:SV-PP zur Vorhersage gewalttätiger Rückfälle (d. h. für alle Entscheidungsregeln der Form: Prognose eines gewalttätigen Rückfalls *cut-off* Wert $\geq c$).

PCL:SV-PP	Spezifität in %	Sensitivität in %	PPP in %	NPP in %	% Fehler insgesamt	Basisrate in %	Selektions- rate in %	ϕ/r
0	0	100	15.6	n.d.	84.4	15.6	100	n.d.
1	8.7	92.1	15.7	85.7	78.3	15.6	91.4	0.01
2	22.3	86.8	17.1	90.2	67.6	15.6	79.1	0.08
3	35.9	81.6	19.0	91.4	56.9	15.6	66.8	0.13
4	45.6	68.4	18.8	88.7	50.8	15.6	56.6	0.10
5	57.8	55.3	19.4	87.5	42.6	15.6	44.3	0.09
6	67.9	39.5	18.5	85.8	36.5	15.6	33.2	0.06
7	74.3	23.7	14.5	84.1	33.6	15.6	25.4	-0.02
8	83.9	15.8	15.4	84.4	26.6	15.6	15.9	-0.00
9	89.3	10.5	15.4	84.4	22.9	15.6	10.7	-0.00
10	95.6	5.3	18.2	84.5	18.4	15.6	4.5	0.02
11	97.1	2.6	14.3	84.4	17.6	15.6	2.9	-0.01
12	99.5	2.6	50.0	84.7	15.6	15.6	0.8	0.09

Tabelle 23: Prognoserelevante Kennwerte im Hinblick auf jeden möglichen *cut-off* Wert der PCL:SV-SD zur Vorhersage genereller Rückfälle (d. h. für alle Entscheidungsregeln der Form: Prognose eines generellen Rückfalls *cut-off* Wert $\geq c$).

PCL:SV-SD	Spezifität in %	Sensitivität in %	PPP in %	NPP in %	% Fehler insgesamt	Basisrate in %	Selektions- rate in %	φ/r
0	0	100	68.4	n.d.	31.6	68.4	100	n.d.
1	12.9	94.6	70.2	52.6	31.1	68.4	92.2	0.13
2	28.6	85.0	72.1	46.8	32.8	68.4	80.7	0.16
3	49.4	78.4	77.1	51.4	30.7	68.4	69.7	0.28
4	61.0	67.1	78.9	46.1	34.8	68.4	58.2	0.26
5	71.4	53.8	80.4	41.7	40.6	68.4	45.9	0.24
6	74.0	44.3	78.7	38.0	46.3	68.4	38.5	0.18
7	83.1	33.5	81.2	36.6	50.8	68.4	28.3	0.17
8	89.6	24.6	83.7	35.4	54.9	68.4	20.1	0.16
9	92.2	17.9	83.3	34.1	58.6	68.4	14.8	0.13
10	96.1	13.2	88.0	33.8	60.7	68.4	10.2	0.14
11	98.7	5.4	90.0	32.5	65.2	68.4	4.1	0.09
12	100	2.4	100	32.1	66.8	68.4	1.6	0.09

Tabelle 24: Prognoserelevante Kennwerte im Hinblick auf jeden möglichen *cut-off* Wert der PCL:SV-SD zur Vorhersage gewalttätiger Rückfälle (d. h. für alle Entscheidungsregeln der Form: Prognose eines gewalttätigen Rückfalls *cut-off* Wert $\geq c$).

PCL:SV-SD	Spezifität in %	Sensitivität in %	PPP in %	NPP in %	% Fehler insgesamt	Basisrate in %	Selektions- rate in %	φ/r
0	0	100	15.6	n.d.	84.4	15.6	100	n.d.
1	9.2	100	16.9	100	76.6	15.6	92.2	0.12
2	21.4	92.1	17.8	93.6	67.6	15.6	80.7	0.12
3	34.5	92.1	20.6	95.9	56.6	15.6	69.7	0.21
4	47.1	86.8	23.2	95.1	46.7	15.6	58.2	0.25
5	59.7	76.3	25.8	93.2	37.7	15.6	45.9	0.26
6	66.0	63.2	25.5	90.7	34.4	15.6	38.5	0.22
7	75.2	47.4	26.1	88.6	29.1	15.6	28.3	0.18
8	82.5	34.2	26.5	87.2	25.0	15.6	20.1	0.15
9	87.4	26.3	27.8	86.5	22.1	15.6	14.7	0.14
10	91.3	18.4	28.0	85.8	20.1	15.6	10.2	0.12
11	96.1	5.3	20.0	84.6	18.0	15.6	4.1	0.03
12	98.5	2.6	25.0	84.6	16.4	15.6	1.6	0.03

4.3.7 Weiterführende Datenanalysen

Im Anschluss an die Auswertungen der zentralen Analysen zur PCL:SV wurde eine logistische Regression mit den Variablen „genereller Rückfall“ als Kriterium und den Variablen „PCL:SV-Gesamt“, „Anzahl der Vorstrafen“ und „Alter bei erster Verurteilung“ als Prädiktoren durchgeführt. Hierbei ging es um die Beantwortung der Frage, ob sich durch Hinzunahme weiterer prognoserelevanter Merkmale die Modellanpassung verbessert. Die Variable „Alter bei erster Verurteilung“ konnte als Prädiktor ausgeschlossen werden, da ihr Einschluss die Vorhersage nicht nennenswert verbesserte. Die Ergebnisse der Modellanpassung für die bei-

den Prädiktoren „PCL:SV-Gesamt“ und „Anzahl der Vorstrafen“ sind in Tab. 25 zusammengefasst.

Tabelle 25: Logistische Regression mit dem (binären) Kriterium genereller Rückfall und den Prädiktoren PCL:SV-Gesamt und Anzahl der Vorstrafen. ^a

N = 244	B	exp(B)	z-Wert	p-Wert
Konstante	-0.690	0.502	-2.20	0.028*
PCL:SV-Gesamt	0.101	1.110	3.28	0.001**
Anzahl der Vorstrafen	0.096	1.100	3.06	0.002**

^aAnalyse ohne Straftäter Nr. 241, da Cooks distance auffällig; *p < .05 ; **p < .01; ***p < .001.

5 Diskussion

Das primäre Ziel der vorliegenden Untersuchung war es, die PCL:SV hinsichtlich ihrer prädiktiven Validität zur Vorhersage krimineller Rückfälle an einer heterogenen deutschen Straftäterpopulation zu evaluieren und die Ergebnisse mit den bisherigen Befunden aus der wissenschaftlichen Literatur zu vergleichen. Dazu wurden 244 Straftäter aus Haftanstalten in Sachsen-Anhalt untersucht.

Die nachfolgenden Interpretationen stehen selbstverständlich unter Vorbehalt der in Abschnitt 3.3 formulierten Beschränkungen bei der Erhebung der Stichprobe, die sich dadurch ergaben, dass nur solche Probanden in die Studie eingeschlossen werden konnten, die ihre Einwilligung zur Auswertung ihrer BZR-Auszüge erteilt hatten und die zudem fünf Jahre nach Abschluss der ersten Datenerhebung von uns erreicht werden konnten. Die Ergebnisse der Untersuchung sollen unter acht Hauptpunkten zusammenfassend erörtert und bewertet werden:

- (1) Prädiktive Validität der PCL:SV,
- (2) PCL:SV-Gesamt und krimineller Rückfall,
- (3) PCL:SV-Gesamt, *Psychopathy* und die Identifikation einer Hochrisikogruppe,
- (4) PCL:SV-SD bzw. PCL:SV-PP und krimineller Rückfall,
- (5) Soziobiographische und psychopathologische Merkmale und krimineller Rückfall,
- (6) Merkmale der strafrechtlichen Vorgeschichte und krimineller Rückfall,
- (7) Komplexere Entscheidungsregeln und
- (8) Limitierungen der Untersuchung.

5.1 Prädiktive Validität der PCL:SV

Um die Vergleichbarkeit mit internationalen Befunden zu ermöglichen, wurden die prädiktiven Validitäten der PCL:SV-Scores sowohl mit Hilfe von ROC-Analysen als auch mit Hilfe von Logistischen Regressionen bestimmt. Zunächst kann positiv festgestellt werden, betrachtet man die ermittelten Größen der AUCs, dass die Ergebnisse weitgehend denen anderer Studien entsprechen. So erzielten die PCL:SV-Gesamt sowie ihre Subskalen (PCL:SV-PP und PCL:SV-SD) ähnliche Vorhersageleistungen, wie sie aus internationalen Publikationen (sowohl für die PCL-R als auch für die PCL:SV) bisher berichtet wurden (Grann et al. 1999, Hare et al. 2000, Laurell & Daderman 2005, Salekin et al. 1996, Urbaniok et al. 2007a, Urbaniok et al. 2007b). Dies ist umso erstaunlicher, wenn man berücksichtigt, dass die in den zitierten Studien untersuchten Populationen recht heterogen sind.

Auf der Grundlage der eigenen statistischen Auswertungen darf davon ausgegangen werden, dass PCL:SV-Gesamt, PCL:SV-PP und PCL:SV-SD mit der Variablen Rückfall einen, wenn auch nicht sehr hohen, Zusammenhang aufweisen. ROC-Analysen – und Logistische Regressionen – kommen, wie zu erwarten konkordant, zu dem Ergebnis, dass sich die Gruppen „genereller Rückfall vs. kein Rückfall“ sowie die Gruppen „gewalttätiger Rückfall vs. kein gewalttätiger Rückfall“ im Prinzip statistisch trennen lassen.

Zur Vorhersage eines generellen Rückfalls lassen sich im Prinzip alle drei PCL:SV-Scores als prognostische „Verfahren“ einsetzen. Zudem sind sie aufgrund ihrer vergleichbaren AUCs in ihrer prognostischen Güte als gleich gut einzuschätzen (PCL:SV-Gesamt: geschätzte AUC = .660; PCL:SV-PP: geschätzte AUC = .609; PCL:SV-SD: geschätzte AUC = .666). Etwas anders liegen die Dinge für die Variable „gewalttätiger Rückfall“. Hier lassen die unterschiedlichen Größen der AUCs vermuten, dass die PCL:SV-SD (geschätzte AUC = .700) hinsichtlich ihrer prognostischen Güte angemessener als die PCL:SV-Gesamt (geschätzte AUC = .660) ist. Die Bedeutung von PCL:SV-PP (geschätzte AUC = .569) für die Vorhersage gewalttätiger Rückfälle hingegen erscheint unklar. Denn sowohl das nicht signifikante Ergebnis der ROC-Analyse bzw. das der Logistischen Regression, legen nahe, dass dieser Faktor zur gewalttätigen Rückfallprognose weniger tauglich ist.²⁰

Die in den Analysen auftretenden statistischen Signifikanzen sind – was deren praktische Bedeutsamkeit angeht – allerdings mit einiger Zurückhaltung zu interpretieren. Denn die geschätzten AUCs und Regressionskoeffizienten zeigen, dass die entsprechenden Tests zwar auf das Vorhandensein von Effekten hinweisen, aber auf Effekte, die aus praktischer Sicht nicht sonderlich groß sind.

Gleichwohl werden die in der Literatur berichteten AUC-Werte, die für den PCL-Gesamtscore für unterschiedliche Straftäterpopulationen und Kriteriumsvariablen (beispielsweise genereller oder gewalttätiger Rückfall) bei Werten zwischen .61 und .75 liegen als zufriedenstellend hoch angesehen. Betrachtet man hierzu noch einmal die Schätzung der AUC für die PCL:SV-Gesamt für die Vorhersage eines generellen Rückfalls, die in der vorliegenden Arbeit 0.66 beträgt, so könnte man, folgt man den in der Literatur verwendeten Standards, auch das Ergebnis dieser Untersuchung als *moderat gut* bewerten. Die Grundlage für diese verhältnismäßig positive Einschätzung der AUC-Werte und damit auch der prädiktiven Vali-

²⁰ Die exp(B)-Werte sollen, da sie im Vergleich zu den AUC-Werten im Prinzip keine wesentlich neuen Informationen liefern, hier nicht noch einmal explizit aufgeführt werden (vgl. dazu Tab. 18).

dität sind allerdings erstaunlicherweise nicht etwa – wie zu erwarten wäre – forensische Überlegungen, wie beispielsweise sorgfältige Güterabwägungen von Straftäterrechten einerseits und Interessen der Allgemeinheit andererseits, sondern ist vielmehr eine mehr oder weniger willkürlich von Cohen (1988) vorgeschlagene Klassifikation von Effektgrößen, die für die meisten Anwendungsfälle eher untauglich ist. Denn unterschiedliche Forschungs- und Anwendungsbereiche erfordern und entwickeln sinnvollerweise ganz eigene Normen zur Bewertung von Effektgrößen. So werden in Forschungsbereichen, in denen der Anteil der Fehlervarianz an der Gesamtvarianz erhobener Daten in der Regel klein ausfällt, auch eher hohe Ansprüche an die Größe der erwarteten Effekte gestellt werden müssen. Auch wird man in jenen Anwendungsbereichen, in denen Fehlentscheidungen bei der Verwendung aktuarischer Prognoseinstrumente zu erheblichen ethischen oder wirtschaftlichen Kosten führen können, die Anforderungen an die Effektgröße (an die Größe der AUC) höher ansetzen als in Fällen, bei denen es lediglich um die Entdeckung von Mittelwertsunterschieden geht. Das Cohensche Klassifikationsschema zur Bewertung der Höhe von AUC-Werten scheint daher – zumindest für die hier intendierten Anwendungsbereiche – ein eher unangemessener und wenig hilfreicher Maßstab zu sein.

Darüber hinaus reicht es für eine Einschätzung der prädiktiven Validität der PCL:SV sicherlich nicht aus, lediglich die Höhe der gefundenen AUC-Maße (bzw. $\exp(B)$ -Werte) zu diskutieren. Vielmehr ist die Wirkungsweise dieses Verfahrens für konkrete Entscheidungsregeln, die durch die Festlegung eines spezifischen *cut-off* Wertes ausgewählt werden, im Einzelnen – unter Einbeziehung der Rolle der Basisrate – zu untersuchen. Zudem ist für eine Evaluation der prognostischen Güte der PCL:SV der direkte Vergleich mit anderen prognoserelevanten Merkmalen hilfreich.

Die weitere Diskussion wird sich daher mit diesen Punkten auseinandersetzen, wobei die Erörterung sich aus Gründen der Übersichtlichkeit der Darstellung in erster Linie auf die PCL:SV-Gesamt und die Vorhersage eines generellen Rückfalls beschränken wird. Die nachfolgende Diskussion gilt in analoger Weise auch für die Vorhersage eines gewalttätigen Rückfalls.

5.2 PCL:SV-Gesamt und krimineller Rückfall

Die (geschätzte) Basisrate für einen generellen Rückfall beträgt in der vorliegenden Untersuchung 68 %. Für eine anwendungsbezogene Bewertung der prognostischen Güte der PCL:SV-Gesamt in einem Kontext mit einer derartigen Basisrate sollen nun die mit möglichen *cut-off*

Werten gegebenen prognostischen Kennwerte diskutiert werden. Primäres Ziel ist dabei die Auswahl einer „vernünftigen“ Entscheidungsregel. Das Kriterium für eine „vernünftige“ Regel soll eine möglichst hohe Korrelation zwischen Prognose und tatsächlichem Verhalten (Realität) bei möglichst kleiner Fehlerrate sein. Wie man Tab. 19 entnehmen kann, erfüllt ein *cut-off* Wert von 6 annähernd diese Bedingungen. Für diesen *cut-off* Wert ist die Korrelation am größten, bei vergleichsweise kleiner Fehlerrate. Eine Rückfallprognose würde nach der mit diesem *cut-off* Wert gegebenen Entscheidungsregel genau dann erfolgen, wenn der PCL:SV-Gesamt größer oder gleich 6 ist. Als Abkürzung für diese Entscheidungsregel soll im Folgenden der Ausdruck „PCL:SV-Gesamt \geq 6“ verwendet werden. Analog soll bei den im Weiteren diskutierten Entscheidungsregeln verfahren werden.

Die Fehlerrate beträgt bei dieser Entscheidungsregel fast 32 %, d. h., es wären 32 von 100 Prognosen falsch. Dies mag auf den ersten Blick akzeptabel erscheinen, allerdings zeigt ein weiterer Blick, dass man aufgrund der hier vorliegenden hohen Basisrate von 68 % bei Anwendung einer Entscheidungsregel, die jeden Straftäter – unabhängig von seinem PCL:SV-Gesamtscore – in uniformer Weise als Rückfälligen klassifiziert, ebenfalls auf eine Fehlerrate von 32 % (= 100 – Basisrate) käme. Die mit einem *cut-off* Wert von 6 verknüpfte Entscheidungsregel – bewertet man sie allein im Hinblick auf die Fehlerrate – bringt somit keinerlei Gewinn gegenüber jener „starren“ Regel, die *jeden* Straftäter als Rückfälligen klassifiziert. Es muss allerdings betont werden, dass die Fehlprognosen bei Anwendung der „starren“ Regel einseitig verteilt sind. Es werden zwar *alle* Rückfälligen korrekt prognostiziert, aber es werden auch *alle* Nicht-Rückfälligen fälschlicherweise als rückfällig eingestuft. Demgegenüber sind die Fehlerraten bei Verwendung der Entscheidungsregel „PCL:SV-Gesamt \geq 6“ auf die beiden Fehlertypen in etwa gleich verteilt. Auch wenn man aus diesem Grund der Entscheidungsregel „PCL:SV-Gesamt \geq 6“ den Vorzug vor der „starren“ Regel geben sollte, ist es kaum einleuchtend, eine Entscheidungsregel, die im Prinzip keinerlei Gewinn gegenüber der „starren“ Regel aufweist, als hinreichend gut zu bewerten. Auch die Kennwerte *PPP* und *NPP* bestärken die Zurückhaltung gegenüber einer zu positiven Bewertung der prädiktiven Güte der Entscheidungsregel „PCL:SV-Gesamt \geq 6“: Von den Personen, die als Rückfällige prognostiziert werden, werden etwa 77 % richtig klassifiziert (*PPP*); aber unter den als Nicht-Rückfällige Prognostizierten befinden sich 50 % Rückfällige (vgl. dazu auch noch einmal Tab. 19).

Es stellt sich die Frage, ob die prognostische Nützlichkeit der PCL:SV-Gesamt bei einer anderen Basisrate höher wäre? Um diese Frage zu beantworten, soll die folgende Situation ange-

nommen werden: Die ROC-Kurve (und damit auch AUC) sei dieselbe, allerdings jetzt bei einer Basisrate von 30 %. In diesem Fall wäre eine günstige Entscheidungsregel „*PCL:SV-Gesamt* ≥ 19 “. Man würde bei dieser Regel also genau dann einen Straftäter als Rückfälligen prognostizieren, wenn dessen *PCL:SV-Gesamt* größer oder gleich 19 ist. Bei dieser Entscheidungsregel betrüge bei einer Basisrate von 30 % die Fehlerrate etwa 28 % und wäre damit etwas geringer als die Fehlerrate der oben diskutierten Regel „*PCL:SV-Gesamt* ≥ 6 “. Allerdings ist auch hier zu bedenken, dass eine „starre“ Regel, die jetzt dadurch definiert ist, dass *jeder* Straftäter als Nicht-Rückfälliger klassifiziert wird, eine Fehlerrate von 30 % besitzt. Der Gewinn der Regel „*PCL:SV-Gesamt* ≥ 19 “ gegenüber der „starren“ Regel beläuft sich also nur auf 2 %. Die *PPP* betrüge 100 %, was bedeutet, dass alle Straftäter, die man als Rückfällige prognostizierte, auch rückfällig würden. Die *NPP* läge bei etwa 72 %, d. h., dass von allen als Nicht-Rückfällige prognostizierten Straftätern 28 % rückfällig würden (zum Vergleich: 50 % bei einer Basisrate von 68 % und bei Verwendung der Regel „*PCL:SV-Gesamt* ≥ 6 “). Es zeigt sich somit, dass bei einer Basisrate von 30 % eine Entscheidungsregel gewählt werden könnte, die leicht akzeptablere Eigenschaften aufweist als die weiter oben diskutierte Regel „*PCL:SV-Gesamt* ≥ 6 “.

5.3 *PCL:SV-Gesamt, Psychopathy* und die Identifikation einer Hochrisikogruppe

In der Literatur wird verschiedentlich hervorgehoben, dass, wenn die *PCL:SV* auch nicht in jeder Hinsicht als Instrument zur Rückfallprognose zu überzeugen vermag, ein *PCL:SV-Gesamt* von ≥ 18 (Diagnose einer *Psychopathy*) sich immerhin zur Identifikation einer hochbrisanten Subgruppe eignet, von der angenommen wird, dass sie zu wiederholt straffälligem Verhalten neigt (Dietiker et al. 2007, Douglas et al. 1997, Hare 1996, Urbaniok et al. 2007a, zusammenfassend z. B. Cooke 1998).

Aus diesem Grund soll die Entscheidungsregel „*PCL:SV-Gesamt* ≥ 18 “ etwas eingehender analysiert werden. In der Tat ist die *PPP* in diesem Fall sehr hoch: Von den als „wird rückfällig“ Prognostizierten werden tatsächlich etwa 94 % rückfällig. Allerdings zeigt die zugehörige *NPP*, dass fast 67 % der als „wird nicht rückfällig“ Klassifizierten rückfällig werden! Die Fehlerrate dieser Entscheidungsregel liegt bei 62 %! Dabei ist die Rate der Falsch-Positiven nahezu Null (0.4 %) und die Rate der Falsch-Negativen fast 62 %. Die hohe *PPP* wird also durch eine sehr hohe Rate von Falsch-Negativen erkauft.

Die Werte der Kenngrößen *PPP* (94%) und *NPP* (33%!) vor Augen könnte man mit guten Gründen die Regel dahingehend abändern, dass – unabhängig vom jeweiligen *PCL:SV-*

Gesamtscore – stets die Prognose „wird rückfällig“ abgegeben wird, da in jedem der beiden möglichen Fälle ($PCL:SV\text{-Gesamt} \geq 18$ und $PCL:SV\text{-Gesamt} < 18$) die Wahrscheinlichkeit dafür, dass der untersuchte Straftäter rückfällig wird, deutlich über 50% liegt. Das wäre gerade die weiter oben im Zusammenhang mit einem *cut-off* Wert von 6 bereits diskutierte „starre“ Regel, die mit einer Fehlerrate von 32 % einhergeht. Die Entscheidungsregel „ $PCL:SV\text{-Gesamt} \geq 18$ “ kann also – zumindest bei der vorliegenden Straftätergruppe mit entsprechender Basisrate für Rückfälligkeit – nicht überzeugen.

Auch die absoluten Zahlen zeigen, was die Regel „ $PCL:SV\text{-Gesamt} \geq 18$ “ praktisch bedeutet: Von 100 zu prognostizierenden Fällen würden 7 (davon 6 korrekt) als Rückfällige klassifiziert; 93 Straftäter würden als Nicht-Rückfällige prognostiziert, davon 31 fälschlicherweise. Es ginge hier also lediglich um zirka 9 % aller Rückfälligen, die man als zur „Hochrisikogruppe“ gehörend erfolgreich „entdecken“ würde, denen aber 91 % von Rückfälligen gegenüberstünden, die fälschlicherweise als Nicht-Rückfällige eingestuft würden. Würde man die Regel „ $PCL:SV\text{-Gesamt} \geq 18$ “ wegen ihrer hohen *PPP* (94 %) dennoch verwenden wollen, so könnte man das eigentlich nur dann tun, wenn diese Regel lediglich Teil eines zweistufigen Prognoseprozesses ist, bei dem in einem zweiten Schritt in der Gruppe der durch die Regel als „wird nicht rückfällig“ Klassifizierten eine möglichst gute Trennung der Nicht-Rückfälligen von den Rückfälligen vorgenommen wird.²¹

Betrachten wir noch einmal die mutmaßlich positive Eigenschaft der $PCL:SV\text{-Gesamt}$, bei Verwendung der Regel „ $PCL:SV\text{-Gesamt} \geq 18$ “, die für die Rückfallprognose relevante „Hochrisikogruppe“ der *Psychopaths* zu detektieren. Hierzu sollte man sich zunächst fragen, unter welchen Bedingungen eine hohe *PPP* zustande kommt. Wenn die *AUC* größer als 0.5 ist (das bedeutet, dass d größer Null ist), ist die Verteilung der Rückfälligen relativ zu der der Nicht-Rückfälligen nach rechts verschoben. Es ist also zu erwarten, dass es im oberen Bereich der Skala einen *cut-off* Wert c gibt, so dass die Wahrscheinlichkeit für das Auftreten von $PCL:SV\text{-Gesamtscores}$, die größer oder gleich diesem Wert c sind, in der Gruppe der Rückfälligen deutlich größer ist als in der Gruppe der Nicht-Rückfälligen. Das heißt, dass die mit einem solchen *cut-off* Wert c verknüpfte Sensitivität größer als der Wert $1 - \text{Spezifität}$ ist. Wenn

²¹ Eine Entscheidungsregel allein wegen ihrer sehr hohen *PPP* zu empfehlen kann riskant sein. Denn es sind Konstellationen mit einer sehr hohen *PPP* aber mit einem ϕ -Koeffizienten von Null denkbar. Man betrachte beispielsweise die nachstehende Tafel.

	R_0	R_1
r_0	8	72
r_1	2	18

Diese besitzt eine *PPP* von 90%, aber der ϕ -Koeffizient ist Null. D. h., die Prognose ist unabhängig vom Ereignis Rückfall/Nicht-Rückfall.

das Verhältnis Sensitivität : (1 – Spezifität) größer wird, wächst auch die *PPP*. Dieser Effekt wird durch die hohe Basisrate noch gesteigert (vgl. Kap. 1.5). Die eben erörterten Gründe, die eine hohe *PPP* garantieren, sollte man in gewissem Maße jedoch von *jedem* akzeptablen Prognoseinstrument erwarten können. Andernfalls wäre ein solches Instrument nämlich kaum in der Lage, die Gruppe der Rückfälligen von der der Nicht-Rückfälligen in befriedigender Weise zu trennen. Es handelt sich hier also nicht um eine Eigenschaft, die in besonderer Weise dem PCL:SV-Gesamtscore zukommt, sondern jedem „Instrument“, mit dem man Rückfallprognosen vornehmen kann.

5.4 PCL:SV-SD bzw. PCL:SV-PP und krimineller Rückfall

Die Variable PCL:SV-SD, die sich ja lediglich aus der Hälfte der Items der PCL:SV-Gesamt zusammensetzt, besitzt eine geschätzte AUC von 0.67 und verhält sich bei der Prognose eines generellen Rückfalls (mit einem nach den oben diskutierten Kriterien ausgewählten *cut-off* Wert von 3) in ähnlicher Weise wie die PCL:SV-Gesamt. Die Variable PCL:SV-SD besitzt sogar für sämtliche Kenngrößen leicht bessere Werte (vgl. Tab. 23). Da die Variable PCL:SV-SD bei einem *cut-off* Wert von 9 ganz ähnliche Eigenschaften aufweist wie die PCL:SV-Gesamt bei der „Entdeckung“ einer „Hochrisikogruppe“, kann hier auf eine entsprechende Diskussion verzichtet werden. Ebenso soll auf eine eigene Analyse der Kennwerte der Variablen PCL:SV-PP (geschätzte AUC = 0.61) verzichtet werden, da diese für die Prognose eines generellen Rückfalls (bei einem *cut-off* Wert von 3) einem ähnlichen Muster folgen wie die Kennwerte der Variablen PCL:SV-Gesamt (mit *cut-off* Wert 6) und PCL:SV-SD (mit *cut-off* Wert 3), wenngleich sämtliche Kennwerte zum Teil deutlich ungünstiger ausfallen (vgl. Tab. 21). Die Wirkungsweise der PCL:SV-PP bei Verwendung eines *cut-off* Wertes von 9 ist vergleichbar der der PCL:SV-Gesamt bei der „Entdeckung“ einer „Hochrisikogruppe“.

5.5 Soziobiographische und psychopathologische Merkmale und krimineller Rückfall

Neben der PCL:SV wurden weitere, in der Rückfallprognostik als relevant erachtete Merkmale erfasst und hinsichtlich ihres Zusammenhanges mit kriminellem Rückfall analysiert. Vorrangiges Ziel dabei war es, die Evaluation der prädiktiven Qualität der PCL:SV mit den prognostischen Eigenschaften dieser Merkmale zu vergleichen.

Insgesamt betrachtet scheinen die Merkmale aus dem Bereich der Soziobiographie und Psychopathologie – zumindest bei der hier untersuchten Population von Straftätern – zur Rück-

fallprognose wenig tauglich zu sein (vgl. dazu auch noch einmal die Tab. 12 und 13). So zeigten soziobiographische Auffälligkeiten nur schwache bis gar keine Zusammenhänge mit genereller Rückfälligkeit. Einzige Ausnahme bildete das Merkmal „Broken-Home“ ($r = .13$). Bei der Vorhersage gewalttätiger Rückfälle erreichte kein einziges Merkmal aus diesem Bereich statistische Relevanz.

Ein ähnliches Bild zeigte sich bei der Analyse der psychiatrischen Diagnosen. Bis auf die Diagnose von mindestens einer psychischen Störung durch psychotrope Substanzen sowie die kombinierte Diagnose von psychischen und Persönlichkeitsstörungen war keine der anderen psychiatrischen Diagnosen mit generellem Rückfall assoziiert. Bei der Untersuchung des Zusammenhangs mit gewalttätigem Rückfall erreichte ausschließlich das Vorliegen einer (oder mehrerer) Persönlichkeitsstörung(en) statistische Signifikanz.

Zusammenfassend kann somit festgehalten werden, dass die Merkmale aus der Soziobiographie sowie psychiatrische Diagnosen – bis auf wenige Ausnahmen – in der vorliegenden Untersuchung keine substanziellen Zusammenhänge mit sowohl genereller als auch gewalttätiger Rückfälligkeit aufwiesen. Es soll daher auf weitergehende Analysen verzichtet werden. Ein anderes Bild zeigt sich bei den Merkmalen zur kriminellen Vorgeschichte, die im Folgenden diskutiert werden sollen.

5.6 Merkmale der strafrechtlichen Vorgeschichte und krimineller Rückfall

In diesem Abschnitt sollen nun die prognostischen Eigenschaften der Merkmale *Alter bei erster Verurteilung (AE)* und *Anzahl der Vorstrafen (AV)* zur Vorhersage genereller Rückfälle ausführlicher analysiert werden. Sie zählen in der kriminologischen Forschung nach wie vor zu den wenigen Faktoren, die die größten und stabilsten Zusammenhänge mit krimineller Rückfälligkeit aufweisen. (z. B. Andrews & Bonta 1998, 2003, Gendreau et al. 1996). Dieser Befund lässt sich auch im Rahmen dieser Untersuchung replizieren.

Unterwirft man diese Variablen einer ROC-Analyse, so stellt man, wenn man die zugehörigen geschätzten d-Werte betrachtet – nicht ganz unerwartet – fest, dass beide eine ähnliche prädiktive Validität aufweisen wie die PCL:SV-Gesamt. Die AUC-Werte beider Variablen bewegen sich in der vorliegenden Stichprobe im selben Größenbereich und liegen sogar leicht über den Werten der Variablen PCL:SV-Gesamt und PCL:SV-SD (vgl. Tab. 26) (Die PCL:SV-PP wurde hier nicht explizit aufgeführt, weil die Kennwerte einem ähnlichen Muster wie die der

PCL:SV-Gesamt bzw. PCL:SV-SD folgen oder schlechter ausfallen). Abb. 8 zeigt die ROC-Kurven der Variablen *Alter bei erster Verurteilung* und *Anzahl der Vorstrafen*.²²

Tabelle 26: AUC-Werte für die Variablen PCL:SV-Gesamt (PCL-G), PCL:SV-SD (PCL-SD), Alter bei erster Verurteilung (AE) und Anzahl der Vorstrafen (AV)

Prognose	PCL-G	PCL-SD	AE	AV
Genereller Rückfall	0.660	0.666	0.679	0.671

In Tab. 27 sind die Ergebnisse der detaillierten ROC-Analysen für diese Variablen wie auch die schon berichteten Resultate für die Variablen PCL:SV-Gesamt und PCL:SV-SD zusammenfassend dargestellt. Für jede Variable wurde ein *cut-off* Wert gewählt, der gleichzeitig einen möglichst hohen ϕ -Koeffizienten mit einer möglichst kleinen Fehlerrate verbindet. Für diesen *cut-off* Wert wurden dann die zugehörigen Kenngrößen bestimmt. Zunächst ist festzuhalten, dass die Werte der Kenngrößen für alle vier Variablen im Prinzip dem gleichen Muster folgen. Auch die Größenordnungen der Kenngrößen sind vergleichbar. Die Variable *Anzahl der Vorstrafen* besitzt dabei den größten ϕ -Koeffizienten ($\phi = 0.403$) und die kleinste Fehler rate (24.2 %). Zudem hat ihre *NPP* als einzige die wünschenswerte Eigenschaft, dass sie deutlich über 0.5 liegt (*NPP* = 0.667).

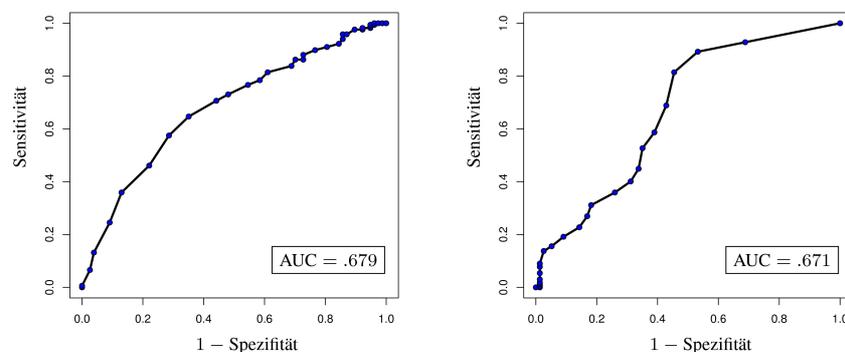


Abbildung 8: Linke Seite: ROC-Kurve der Variablen *Alter bei erster Verurteilung*. Rechte Seite: ROC-Kurve der Variablen *Anzahl Vorstrafen*.

²² Die Verteilung der Variablen „Alter bei erster Verurteilung“ in der Gruppe der Rückfälligen ist relativ zu der der Gruppe der Nicht-Rückfälligen nach links, und nicht, wie bei den anderen hier betrachteten Variablen, nach rechts verschoben. Dass heißt, dass *kleine* Werte dieser Variablen auf Rückfälligkeit hinweisen. Um in der ROC-Analyse diese Variablen vergleichbar mit den anderen zu machen, wurden alle Altersangaben von der Zahl 100 abgezogen, wodurch die Anordnung der beiden Verteilungen vertauscht wurde.

Tabelle 27: Prognosekennwerte der Variablen PCL:SV-Gesamt (PCL-Gesamt), PCL:SV-SD (PCL-SD), Alter bei erster Verurteilung (AE) und Anzahl der Vorstrafen (AV) bei der Vorhersage eines generellen Rückfalls, jeweils unter Verwendung eines *cut-off* Wertes c , der mit möglichst hohem ϕ -Koeffizienten und möglichst niedriger Fehlerrate einhergeht.

Kenngröße	PCL-Gesamt	PCL-SD	AE	AV
	mit $c = 6$	mit $c = 3$	mit $c = 22^a$	mit $c = 3$
Spezifität	0.481	0.494	0.649	0.468
Sensitivität	0.778	0.784	0.647	0.892
PPP	0.765	0.771	0.800	0.784
NPP	0.500	0.514	0.459	0.667
Fehlerrate	0.316	0.307	0.352	0.242
Basisrate	0.684	0.684	0.684	0.684
Selektionsrate	0.697	0.697	0.553	0.779
ϕ -Koeffizient	0.262	0.281	0.277	0.403

^a Entscheidungsregel hier: Prognostiziere „wird rückfällig“, falls $AE \leq 22$.

Diese Analyse zeigt, dass die beiden Variablen *AE* und *AV* im Hinblick auf ihre prognostische Güte mit den Variablen PCL:SV-Gesamt und PCL:SV-SD durchaus mithalten können. Die Kenngrößen der Variablen *AV* fallen sogar deutlich günstiger aus als die der anderen Variablen (dies betrifft vor allem die Fehlerrate und die *NPP*). Zur Vorhersage eines generellen Rückfalls sollte man nach diesen Ergebnissen der Variablen PCL:SV-Gesamt wohl die Variable *AV* (Anzahl der Vorstrafen) vorziehen.

Auch wenn es um die „Entdeckung“ einer sogenannten „Hochrisikogruppe“ unter den Straftätern geht, verhalten sich diese beiden Variablen ähnlich wie die betrachteten PCL:SV Variablen. Die prognostischen Kenngrößen hierfür (für geeignet gewählte *cut-off* Werte) werden in Tab. 28 zusammengefasst.

Tabelle 28: Prognosekennwerte der Variablen PCL:SV-Gesamt (PCL-Gesamt), PCL:SV-SD (PCL-SD), Alter bei erster Verurteilung (AE) und Anzahl der Vorstrafen (AV) bei der Vorhersage eines generellen Rückfalls, jeweils unter Verwendung eines *cut-off* Wertes c , der zur Aufdeckung einer „Hochrisikogruppe“ führt.

Kenngröße	PCL-Gesamt	PCL-SD	AE	AV
	mit $c = 18$	mit $c = 9$	mit $c = 18^a$	mit $c = 17$
Spezifität	0.987	0.922	0.909	0.987
Sensitivität	0.096	0.180	0.246	0.090
PPP	0.941	0.833	0.854	0.938
NPP	0.335	0.341	0.357	0.333
Fehlerrate	0.623	0.586	0.545	0.627
Basisrate	0.684	0.684	0.684	0.684
Selektionsrate	0.070	0.148	0.197	0.066
ϕ -Koeffizient	0.151	0.133	0.181	0.144

^a Entscheidungsregel hier: Prognostiziere „wird rückfällig“, falls $AE \leq 18$.

Wählt man beispielsweise für die Anzahl der Vorstrafen (*AV*) einen *cut-off* Wert von 17, so erhält man prognostischen Kennwerte, die fast identisch mit denen der PCL:SV-Gesamt bei einem *cut-off* Wert von 18 sind. Die in diesem Zusammenhang weiter oben vorgetragenen kritischen Anmerkungen gelten hier natürlich in analoger Weise.

5.7 Komplexere Entscheidungsregeln

Die Diskussion des letzten Abschnitts wirft die Frage auf, ob eine geeignete *Kombination* der Variablen *PCL:SV-Gesamt*, *Alter bei erster Verurteilung* und *Anzahl der Vorstrafen* zu einer Verbesserung der prognostischen Güte führt. Wegen der Ähnlichkeit der mit diesen Variablen assoziierten prognostischen Kenngrößen könnte man vermuten, dass die drei diskutierten Entscheidungsregeln zu weitgehend übereinstimmenden Prognosen führen und daher eine Kombination dieser drei Variablen bei der Prognose eines generellen Rückfalls nicht lohnenswert ist. Hohe Übereinstimmungen liegen aber, wie Tab. 29 zeigt, nicht vor. Bei hohen Übereinstimmungen sollte nämlich die Summe der Werte in der Nebendiagonalen klein, bzw. der φ -Koeffizient groß sein.

Tabelle 29: Paarweise Prognoseübereinstimmungen zwischen den Entscheidungsregeln „PCL:SV-Gesamt ≥ 6 “ (PCL ≥ 6); „Anzahl Vorstrafen ≥ 3 “ (AV ≥ 3) und „Alter bei erster Verurteilung ≤ 22 “ (AE ≤ 22).

		$\varphi = .21$		$\varphi = .36$		$\varphi = .30$			
		AE ≤ 22		ZR ≥ 3		ZR ≥ 3			
		0	1	0	1	0	1		
PCL ≥ 6	0	0.20	0.10	0	0.14	0.17	0	0.18	0.34
	1	0.31	0.39	1	0.09	0.61	1	0.05	0.44

Tab. 29 zeigt, dass die größte Übereinstimmung zwischen den beiden Regeln „PCL:SV-Gesamt ≥ 6 “ und „Anzahl der Vorstrafen ≥ 3 “ besteht. Jedoch werden auch hier noch 26 % der Straftäter nicht in übereinstimmender Weise klassifiziert. Es soll daher untersucht werden, ob Kombinationen der drei betrachteten Variablen zu geeigneteren Prognoseregeln führen.

5.7.1 Lineare Kombinationen der Prädiktoren

In Abschnitt 4.3.7 wurde eine Logistische Regressionsanalyse mit der Variablen „genereller Rückfall“ als Kriterium und den Variablen „PCL:SV-Gesamt“ und „Anzahl der Vorstrafen“ durchgeführt (vgl. Tab. 25; die Variable „Alter bei erster Verurteilung“ konnte als Prädiktor ausgeschlossen werden, da ihr Einschluss zu keiner nennenswerten Verbesserung der Modell-

anpassung führte). Bestimmt man nun mit Hilfe der Logistischen Regression die Schätzungen der (bedingten) Rückfallwahrscheinlichkeiten und stellt genau dann die Prognose „wird rückfällig“, wenn die (bedingte) Wahrscheinlichkeit größer als 0.5 ist, dann liefert diese Entscheidungsregel die in Tab. 30 aufgeführten prognostischen Kennwerte. Vergleicht man diese Kennwerte mit denen aus Tab. 27, so stellt man fest, dass sie etwas besser ausfallen als die der Variablen „PCL:SV-Gesamt“ mit einem *cut-off* Wert von 6, aber nicht besser als die der Variablen „Anzahl der Vorstrafen“ mit *cut-off* Wert 3. Die lineare Kombination dieser Prädiktoren führt somit zu keiner Prognoseregeln, die gegenüber den vorher berichteten einen prognostischen Vorteil bringt.

Tabelle 30: Kenngrößen für die Prognose eines generellen Rückfalls mit Hilfe der in Tab. 25 berichteten Logistischen Regression

Spez.	Sens.	PPP	NPP	Fehlerr.	Basisr.	Selektionsr.	φ -Koeff.
0.303	0.928	0.745	0.657	0.267	0.687	0.856	0.305

5.7.2 Nicht lineare Kombinationen der Prädiktoren

Es sollen nun komplexere Alternativen zu den einfachen Entscheidungsregeln, die in den vorangegangenen Abschnitten analysiert wurden, diskutiert werden. Die betrachteten Alternativen beruhen auf Methoden zur Konstruktion und Evaluation von sogenannten Entscheidungsbäumen, die verschiedentlich als methodische Alternativen in der Literatur Erwähnung finden (Nedopil 2005, Dahle 2005).²³ Da ein umfassendes Studium dieser Methoden den Rahmen dieser Arbeit sprengen würde, sollen sie hier nur exemplarisch skizziert werden, um das Grundprinzip und die Vorteile der sogenannten *tree based methods* zu verdeutlichen.

Tabelle 31: Absolute und relative Häufigkeiten aller *pattern*

Pattern:	000	001	010	011	100	101	110	111
abs. Häuf.:	28	19	5	22	13	49	8	100
rel. Häuf.:	0.115	0.078	0.020	0.090	0.053	0.201	0.033	0.410

Zunächst soll eine Methode diskutiert werden, die als ein Spezialfall der *tree based methods* aufgefasst werden kann. Dazu werden die drei Aussagen „PCL:SV-Gesamt ≥ 6 “, „Anzahl Vorstrafen ≥ 3 “ sowie „Alter bei erster Verurteilung ≤ 22 “ betrachtet. Für jede der drei Aus-

²³ Vgl. für weitere Informationen z. B. Breiman et al. (1984).

sagen erhält ein Straftäter eine Eins, falls die Aussage zutrifft, ansonsten eine Null. Es werden nun alle möglichen Ergebnispattern betrachtet (beispielsweise ist das pattern „011“ zu lesen als: PCL:SV-Gesamt liegt unter 6, Anzahl der Vorstrafen ist größer oder gleich 3 und Alter bei erster Verurteilung ist kleiner oder gleich 22).

Die absolute und relative Häufigkeitsverteilung dieser pattern in der vorliegenden Straftäterstichprobe liefert Tab. 31. Die acht Merkmalskombinationen zerlegen die Straftäterstichprobe in acht Teilgruppen und in jeder dieser acht Teilgruppen kann nun die Verteilung der Rückfallvariablen bestimmt werden (kein Rückfall: R_- / Rückfall: R_+). Tab. 32 zeigt für jedes der acht pattern die zugehörige Verteilung der Rückfallvariablen.

Tabelle 32: Bedingte Verteilung der Rückfallvariablen unter jedem der acht möglichen pattern sowie die von diesen Verteilungen abhängigen Rückfallprognosen (letzte Zeile). Der Ausdruck $p(R_+)$ bezeichnet die relativen Häufigkeiten für einen Rückfall, $p(R_-) = 1 - p(R_+)$

Pattern:	000	001	010	011	100	101	110	111
$p(R_-)$:	0.750	0.421	0.400	0.273	0.615	0.265	0.625	0.140
$p(R_+)$:	0.250	0.579	0.600	0.727	0.385	0.735	0.375	0.860
Prognose:	0	1	1	1	0	1	0	1

Mit Hilfe dieser bedingten Verteilungen wird nun die folgende Entscheidungsregel konstruiert: Prognostiziere für einen Straftäter genau dann Rückfälligkeit, wenn er zu einer Pattern-Gruppe gehört, bei der die Rückfallhäufigkeit größer als die Nicht-Rückfallhäufigkeit ist. Die letzte Zeile in Tab. 32 zeigt, dass genau die Straftäter mit den pattern „000“, „100“ und „110“ als Nicht-Rückfällige klassifiziert werden. Die Kenngrößen für diese Entscheidungsregel liefert Tab. 33.

Tabelle 33: Kenngrößen für die Prognose eines generellen Rückfalls mit Hilfe der in Tab. 32 gegebenen Entscheidungsregel

Spez.	Sens.	PPP	NPP	Fehlerr.	Basisr.	Selektionsr.	φ -Koeff.
0.442	0.910	0.779	0.694	0.238	0.684	0.799	0.408

Diese Kenngrößen zeigen, dass die zugehörige Prognoseregeln in ihrer prognostischen Güte mit der einfachen Regel „Anzahl der Vorstrafen ≥ 3 “ vergleichbar ist und keinen wirklichen Gewinn bringt. Allerdings stellt sich die Frage, ob eine andere Wahl der jeweiligen *cut-off* Werte bei der Definition der (binären) Variablen zu besseren Resultaten führen würde. Es ist ja denkbar, dass bei kombinierter Betrachtung andere *cut-off* Werte günstiger sind als bei Ein-

zelbetrachtung der Aussagen. Durchsucht man alle möglichen Kombinationen von *cut-off* Werten der betrachteten Variablen, so findet man für die Kombination „PCL:SV-Gesamt ≥ 7 “, „Anzahl der Vorstrafen ≥ 3 “ sowie „Alter bei erster Verurteilung ≤ 21 “ unter Anwendung der soeben beschriebenen Methode tatsächlich eine Prognoseregeln, die deutlich bessere Werte besitzt: Die Fehlerrate beträgt nun nur noch 22.5 % und der φ -Koeffizient steigt auf den Wert 0.44.

Eine systematische Durchmusterung aller möglichen *cut-off* Wert Kombinationen lässt sich allerdings bei einer größeren Zahl von Variablen nicht mehr ökonomisch durchführen. Zudem könnte die vorherige Festlegung auf genau einen *cut-off* Wert für jeden Prädiktor eine möglicherweise zu rigide Strategie sein. Bei der Konstruktion von Entscheidungsbäumen werden derartige Beschränkungen vermieden. Auch diese Methode hat zum Ziel, eine „günstige“ Partition der Straftäter zu ermitteln, die hier durch die Endknoten des Baumes (die *Blätter*) repräsentiert wird. In den durch die Endknoten gegebenen Partitionsklassen wird nun jeweils die Verteilung der Zielvariablen (hier die binäre Variable „genereller Rückfall“) betrachtet und für alle Personen dieser Klasse wird nun jenes Ereignis (Rückfälliger/Nicht-Rückfälliger) prognostiziert, das dort die höchste Auftretenshäufigkeit besitzt. Jeder Entscheidungsbaum liefert also wieder eine Vier-Felder-Tafel *Prognose x Realität*, für die sich dann die Prognosekennwerte bestimmen lassen.

Abb. 9 zeigt einen Entscheidungsbaum, der für die Variablen „PCL:SV-Gesamt“, „Alter bei erster Verurteilung“ und „Anzahl der Vorstrafen“ und die Zielvariable „genereller Rückfall“ konstruiert wurde. In dieser Graphik sind die Endknoten – anders als die restlichen Knoten – ohne Umrandung dargestellt. Die Zahlen in den Klammern stehen für die Häufigkeitsverteilungen der Variablen „genereller Rückfall“ im entsprechenden Knoten, wobei die erste Zahl für die Häufigkeit der Nicht-Rückfälligen und die zweite Zahl für die Häufigkeit der Rückfälligen steht. Wenn die Zahl der Nicht-Rückfälligen größer ist als die der Rückfälligen, dann lautet die Prognose für alle Straftäter dieses Knotens Nicht-Rückfälliger (R_-), sonst Rückfälliger (R_+).

Tabelle 34: Kenngrößen für die Prognose eines generellen Rückfalls mit Hilfe des in Abb. 9 wiedergegebenen Entscheidungsbaumes

Spez.	Sens.	PPP	NPP	Fehlerr.	Basisr.	Selektionsr.	φ -Koeff.
0.416	0.952	0.779	0.800	0.217	0.684	0.836	0.462

Die eingerahmten Knoten enthalten darüber hinaus (in der ersten Zeile) die Information über die im nächsten Schritt vorgenommene Zerteilung der zum Knoten gehörenden Straftätergruppe: Die Aussage formuliert dabei die Bedingung, unter welcher Straftäter des Knotens dem linken Verzweigungsknoten zugeordnet werden. Die Endknoten liefern schließlich die Entscheidungsregel: Alle Personen, die zu den beiden linken Endknoten gehören, werden als Nicht-Rückfällige, alle anderen als Rückfällige prognostiziert. Die Kennwerte dieser Entscheidungsregel sind in Tab. 34 wiedergegeben. Sie zeigen, dass durch die Verwendung einer *tree based method* die Kennwerte nochmals verbessert werden konnten.

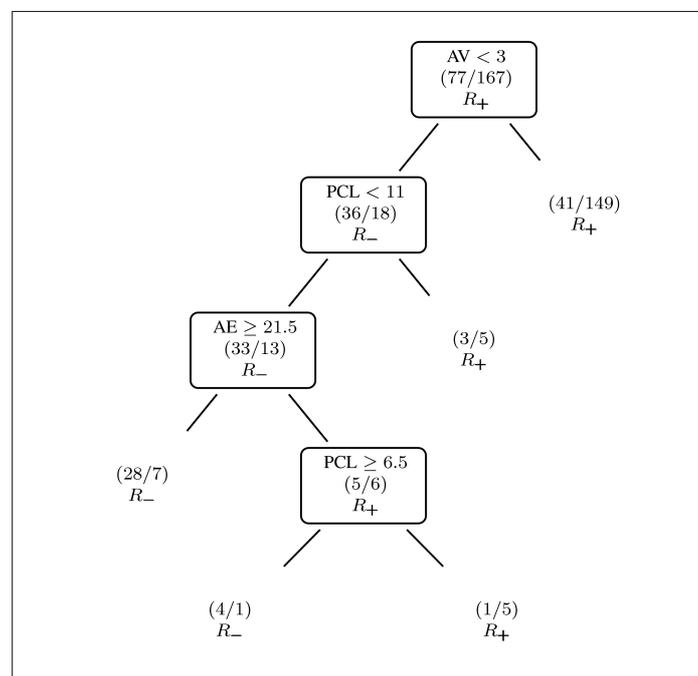


Abbildung 9: Entscheidungsbaum für die Variablen *PCL:SV-Gesamt (PCL)*, *Alter bei erster Verurteilung (AE)* und *Anzahl der Vorstrafen (AV)* sowie die Zielvariable genereller Rückfall (die Berechnungen wurden mit dem R-Paket „rpart“ durchgeführt).

5.8 Limitierungen

Am Ende der Diskussion sei darauf hingewiesen, dass die kritischen Ausführungen selbstverständlich unter dem Vorbehalt der nachfolgenden Einschränkungen zu lesen sind. So ist man bei Untersuchungen wie dieser stets mit dem Problem der Vernachlässigung des Dunkelfeldes konfrontiert, da nur Daten von Straftätern ausgewertet werden konnten, deren Taten offiziell registriert und im BZR aufgeführt worden waren.

Zudem konnten in diese Studie nur solche Probanden eingeschlossen werden, die ihre Einwilligung zur Auswertung ihrer BZR-Auszüge erteilt hatten und die zudem fünf Jahre nach Ab-

schluss der Ausgangsstudie von uns erreicht werden konnten, was zu Selektionseffekten geführt haben kann und die Repräsentativität der Stichprobe einschränkt.

Aus den in Kapitel 3 aufgeführten Gründen, wurde auf eine „*Time-at-risk*“-Auswertung verzichtet. Somit fehlen Informationen über die tatsächlich in Freiheit verbrachte Zeit, die wiederum in ihrem Einfluss auf die Rückfallfrequenz nicht untersucht werden konnte. Schließlich sind die Ergebnisse zu kriminellen Rückfällen immer von der Wahl des Kriteriums, d. h., von der verwendeten Rückfalldefinition abhängig. In der vorliegenden Studie wurde „krimineller Rückfall“ eher konservativ definiert, d. h., es wurde jeder erneute Eintrag ins BZR als Rückfall gewertet, ganz gleich, ob es sich dabei um Delikte wie „Fahren ohne Fahrerlaubnis“ oder „schwere Körperverletzung“ gehandelt hatte.

6 Zusammenfassung

Ein zentrales Anliegen dieser Studie war es, die prognostische Validität der PCL:SV zur Vorhersage krimineller Rückfälle bei Gefangenen aus deutschen Haftanstalten zu ermitteln. Obwohl die PCL:SV ursprünglich nicht als Prognoseinstrument konstruiert wurde, wurde vielfach eine moderate prognostische Validität für die Vorhersage krimineller Rückfälle berichtet. Die überwiegende Zahl einschlägiger Studien stammt dabei aus dem englischsprachigen Raum. Systematische Untersuchungen an deutschen Straftäterpopulationen sind bislang noch rar, beziehungsweise liegen ausschließlich für die PCL-R vor.

Die Stichprobe umfasste 244 Straftäter, die von Amts- und Landesgerichten Sachsen-Anhalts aufgrund verschiedener Delikte rechtskräftig zu Haftstrafen verurteilt worden waren. Die Auswertung der kriminellen Rückfälle erfolgte anhand von BZR-Einträgen.

Im Hinblick auf das eigentliche Hauptziel der Studie kann zunächst positiv festgehalten werden, dass die beiden Subskalen der PCL:SV sowie die PCL:SV-Gesamt mit der Variablen Rückfall einen gewissen Zusammenhang aufweisen. Sowohl die ROC-Analysen – als auch die Logistischen Regressionen – kommen zu dem Ergebnis, dass sich die Gruppen „genereller Rückfall vs. kein Rückfall“ sowie die Gruppen „gewalttätiger Rückfall vs. kein gewalttätiger Rückfall“ statistisch trennen lassen. Allerdings sind die in den Analysen aufgetretenen statistischen Signifikanzen – was deren praktische Bedeutsamkeit angeht – mit einiger Zurückhaltung zu interpretieren. Denn die geschätzten AUCs und Regressionskoeffizienten zeigen, dass die entsprechenden Tests zwar auf das Vorhandensein von Effekten hinweisen, aber auf Effekte, die aus praktischer, ethischer und rechtspolitischer Sicht nicht sonderlich groß sind. Dieser Tatbestand spiegelt sich auch in den für jeden möglichen *cut-off* Wert der PCL:SV-Gesamt ermittelten Größen Sensitivität, Spezifität, *PPP* und *NPP* wider. Nachdenklich müssen auch die hohen Fehlerraten der PCL:SV-Gesamt sowie ihrer Subskalen (PCL:SV-SD und PCL:SV-PP) stimmen. Die PCL:SV-SD besitzt bei der Vorhersage gewalttätiger Rückfälle – obgleich sie sich lediglich aus der Hälfte der Items der PCL:SV-Gesamt zusammensetzt – im Vergleich zu dieser sogar leicht bessere Werte.

Um zu einer angemessenen Bewertung der Wirkungsweise eines Prognoseinstrumentes in einer bestimmten Straftäterpopulation zu gelangen, reicht es nicht hin, lediglich den ermittelten AUC-Wert zu berichten. Auch sind die häufig auf der Basis des Cohenschen Klassifikationsschemas vorgenommenen Einschätzungen der Höhe solcher Werte nicht hilfreich. Denn dieses Schema wird prognostischen Situationen, wie sie hier untersucht werden, nicht gerecht. Neben der Angabe der AUC und der Einschätzung ihrer Höhe sollte sich die Beurteilung des

Prognoseinstrumentes daher vor allem auch auf die Analyse konkreter basisabhängiger Entscheidungsregeln stützen.

Für die Vorhersage eines generellen Rückfalls weist die PCL:SV-Gesamt in der untersuchten Straftätergruppe eine geschätzte AUC von 0.66 und einen geschätzten d-Wert von 0.57 auf. Die prognostischen Kennwerte der auf der PCL:SV-Gesamt fußenden günstigen Entscheidungsregeln (möglichst geringe Fehlerrate bei möglichst großem ϕ -Koeffizienten) sind – zumindest bezogen auf die hier untersuchte Gruppe von Straftätern – alles andere als zufriedenstellend. Sie besitzen hohe Fehlerraten von etwa 30 %. Bei einer Basisrate für Rückfälligkeit von fast 70 % sind diese nicht geringer als die Fehlerrate derjenigen Regel, die jeden Straftäter „blind“ als „wird rückfällig“ klassifiziert. Darüber hinaus ist die *NPP* mit Werten um 50 % für eine Entscheidungsregel alles andere als wünschenswert.

Auch die verschiedentlich in der Literatur hervorgehobene Fähigkeit der PCL:SV-Gesamt, zumindest die Hochrisikogruppe der *Psychopaths* zu detektieren, erweist sich bei genauer Betrachtung als nicht so beeindruckend, wie es zunächst den Anschein hat. Denn die bei einem *cut-off* Wert von 18 in der Tat sehr hohe *PPP* wird erkaufte durch eine sehr hohe Rate von Falsch-Negativen, die eine sehr kleine *NPP* zur Folge hat. Dabei werden 67 % der als „wird nicht rückfällig“ Prognostizierten falsch prognostiziert und bei den als erfolgreich als „wird rückfällig“ Klassifizierten handelt es sich lediglich um 7 % aller Straftäter. Im Übrigen ist hervorzuheben, dass es sich bei dieser Fähigkeit nicht etwa um eine Besonderheit der PCL:SV-Gesamt handelt. Vielmehr ist jeder Test, mit dem man Rückfällige von Nicht-Rückfälligen einigermaßen trennen kann, in der Lage, (s)eine spezifische „Hochrisikogruppe“ zu selektieren.

Die besondere Stellung der PCL:SV-Gesamt als Instrument zur Vorhersage eines generellen Rückfalls wird auch dadurch geschwächt, dass die Variablen „Alter bei erster Verurteilung“ und „Anzahl der Vorstrafen“ Entscheidungsregeln ermöglichen, die in ihren prognostischen Eigenschaften nicht schlechter sind als die der PCL:SV-Gesamt. Die prognostischen Kennwerte der Variablen „Anzahl der Vorstrafen“ fallen sogar deutlich besser aus.

Man wird wohl kaum erwarten können, dass ein Ereignis, das so komplex determiniert ist wie das eines kriminellen Rückfalls, sich ohne größere Fehlerrate durch nur *eine* Variable vorherzusagen lässt. Es liegt daher nahe, dieses Ereignis mit einer Reihe von einschlägigen, unterschiedliche Bereiche erfassenden Prädiktoren vorherzusagen. Im vorliegenden Datensatz waren Vorhersageregeln, die auf linearen Modellen beruhten, weniger erfolgreich als eine Regel, die ein entscheidungsbaumbasiertes, nicht-lineares Modell zur Grundlage hatte (mit den Prä-

dikator-Variablen „PCL:SV-Gesamt“, „Alter bei erster Verurteilung“ und „Anzahl der Vorstrafen“). Diese Regel erwies sich als die erfolgreichste aller hier diskutierten Entscheidungsregeln: Sie besitzt die kleinste Fehlerrate, den größten φ -Koeffizienten und hat die wünschenswerte Eigenschaft, dass sie neben einer hohen *PPP* auch eine hohe *NPP* aufweist.

Zusammenfassend zeigen die Ergebnisse der vorliegenden Arbeit, dass – gemessen an den hohen Anforderungen, die berechtigterweise an Kriminalprognosen gestellt werden – die prädiktive Validität der PCL:SV kaum befriedigen kann. Es bleibt abzuwarten, ob weitere Untersuchungen diesen Befund replizieren. Bis dahin scheint ein eher zurückhaltender Umgang mit der PCL:SV als Prognoseinstrument geboten.

7 Literaturverzeichnis

1. Andrews, D.A. & Bonta, J. (1995). *LSI-R: The Level of Service Inventory-Revised*. Toronto: Multi-Health Systems.
2. Andrews, D.A. & Bonta, J. (1998). *The psychology of criminal conduct* (2nd ed.). Cincinnati, Ohio: Anderson.
3. Andrews, D.A. & Bonta, J. (2003). *The psychology of criminal conduct* (3rd ed.). Cincinnati, Ohio: Anderson.
4. Babiak, P. (1995). When psychopaths go to work: A case study of an industrial psychopath. *Applied Psychology: An International Review*, 44, 171-188.
5. Bonta, J. (1996). Risk-needs assessment and treatment. In: A.T. Harland (Ed.), *Choosing correctional options that work: Defining the demand and evaluating the supply* (pp. 18-32). Thousand Oaks, CA: Sage.
6. Bonta, J., Law, M. & Hanson, K. (1998). The prediction of criminal and violent recidivism among mentally disordered offenders: A meta-analysis. *Psychological Bulletin*, 123, 123-142.
7. Breiman, L., Friedman, J.H., Olsen, R.A. & Stone, C.J. (1984). *Classification and Regression Trees*. New York.
8. Burges, E.W. (1928). Factors determining success or failure on parole. *Journal criminal Law and Criminology*, 19, 241.
9. Burges, E.W. (1968). Factors determining success or failure on parole. In: A.A. Bruce, A.J. Harno, E.W. Burges & J. Landesco (Eds.), *The workings of the indeterminate-sentence law and the parole system in Illinois* (pp. 205-234). Montclair, NJ: Patterson Smith (original work published 1928).
10. Claix, A. & Pham, T.H. (2004). Evaluation of the HCR-20 Violence Risk Assessment Scheme in Belgian forensic population. *Encephale*, 30, 447-453.
11. Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
12. Cohen, J. (1992). A power primer. *Psychological Bulletin* 112, 155-159.
13. Cooke, D.J. (1998). Psychopathy across cultures. In: D.J. Cooke, A.E. Forth & R.D. Hare (Eds.), *Psychopathy: Theory, research, and implications for society* (pp. 13-35). Dordrecht: Kluwer.
14. Cooke, D.J. & Michie, C. (1997). An item response theory evaluation of Hare's Psychopathy Checklist-Revised. *Psychological Assessment*, 9, 2-13.
15. Cooke, D.J. & Michie, C. (2001). Refining the construct of Psychopathy: Towards a hierarchical model. *Psychological Assessment*, 13, 171-188.
16. Cooke, D.J., Kosson, D.S. & Michie, C. (2001). Psychopathy and ethnicity: Structural, item and test generalizability of the Psychopathy Checklist-Revised (PCL-R) in Caucasian and African-American participants. *Psychological Assessment*, 13, 531-542.
17. Cooke, D.J., Michie, C., Hart, S.D. & Hare, R.D. (1999). Evaluating the screening version of the Hare Psychopathy Checklist-Revised (PCL:SV): An item response theory analysis. *Psychological Assessment*, 11, 3-13.
18. Cottle, C.C., Lee, R.J. & Heilbrun, K. (2001). The predicting of criminal recidivism in juveniles. A meta-analysis. *Criminal Justice and Behavior*, 28, 367-394.
19. Dahle, K.-P. (2004). *Abschlussbericht der Berliner CRIME-Studie an die Deutsche Forschungsgemeinschaft*. Institut für Forensische Psychiatrie der Charité, Universität Berlin.

20. Dahle, K.-P. (2005). *Psychologische Kriminalprognose: Wege zu einer integrativen Methodik für die Beurteilung der Rückfallwahrscheinlichkeit bei Strafgefangenen*. Herbolzheim: Centaurus.
21. Dahle, K.-P. (2006). Strengths and limitations of actuarial prediction of criminal reoffence in a German prison sample: A comparative study of LSI-R, HCR-20 and PCL-R. *International Journal of Law and Psychiatry*, 29, 431-442.
22. Dahle, K.-P., Janka, C., Gallasch, F. & Lehmann, R. (2008). Jugendliche Sexualstraftäter: Spezialfälle in der Prognoseerstellung? *Forensische Psychiatrie, Psychologie und Kriminologie*, 4, 213-221.
23. Dahle, K.-P., Schneider, V. & Ziethen, F. (2007). Standardisierte Instrumente zur Kriminalprognose. *Forensische Psychiatrie, Psychologie und Kriminologie*, 1, 15-26.
24. Dawes, R.M., Faust, D. & Meehl, P.E. (1989). Clinical versus actuarial judgment. *Science*, 243, 1668-1674.
25. Dietiker, J., Dittmann, V. & Graf, M. (2007). Gutachterliche Risikoeinschätzung bei Sexualstraftätern. *Nervenarzt*, 78, 53-61.
26. Douglas, K.S., Ogloff, J.R.P. & Hart, D. (2003). Evaluation of a model of violence risk assessment among forensic psychiatric patients. *Psychiatric Services*, 54, 1372-1379.
27. Douglas, K.S., Ogloff, J.R.P. & Nicholls, T.L. (1997). The assessment of risk for violence in instrumental and reactive violent offenders. *Canadian Psychology*, 38, 111.
28. Douglas, K.S., Ogloff, J.R.P., Nicholls, T.L. & Grant, I. (1999). Assessing risk for violence among psychiatric patients: The HCR-20 violence risk assessment scheme and the PCL:SV. *Journal of Consulting and Clinical Psychology*, 6, 917-930.
29. Dünkel, F. & Geng, B. (2003). Rückfall und Bewährung von Karrieretätern nach Entlassung aus dem sozialtherapeutischen Behandlungsvollzug und aus dem Regelvollzug. In: M. Steller, K.-P. Dahle & M. Basqué (Hrsg.), *Straftäterbehandlung – Argumente für eine Revitalisierung in Forschung und Praxis* (2. Aufl.) (S. 35-59). Herbolzheim: Centaurus.
30. Egg, R. (1999). Zur Rückfälligkeit von Sexualstraftätern. *Kriminalistik*, 6, 367-373.
31. Egg, R. (2001). *Rückfälligkeit nach sexuellen Gewalt- und Missbrauchsdelikten – empirische Erkenntnisse und praktische Konsequenzen*. Asept: Königslutter.
32. Endres, J. (2000). Die Kriminalprognose im Strafvollzug: Grundlagen, Methoden und Probleme der Vorhersage von Straftaten. *Zeitschrift für Strafvollzug und Straffälligenhilfe*, 2, 67-83.
33. Endres, J. (2002). Gutachten zur Gefährlichkeit von Strafgefangenen: Probleme und aktuelle Streitfragen der Kriminalprognose. *Praxis der Rechtspsychologie*, 2, 161-182.
34. Freedman, D. (2001). False prediction of future dangerousness: Error rates of the Psychopathy Checklist-Revised. *Journal of the American Academy of Psychiatry and Law*, 29, 89-95.
35. Freese, R. (1999). Deutschsprachige Handbuchbeilage (Supplement) zu S.D. Hart, D.N. Cox & R.D. Hare (Eds.), *The Hare PCL:SV. Psychopathy Checklist: Screening Version*. North Tonawanda-Toronto, Ontario. Multi Health Systems.
36. Gendreau, P., Little, T. & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology*, 34, 575-607.
37. Grann, M., Langström, N., Tengström, A. & Kullgren, G. (1999). Psychopathy (PCL-R) predicts violent recidivism among criminal offenders with personality disorders in Sweden. *Law and Human Behavior*, 23, 205-217.

38. Groß, G. (2004). *Deliktbezogenen Rezidivraten im internationalen Vergleich*. Dissertation, LMU München, Medizinische Fakultät.
39. Grove, W.M. & Meehl, P.E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy and Law*, 2, 293-323.
40. Hanson, R.K. & Bussière, M.T. (1998). Predicting relapse : A meta-analysis of sexual offender recidivism studies. *Journal of Clinical and Consulting Psychology*, 66, 348-326.
41. Hanson, R.K., Scott, H. & Steffy, R.A. (1995). A comparison of child molesters and non-sexual criminals: Risk predictors and long-term recidivism. *Journal of Research in Crime and Delinquency*, 32, 325-337.
42. Hanson, R.K. & Thornton, D. (1999). *Static 99: Improving actuarial risk assessments for sex offenders*. Ottawa: Department of the Solicitor General of Canada.
43. Hare, R.D. (1985). A comparison of procedures for the assessment of Psychopathy. *Journal of Consulting and Clinical Psychology*, 53, 7-16.
44. Hare, R.D. (1991). *The Hare Psychopathy Checklist-Revised: Manual*. Toronto: Multi-Health Systems.
45. Hare, R.D. (1996). Psychopathy: A clinical construct whose time has come. *Criminal Justice and Behavior*, 23, 25-54.
46. Hare, R.D. (1998a). Psychopaths and their nature: Implications for the mental health and the criminal justice system. In: T. Millon, E. Simonsen, M. Burket-Smith & R. Davis (Eds.), *Psychopathy: Antisocial, criminal, and violent behavior* (pp. 188-212). New York: Guilford Press.
47. Hare, R.D. (1998b). *Without Conscience: The disturbing world of the Psychopaths among Us*. New York: Guilford Press.
48. Hare, R.D. (1999). Psychopathy as a risk factor for violence. *Psychiatric Quarterly*, 70, 181-197.
49. Hare, R.D. (2001). Psychopaths and their nature: Some implications for understanding human predatory violence. In: A. Raine & J. Sanmartin (Eds.), *Violence and psychopathy* (pp. 5-34). New York: Kluwer Academic/Plenum Publisher.
50. Hare, R.D. (2003). *Manual for the Psychopathy Checklist-Revisited, 2nd ed.* Toronto, Ontario, Canada: Multi-Health Systems.
51. Hare, R.D., Clark, D., Grann, M. & Thornton, D. (2000). Psychopathy and the predictive validity of the PCL-R: An international perspective. *Behavioral Sciences and the Law*, 18, 623-645.
52. Hare, R.D., Harpur, T.J., Hakstian, A.R., Forth, A.E., Hart, S.D. & Newman, J.P. (1990). The revised Psychopathy-Checklist: Reliability and factor structure. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 1, 6-17.
53. Harris, G.T., Rice, M.E. & Camilleri, J.A. (2004). Applying a forensic actuarial assessment (the Violence Risk Appraisal Guide) to non forensic patients. *Journal of Interpersonal Violence*, 19, 1063-1074.
54. Harris, G.T., Rice, M.E. & Cormier, C.A. (1991). Psychopathy and violent recidivism. *Law and Human Behavior*, 15, 625-637.
55. Harris, G.T., Rice, M.E. & Quinsey, V.L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior*, 20, 315-335.

56. Harris, G.T., Rice, M.E. & Quinsey, V.L. (1994). Psychopathy as a taxon: Evidence that psychopaths are discrete class. *Journal of Consulting and Clinical Psychology*, 62, 387-397.
57. Hart, S.D., Cox, D.N. & Hare, R.D. (1995). *The Hare Psychopathy Checklist: Screening Version*. Toronto: Multi-Health Systems.
58. Hart, S.D. & Hare, R.D. (1997). Psychopathy: Assessment and association with criminal conduct. In: D.M. Stoff, J. Maser & J. Brieling (Eds.), *Handbook of antisocial behavior* (pp. 22-35). New York: Wiley.
59. Hart, S.D., Hare, R.D. & Forth, A.E. (1994). Psychopathy as a risk marker for violence: Development and validation of a screening version of the Revised Psychopathy Checklist. In: J. Monahan & H. Steadman (Eds.), *Violence and Mental Disorder: Developments in Risk Assessment* (pp. 81-98). University of Chicago Press: Chicago.
60. Heilbrun, K., Hart, S.D., Hare, R.D., Gustafson, D., Nunez, C. & White, A.J. (1998). Inpatient and post discharge aggression in mentally disordered offenders: The role of psychopathy. *Journal of Interpersonal Violence*, 13, 514-527.
61. Hemphill, J.F., Hare, R.D. & Wong, S. (1998a). Psychopathy and recidivism: A review. *Legal and Criminological Psychology*, 3, 139-170.
62. Hemphill, J.F., Templeman, R., Wong, S. & Hare, R.D. (1998b). Psychopathy and crime: Recidivism and criminal careers. In: D.J. Cooke, A.E. Forth & R.D. Hare (Eds.), *Psychopathy: Theory, research and implications for society* (pp. 375-399). Dordrecht: Kluwer.
63. Hill, C.D., Rogers, R. & Bickford, M.E. (1996). Predicting aggressive and socially disruptive behavior in a maximum security forensic hospital. *Journal of Forensic Sciences*, 41, 56-59.
64. Kröber, H.-L. (2006). Kriminalprognostische Begutachtung. In: H.-L. Kröber, D. Dölling, N. Leygraf & H. Sass (Hrsg.), *Handbuch der forensischen Psychiatrie – Psychiatrische Kriminalprognose und Kriminaltherapie*. Darmstadt: Steinkopf.
65. Kühl, J. & Schumann, K.F. (1989). Prognosen im Strafrecht – Probleme der Methodologie und Legitimation. *Recht und Psychiatrie*, 7, 126-148.
66. Laurell, J. & Daderman, A.M. (2005). Recidivism is related to psychopathy (PCL-R) in a group of men convicted of homicide. *International Journal of Law and Psychiatry*, 28, 255-268.
67. Leistico, A.-M.R., Salekin, R.T., DeCoster, J. & Rogers, R. (2008). A large-scale meta-analysis relating the Hare measures of Psychopathy to antisocial conduct. *Law and Human Behavior*, 32, 28-45.
68. Lewis, A.H.O. & Webster, C.D. (2004). General instruments for risk assessment. *Current Opinion in Psychiatry*, 17, 401-405.
69. Leygraf, N. (2004). Die Begutachtung der Gefährlichkeitsprognose. In: U. Venzlaff & K. Foerster (Hrsg.), *Psychiatrische Begutachtung (4.Aufl.)* (S. 437-450). München: Urban & Fischer.
70. Leygraf, N. & Nowara, S. (1992). Prognosegutachten. Klinisch-psychiatrische und psychologische Beurteilungsmöglichkeiten der Kriminalprognose. *Forensia-Jahrbuch*, 3, 43-53.
71. Meehl, P.E. (1996). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Northvale, NJ: Jason Aronson (original work published 1954).
72. Monahan, J. (1978). Prediction research and the emergency commitment of mentally ill offenders. *American Journal of Psychiatry*, 135, 198-201.

73. Monahan, J. (1981). *Predicting violent behavior: An assessment of clinical techniques*. Beverly Hills: Sage.
74. Monahan, J. (1984). The Prediction of Violent Behavior: Toward a Second Generation of Theory and Policy. *American Journal of Psychiatry*, 141, 10-15.
75. Monahan, J. (1996). Violence Prediction. The Past Twenty and the Next Twenty Years. *Criminal Justice and Behavior*, 23, 107-120.
76. Monahan, J., Steadman, H.J., Silver, E., Appelbaum, P.S., Robbins, P.C., Mulvey, E.P., Roth, L.H., Grisso, T. & Banks, S. (2001). *Rethinking risk assessment: The MacArthur study of mental disorder and violence*. New York: Oxford University Press.
77. Mossmann, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology*, 62, 783-792.
78. Müller-Isberner, R., Jöckel, D. & Cabeza, S.G. (1998). *Die Vorhersage von Gewalttaten mit dem HCR-20*. Haina: Institut für Forensische Psychiatrie.
79. Nedopil, N. (1998). Möglichkeiten und Grenzen bei Entlassungsprognosen von Rechtsbrechern. *Forensische Psychiatrie und Psychotherapie* 4, 7-21.
80. Nedopil, N. (2005). *Prognosen in der Forensischen Psychiatrie – Ein Handbuch für die Praxis*. Berlin: Pabst Science Publisher.
81. Nedopil N. & Stadtland, C. (2005). Ergebnisse des Münchner Prognoseprojektes. In: N. Nedopil, *Prognosen in der Forensischen Psychiatrie – Ein Handbuch für die Praxis*. Berlin: Pabst Science Publisher.
82. Neuman, C.S., Kosson, D.S. & Salekin, R.T. (2007). Exploratory and confirmatory factor analysis of the psychopathy construct: Methodological and conceptual issues. In: H.Hervé & J.C. Yuille (Eds.), *The psychopath: Theory, Research, and Practice*, (pp. 79-104). Mahwah: New York.
83. Nygaard, D. (2005). *Die prädiktive Validität des HCR-20/2, der PCL-R und der ILRV für die Vorhersage gewalttätiger krimineller Rückfälle – eine Langzeituntersuchung von 185 Gutachtenprobanden*. Dissertation, LMU München, Medizinische Fakultät.
84. Ogloff, J., Wong, S. & Greenwood, A. (1990). Treating criminal psychopaths in a therapeutic community program. *Behavioral Sciences and the Law*, 8, 81-90.
85. Prentky R.A. & Burgess, A.W. (2000). *Forensic management of sexual offenders*. New York.
86. Quinsey, V.L., Harris, G.T., Rice, M.E. & Cormier, C.A. (1998). *Violent offenders: Appraising and managing risk*. Washington DC: American Psychological Association.
87. Quinsey, V.L., Rice, M.E. & Harris, G.T. (1995). Actuarial prediction of sexual recidivism. *Journal of Interpersonal Violence*, 10, 85-105.
88. Rasch, W. (1984). Zur Praxis des Maßregelvollzugs. Verhalten in der Institution als Basis der Prognosebeurteilung. In: I. Eisenbach-Stangl und W. Stangl (Hrsg.), *Grenzen der Behandlung, soziale Kontrolle und Psychiatrie*. Westdeutscher Verlag.
89. Rasch, W. (1999). *Forensische Psychiatrie* (2. überarbeitete Aufl.). Stuttgart: Kohlhammer.
90. Rice, M.E. & Harris, G.T. (1992). A comparison of criminal recidivism among schizophrenic and non schizophrenic offenders. *International Journal of Law and Psychiatry*, 15, 397-408.
91. Rice, M.E. & Harris, G.T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology*, 63, 737-748.
92. Rice, M.E. & Harris, G.T. (2005). Comparing Effect sizes in follow-up studies: ROC area, Cohen's *d*, and *r*. *Law and Human Behavior*, 29, 615-620.

93. Ross, T. & Pfäfflin, F. (2005). Risk Assessment im Maßregelvollzug: Grenzen psychometrischer Gefährlichkeitsprognose im therapeutischen Umfeld. *Monatszeitschrift für Kriminologie und Strafrechtsreform*, 1, 1-11.
94. Rusche, S. (2004). In *Freiheit gefährlich? Eine Untersuchung zu Häufigkeit und Gründen falscher Kriminalprognosen bei psychisch kranken Gewaltverbrechern*. Regensburg: Roderer.
95. Salekin, R.T., Rogers, R. & Sewell, K.W. (1996). A review and meta-analysis of the Psychopathy Checklist and Psychopathy Checklist-Revised: Predictive validity of dangerousness. *Clinical Psychology: Science and Practice*, 3, 203-215.
96. Serin, R.C. (1991). Psychopathy and violence in criminals. *Journal of Interpersonal Violence*, 6, 423-431.
97. Serin, R.C. (1996). Violent recidivism in criminal psychopaths. *Law and Human Behavior*, 20, 207-217.
98. Serin, R.C. & Amos, N.L. (1995). The role of psychopathy in the assessment of dangerousness. *International Journal of Law and Psychiatry*, 18, 231-238.
99. Serin, R.C. & Brown, S.L. (2000). The clinical use of the Hare Psychopathy Checklist-Revised in contemporary risk assessment. In: C. Gacono (Ed.), *The clinical and forensic assessment of psychopathy* (pp. 251-268). Mahwah, NJ: Lawrence Erlbaum.
100. Shine, J. & Hobson, J. (1997). Construct validity of the Hare Psychopathy Checklist-Revised, in a UK prison population. *Journal of Forensic Psychiatry*, 8, 545-561.
101. Skeem, J.L. & Mulvey, E.P. (2001). Psychopathy and community violence among civil psychiatric patients: Results from the MacArthur Violence Risk Assessment Study. *Journal of Consulting and Clinical Psychology*, 69, 358-374.
102. Stadtland, C., Kleindienst, N., Kroner, C., Eidt, M. & Nedopil, N. (2005). Psychopathic traits and risk of criminal recidivism in offenders with and without mental disorders. *International Journal of Forensic Mental Health*, 4, 89-97.
103. Stadtland, C. & Nedopil, N. (2005). Psychiatrische Erkrankungen und die Prognose krimineller Rückfälligkeit. *Nervenarzt*, 76, 1402-1411.
104. Steadman, H.J. (1973). Follow-up on Baxtrom patients returned to hospitals for the criminal insane. *American Journal of Psychiatry*, 130, 317-319.
105. Steadman, H.J. & Cocozza, J. (1974). *Careers of the Criminal Insane*. Lexington, Mass: Lexington Books.
106. Steadman, H.J., Monahan, J., Appelbaum, P.S., Grisso, T., Mulvey, E.P., Roth, L.H., Robbins, P. & Klassen, D. (1994). Designing a new generation of risk assessment research. In: J. Monahan & H.J. Steadman (Eds.), *Violence and Mental Disorder: Developments in Risk Assessment* (pp. 287-318). Chicago: University of Chicago Press.
107. Stone, A.A. (1985). The new legal standard of dangerousness: Fair in theory, unfair in practice. In: C.D. Webster, M.H. Ben-Aron & S.J. Hucker (Eds.), *Dangerousness*. Cambridge, London: Cambridge University Press.
108. Ullrich, S., Paelecke, M., Kahle, I. & Marneros, A. (2003). Kategoriale und dimensionale Erfassung von „Psychopathy“ bei deutschen Straftätern. Prävalenz, Geschlechts- und Alterseffekte. *Nervenarzt*, 74, 1002-1008.
109. Urbaniok, F., Endrass, J., Rossegger, A., Noll, T. (2007a). Violent and sexual offences: A validation of the predictive quality of the PCL:SV in Switzerland. *International Journal of Law and Psychiatry*, 30, 147-152.

110. Urbaniok, F., Noll, T., Rossegger, A. & Endrass, J. (2007b). Die prädiktive Validität der Psychopathy Checklist-Revised (PCL-R) bei Gewalt- und Sexualstraftätern in der Schweiz. *Fortschritte in Neurologie und Psychiatrie*, 75, 155-159.
111. Urbaniok, F., Rossegger, A., Kherfouche, C. & Endrass, J. (2007c). Validität von fokalen Risikoeinschätzungen und Interventionsempfehlungen bei Personen mit Anzeichen für kurz- bis mittelfristige Gefährlichkeit – eine Evaluationsstudie des Züricher Kurzgutachtenprojekts. *Schweizer Archiv für Neurologie und Psychiatrie*, 3, 107-114.
112. Volckart, B. (2002). Zur Bedeutung der Basisrate in der Kriminalprognose. *Recht und Psychiatrie*, 20, 105-114.
113. Walters, G.D. (2003). Predicting institutional adjustment and recidivism with the Psychopathy Checklist factor scores: A meta-analysis. *Law and Human Behaviour*, 27, 541-558.
114. Webster, C.D. (1997). A guide for conducting risk assessment. In: C.D. Webster & J. Jackson (Eds.), *Impulsivity: Theory, Assessment and Treatment*. New York: Guilford.
115. 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.
116. Webster, C.D., Harris, G.T., Rice, M.E., Cormier, C.A. & Quinsey, V.L. (1994). *The violence risk prediction scheme: Assessing dangerousness in high risk men*. Toronto, Canada: University of Toronto.
117. WHO (1993). *Internationale Klassifikation psychischer Störungen. ICD-10 Kapitel V (F). Forschungskriterien*. H. Dilling, W. Mombour, E. Schmidt & E. Schulte-Markwort (Hrsg.). Bern: Huber.
118. WHO (1995). *Schedules for Clinical Assessment in Neuropsychiatry*. M. VanGulick-Bailer, K. Maurer & H. Häfner (Hrsg.). Bern: Huber.
119. WHO (1996). *International Personality Disorder Examination. ICD-10 Modul*. W. Mombour, M. Zaudig, P. Berger, K. Gutierrez, W. Berner, K. Berger, M. vanCranach (Hrsg.). Bern: Huber.
120. Wickens, T.D. (2002). *Elementary Signal Detection Theory*. Oxford.
121. Wilson, N.J. (2003). *The utility of the Psychopathy Checklist Screening-Version for predicting serious violent recidivism in a New Zealand offender sample*. University of Waikato.

8 Thesen

1. Die Evaluation der prognostischen Validität der PCL:SV zur Vorhersage krimineller Rückfälle an einer Stichprobe von 244 Straftätern aus dem Regelvollzug stellt eine der ersten Untersuchungen für den deutschen Sprachraum dar.
2. Die prognostischen Validitäten der PCL:SV-Skalen entsprechen sowohl für die Prognose genereller wie auch für die Prognose gewalttätiger Rückfälle, betrachtet man lediglich die ermittelten Größen der AUCs, weitestgehend denen internationaler Publikationen.
3. Auch wenn die *statistische* Bedeutsamkeit der prognostischen Validität der PCL:SV-Gesamt und ihrer Subskalen zur Vorhersage krimineller Rückfälle in dieser Studie belegt werden konnte, überzeugt dagegen die *praktische* Bedeutsamkeit der prognostischen Validitäten nicht, denn die geschätzten Effektgrößen sind durchweg sehr gering. Dies bedeutet, dass die Überlappung der Verteilungen der PCL:SV-Gesamt (bzw. die der Subskalen) in den Teilgruppen der Rückfälligen bzw. der Nicht-Rückfälligen für prognostische Belange zu groß ist.
4. Für konkrete Prognoseregeln, die auf der Grundlage der PCL:SV-Gesamt (oder einer der beiden Subskalen) formuliert werden, sind die zugehörigen prognostischen Kennwerte wie Sensitivität, Spezifität, *NPP* und *PPP* zu gering, insbesondere sind die Fehlerraten für die Prognose eines kriminellen Rückfalls zu hoch. Fehlerraten dieser Größenordnung sind im forensischen Kontext, in dem sowohl das Schutzinteresse der Allgemeinheit als auch die Freiheitsrechte eines jeden einzelnen Straftäters zu berücksichtigen sind, kaum zu verantworten.
5. Betrachtet man konkrete Prognoseregeln, so ist der Zugewinn an Trefferwahrscheinlichkeit gegenüber *prognostischen Zufallsverfahren* zu gering.
6. Das Cohensche Klassifikationsschema, das häufig der Evaluation von AUC-Werten zugrunde gelegt wird, kann prognostischen Situationen, wie sie hier untersucht wurden, nicht gerecht werden.

7. Das Studium basisratenabhängiger Kennwerte von konkreten Prognoseregeln zur Einschätzung der prognostischen Brauchbarkeit eines Entscheidungsverfahrens ist nicht minder wichtig wie die Untersuchung globaler Merkmale wie die AUC.
8. Die verschiedentlich hervorgehobene Eigenschaft der PCL:SV-Gesamt, eine hoch rückfallgefährdete Risikogruppe zu detektieren (hier: die der *Psychopaths*) erweist sich bei formaler Analyse als eine Eigenschaft, die notwendigerweise *jedes* brauchbare Prognoseinstrument besitzen muss.
9. Die besondere Rolle der PCL:SV-Gesamt für die Prognose eines kriminellen Rückfalls wird durch die Tatsache relativiert, dass wesentlich einfacher feststellbare Merkmale, nämlich *Alter bei erster Verurteilung* und *Anzahl der Vorstrafen*, in ihren prognostischen Eigenschaften nicht schlechter als die der PCL:SV-Gesamt sind.

Anhang

Tabellarischer Lebenslauf

Name:	Renate Reichel
Geburtsdatum und -ort:	06.12.1962 in Magdeburg
Staatsangehörigkeit:	Bundesrepublik Deutschland
Familienstand:	ledig eine Tochter, geboren am 15.12.1987
Schulbildung:	1969-1981 Schulausbildung Allgemeinbildende Schule und Gymnasium Abschluss: Abitur „sehr gut“
Studium:	1984-1989 erstes Studium: klassischer Gesang Hochschule für Musik „Felix Mendelssohn Bartholdy“, Leipzig Abschluss: Staatsexamen „gut“ 1996-2003 zweites Studium: Psychologie Martin-Luther-Universität, Halle Abschluss: Diplom „sehr gut“
Beruf:	1981-1984: Regieassistentin an den Bühnen der Stadt Magdeburg seit 1989: Mitglied des Opernchores in der Oper Halle 2004-2005: wissenschaftliche Mitarbeiterin am MPI für Strafrecht und Kriminologie in Freiburg seit 2005: wissenschaftliche Mitarbeiterin an der Klinik und Poliklinik für Psychiatrie, Psychotherapie und Psychosomatik in Halle

Halle, den 14.04.2009

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorgelegte Dissertation selbstständig und ohne unzulässige Hilfe Dritter verfasst habe. Ich habe mich keiner anderen als der ausdrücklich erwähnten Quellen bedient.

Halle, den 14.04.2009

Erklärung über frühere Promotionsversuche

Hiermit erkläre ich, dass keine Promotionsversuche mit der gleichen oder einer anderen Dissertation erfolgt sind.

Halle, den 14.04.2009

Danksagung

Ganz besonders herzlich möchte ich Herrn Prof. Dr. Dr. Andreas Marneros für die wissenschaftliche Begleitung dieser Arbeit, seine konstruktiven Rückmeldungen und nicht zuletzt für seine außerordentliche Geduld bei der Fertigstellung der Arbeit danken.

Herrn Prof. Dr. Dieter Heyer danke ich für seine kritischen und hilfreichen Anmerkungen zu methodischen und mathematischen Fragen sowie bei inhaltlichen Diskussionen.

Besonderer Dank gilt auch Frau Heidi Baumbach für das Korrekturlesen dieser Arbeit und ihren engagierten Einsatz beim Layout.

Ein liebevolles Dankeschön gebührt meiner Tochter Anna, meinen Eltern sowie allen Verwandten und Freunden, die mich in allen Lebenslagen unterstützt und verständnisvoll begleitet haben.