



ulm university universität
uulm

Universitätsklinik für Psychosomatische Medizin und Psychotherapie

Sektion Forensische Psychotherapie

Leiter: Professor Dr. med. Friedemann Pfäfflin

Kriminalprognose und Sexualdelinquenz – Möglichkeiten und Grenzen standardisierter Kriminalprognosemethoden bei Sexualstraftätern

Dissertation zur Erlangung des Doktorgrades der Humanbiologie der Medizinischen
Fakultät der Universität Ulm

Dipl.-Psych. Martin Rettenberger

aus Passau

2009

Amtierender Dekan: Prof. Dr. Klaus-Michael Debatin

1. Berichterstatter: PD Dr. Reinhard Eher

2. Berichterstatter: Prof. Dr. Klaus Hoffmann

Tag der Promotion: 29. Mai 2009

Inhaltsverzeichnis

ABKÜRZUNGSVERZEICHNIS	II
1. EINLEITUNG	1
2. MATERIAL UND METHODEN	3
3. ERGEBNISSE	8
4. DISKUSSION	12
5. ZUSAMMENFASSUNG	17
6. LITERATURVERZEICHNIS	18
ANHANG	26
DANKSAGUNG	

Abkürzungsverzeichnis

AUC	Area Under Curve
BEST	Begutachtungs- und Evaluationsstelle für Gewalt- und Sexualstraftäter
et al.	et alii, et aliae, und andere
FTZW	Forensisch Therapeutisches Zentrum Wien
ICC	Intraclass-Korrelationskoeffizient
n	Stichprobenumfang
p	Maßzahl der statistischen Signifikanz
PCL-R	Psychopathy Checklist-Revised
r	Produkt-Moment-Korrelationskoeffizient
RRASOR	Rapid Risk Assessment for Sexual Offense Recidivism
ROC	Receiver Operating Characteristic
SACJ-Min	Structured Anchored Clinical Judgement
SORAG	Sexual Offender Risk Appraisal Guide
SPSS	Statistical Package for the Social Sciences
SVR-20	Sexual Violence Risk-20
VRAG	Violence Risk Appraisal Guide

1. Einleitung

Kriminalprognostischen Einschätzungen kommt im internationalen und somit auch im deutschen und österreichischen strafrechtlichen Sanktionssystem eine große Bedeutung zu (z.B. Dahle, 1997, 2000, 2005; Streng, 1995). Neben der richterlichen Abwägung erfolgt seit jeher bei besonders problematischen Fällen eine wissenschaftlich fundierte kriminalprognostische Einschätzung durch akademisch ausgebildete Sachverständige (Deutsch, 2005). Darüber hinaus kam es in den letzten Jahren zu mehreren Gesetzesnovellierungen, die die Bedeutung wissenschaftlich begründeter Kriminalprognosen weiter stärkten (z.B. Prittwitz, 2000; Kröber, 1998; Bogensberger, 2008). Der forensisch ausgebildete Psychologe und Psychiater bedient sich bei der Erstellung einer kriminalprognostischen Stellungnahme unterschiedlicher Methoden (für einen Überblick z.B. Nedopil, 2000, 2005; Dahle, 1997, 2000, 2005), wobei in der vorliegenden Arbeit das Hauptaugenmerk auf der Methodengruppe der standardisierten Kriminalprognoseinstrumente liegt, worunter in erster Linie statistisch-nomothetische oder „aktuarische“ Instrumente, teilweise aber auch standardisierte klinische Verfahren subsumiert werden.

Ursprünglich im angloamerikanischen Sprachraum entwickelt, kommt diesen Instrumenten mittlerweile auch hierzulande eine immer größere Bedeutung zu (Dahle et al., 2007). So finden sie auch in den kürzlich von einer interdisziplinären Expertengruppe veröffentlichten Mindeststandards für Prognosegutachten Beachtung (Boetticher et al., 2007). Wie bereits anhand der Begriffe „standardisiert“, „statistisch-nomothetisch“ oder „aktuarisch“¹ zu erkennen ist, handelt es sich hierbei um den methodischen Idealtypus einer vollständig regelgeleiteten Vorgehensweise (Dahle, 1997, 2000). Die größte Stärke dieser Methoden stellt dabei die – vom Gesetzgeber explizit geforderte und in den eben genannten Mindeststandards festgeschriebene – wissenschaftliche Fundierung dar: Diese drückt sich in einem transparenten Vorgehen, dem Folgen einer nachvollziehbaren Logik sowie der Basierung auf systematisch empirisch gut gesicherten Erkenntnissen aus (Dahle, 2005). Darüber hinaus weisen die meisten Instrumente vielfach empirisch

¹ Der Begriff „aktuarisch“ stammt ursprünglich aus dem versicherungswirtschaftswissenschaftlichen Bereich und beschreibt eine wahrscheinlichkeitstheoretisch und statistisch fundierte Bearbeitung versicherungstechnischer und finanzmarktbezogener Fragestellungen. In der forensischen Prognoseforschung meint „aktuarisch“ eine in den letzten Jahren vorgenommene Weiterentwicklung der statistischen Prognoseinstrumente (Nedopil, 2005).

abgesicherte gute Vorhersagewerte auf, die zum Teil erheblich über der Vorhersageleistung anderer Prognosemethoden liegen (für einen Überblick z.B. Hanson & Morton-Bourgon, 2007; Dahle et al., 2007). Aufgrund methodenimmanenter Nachteile, wie beispielsweise der fehlenden Individuumszentrierung² oder der sogenannten „Mittelfeldproblematik“³, ist allerdings in der gutachterlichen Praxis bei der Verwendung der jeweiligen kriminalprognostischen Ergebnisse Vorsicht geboten. In der Regel wird eine fundierte Ausbildung und Auseinandersetzung mit Stärken und Schwächen der jeweiligen Verfahren vorausgesetzt (siehe ausführlich z.B. bei Dahle, 2005).

Nach Hanson und Morton-Bourgon (2004) zählen die Verfahren *Rapid Risk Assessment for Sexual Offense Recidivism* (RRASOR; Hanson, 1997), *Static-99* (Hanson & Thornton, 1999, 2000), *Sexual Offender Risk Appraisal Guide* (SORAG; Quinsey et al., 2006), *Sexual Violence Risk-20* (SVR-20; Boer et al., 1997) und die *Psychopathy Checklist-Revised* (PCL-R; Hare, 1991)⁴, die Gegenstand der vorliegenden Arbeit sind, zu den wichtigsten Instrumenten dieser Methodengruppe bei kriminalprognostischen Einschätzungen bei Sexualstraftätern⁵. In der vorliegenden Arbeit werden eine Reihe von Reliabilitäts- und Validitätsstudien deutschsprachiger Versionen und Adaptationen dieser Instrumente vorgestellt, um einerseits dem praktisch tätigen Gutachter ein wissenschaftlich fundiertes und anwendungsorientiertes Instrumentarium für die Bearbeitung prognostischer Fragestellungen zur Verfügung zu stellen und um andererseits die Grundlage für zukünftige Forschungsprojekte zur wissenschaftlich fundierten Weiterentwicklung dieser Methoden zu legen.

² Da die Erkenntnisse standardisierter Prognoseinstrumente auf empirisch ermittelte statistische Durchschnittswerte beruhen, sind sie per definitionem nicht in der Lage, individuelle Besonderheiten zu berücksichtigen, so dass eine allein auf standardisierte Verfahren beruhende Prognose-schätzung keine Individualprognose – wie sie vom Gesetzgeber gefordert wird – darstellen kann.

³ Standardisierte Prognoseinstrumente weisen die Eigenschaft auf, dass sich ihre Scorewerte bei wiederholter Anwendung entsprechend der Gaußschen Glockenkurve verteilen. Dies bedeutet, dass überproportional viele Probanden einen durchschnittlichen Summenscore nahe dem Mittelwert aufweisen und somit einer Risikokategorie mit einer „mittleren“ Rückfallwahrscheinlichkeit zugeordnet werden. Diese Zuordnung impliziert in der Folge interpretatorische Schwierigkeiten.

⁴ Bei der PCL-R handelt es sich – im Gegensatz zu den anderen genannten Instrumenten – nicht um ein statistisch-aktuarisches Instrument, sondern um ein standardisiertes klinisches Diagnose- und Prognosverfahren. Aufgrund des Operationalisierungsgrades und des Forschungsstandes kann die PCL-R jedoch auch unter die statistischen Verfahren subsumiert werden (vgl. Dahle, 2005). Als Beleg hierfür kann unter anderem angeführt werden, dass der PCL-R-Score bei manchen aktuarischen Instrumenten (wie z.B. SORAG und SVR-20) integriert wurde.

⁵ Für eine differenziertere Darstellung sowie Auflistung bisheriger empirischer Ergebnisse sei auf die Originalartikel, die der vorliegenden Arbeit zugrunde liegen, sowie auf die einschlägigen Lehrbücher (z.B. Dahle, 2005; Nedopil, 2005) verwiesen.

Aufgrund der angloamerikanischen Herkunft dieser Instrumente bestand zunächst teilweise – nämlich bei RRASOR, Static-99 und SORAG – die Notwendigkeit, deutschsprachige Übersetzungen oder Adaptationen der jeweiligen Manuale und Instrumente zu erstellen. Mittlerweile liegen diese für die bekanntesten Vertreter der standardisierten Prognosemethodik vor (Rettenberger & Eher, 2006a, 2006b, 2007a, Müller-Isberner et al., 2000; Nedopil, 2000). Aufgrund zahlreicher sozialer, kultureller und juristischer Unterschiede zwischen dem deutschen und angloamerikanischen Sprachraum ist eine ungeprüfte Übertragung und Anwendung dieser Methoden hierzulande jedoch nicht zulässig. Die Forderung nach wissenschaftlich fundierten Prognosemethoden schließt vielmehr den Nachweis der Reliabilität und Validität an einer hinreichend großen Stichprobe aus der jeweiligen Jurisdiktion mit ein (sog. „Kreuzvalidierung“; siehe hierzu z.B. Dahle, 2005). Deshalb stand zunächst die Überprüfung der Beurteilerübereinstimmung oder Interraterreliabilität im Fokus des Interesses, anschließend wurde die konvergente sowie die prädiktive Validität der einzelnen Instrumente überprüft, wobei das Hauptaugenmerk auf den eigens übersetzten Instrumenten (d.h. RRASOR, Static-99 und SORAG) lag. Eine in der forensischen Prognoseforschung oft vernachlässigte Fragestellung besteht darüber hinaus in der Überprüfung der differentiellen Validität (Dahle, 2007), durch die relevante Moderatorvariablen identifiziert werden sollen, anhand derer Validitätsunterschiede in der Gesamtpopulation festgestellt werden können und somit der Heterogenität der Gesamtgruppe der Sexualstraftäter (vgl. z.B. Fiedler, 2004) Rechnung getragen wird. Konkret wurde hierfür die Prädiktivität der genannten Instrumente bei unterschiedlichen Subgruppen und für unterschiedliche Rückfallkriterien überprüft.

2. Material und Methoden

Gegenstand der vorliegenden Arbeit sind die standardisierten Rückfallprognoseinstrumente RRASOR, Static-99, SORAG und SVR-20 sowie das spezifisch für den forensischen Kontext entwickelte standardisierte Diagnoseverfahren PCL-R, deren Anwendung gegenwärtig ein fester Bestandteil der kriminalprognostischen Einschätzung bei Sexualstraftätern ist.

Der RRASOR wurde bereits 1997 von R. Karl Hanson veröffentlicht und besteht aus lediglich vier einfach zu erhebenden statischen Items (Hanson, 1997), die anhand einer umfangreichen Meta-Analyse von Verlaufsuntersuchungen aus dem angloamerikanischen Sprachraum extrahiert wurden (Hanson & Bussière, 1996).

Mit dem Ziel, die Vorhersagegüte zu erhöhen, wurde 1999 der Static-99 anhand einer Synthese aus RRASOR und dem *Structured Anchored Clinical Judgement-Minimum* (SACJ-Min; Grubin, 1998) konzipiert. Der Static-99 gilt gegenwärtig aufgrund seines umfangreichen Manuals und der vielfach replizierten moderaten bis guten Validitätsindizes als das am meisten verwendete und am besten validierte standardisierte Prognoseverfahren für Sexualstraftäter (Anderson & Hanson, zur Publikation angenommen).

Der SORAG ist eine Modifikation des *Violence Risk Appraisal Guide* (VRAG; Quinsey et al., 2006) und wurde entwickelt, um gewalttätige (inklusive sexuell motivierte) Rückfälle entlassener Sexualstraftäter vorherzusagen. Trotz einer – im Vergleich zu RRASOR und Static-99 – relativ kleinen Konstruktionsstichprobe und einer weit weniger strukturierten Manualisierung wies auch dieses Instrument sowohl in englischsprachigen als auch in europäischen Kreuzvalidierungsstudien eine hohe Reliabilität und Validität auf (z.B. Nunes et al., 2002; Hanson & Morton-Bourgon, 2007; Barbaree et al., 2001; Ducro & Pham, 2006).

Beim SVR-20 handelt es sich ursprünglich um eine Sammlung von Richtlinien zur systematischen Risikobeurteilung bei Sexualstraftätern, die den klinisch-ideographischen Begutachtungsprozess anleiten sollten (Nedopil, 2005). Bei einer Addition der Itemscorewerte ist jedoch das SVR-20 konzeptionell mit aktuarischen Instrumenten vergleichbar (Hanson & Morton-Bourgon, 2007).

Die PCL-R hingegen stellt zwar in erster Linie eine Persönlichkeitsskala zur diagnostischen Erfassung des Persönlichkeitskonstrukts „psychopathy“ – und deshalb strenggenommen kein Prognoseinstrument – dar (Hare, 1991; Dahle, 2005), wird jedoch aufgrund zahlreicher Befunde, die die Vorhersagegüte des Instruments bei unterschiedlichen delinquenten Subgruppen belegen, in der Regel (auch) als ein standardisiertes Prognoseinstrument verwendet (z.B. Ross & Pfäfflin, 2005; Quinsey et al., 1995; Salekin et al., 1996; Rice & Harris,

1995; Dahle, 2005; Hare et al., 2000). Auch wenn sich bis dato bereits eine Vielzahl internationaler Studien über die Bedeutung der PCL-R und des Persönlichkeitskonstrukts „psychopathy“ im Allgemeinen (für einen Überblick z.B. Nedopil, 2005; Dahle, 2005; Hare, 1991) und über die Bedeutung für die kriminalprognostische Begutachtung von Sexualstraftätern im Besonderen (z.B. Habermeyer & Herpertz, 2005; Hare, 2000) beschäftigt haben, bleiben nach wie vor eine Reihe von ungelösten Forschungsfragen (Nedopil, 2005; Ross & Pfäfflin, 2005), zu deren Beantwortung die vorliegende Arbeit einen Beitrag leisten möchte.

Um die genannten Fragestellungen zu beantworten, wurde fast ausschließlich Datenmaterial der *Begutachtungs- und Evaluationsstelle für Gewalt- und Sexualstraftäter* (BEST; Gerichtsgasse 6, A-1210 Wien; siehe Eher et al., 2007; früher: *Dokumentations- und Koordinationsstelle für Sexualstraftäter im Österreichischen Strafvollzug*) verwendet, das im Sinne eines prospektiven längsschnittlichen Forschungsdesigns analysiert wurde⁶. Dies stellt in methodischer Hinsicht eine qualitative Besonderheit dar, da in der forensischen Prognoseforschung in der Regel retrospektive Designs verwendet werden, die lediglich eine Annäherung an reale Gegebenheiten darstellen und somit eine reduziertere Aussagekraft besitzen (z.B. Helmus & Hanson, 2007). In diesem Zusammenhang ist darauf hinzuweisen, dass in der forensischen Psychiatrie und Psychologie die Prognoseerstellung – insbesondere wenn sie ungünstig ist – Maßnahmen anstößt bzw. anstoßen sollte, die dieser Prognose entgegenarbeiten. Prospektive Forschungsdesigns sind deshalb von derart großem Interesse für die forensische Forschung und Praxis, weil sie Auskunft darüber geben, ob ein Prognoseverfahren trotz gleichzeitig stattfindender Bemühungen, das vorherzusagende Ereignis – den Rückfall – zu verhindern, dennoch valide vorhersagt.

Im Zuge des Dokumentations- und Evaluationsauftrags wurden seit Implementierung der BEST systematisch kriminologische Basisdaten aller seit 01. Januar 2002 zu einer un- bzw. teilbedingt ausgesprochenen Freiheitsstrafe verurteilten Sexualstraftäter erhoben und archiviert. Straftäter, die nach dem Eindruck eines aktenbasierten Auswahlverfahrens besonderer Fachlichkeit bedurften, wurden

⁶ Eine Ausnahme stellt hier lediglich die erste Untersuchung aus dem Jahre 2006 dar (Rettenberger & Eher, 2006c, 2006d), da hier aus methodischen Gründen auf retrospektive Daten, die zum Teil nicht an der BEST generiert wurden, zurückgegriffen werden musste.

zu einer ausführlichen forensisch-psychologischen und -psychiatrischen Begutachtung an die BEST einberufen, wobei unter anderem die bereits genannten standardisierten Kriminalprognoseverfahren Static-99, SVR-20, SORAG, RRA-SOR und PCL-R zur Anwendung kamen.

Das Ziel der vorliegenden Arbeit bestand nun darin, die Interraterreliabilität sowie die konvergente und prädiktive Validität dieser standardisierten Kriminalprognosemethoden zu überprüfen. Die in den einzelnen Studien verwendeten Stichprobendaten stammten mit einer Ausnahme⁷ alle aus dem Datenpool der BEST. Es handelte sich also ausschließlich um männliche Sexualstraftäter, die zwischen 01.01.2002 und dem Stichtag der Vorstrafenregisteranalyse der jeweiligen Studie an der BEST gemeldet und in der Folge mithilfe von aktenbasierten Screening Verfahren analysiert oder direkt begutachtet wurden. Alle Probanden waren zum Zeitpunkt der Erfassung der prädiktiven Validität der Instrumente bereits einen zuvor definierten Zeitraum „on risk“, d.h. sie waren bereits wieder entlassen und somit theoretisch in der Lage, erneut eine (sexuell motivierte) Straftat zu begehen⁸.

Zur Berechnung der Interraterreliabilität und der konvergenten Validität wurden unterschiedliche korrelative Maße verwendet: Zur Quantifizierung der Beurteilerübereinstimmung wurden Intraclass-Korrelationskoeffizienten (ICC) ermittelt, deren Interpretation sich zum einen an denen von unterschiedlichen Autoren vorgeschlagenen Bewertungskriterien (Fleiss, 1986; Greve & Wentura, 1997)⁹, zum anderen an bisherigen angloamerikanischen Ergebnissen (z.B. Harris et al., 2003; Barbaree et al., 2001) orientierte. Im Bezug auf die Interraterreliabilität sprechen Greve und Wentura (1997) ab einem Wert von ICC = .70 von einem zufriedenstellenden Ergebnis, wohingegen Fleiss (1986) die folgenden Interpretations-

⁷ Die bereits angesprochene retrospektive Datenerhebung (Rettenberger & Eher, 2006c) erfolgte am *Forensisch Therapeutischen Zentrum Wien* (FTZW; Novaragasse 9, A-1020 Wien; siehe z.B. Eher et al., 2006), einer seit 1992 bestehenden forensischen Fachambulanz zur psychotherapeutischen und psychiatrischen Nachbetreuung von Straftätern.

⁸ Eine differenziertere Darstellung der jeweils verwendeten Stichproben ist den entsprechenden Originalartikeln zu entnehmen.

⁹ Die Orientierung an allgemein formulierten Richtlinien wird allerdings sogar zum Teil von den genannten Autoren selbst kritisiert, da es in der Regel nicht bzw. nur sehr eingeschränkt möglich ist, allgemeingültige Qualitätskriterien für die Interpretation von Reliabilitätskoeffizienten festzulegen (Wirtz & Casper, 2002). Es erscheint stattdessen sinnvoller, individuell für jede Untersuchung entweder aufgrund inhaltlicher Überlegungen oder anhand der Ergebnisse aus vergleichbaren Studien kritische Werte festzulegen, ab denen die Reliabilität als zufriedenstellend bezeichnet werden kann.

richtlinien vorschlägt: $ICC \geq .75$ = exzellent, $.60 < ICC < .75$ = gut, $.40 < ICC < .60$ = moderat und $ICC \leq .40$ = schwach.

Die konvergente Validität wird überprüft, indem die Ergebnisse verschiedener Instrumente, die vorgeben, das gleiche Konstrukt zu messen, einander gegenübergestellt werden. Konkret wurden – nach Überprüfung des Vorliegens normalverteilter Daten – Produkt-Moment-Korrelationen der prognostischen Einschätzungen der verschiedenen Instrumente, die zur Rückfallprognose bei Sexualstraftätern konzipiert wurden, erhoben. Die Festlegung allgemeingültiger Interpretationsrichtlinien ist hierbei nicht zweckmäßig, die konvergente Validität wird vielmehr dann als gegeben erachtet, wenn sich die entsprechenden Korrelationen signifikant von Null unterscheiden und eine substantielle Höhe erreichen, die je nach Gegenstand unterschiedlich sein kann (vgl. Fisseni, 2004).

Zur Überprüfung der prädiktiven Validität, der gemeinhin das Hauptaugenmerk bei der Qualitätsüberprüfung eines Prognoseverfahrens gilt (Witt, 2000), wurden nun die Prädiktoren – die zum Begutachtungszeitpunkt anhand der standardisierten Prognoseinstrumente vorgenommenen kriminalprognostischen Einschätzungen – dem relevanten Kriterium, nämlich der Rückfälligkeit von Sexualstraftätern, gegenübergestellt. Hierfür wurden jeweils aktuelle Vorstrafenregister von einem für die Prognosedaten „blinden“ Mitarbeiter analysiert und in der Folge mit den Prognosedaten verglichen. Um das lediglich auf den ersten Blick einfache, tatsächlich aber sehr komplexe Konstrukt „Rückfälligkeit“ (z.B. Harris & Hanson, 2004) weiter zu differenzieren, wurden unterschiedliche Rückfallkriterien verwendet (ausführlich z.B. bei Nedopil, 2005), die zum einen unterschiedliche Delikttypen (z.B. allgemeine vs. sexuell motivierte Rückfälligkeit), zum anderen unterschiedliche Schweregrade (z.B. Verurteilung vs. Inhaftierung) erfassten. Zudem wurde versucht, den Fokus auf kriminalpolitisch besonders relevante Rückfallereignisse, wie zum Beispiel die „schwere gewalttätige Rückfälligkeit“, zu legen. Zur Berechnung der prädiktiven Validität wurden die aus den ROC-Analysen (Analysen der *Receiver Operating Characteristics*) resultierenden AUC-Werte (Maßzahlen für die *Area Under Curve*) unter Verwendung des Programms *Statistical Package for the Social Sciences* (SPSS) 11.0 berechnet. Wenn notwendig wurden die Unterschiede zwischen den jeweiligen AUC-Werten unter Zuhilfenahme des Programms *MedCalc* 9.3.1.0 auf Signifikanz hin überprüft. AUC-Werte weisen im Vergleich zu anderen statistischen

Verfahren (z.B. zu korrelativen Maßen) eine Reihe von Vorteilen auf (Hanley & McNeil, 1982; Rice & Harris, 1995, 2005), weshalb sie bei der Überprüfung von prognostischen Fragestellungen im Bereich der klinischen und forensischen Forschung gemeinhin das Mittel der Wahl darstellen (Mossmann, 1994). Zur Interpretation der AUC-Kennwerte wurden kürzlich von Dahle (2007) unter Bezugnahme auf die von Cohen (1992) formulierten Kriterien die folgenden allgemeinen Standards für die Bewertung der Vorhersagegüte von Prognoseinstrumenten postuliert: Werte von $AUC \geq .72$ werden demnach als „hoch“, Werte von $AUC \geq .64$ als „moderat“ und darunter liegende, aber signifikante Werte als „schwach“ eingestuft.

3. Ergebnisse

Zur Überprüfung der Reliabilität und Validität der genannten Verfahren wurde eine Reihe von Einzeluntersuchungen durchgeführt, deren zentrale Ergebnisse im Folgenden überblicksartig dargestellt werden.

Da das Vorliegen einer zufriedenstellenden Reliabilität eine notwendige Voraussetzung für die Überprüfung der Validität eines Messverfahrens ist, bestand zunächst die Notwendigkeit des Nachweises einer zufriedenstellenden Beurteilerübereinstimmung. Gemäß den genannten Interpretationskriterien (Fleiss, 1986; Greve & Wentura, 1997) wurden dabei für alle in der vorliegenden Arbeit verwendeten Verfahren eine exzellente Interraterreliabilität ermittelt (Rettenberger & Eher, 2006c; Rettenberger & Eher, 2007b): Der Static-99 wies eine Interraterreliabilität von $ICC = .98$ auf, der RRASOR von $ICC = .78$, SVR-20 von $ICC = .84$ sowie SO-RAG und PCL-R von jeweils $ICC .93$.

Die zur Überprüfung der konvergenten Validität ermittelten Produkt-Moment-Korrelation waren alle statistisch hoch signifikant, allerdings wiesen die Validitätsindizes zum Teil beträchtliche Unterschiede im Hinblick auf die Stärke des Zusammenhangs auf: So betrug die Korrelation zwischen SO-RAG und SVR-20 beispielsweise $r = .78$, wohingegen die Korrelation zwischen RRASOR und PCL-R lediglich $r = .24$ betrug, was für die Erfassung unterschiedlicher Konstrukte durch die einzelnen Instrumente spricht.

Die ersten Studien zur prädiktiven Validität standardisierter Kriminalprognosemethoden bei Sexualstraftätern bezogen sich auf Static-99 und RRASOR, da diese beiden Instrumente am Beginn des gesamten Arbeitsprozesses standen (Rettenberger & Eher, 2006c, 2006d). Wie bereits dargestellt, wiesen diese Studien methodische Limitierungen in Form eines retrospektiven Datenerhebungsverfahrens und vergleichsweise kleiner Stichproben (jeweils $n = 81$) auf. Dennoch konnten neben den genannten zufriedenstellenden Ergebnissen im Hinblick auf die Interraterreliabilität und konvergente Validität auch erste Belege für die prädiktive Validität beider Instrumente gewonnen werden. Der RRASOR wies dabei für sämtliche Rückfallkriterien (allgemein, sexuell motiviert und gewalttätig) eine moderate Vorhersagegüte auf ($AUC = .66$ bis $.68$), wohingegen die Validitätsindizes des Static-99 Werte zwischen $AUC = .74$ und $.76$ annahmen und somit als „gut“ zu klassifizieren sind.

Im Jahr 2007 wurde schließlich die deutschsprachige Adaptation des SORAG fertiggestellt, weshalb die Überprüfung dieses Instruments nun im Mittelpunkt stand (Rettenberger & Eher, 2007b, 2007c): Anhand eines prospektiven Designs und anhand deutlich größerer Stichproben ($n = 178$ bzw. 254), die auch eine Differenzierung nach relevanten Subgruppen erlaubte, konnte zwar einerseits die prädiktive Validität des Instruments belegt werden, indem für die Gesamtgruppe hoch signifikante Validitätsindizes in Form von AUC-Werten zwischen $AUC = .72$ und $.82$ ermittelt wurden. Andererseits ergaben sich bereits erste Hinweise darauf, dass die Vorhersagegüte je nach Subgruppe und Rückfallkriterium zum Teil beträchtlich differierte: So konnten beispielsweise – trotz überwiegend guter Werte in den restlichen Rückfallkategorien – bei der Subgruppe der Vergewaltigungstäter keine signifikanten Validitätsindizes bezüglich der sexuell motivierten Rückfälligkeit ermittelt werden. Gleiches gilt für die gewalttätige (exklusive der sexuell motivierten) Rückfälligkeit bei der Subgruppe der extrafamiliären pädosexuellen Täter. Aufgrund zu niedriger Rezidivraten – in manchen Rückfallkategorien gab es überhaupt keine Rückfälle in Form von Wiederverurteilungen – konnten für die Subgruppe der innerfamiliären pädosexuellen Täter zum Teil keine AUC-Werte berechnet werden.

Nachdem es nunmehr erste Hinweise auf das Vorliegen der prädiktiven Validität standardisierter Prognoseinstrumente bei deutschsprachigen Stichproben gab, standen zwei weitere Fragestellungen im Raum, die eng mit der Beobach-

tung zusammenhängen, dass die Rezidivraten in den kriminalpolitisch besonders relevanten Rückfallkategorien (schwere sexuell motivierte und/oder gewalttätige Rückfälle) zum Teil unerwartet niedrig ausfielen: So stellte sich zum einen die Frage nach der praktischen Relevanz dieser Ergebnisse (Eher et al., 2008a, 2008b), zum anderen bestand die Notwendigkeit einer systematischen Überprüfung aller standardisierter Kriminalprognoseverfahren bei unterschiedlichen Subgruppen und Rückfallkriterien, die eine direkte Gegenüberstellung der Validitätsindizes erlaubte (Rettenberger et al., zur Publikation angenommen). Die Beantwortung der zweitgenannten Forschungsfrage sollte die Voraussetzung dafür schaffen, die Vorhersagegüte bei unterschiedlichen Subgruppen und Rückfallkriterien direkt vergleichen zu können und somit eine zielgerichtete Auswahl der Verfahren zu ermöglichen.

Bezogen auf das erstgenannte Ziel zeigte sich als wichtigstes Ergebnis, dass eine Diskrepanz zwischen den hohen Validitätsindizes einerseits, die für die jeweiligen Instrumente – in diesem Fall für Static-99 und SORAG – ermittelt wurden, und dem eingeschränkten praktischen Nutzen andererseits bestand (Eher et al., 2008a, 2008b). Der eingeschränkte praktische Nutzen wurde darin verdeutlicht, dass selbst in den durch valide Prognoseinstrumente definierten höchsten Risikogruppen oft weit weniger als 50% der entlassenen Straftäter erneut wieder verurteilt wurden. Es zeigte sich zwar die hypothesenkonforme Tendenz, dass mit Zugehörigkeit zu einer höheren Risikokategorie auch die Rückfallwahrscheinlichkeit anstieg, gleichzeitig wies diese Normwertetabelle aber auch einige Ausreißer auf, die auch zur Vorsicht vor einer allzu unkritischen Übernahme der Interpretationskriterien des angloamerikanischen Originals mahnen sollten.

In einer aktuellen Studie aus dem Jahre 2009 (Rettenberger et al., zur Publikation angenommen) wurden schließlich die bereits genannten fünf Instrumente einer systematischen Prüfung unterzogen, wobei die bereits dargestellten Ergebnisse auch in dieser Untersuchung Bestätigung fanden: So werden zwar wiederum viele moderate bis gute Vorhersagewerte berichtet, insgesamt variieren die Ergebnisse – abhängig von Prognoseverfahren, Rückfallkriterium und Subgruppe – jedoch stark. Bezogen auf relevante Subgruppen von Sexualstraftätern zeigte sich beispielweise, dass die für die Gesamtgruppe ermittelte fast ausschließlich als moderat oder gut zu klassifizierende Vorhersagegüte bei den drei berichteten Untergruppen stark variierte: Bei den Vergewaltigungstätern, bei denen aufgrund zu

niedriger Rezidivraten lediglich die allgemeine und gewalttätige (inklusive sexuell motivierte) Rückfälligkeit untersucht werden konnte, wiesen z.B. nur RRASOR, Static-99 und SORAG zufriedenstellende Werte auf, wohingegen bei den extrafamiliären pädosexuellen Tätern SORAG, SVR-20 und PCL-R die besten Prognosewerte besaßen. Im Hinblick auf die innerfamiliären pädosexuellen Täter konnte aufgrund zu niedriger Rückfallraten lediglich für die allgemeine Rückfälligkeit zufriedenstellende AUC-Werte ermittelt werden, wobei hier der RRASOR, der einen nicht signifikanten Wert aufwies, eine Ausnahme bildete. Kein Instrument war hingegen in der Lage, das kriminal- und vollzugspolitisch besonders relevante Rückfallkriterium der sexuell motivierten Hands-on-Delikte vorherzusagen – weder bei der Gesamtstichprobe noch bei einer der Subgruppen.

Die Notwendigkeit einer weiteren Differenzierung nach relevanten Subgruppen und Rückfallkriterien wurde in einer weiteren aktuellen Studie, in deren Mittelpunkt wiederum der Static-99 stand, nochmals eindrucksvoll nachgewiesen (Eher et al., 2009): Anhand einer für forensische Studien äußerst umfangreichen Stichprobe ($n = 785$), die als repräsentativ für eine inhaftierte Sexualstrafäterpopulation in Österreich bezeichnet werden kann, wurde anhand eines prospektiven Studiendesigns erneut die prädiktive Validität überprüft und der Versuch unternommen, deutschsprachige Normwerte zu generieren. Zunächst wurden allerdings auch bei dieser repräsentativen Sexualstrafäterpopulation die bereits zuvor mehrmals ermittelten und im Vergleich zu internationalen Untersuchungen niedrigeren Rezidivraten bestätigt. Auch bei Verwendung dieser Stichprobe ergaben sich die bereits bekannten moderaten bis guten AUC-Werte für die Gesamtstichprobe, die allerdings bei weiterer Differenzierung nach unterschiedlichen Deliktgruppen und Rückfallkategorien wiederum eine große Spannbreite aufwiesen: Die ermittelten AUC-Werte lagen zwischen dem statistisch signifikanten Wert von $AUC = .94$ (Vorhersage des „Rückfall ‚violent‘ – Haft“ bei der Subgruppe der innerfamiliären Kindesmissbraucher) und AUC-Werten, die bei mehreren Rückfallkategorien maximal eine schwache Prognosegüte indizierten (z.B. betrug der Validitätsindex für die Vorhersagegüte der allgemeinen Rückfälligkeit bei der Gruppe der außerfamiliären Kindesmissbraucher lediglich $AUC = .63$, $p < 0,01$) und zum Teil auch gar keine statistische Signifikanz erreichten (hier sei beispielsweise auf die Prognose sexuell motivierter Straftaten bei Vergewaltigungstäter verwiesen). Die deutschsprachigen Normwertetabellen wiesen zwar – zumindest tendenziell – den

hypothesekonformen kontinuierlichen Anstieg der Rückfallereignisse mit Zunahme des Static-99-Wertes auf. Andererseits aber zeigten sich – wiederum abhängig von Subgruppe und Rückfallkriterium – einige Abweichungen und Unregelmäßigkeiten (stellvertretend hierfür seien die Normwerte der Gruppe der innerfamiliären Kindesmissbraucher genannt).

4. Diskussion

Die Ergebnisse der durchgeführten Studien belegen weitgehend den prognostischen Wert standardisierter Prognoseinstrumente bei Sexualstraftätern. Sie mahnen jedoch gleichzeitig zur Vorsicht im Hinblick auf die Interpretation und praktische Verwendung der entsprechenden Prognosedaten, da die Ergebnisse vor allem der prädiktiven Validität je nach verwendetem Prognoseinstrument, Rückfallkriterium und Sexualstraftätersubgruppe zum Teil stark variieren. Es wird somit deutlich, dass ein differenziertes Vorgehen, sowohl bei der Anwendung des Instruments selbst, als auch bei der Interpretation der Prognoseergebnisse, im Einzelfall unerlässlich und eine genaue Kenntnis methodenimmanenter Stärken und Schwächen unabdingbar ist. Mit anderen Worten, um die Möglichkeiten dieser Verfahren optimal zu nutzen, ist es notwendig, deren Grenzen genau zu kennen.

Zunächst stand die Überprüfung der Interraterreliabilitätsindizes im Fokus des Interesses. Diesbezüglich weisen sämtliche Prognoseverfahren auch in den deutschen Versionen und Adaptationen exzellente Werte auf, weshalb davon ausgegangen werden kann, dass es sich hierbei um reliable Prognoseinstrumente handelt. Es muss allerdings in diesem Zusammenhang nochmals darauf hingewiesen werden, dass dies nur für forensisch erfahrene sowie in den jeweiligen Instrumenten ausgebildete Rater gilt. Es ist davon auszugehen, dass bei weniger erfahrenen und weniger gut ausgebildeten Ratern die Beurteilerübereinstimmung – und damit in weiterer Konsequenz auch die Vorhersagegüte – deutlich niedriger ist (vgl. Rettenberger, 2006).

Auch die Ergebnisse der konvergenten Validität können insgesamt als zufriedenstellend bezeichnet werden, da die hypothetisierten Zusammenhänge durch hoch signifikante positive Korrelationen bestätigt wurden (z.B. Rettenberger & Eher, 2007b; Rettenberger et al., zur Publikation angenommen). Mögliche Erklä-

rungsansätze für die berichteten Ergebnisse liefern Studien und theoretische Abhandlungen über die Dimensionen und die entsprechenden Faktoren, die dem Konstrukt „Rückfälligkeit“ bei Sexualstraftätern zugrunde liegen (z.B. Doren, 2004; Barbaree et al., 2006a, 2006b; Eher et al., 2003). Die Kernaussage dieser Untersuchungen besteht darin, dass es zumindest zwei voneinander getrennte Dimensionen bei der Rückfallvorhersage bei Sexualstraftätern gibt: Auf der einen Seite die Dimension, die auf die sexuelle Devianz von Sexualdelinquenten fokussiert, und auf der anderen Seite eine Dimension, die den Bereich eines allgemeinen impulsiven und dissozialen Lebensstils abdeckt. Aktuelle faktorenanalytische Studien (Barbaree et al., 2006b) untersuchten diese hypothetisierten Zusammenhänge empirisch, wobei sie einerseits Belege für die postulierten multidimensionalen Konzeptionen fanden, andererseits ein differenzierteres Bild zeichneten, als man zuvor anhand der beiden genannten Komponenten (Dissozialität vs. sexuelle Devianz) angenommen hatte. Zusammenfassend belegen die Resultate somit zum einen, dass es sich bei den hier verwendeten Instrumenten um konvergent valide Verfahren handelt. Zum anderen liefern sie Ansatzpunkte und Hypothesen für weitere Studien zur Faktorenstruktur standardisierter Prognoseverfahren.

Nach ersten Hinweisen aus retrospektiven Untersuchungen konnten nachfolgende Studien die prädiktive Validität der deutschsprachigen Versionen und Adaptationen der bekanntesten standardisierten Prognoseinstrumente auch innerhalb eines prospektiven Forschungsdesigns belegen. Durch die Anwendung eines prospektiven Forschungsdesigns konnte somit eine – von wenigen Ausnahmen abgesehen (z.B. Helmus & Hanson, 2007) – bis dahin bestehende Lücke der nationalen und internationalen forensischen Prognoseforschung geschlossen werden. Die Resultate belegen zusammenfassend den Wert standardisierter Prognoseverfahren für die forensische Prognostik, verdeutlichen jedoch andererseits auch die Notwendigkeit einer differenzierten Betrachtung sowie einer vorsichtigen Interpretation der Ergebnisse im Zuge der praktischen Anwendung. So variierte die Vorhersagegüte von RRASOR, Static-99, SORAG, SVR-20 und PCL-R je nach Rückfallkriterium und Subgruppe stark, was bei der Anwendung dieser Instrumente in der Praxis zu berücksichtigen ist. Außerdem ist zu beachten, dass es sich bei kriminalpolitisch besonders relevanten Rückfallereignissen – wie zum Beispiel gravierende Sexual- oder Gewaltverbrechen – nach wie vor um sehr seltene Ereignisse handelt, was zwangsläufig zu (zum Teil zu) niedrigen Rezidivraten führt.

Dies macht nicht nur eine Überprüfung von Gütekriterien teilweise unmöglich, weshalb der entsprechende Nachweis nach wie aussteht, sondern geht auch mit der ganz allgemeinen Frage einher, ob derart seltene, vollzugs- und gesellschaftspolitisch allerdings höchst relevante Ereignisse überhaupt empirisch vorhergesagt werden können (vgl. hierzu auch Rice et al., 2006). Damit im Zusammenhang zu sehen ist der nach wie vor fehlende Beleg dafür, dass standardisierte Prognoseinstrumente den sexuell motivierten Rückfall bei Vergewaltigungstätern (einer repräsentativen Stichprobe entlassener Strafgefangenen) vorhersagen können. Auch bei der Subgruppe der innerfamiliären pädosexuellen Täter indizieren die bisherigen Ergebnisse lediglich eine zufriedenstellende Prognosegüte im Bezug auf die allgemeine Rückfälligkeit, die – im Vergleich zu den anderen Rückfallkriterien, die in den vorliegenden Untersuchungen verwendet wurden – aus kriminalpolitischer Sicht eher von untergeordneter Bedeutung ist. Es bleibt als vorläufiges Fazit, dass lediglich für die Gruppe der ausserfamiliären sexuellen Kindesmissbraucher zufriedenstellende Vorhersageleistungen betreffend einer sexuell motivierten Wiederholungstat geleistet werden können.

Diese Einschätzung gilt – wie oben erwähnt – auch bei Verwendung sehr hoher Fallzahlen einer repräsentativen Stichprobe inhaftierter Sexualstraftäter (Eher et al., 2009). Trotz Verwendung einer derart umfangreichen Stichprobe variierte die Prognosegüte je nach Subgruppe und Rückfallkriterium stark, so dass der kritische Umgang mit diesen Verfahren in der Praxis durch ausschließlich gut ausgebildete Forensiker nachdrücklich zu fordern ist und insgesamt von der alleinigen Anwendung dieser Verfahren bei kriminalprognostischen Stellungnahmen abgeraten werden muss. Allerdings muss hierbei einschränkend darauf hingewiesen werden, dass in der zuletzt genannten Untersuchung lediglich der Static-99 untersucht wurde, da für die anderen Instrumente keine prospektiven Daten für die gesamte Stichprobe vorlagen.

Mit Rückgriff auf das Konzept der differentiellen Validität ist zu sagen, dass es sich bei Index-Delikt und Rückfallvariable um prognoserelevante Moderatorvariablen handelt, die in zukünftigen Untersuchungen in jeden Fall zu berücksichtigen sind. So ergibt eine Differenzierung der Gesamtpopulation nach Index-Delikt Subgruppen, für die entweder unterschiedliche Validitätsindizes zu erwarten sind (z.B. Vergewaltigungstäter sowie extra- und innerfamiliäre Täter) oder bei denen nach wie vor der Nachweis der prädiktiven Validität noch aussteht, wie dies bei-

spielsweise bei jugendlichen Sexualstraftätern (Dahle et al., 2008), bei sogenannten Hands-off-Tätern (Barbaree et al., 2006a; Helmus & Hanson, 2007) sowie bei Tätern mit sexuell assoziierten Tötungsdelikten (Berner et al., 2008) der Fall ist.

Im Zusammenhang mit diesen Ergebnissen sind die folgenden methodischen Einschränkungen der vorliegenden Arbeit zu berücksichtigen: Aufgrund des prospektiven Forschungsdesigns und der damit zwangsläufig verbundenen Restriktionen bei der Festlegung des Nachbeobachtungszeitraums war eine weitere Verlängerung des Untersuchungszeitraums zu den jeweiligen Datenauswertungszeitpunkten nicht möglich. Die unter anderem deshalb niedrigen Rezidivraten verhinderten zum Teil eine sinnvolle Datenanalyse. Es ist zu erwarten, dass bei längeren Nachbeobachtungszeiträumen die Rezidivraten höher werden, wodurch zum Beispiel eine (derzeit noch ausstehende) Überprüfung der Vorhersagegüte bei gravierenden Sexual- und/oder Gewaltstraftaten in manchen Untergruppen erst ermöglicht werden könnte. Auch sind bei Verlängerung der Nachbeobachtungszeiträume mehr Möglichkeiten bezüglich der Ermittlung von praxisrelevanten Normwerttabellen gegeben. Zusätzlich werden in zukünftigen Untersuchungen noch größere Stichproben zur Verfügung stehen, die zum einen eine bessere Absicherung und klarere Interpretation der ermittelten Ergebnisse erlauben und zum anderen eine immer feinere Differenzierung in psychologisch, psychiatrisch und kriminologisch relevante Subgruppen zulassen.

Neben der Ermittlung zusätzlicher prognoserelevanter Moderatorvariablen gibt es eine Reihe weiterer derzeit noch bestehender Forschungsdesiderate, worunter unter anderem eine eingehendere Untersuchung der Faktorenstruktur der Variablen, die bei der Verwendung standardisierter Prognoseinstrumente eingesetzt werden, zu nennen ist (Barbaree et al., 2006b; Doren, 2004). Von besonderem Interesse ist auch die Frage nach einem prognoserelevanten Informationsgewinn bei der gleichzeitigen Verwendung von mehreren der genannten standardisierten Verfahren (Seto, 2005) sowie der Frage nach einer systematischen Verbindung aktuarisch-statistischer und klinisch-ideographischer Prognoseansätze (Boer, 2006). Zuletzt explizit genannt sei schließlich die Identifizierung, Operationalisierung und Integration (stabil und akut) dynamischer Risikovariablen, die im Verbund mit den hier verwendeten Prognosemethoden zu einer weiteren Verbesserung kriminalprognostischer Urteile beitragen könnten (vgl. Hanson & Harris, 2001; Hanson et al., 2007).

Die Ergebnisse der vorliegenden Arbeit verdeutlichen somit einerseits den potentiellen Nutzen, der aus der Anwendung standardisierter Prognoseverfahren hervorgeht, zeigen aber andererseits gleichzeitig die Grenzen der praktischen Anwendbarkeit auf, wodurch die Notwendigkeit von theoretischem Wissen und praktischer Erfahrung beim Anwender umso deutlicher wird. Zusammenfassend ist der Einsatz dieser Methoden bei kriminalprognostischen Einschätzungen von Sexualstraftätern in jedem Fall zu empfehlen – sofern ein verantwortungsvoller Umgang gewährleistet und die Möglichkeiten und Grenzen der Methoden berücksichtigt werden.

5. Zusammenfassung

Gegenstand der vorliegenden Arbeit sind standardisierte Kriminalprognosemethoden zur Vorhersage von Rückfällen bei entlassenen Sexualstraftätern. Ursprünglich im angloamerikanischen Sprachraum entwickelt kommen diesen Instrumenten, die in der Regel als statistisch-nomothetische oder „aktuarische“ Prognoseverfahren bezeichnet werden, nicht zuletzt aufgrund aktueller Gesetzesnovellierungen auch im deutschen Sprachraum zunehmend mehr Gewicht zu. Anhand eines prospektiven längsschnittlichen Forschungsdesigns wird die Reliabilität in Form der Beurteilerübereinstimmung sowie die konvergente und prädiktive Validität der deutschen Versionen und Adaptationen des *Rapid Risk Assessment for Sexual Offense Recidivism* (RRASOR), des *Static-99*, des *Sexual Offender Risk Appraisal Guide* (SORAG), des *Sexual Violence Risk-20* (SVR-20) und der *Psychopathy Checklist-Revised* (PCL-R) überprüft. Die Ergebnisse mehrerer Untersuchungen, die Teil des vorliegenden Forschungsprojektes waren, belegen zwar einerseits anhand zufriedenstellender Reliabilitäts- und Validitätsindizes den potentiellen prognostischen Nutzen dieser Verfahren bei der Vorhersage von Rückfällen bei Sexualstraftätern, andererseits variiert vor allem die prädiktive Validität je nach Prognoseinstrument, Rückfallkriterium und Sexualstraftätersubgruppe zum Teil erheblich, weshalb bei der Interpretation der Prognoseergebnisse die Grenzen der standardisierten Kriminalprognosemethodik berücksichtigt werden müssen.

6. Literaturverzeichnis

- (1) Anderson, D. & Hanson, R. K. (zur Publikation angenommen). Static-99: An Actuarial Tool to Assess Risk of Sexual and Violent Recidivism Among Sexual Offenders. In R.K.Otto & K. Douglas (Hrsg.), *Handbook of Violence Risk Assessment Tools*. Milton Park: Routledge.
- (2) Barbaree, H. E., Seto, M. C., Langton, C. M. & Peacock, E. J. (2001). Evaluating the predictive Accuracy of six risk assessment instruments for adult sex offenders. *Criminal Justice and Behavior*, 28, 490-521.
- (3) Barbaree, H. E., Langton, C. M. & Peacock, E. J. (2006a). Different Actuarial Risk Measures Produce Different Risk Rankings for Sexual Offenders. *Sexual Abuse: A Journal of Research and Treatment*, 18, 423-440.
- (4) Barbaree, H. E., Langton, C. M. & Peacock, E. J. (2006b). The Factor Structure of Static Actuarial Items: Its Relation to Prediction. *Sexual Abuse: A Journal of Research and Treatment*, 18, 207-226.
- (5) Berner, W., Briken, P., Habermann, N. & Hill, A. (2008). Verläufe bei sexuellen Tötungen. *Forensische Psychiatrie, Psychologie, Kriminologie*, 2, 105-111.
- (6) Boer, D. P., Hart, S. D., Kropp, P. R. & Webster, C. D. (1997). *Manual for the Sexual Violence Risk - 20: Professional Guidelines for Assessing Risk of Sexual Violence*. Vancouver: The Mental Health, Law, & Policy Institute.
- (7) Boer, D. P. (2006). Sexual Offender Risk Assessment Strategies: Is There a Convergence of Opinion Yet? *Sexual Offender Treatment*, 1, 1-4.
- (8) Boetticher, A., Kröber, H.-L., Müller-Isberner, R., Böhm, K. M., Müller-Metz, R. & Wolf, T. (2007). Mindestanforderungen für Prognosegutachten. *Forensische Psychiatrie, Psychologie, Kriminologie*, 1, 90-100.
- (9) Bogensberger, W. (2008). *Neuordnung der bedingten Entlassung durch das Strafrechtsänderungsgesetz 2008*. Vortrag auf der 15. Forensisch-psychiatrischen Tagung, Wien.

- (10) Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- (11) Dahle, K. P. (1997). Kriminalprognosen im Strafrecht. Psychologische Aspekte individueller Verhaltensvorhersagen. In M. Steller & R. Volbert (Hrsg.), *Psychologie im Strafverfahren* (S. 119-140). Bern: Huber.
- (12) Dahle, K.-P. (2000). Psychologische Begutachtung zur Kriminalprognose. In H.-L.Krüber & M. Steller (Hrsg.), *Psychologische Begutachtung im Strafverfahren* (pp. 77-111). Darmstadt: Steinkopff.
- (13) Dahle, K.-P. (2005). *Psychologische Kriminalprognose*. Herbolzheim: Centaurus.
- (14) Dahle, K.-P. (2007). Methodische Grundlagen der Kriminalprognose. *Forensische Psychiatrie, Psychologie, Kriminologie*, 1, 101-110.
- (15) Dahle, K.-P., Schneider, V. & Ziethen, F. (2007). Standardisierte Instrumente zur Kriminalprognose. *Forensische Psychiatrie, Psychologie, Kriminologie*, 1, 15-26.
- (16) Dahle, K.-P., Janka, C., Gallasch, F. & Lehmann, R. (2008). Jugendliche Sexualstraftäter: Spezialfälle der Prognoseerstellung? *Forensische Psychiatrie, Psychologie, Kriminologie*, 2, 213-221.
- (17) Deutsch, C. C. (2005). Prognosegutachten bei Strafaussetzung. In D. Schläfke, F. Häßler & J. M. Fegert (Hrsg.), *Sexualstraftaten* (S. 33-45). Stuttgart: Schattauer.
- (18) Doren, D. M. (2004). Toward a Multidimensional Model for Sexual Recidivism Risk. *Journal of Interpersonal Violence*, 19, 835-856.
- (19) Ducro, C. & Pham, T. (2006). Evaluation of the SORAG and the Static-99 on Belgian Sex Offenders Committed to a Forensic Facility. *Sexual Abuse: A Journal of Research and Treatment*, 18, 15-26.
- (20) Eher, R., Neuwirth, W., Frühwald, S. & Frottier, P. (2003). Sexualization and Lifestyle Impulsivity: Clinically Valid Discriminators in Sexual Of-

fenders. *International Journal of Offender Therapy and Comparative Criminology*, 47, 452-467.

- (21) Eher, R., Lackinger, F., Frühwald, S. & Frottier, P. (2006). Beziehungsorientierte Psychotherapie bei entlassenen Straftätern und einer Gruppe von Sexualstraftätern – Ergebnisse einer 7-Jahreskatamnese. *Recht & Psychiatrie*, 24, 83-90.
- (22) Eher, R., Frühwald, S. & Frottier, P. (2007). Die standardisierte Begutachtung von Sexualstraftätern im Österreichischen Strafvollzug - Erste Ergebnisse und Ausblicke. *Neuropsychiatrie*, 20, 50-55.
- (23) Eher, R., Rettenberger, M., Schilling, F. & Pfäfflin, F. (2008a). Validität oder praktischer Nutzen? Rückfallvorhersagen mittels Static-99 und SORAG. Eine prospektive Rückfallstudie an 275 Sexualstraftätern. *Recht & Psychiatrie*, 26, 79-88.
- (24) Eher, R., Rettenberger, M., Schilling, F. & Pfäfflin, F. (2008b). Failure of Static-99 and SORAG to Predict Relevant Reoffense Categories in Relevant Sexual Offender Subtypes: A Prospective Study. *Sexual Offender Treatment*, 3, 1-14.
- (25) Eher, R., Rettenberger, M. & Matthes, A. (2009). Möglichkeiten und Grenzen der aktuarischen Prognose bei Sexualstraftätern: Ergebnisse einer prospektiven Studie an 785 nachuntersuchten Tätern. *Monatsschrift für Kriminologie und Strafrechtsreform*, 92, 18-27.
- (26) Fiedler, P. (2004). *Sexuelle Orientierung und sexuelle Abweichung*. Weinheim: Beltz.
- (27) Fisseni, H.-J. (2004). *Lehrbuch der psychologischen Diagnostik*. Göttingen: Hogrefe.
- (28) Fleiss, J. L. (1986). *The Design and Analysis of Clinical Experiments*. New York: Wiley.
- (29) Greve, W. & Wentura, D. (1997). *Wissenschaftliche Beobachtung in der Psychologie*. München: Quintessenz.

- (30) Grubin, D. (1998). *Sex offending against children: Understanding the risk*. Police Research Series Paper 99. London: Home Office.
- (31) Habermeyer, E. & Herpertz, S. (2005). Bedeutung des "psychopathy"-Konzepts von Hare für Sexualstraftaten. In D. Schläfke, F. Häßler & J. M. Fegert (Hrsg.), *Sexualstraftaten* (S. 131-141). Stuttgart: Schattauer.
- (32) Hanley, J. A. & McNeil, B. J. (1982). The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. *Radiology*, 143, 29-36.
- (33) Hanson, R. K. & Bussière, M. T. (1996). *Predictors of Sexual Offender Recidivism: A Meta-Analysis*. User Report 1996-04. Ottawa: Department of the Solicitor General of Canada.
- (34) Hanson, R. K. (1997). *The Development of a Brief Actuarial Risk Scale for Sexual Offense Recidivism*. (User Report No. 1997-04). Ottawa: Department of the Solicitor General of Canada.
- (35) Hanson, R. K. & Thornton, D. (1999). *Static-99: Improving actuarial risk assessments for sex offenders*. User Report 1999-02. Ottawa: Department of the Solicitor General of Canada.
- (36) Hanson, R. K. & Thornton, D. (2000). Improving Risk Assessment for Sex Offenders: A Comparison of Three Actuarial Scales. *Law and Human Behavior*, 24, 119-136.
- (37) Hanson R. K. & Harris A. J. R (2001). A Structured Approach to Evaluating Change Among Sexual Offenders. *Sexual Abuse: A Journal of Research and Treatment*, 13, 105-122.
- (38) Hanson, R. K. & Morton-Bourgon, K. (2004). *Predictors of Sexual Recidivism: An Updated Meta-Analysis*. (User Report No. 2004-02). Canada: Public Safety and Emergency Preparedness.
- (39) Hanson R. K., Harris A. J. R., Scott T.-L. & Helmus L. M. D. (2007). *Assessing the Risk of Sexual Offenders on Community Supervision: The*

Dynamic Supervision Project. (User Report No. 2007-05). Canada: Public Safety and Emergency Preparedness.

- (40) Hanson, R. K. & Morton-Bourgon, K. (2007). *The Accuracy of Recidivism Risk Assessment for Sexual Offenders: A Meta-Analysis. (User Report No. 2007-01).* Canada: Public Safety and Emergency Preparedness.
- (41) Hare, R. D. (1991). *Manual for the Psychopathy-Checklist-Revised.* Toronto: Multi Health Systems.
- (42) Hare, R. D. (2000). Eigenschaften von antisozialen Borderline-Patienten und Psychopathen: Konsequenzen für das Gesundheitswesen und das Strafrecht. In O. F. Kernberg, B. Dulz & U. Sachsse (Hrsg.), *Handbuch der Borderline-Störungen* (S. 393-411). Stuttgart: Schattauer.
- (43) 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.
- (44) Harris, A.J.R. & Hanson, R. K. (2004). *Sex Offender Recidivism : A Simple Question (User Report No. 2004-03).* Canada: Public Safety and Emergency Preparedness.
- (45) Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumière, M. L., Boer, D. & Lang, C. (2003). A Multisite Comparison of Actuarial Risk Instruments for Sex Offenders. *Psychological Assessment*, 15, 413-425.
- (46) Helmus, L. M. D. & Hanson, R. K. (2007). Predictive Validity of the Static-99 and Static-2002 for Sex Offenders on Community Supervision. *Sexual Offender Treatment*, 2, 1-14.
- (47) Kröber, H.-L. (1998). Die Strafrechtsreformen zur Sexual- und Gewaltdelinquenz. *Zeitschrift für Sexualforschung*, 11, 59-66.
- (48) Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology*, 62, 783-792.

- (49) Müller-Isberner, R., Cabeza, S. G. & Eucker, S. (2000). *Die Vorhersage sexueller Gewalttaten mit dem SVR-20*. Haina: Insitut für Forensische Psychiatrie.
- (50) Nedopil, N. (2000). *Forensische Psychiatrie*. Stuttgart: Thieme.
- (51) Nedopil, N. (2005). *Prognosen in der Forensischen Psychiatrie. Ein Handbuch für die Praxis*. Lengerich: Pabst Science Publishers.
- (52) Nunes, K. L., Firestone, P., Bradford, J. M., Greenberg, D. M. & Broom, I. (2002). A Comparison of Modified Versions of the Static-99 and the Sex Offender Risk Appraisal Guide (SORAG). *Sexual Abuse: A Journal of Research and Treatment*, 14, 253-269.
- (53) Prittwitz, C. (2000). Reform des Sexualstrafrechts - Verantwortbare Risiken? In J. M. Fegert & F. Häßler (Hrsg.), *Qualität forensischer Begutachtung, insbesondere bei Jugenddelinquenz und Sexualstraftaten* (S. 107-135). Herbolzheim: Centaurus.
- (54) Quinsey, V. L., Rice, M. E. & Harris, G. T. (1995). Actuarial Prediction of Sexual Recidivism. *Journal of Interpersonal Violence*, 10, 85-105.
- (55) Quinsey, V. L., Harris, G. T., Rice, M. E. & Cormier, C. (2006). *Violent Offenders: Appraising and Managing Risk*. Washington, DC: American Psychological Association.
- (56) Rettenberger, M. (2006). *Deutschsprachige Adaption und Validierung der revidierten Version des Static-99*. Unveröffentlichte Diplomarbeit an der Freien Universität Berlin.
- (57) Rettenberger, M. & Eher, R. (2006a). *Die revidierten Kodierungsrichtlinien des Static-99*. Wien: Verein für forensische Forschung und Weiterbildung.
- (58) Rettenberger, M. & Eher, R. (2006b). *Screening-Instrument zur Einschätzung der Gefährlichkeit und Rückfallwahrscheinlichkeit bei Sexualstraftätern: Die deutsche Version des RRASOR*. Wien: Verein für forensische Forschung und Weiterbildung.

- (59) Rettenberger, M. & Eher, R. (2006c). Actuarial Assessment of Sex Offender Recidivism Risk: A Validation of the German Version of the Static-99. *Sexual Offender Treatment, 1*, 1-11.
- (60) Rettenberger, M. & Eher, R. (2006d). Die deutsche Übersetzung des Static-99 zur aktuarischen Kriminalprognose verurteilter Sexualstraftäter. *Monatsschrift für Kriminologie und Strafrechtsreform, 89*, 352-365.
- (61) Rettenberger, M. & Eher, R. (2007a). *Aktuarische Kriminalprognose bei Sexualstraftägern: Die deutsche Version des Sex Offender Risk Appraisal Guide (SORAG)*. Wien: Verein für forensische Forschung und Weiterbildung.
- (62) Rettenberger, M. & Eher, R. (2007b). Aktuarische Kriminalprognosemethoden und Sexualdelinquenz: Die deutsche Version des SORAG. *Monatsschrift für Kriminologie und Strafrechtsreform, 90*, 484-497.
- (63) Rettenberger, M. & Eher, R. (2007c). Predicting Reoffence in Sexual Offender Subtypes: A Prospective Validation Study of the German Version of the Sexual Offender Risk Appraisal Guide (SORAG). *Sexual Offender Treatment, 1*, 1-12.
- (64) Rettenberger, M., Matthes, A., Boer, D. & Eher, R. (zur Publikation angenommen). Prospective Actuarial Risk Assessment: A Comparison of Five Risk Assessment Instruments in Different Sexual Offender Subtypes. *International Journal of Offender Therapy and Comparative Criminology*.
- (65) Rice, M. E. & Harris, G. T. (1995). Violent Recidivism: Assessing Predictive Validity. *Journal of Consulting and Clinical Psychology, 63*, 737-748.
- (66) 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.
- (67) Rice, M. E., Harris, G. T., Lang, C. & Cormier, C. (2006). Violent Sex Offenses: How are They Best Measured from Official Records? *Law and Human Behavior, 30*, 525-541.

- (68) Ross, T. & Pfäfflin, F. (2005). Risk Assessment im Maßregelvollzug: Grenzen psychometrischer Gefährlichkeitsprognose im therapeutischen Umfeld. *Monatsschrift für Kriminologie und Strafrechtsreform*, 88, 1-11.
- (69) Salekin, R. T., Rogers, R. & Sewell, K. W. (1996). A Review and Meta-Analysis of the Psychopathy Checklist and Psychopathy Checklist – Revised. *Clinical Psychology: Science and Practice*, 3, 203-215.
- (70) Seto, M. C. (2005). Is More Better? Combining Actuarial Risk Scales to Predict Recidivism Among Adult Sex Offenders. *Psychological Assessment*, 17, 156-167.
- (71) Streng, F. (1995). Strafrechtliche Folgenorientierung und Kriminalprognose. In D. Dölling (Hrsg.), *Die Täter-Individualprognose* (S. 97-127). Heidelberg: Kriminalistik.
- (72) Wirtz, M. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe.
- (73) Witt, P. H. (2000). A Practitioner`s View of Risk Assessment: The HCR-20 and SVR-20. *Behavioral Sciences and the Law*, 18, 791-798.

Anhang (Originalartikel)

- (1) Eher, R., Rettenberger, M., Schilling, F., & Pfäfflin, F. (2008a). Validität oder praktischer Nutzen? Rückfallvorhersagen mittels Static-99 und SORAG. Eine prospektive Rückfallstudie an 275 Sexualstraftätern. *Recht & Psychiatrie*, 26, 79-88.
- (2) Eher, R., Rettenberger, M., Schilling, F., & Pfäfflin, F. (2008b). Failure of Static-99 and SORAG to Predict Relevant Reoffense Categories in Relevant Sexual Offender Subtypes. A Prospective Study. *Sexual Offender Treatment*, 3, 1-14.
- (3) Eher, R., Rettenberger, M., & Matthes, A. (2009). Aktuarische Prognose bei Sexualstraftätern: Ergebnisse einer prospektiven Studie an 785 Tätern unter besonderer Berücksichtigung von relevanten Tätergruppen und Rückfallkategorie. *Monatsschrift für Kriminologie und Strafrechtsreform* 92, 18-27.
- (4) Rettenberger, M. & Eher, R. (2006a). Die deutsche Übersetzung des Static-99 zur aktuarischen Kriminalprognose verurteilter Sexualstraftäter. *Monatsschrift für Kriminologie und Strafrechtsreform*, 89, 352-365.
- (5) Rettenberger, M. & Eher, R. (2006b). Actuarial Assessment of Sex Offender Recidivism Risk: A Validation of the German Version of the Static-99. *Sexual Offender Treatment*, 1, 1-11.
- (6) Rettenberger, M. & Eher, R. (2007a). Predicting Reoffense in Sexual Offender Subtypes: A Prospective Validation Study of the German Version of the Sexual Offender Risk Appraisal Guide (SORAG). *Sexual Offender Treatment*, 2, 1-12.
- (7) Rettenberger, M. & Eher, R. (2007b). Aktuarische Kriminalprognosemethoden und Sexualdelinquenz: Die deutsche Version des SORAG. *Monatsschrift für Kriminologie und Strafrechtsreform*, 90, 484-497.
- (8) Rettenberger, M., Matthes, A., Boer, D. & Eher, R. (zur Publikation angenommen). Prospective Actuarial Risk Assessment: A Comparison of Five Risk Assessment Instruments in Different Sexual Offender Subtypes. *International Journal of Offender Therapy and Comparative Criminology*.

R & P

Recht und Psychiatrie

Reinhard Eher, Martin Rettenberger, Frank Schilling und Friedemann Pfäfflin

Validität oder praktischer Nutzen? Rückfallvorhersagen mittels Static-99 und SORAG. Eine prospektive Rückfallstudie an 275 Sexualstraftätern

Aktuarische Prognoseinstrumente spielen bei der Einschätzung des Rückfallrisikos insbesondere bei Sexualstraftätern eine immer größere Rolle. Sie werden häufig anderen Prognosemethoden als überlegen dargestellt. Hier werden die Daten von 275 zu einer Straftat verurteilten Sexualtätern präsentiert, die während der Haft ausführlich und standardisiert begutachtet worden waren. Begutachtungsdaten, insbesondere die Risikoeinschätzungen mittels Static-99 und des Sexual Offender Risk Appraisal Guide (SORAG), werden Rückfallzahlen nach durchschnittlich 3,6 Jahren gegenübergestellt.

Die Ergebnisse zeigen, dass die Instrumente nach herkömmlicher Definition bestimmte Rückfallkategorien sehr gut vorhersagen. Dennoch ist der unmittelbare Nutzen solcher Vorhersagen für den praktizierenden Forensiker kritisch zu diskutieren. Mit Static-99 und SORAG lässt sich nämlich der »Nicht-Rückfall« besser vorhersagen als der Rückfall. Selbst Täter mit höchsten Werten auf diesen Risikoinstrumenten werden innerhalb des 3,6 Jahres-Nachbeobachtungszeitraums in den relevanten Deliktgruppen in weit weniger als 50 % der Fälle wegen eines neuerlichen Sexual- oder Gewaltdelikts verurteilt. Das Spannungsfeld zwischen statistischen Gütekriterien (Validität) und praktischem Nutzen wird anhand der gewonnenen Daten kritisch diskutiert.

Schlüsselwörter: Static-99, SORAG, Prognose, Rückfälligkeit, Validität, Sexualstraftäter

Herausgeber: Redaktion Recht & Psychiatrie

Redaktion: Helmut Pollähne, Bremen (verantwortlich); Martin Zinkler, London (verantwortlich); Uwe Dönisch-Seidel, Düsseldorf; Heinfried Duncker, Moringen; Dirk Fabricius, Frankfurt; Birgit Hoffmann, Freiburg; Heinz Kammeier, Münster; Norbert Konrad, Berlin; Wolfgang Lesting, Oldenburg; Rolf Marschner, München; Sabine Nowara, Waltrop; Friedemann Pfäfflin, Ulm; Dorothea Rzepka, Bielefeld/Frankfurt a. M.; Norbert Schalast, Essen; Herbert Steinböck, Haar; Birgit Völlm, Manchester; Helga Wullweber, Berlin

Redaktionsanschrift: Marina Broll, Lange Straße 17, 44137 Dortmund; Tel.: 0231/1505460, Fax: 0231/1505461

E-Mail: rp@psychiatrie.de

Verlag: Psychiatrie-Verlag GmbH, Thomas-Mann-Straße 49a, 53111 Bonn
verlag@psychiatrie.de; www.psychiatrie.de/verlag



Sonderdruck

2008, 26. Jahrgang, 2. Vierteljahr, Seite 79 – 88

Recht und Psychiatrie is regularly indexed in:

Embase, Journal Citation Reports/Social Sciences Edition, Juris, Journal Citation Reports/Science Edition, KJB, PsycInfo, Science Citation Index Expanded (SciSearch)®, Social Sciences Citation Index (SSCI)®, Social Scisearch®

Reinhard Eher^{1,2}, Martin Rettenberger¹, Frank Schilling¹ und Friedemann Pfäfflin²

Validität oder praktischer Nutzen? Rückfallvorhersagen mittels Static-99 und SORAG. Eine prospektive Rückfallstudie an 275 Sexualstraftätern

Aktuarische Prognoseinstrumente spielen bei der Einschätzung des Rückfallrisikos insbesondere bei Sexualstraftätern eine immer größere Rolle. Sie werden häufig anderen Prognosemethoden als überlegen dargestellt. Hier werden die Daten von 275 zu einer Straftat verurteilten Sexualtätern präsentiert, die während der Haft ausführlich und standardisiert begutachtet worden waren. Begutachtungsdaten, insbesondere die Risikoeinschätzungen mittels Static-99 und des Sexual Offender Risk Appraisal Guide (SORAG), werden Rückfallzahlen nach durchschnittlich 3,6 Jahren gegenübergestellt.

Die Ergebnisse zeigen, dass die Instrumente nach herkömmlicher Definition bestimmte Rückfallkategorien sehr gut vorhersagen. Dennoch ist der unmittelbare Nutzen solcher Vorhersagen für den praktizierenden Forensiker kritisch zu diskutieren. Mit Static-99 und SORAG lässt sich nämlich der »Nicht-Rückfall« besser vorhersagen als der Rückfall. Selbst Täter mit höchsten Werten auf diesen Risikoinstrumenten werden innerhalb des 3,6 Jahres-Nachbeobachtungszeitraums in den relevanten Deliktgruppen in weit weniger als 50% der Fälle wegen eines neuerlichen Sexual- oder Gewaltdelikts verurteilt. Das Spannungsfeld zwischen statistischen Gütekriterien (Validität) und praktischem Nutzen wird anhand der gewonnenen Daten kritisch diskutiert.

Schlüsselwörter: Static-99, SORAG, Prognose, Rückfälligkeit, Validität, Sexualstraftäter

Validity or benefit? The prediction of relapses with Static-99 and SORAG. A prospective study with 275 sex offenders

When evaluating sex offenders' risk of re-offending, actuarial risk assessment is getting more and more important and is often seen as superior to other methods. This paper presents data from a prospective study of 275 incarcerated sex offenders who had been thoroughly evaluated clinically as well as with actuarial instruments including the Static-99 and the Sexual Offender Risk Appraisal Guide (SORAG) before being released from prison. Their relapse rates after an average time »at risk« of 3.6 years are correlated with their previous scores on the Static-99 and the SORAG.

The results show that both prognostic instruments predict certain types of relapse quite well. However, their usefulness for the courts and for forensic clinicians needs further discussion. Static-99 and the SORAG are better in predicting a »non-relapse« than a relapse. This is mainly due to the low base rates of certain types of relapses. Even of those individuals with high scores in the Static-99 and the SORAG far less than 50% relapsed during the 3.6 years follow-up period and were again convicted for a new sexual or violent crime. The authors critically review the conflict between statistic validity and practical usefulness of actuarial risk assessments.

Key Words: Static-99, SORAG, risk assessment, relapse, validity, sex offenders

Einleitung

Rückfallprognosen lassen sich nach DAHLE et al. (2007) grundsätzlich entweder individuumszentriert oder gruppenstatistisch (»aktuarisch«) formulieren. Die zuerst genannte Methode stellt die inhaltliche Aufklärung des Einzelfalls in den Vordergrund. Die zweite orientiert sich an Erfahrungen, die an größeren Stichproben rückfällig gewordener Täter gewonnen wurden, nämlich Erkenntnissen über Tat- und Tätermerkmale, Behandlungseffekten oder anderen, die Rückfallwahrscheinlichkeit beeinflussenden Faktoren, sowie der Bewertung von Kumulationseffekten derartiger Faktoren. Ein dritter Typus von Prognoseinstrumenten, und somit eine Weiterentwicklung der zweiten Gruppe, stellen solche dar, die neben unveränderbaren statischen auch dynamische, also potenziell veränderbare Variablen einbeziehen.

Sowohl der Static-99 (HANSON & THORNTON 2000) als auch der SORAG (Sexual Offender Risk Appraisal Guide) (QUINSEY et al. 2006) gehören zur Gruppe der standardisierten aktuarischen Prognoseinstrumente, die jeweils über die Verrechnung eines Gesamtsummenscores eine statistische Wahrscheinlichkeit für ein bestimmtes Rückfallereignis einer bestimmten Tätergruppe ausgeben. Ihr Vorteil liegt darin, dass sie relativ einfach zu handhaben sind. Der Static-99 ist mittlerweile das verbreitetste Rückfallprognoseinstrument für Sexualstraftäter, obwohl seine Vorhersageleistungen trotz oftmals hoher Validitätswerte nur mittelmäßig sind – worüber im Folgenden zu handeln sein wird –, und obwohl sein konzeptueller Hinter-

¹ Begutachtungs- und Evaluationsstelle für Gewalt- und Sexualstraftäter, Wien

² Sektion Forensische Psychotherapie, Universitätsklinik für Psychosomatische Medizin und Psychotherapie, Ulm

grund eher dürftig ist – worauf wir hier aber nicht näher eingehen (HELMUS & HANSON 2007).

Metaanalytisch (HANSON & MORTON-BOURGON 2007) wurde gezeigt, dass die Vorhersage eines sexuell motivierten Rückfalls bei Sexualstraftätern durch strukturierte aktuarische Prognoseinstrumente wesentlich zuverlässiger ist als durch klinische Verfahren (Cohen's d von 0,70 im Vergleich zu 0,40). Auch die Vorhersage eines gewalttätigen Rückfalls (inklusive sexuell motivierter gewalttätiger Straftaten) war mithilfe der aktuarischen, für diesen Zweck entwickelten Instrumente der klinischen Einschätzung weit überlegen (Cohen's d 0,85 gegenüber 0,31). Insbesondere in Studien mit vergleichbaren Zeiten in Freiheit konnten hohe Validitätsscores ermittelt werden (HARRIS & RICE 2003).

Static-99 und SORAG gehören nach HANSON und MORTON-BOURGON (2007) zu den empirisch validierten, aktuarischen Prognoseinstrumenten. Ein empirischer Nachweis der Vorhersagegüte derartiger Instrumente ist vor allem dann anzunehmen, wenn die Beobachtungszeit der einzelnen Probanden möglichst konstant und die Reliabilität der Instrumente gewährleistet ist und schließlich, wenn in den Stichproben wenig »missing data« sind (HARRIS & RICE 2003). Die Item-Selektion erfolgte entsprechend dem empirischen Wissen über Zusammenhänge zwischen bestimmten Items und Rückfälligkeit. Die Bewertung und Verrechnung der einzelnen Item-Werte sowie deren In-Bezug-Setzung zum Risiko sind standardisiert.

Derartige – aktuarische bzw. statistische – Prognoseverfahren sind allerdings nicht unumstritten (CRAIG et al. 2006). Ein wesentlicher Kritikpunkt ist, dass ein Summenscore für das Verständnis und Management potenzieller Rückfalltäter von beschränktem Wert ist (HART 1998). Außerdem wird kritisiert, dass aktuarische Instrumente nicht ohne Weiteres mit der gleichen Zuverlässigkeit auf unterschiedliche Subgruppen von Tätern angewandt werden können. So waren z. B. weder Static-99 noch SORAG in der Lage, den sexuell motivierten Rückfall in einer Gruppe von Vergewaltigern vorherzusagen (HARRIS et al. 2003 b). Häufig wurde auch beschrieben, dass sie sich besser zur Vorhersage der allgemeinen Rückfälligkeit eignen als zur Vorhersage spezifischer, z. B. sexueller, Rückfälle. CRAIG et al. (2003) zeigten in einer Übersichtsarbeit, dass zehn der zwölf untersuchten aktuarischen Rückfallprognoseinstrumente besser den Rückfall in allgemeine Kriminalität als den sexuellen Rückfall vorhersagten.

Relevant für die Aussagekraft aktuarischer Instrumente sind schließlich die Basisrückfallraten, die sich für Untergruppen von Sexualstraftätern unterscheiden. Dies gilt insbesondere für die Untergruppen der Vergewaltigter sowie der extra- und intrafamiliären Kindesmissbraucher (HANSON 2002). Basisrückfallraten nehmen mit Fortdauer des Beobachtungszeitraumes zu. Andererseits gilt, dass die Wahrscheinlichkeit, das Risiko zu überschätzen (»falsch Positive«), umso höher ist, je niedriger die Basisraten sind. Nach HOOD et al. (2002) würde der Static-99 bei einer Basisrate von 4 % in 49 von 50 Fällen das Risiko überschätzen.

Bisherige Ergebnisse

Validierungsstudien für den Static-99 (HARRIS et al. 2003 a) und den SORAG (QUINSEY et al. 1998) sprechen für die prädiktive Validität dieser Instrumente. Die statistischen Kennwerte werden mit ROC-Analysen (Receiver Operating Characteristics) ermittelt und in AUC-Werten (area under the curve) quantifiziert (zur Erklärung von ROC-Kurven bzw. AUC-Werten s. u. im Abschnitt Statistik sowie bei NEDOPIL 2005 und bei DAHLE et al. 2007). AUC-Werte größer als 0,71 sprechen für hohe, AUC-Werte zwischen 0,63 und 0,70 für moderate und signifikante AUC-Werte kleiner als 0,63 für schwache prädiktive Validität. Gemessen an den AUC-Werten ist die prädiktive Validität für die Vorhersage späterer sexueller (vgl. Tab. 1) und gewalttätiger Rückfälle (vgl. Tab. 2) in bisherigen Studien moderat bis hoch. Wie später anhand der eigenen Daten noch zu zeigen sein wird, ist sie aber unter klinischen Gesichtspunkten oft nur schwach. In vielen Studien wird unterschieden zwischen »contact-« bzw. »hands-on«-Delikten einerseits und »non-contact-« bzw. »hands-off«-Delikten andererseits. Mit Ersteren werden alle gewalttätigen (Sexual-)Delikte bezeichnet, mit Letzteren alle Delikte, bei denen es nicht zu körperlichen Übergriffen kommt (z. B. Exhibitionismus).

Der Static-99 erzielte in einer früheren retrospektiven Studie anhand von 1208 nachuntersuchten Sexualstraftätern aus England und Kanada für das Kriterium »sexuell motivierter Rückfall« (jede sexuell motivierte neuerliche Straftat ohne Unterscheidung zwischen »contact-« und »non-contact«-Delikt) einen AUC-Wert von 0,71, und für das Kriterium »gewalttätiger Rückfall« (jedes gewalttätige Delikt, also auch alle sexuellen »hands-on«-Delikte) einen AUC-Wert von 0,69 (HANSON & THORNTON 2000). Für die Gruppe der Kindesmissbraucher wurde in dieser Studie die Vorhersageleistung insgesamt geringfügig höher gesehen als für Vergewaltigter (AUC-Werte für sexuellen Rückfall: 0,72 gegenüber 0,71, AUC-Werte für gewalttätigen Rückfall: 0,71 gegenüber 0,69). Nach 15 Jahren Nachbeobachtungszeit waren allerdings selbst in der höchsten Risikogruppe des Static-99 (Summenscore des Static-99 von 6 oder 6+: insgesamt 12 % der Gesamtgruppe) nur knapp über 50 Prozent der Täter einschlägig rückfällig geworden. Die Autoren ziehen daraus den Schluss, dass der Static-99 immerhin in der Lage war, eine nicht unwesentliche Subgruppe von Sexualstraftätern mit einem über 50 %igen Rückfallrisiko zu identifizieren (vgl. Tab. 1 und 2).

Eine Forschergruppe um Barbaree verglich anhand von 215 entlassenen Straftätern, die im Durchschnitt 4,5 Jahre nachbeobachtet wurden, verschiedene Risikoprognoseinstrumente für Sexualstraftäter (BARBAREE et al. 2001). Sie fanden, dass der SORAG ähnlich gut wie der Static-99 den Rückfall in ein weiteres Sexualdelikt, aber auch den in ein schwerwiegendes Gewaltdelikt (»serious offense«, gewalttätige Delikte inkl. sexuell motivierte »hands-on«-Sexualdelikte) voraussagen konnte. Bei der Vorhersage des »schwerwiegenden« Delikts zeigte sich für den SORAG ein AUC-Wert von 0,73 und für den Static-99 ein AUC-Wert von 0,70. Hinsichtlich des Rückfallkriteriums »Sexualdelikt« (grundsätzlich wurden alle sexuell motivierten Rückfallereignisse erhoben, allerdings lagen nur »contact«-Delikte vor) erwiesen sich beide Instrumente als gleich gut (AUC = 0,70). Eine weitere Differenzierung der Population nach der Indexdelikt-Kategorie (Kindesmissbrauch oder Vergewaltigung) wurde nicht vorgenommen (vgl. Tab. 1 und 2).

Tab. 1: Validierungsstudien für Static-99 und SORAG für sexuellen Rückfall (AUC-Werte)

	Sexueller Rückfall					
	Gesamtgruppe	Static-99		Gesamtgruppe	SORAG	
		Kindesmissbraucher	Vergewaltiger		Kindesmissbraucher	Vergewaltiger
HANSON & THORNTON 2000+	0,71	0,72	0,71	–	–	–
BARBAREE et al. 2001***	0,70	–	–	0,70	–	–
NUNES et al., 2002+	0,70	–	–	0,65	–	–
BARTOSH et al., 2003+	0,64	0,65/0,74*	0,71	0,58	0,70/0,72*	0,71
HARRIS et al., 2003 b**	0,62	0,65	0,59	0,66	0,70	0,62
DUCRO & PHAM, 2006+	0,66	0,70	0,71	0,64	0,65	0,64

AUC-Wert > 0,71 = hoch, AUC-Wert > 0,63 = moderat, AUC-Wert < 0,63 = schwach; * Die Autoren geben getrennte Werte für extra- und intrafamiliäre Kindesmissbrauchstäter an; ** nur »contact offenses« als sexuelle Rückfallereignisse; + »contact« und »non-contact-offenses« als Rückfallereignisse

BARTOSH et al. (2003) untersuchten an 186 aus einer Haftstrafe entlassenen Sexualstraftätern Wiederinhaftierungsraten aufgrund von verschiedenen Delikten nach 5 bis 5,5 Jahren. Darüber hinaus ermittelten sie Validitätsindizes für vier Rückfallprognoseinstrumente, unter ihnen auch Static-99 und (eine etwas modifizierte Form des) SORAG. Sie unterschieden zwischen Vergewaltigern, intra- und extrafamiliären Kindesmissbrauchern und »non-contact«-Tätern. In der Gesamtgruppe konnte der Static-99 die Wiederinhaftierung aufgrund eines Sexualdelikts nur mit moderater Güte vorhersagen (AUC = 0,64), der SORAG gar nicht (AUC = 0,58). Das gewalttätige Delikt (Wiederinhaftierung aufgrund von sexueller oder nicht sexueller Gewalt) konnte mithilfe des Static-99 mit hoher Güte (AUC = 0,73) und mithilfe des SORAG mit gerade noch moderater Güte (AUC = 0,64) vorhergesagt werden. In der Gruppe der außerfamiliären Kindesmissbraucher zeigten Static-99 und SORAG bei der Vorhersage der Wiederinhaftierung aufgrund eines sexuell motivierten Rückfalls moderate Validität (AUC: 0,65 bzw. 0,69). Bei der Vorhersage der Inhaftierung aufgrund eines gewalttätigen Delikts (sexuell und nicht sexuell) fanden sich dagegen hohe AUC-Werte (AUC: 0,94 bzw. 0,93). In der Gruppe der intrafamiliären Täter war der Trend ähnlich, in der Gruppe der Vergewaltiger konnte keine der angeführten Rückfallkategorie signifikant vorhergesagt werden (vgl. Tab. 1 u. 2).

Auch NUNES et al. (2002) fanden beim Vergleich von Static-99 und SORAG keine Unterschiede in der prädiktiven Validität, differenzierten aber ebenfalls nicht zwischen den verschiedenen Untergruppen von Sexualstraftätern. Die AUC-Werte für ein sexuelles Rückfalldelikt (keine Unterscheidung zwischen »contact«- und »non-contact«-Delikt) wurden mit 0,70 (Static-99) bzw. 0,65 (SORAG) angegeben. Betreffend der Vorhersage eines gewalttätigen Delikts (inklusive Sexualdelikt) unterschieden sich die Leistungen der beiden Instrumente nicht (AUC = 0,69) (vgl. Tab. 1 und 2).

HARRIS et al. (2003 b) verglichen in einer retrospektiven Studie Static-99, SORAG und andere Risikoeinschätzungsinstrumente. Dazu diente eine Stichprobe von 396 männlichen Sexualstraftätern. Die Nachbeobachtungszeit lag durchschnittlich bei etwa fünf Jahren (»time of opportunity«). In der Gesamtgruppe zeigte der Static-99 für sexuell motivierte Rückfalligkeit (nur »contact«-Delikte, sexuelle Rückfalligkeit wurde

als Untergruppe der Gewaltdelikte erfasst) einen AUC-Wert von 0,62 (niedrig). Auch die differenzierten Werte für die Untergruppe der Vergewaltiger (AUC = 0,59) und Kindesmissbraucher (AUC = 0,65) waren niedrig. Gewalttätige Rückfalligkeit konnte sowohl in der Gesamtgruppe (AUC = 0,63) als auch in der Gruppe der Vergewaltiger (AUC = 0,58) und Kindesmissbraucher (AUC = 0,64) durch den Static-99 nur mit schwacher bzw. moderater Güte oder gar nicht vorhergesagt werden. Der SORAG schnitt in dieser Studie hinsichtlich beider Rückfallkriterien (sexuell und gewalttätig) besser ab. Bezüglich des sexuellen Rückfalls zeigte er in der Gesamtgruppe moderate Vorhersageleistungen (AUC = 0,66), bezüglich der Vorhersage eines Gewaltdelikts eine hohe Prognosegüte (AUC = 0,73). Sowohl für die Gruppe der Vergewaltiger als auch die der Kindesmissbraucher prognostizierte er Rückfallereignisse besser als der Static-99, wobei er den unspezifischen gewalttätigen Rückfall besser vorhersagte als den sexuellen (sexueller Rückfall: Vergewaltiger AUC = 0,62, Kindesmissbraucher AUC = 0,70, gewalttätiger Rückfall: Vergewaltiger AUC = 0,70, Kindesmissbraucher AUC = 0,72) (vgl. Tab. 1 und 2).

In einer belgischen Vergleichsstudie mit durchschnittlich 4,2 Jahren Nachbeobachtungszeitraum (DUCRO & PHAM 2006) fanden sich bezüglich des sexuell motivierten Rückfalls in der Gruppe der Kindesmissbraucher und Vergewaltiger bessere Werte für den Static-99 (AUC = 0,70 und AUC = 0,71) als für den SORAG (AUC = 0,65 und AUC = 0,64). Insbesondere in der Gruppe der Vergewaltiger sagte aber der SORAG das gewalttätige Rückfallereignis besser vorher als der Static-99 (AUC = 0,77 bzw. AUC = 0,65) (vgl. Tab. 1 und 2).

Bisherige Validierungsstudien aus deutschsprachigen Ländern zeigen für den Static-99 bei Untersuchungen von Sexualstraftäterstichproben Werte von AUC = 0,72 für den gewalttätigen Rückfall (STADTLAND et al. 2006) – diese Studie beschränkte sich auf das angeführte Rückfallkriterium –, sowie AUC = 0,74 für den sexuell motivierten und AUC = 0,76 für den gewalttätigen Rückfall (RETTENBERGER & EHER 2006 b). Auch die deutsche Übersetzung des SORAG ergab bislang hohe Validitätsindizes in Bezug auf das sexuelle und gewalttätige Rückfallereignis (RETTENBERGER & EHER 2007 a, b). Die AUC-Werte lagen für die relevanten Rückfallkategorien (mit Ausnahme des sexuell motivierten Rückfalls bei Vergewaltigern) durchgängig über 0,71.

Tab. 2: Validierungsstudien für Static-99 und SORAG für gewalttätigen Rückfall (AUC-Werte)

	Gewalttätiger Rückfall					
	Gesamtgruppe	Static-99		Gesamtgruppe	SORAG	
		Kindesmissbraucher	Vergewaltiger		Kindesmissbraucher	Vergewaltiger
HANSON & THORNTON 2000	0,69	0,71	0,69	–	–	–
BARBAREE et al. 2001	0,70	–	–	0,73	–	–
NUNES et al., 2002	0,69	–	–	0,69	–	–
BARTOSH et al., 2003	0,73	0,94/0,86*	0,53	0,64	0,93/0,82*	0,57
HARRIS et al., 2003 b**	0,63	0,64	0,58	0,73	0,72	0,70
DUCCRO & PHAM, 2006	0,68	0,72	0,65	0,72	0,70	0,77

AUC-Wert > 0,71 = hoch, AUC-Wert > 0,63 = moderat, AUC-Wert < 0,63 = schwach; * Die Autoren geben getrennte Werte für extra- und intrafamiliäre Kindesmissbrauchstäter an

Fragestellung

Wenngleich in Zusammenschau der bisherigen Daten die Prognosegüte nach statistischen Kennwerten in den meisten Fällen zumindest als moderat, in manchen Fällen sogar als hoch bezeichnet werden kann, stellt sich für den Rechtsanwender und Kliniker in der Praxis dennoch die Frage, mit welcher Verlässlichkeit er im Einzelfall die angeführten Prognoseverfahren anwenden bzw. interpretieren kann, bzw. was diese nun tatsächlich auszusagen imstande sind (siehe auch Ergebnisse über »Falsch Positive« bei HOOD et al. 2002). Vor dem Hintergrund eingeschränkter Ressourcen ist die Beantwortung dieser Frage von entscheidender Bedeutung, insbesondere wenn Verantwortliche über die Finanzierung und die konkrete Form des Risikomanagements im Strafvollzug und danach entscheiden müssen.

Um diese Frage beantworten zu können, war es zunächst erforderlich, validierte deutsche Versionen dieser Instrumente zur Verfügung zu stellen. Nachdem diese nunmehr vorliegen (RETTENBERGER & EHER, 2006 a, b, 2007 a, b), sollten die Validitätsindizes der Instrumente einer quantitativen Analyse des tatsächlichen Rückfallgeschehens gegenübergestellt werden.

Methode

Prognoseinstrumente Static-99 und SORAG

Der Static-99 ist eine Synthese aus zwei früheren Prognoseinstrumenten, dem Rapid Risk Assessment for Sex Offender Recidivism (RRASOR; HANSON 1997) und dem Structured Anchored Clinical Judgement – Minimum (SACJ – Min; GRUBIN 1998). Alle Merkmale wurden empirisch aus früheren Rückfallstudien extrahiert (HANSON & BUSSIÈRE 1998). Die Kurzform des SACJ und der RRASOR wurden schließlich in den zehn Items des Static-99 zusammengefasst. Eine aufwendige Revision wurde 2003 veröffentlicht (HARRIS et al. 2003 a). Der SORAG ist eine Modifikation des Violence Risk Appraisal Guide (VRAG; QUINSEY et al. 1998), entwickelt mit dem vorrangigen Ziel, gewalttätige (inklusive sexuell motivierte gewalttätige) Rückfälle entlassener Sexualstraftäter vorherzusagen. Der SORAG wurde an nur 178 entlassenen Sexualstraftäter validiert. Die 14 vorhersagestärksten Variablen wurden – in gewichteter Form – zum SORAG zusammengefasst.

Deutschsprachige Versionen des Static-99 und des SORAG wurden in den Jahren 2002 bis 2007 im Kontext von Begutachtungen an der Begutachtungs- und Evaluationsstelle für Gewalt- und Sexualstraftäter (BEST) (früher: »Zentrale Dokumentations- und Koordinationsstelle für Sexualstraftäter«) im Österreichischen Strafvollzug angewandt. Während des genannten Zeitraumes wurden dort 714 zu Freiheitsstrafen verurteilte Sexualstraftäter begutachtet. Dabei wurden 629 vollständige Static-99- und 614 vollständige SORAG-Datensätze gewonnen. Beide Instrumente wiesen in früheren Studien eine gute Interraterreliabilität und prädiktive Validität auf (RETTENBERGER & EHER 2006 a, b, 2007 a, b)³.

Stichprobe

Seit dem Jahr 2002 müssen in Österreich zu einer Freiheitsstrafe verurteilte Sexualstraftäter an die Begutachtungs- und Evaluationsstelle für Gewalt- und Sexualstraftäter gemeldet werden. In den Jahren 2002 bis zum Stichtag 17. September 2007 wurden etwa 60 % aller in Österreich in dieser Zeit in Strafhaf genommenen Sexualstraftäter ausführlich klinisch und kriminologisch begutachtet (EHER et al. 2006). Von den 714 begutachteten Sexualstraftätern lagen zum Stichtag von 275 Tätern Informationen über Rückfälle in Form von Strafregisterauskünften des Bundesministerium für Inneres der Republik Österreich vor. Der frühestmögliche Nachuntersuchungszeitpunkt für jeden Täter war 27 Monate nach Entlassung aus der Strafhaf. Die mittlere Zeit »at risk« betrug 3,6 Jahre (2,3–5,9 Jahre). Das Alter bei Entlassung betrug im Durchschnitt 40,9 Jahre (16,1–71,4). Von den 275 Tätern hatten 136 ein Indexdelikt an Minderjährigen und 132 ein Indexdelikt an Erwachsenen begangen. Bei sieben Tätern lag kein »hands-on«-Delikt vor. Die allgemeine Rückfallrate betrug für die Gesamtgruppe nach durchschnittlich 3,6 Jahren 30,2 %, 17,5 % waren im Nachbeobachtungszeitraum wieder inhaftiert worden, 6,2 % waren mit einem Sexualdelikt rückfällig geworden, allerdings handelte es sich hierbei nur in 3,3 % um ein »hands-on«-Delikt. Fast gleich viele Rückfälle (2,9 %) waren »hands-off«-Sexualdelikte. Gewalttätige (»violent«) Delikte – eine Deliktategorie, die sexuelle »hands-on«- und nicht sexu-

³ Informationen über deutschsprachige Handbücher und Schulungen sind über den Verein für forensische Forschung und Weiterbildung, www.vffw.org erhältlich.

ell motivierte Gewaltdelikte umfasst –, fanden sich in 14,6 % der Fälle, Haftstrafen infolge derartiger Delikte wurden in 8,36 % verhängt (vgl. Tab. 3).

Tab. 3: Rückfallraten der Gesamtstichprobe und jeweils getrennt nach Indexdelikt (nach 3,6 Jahren Zeit »at risk«)

	Gesamt (N = 275)	Indexdelikt an Minderjäh- rigen (n = 136)	Indexdelikt an Erwachse- nen (n = 132)
Rückfall allgemein	83 (30,2%)	30 (22,1%)	48 (36,4%)
Rückfall allgemein plus Haft	48 (17,5%)	18 (13,2%)	28 (21,2%)
Rückfall Gewalt (nicht sexuell motiviert)	40 (14,6%)	8 (5,9%)	31 (23,5%)
Rückfall Gewalt plus Haft (nicht sexuell motiviert)	24 (8,7%)	3 (2,2%)	20 (15,2%)
Rückfall sexuell	17 (6,2%)	12 (8,8%)	3 (2,3%)
Rückfall sexuell plus Haft	16 (5,8%)	12 (8,8%)	3 (2,3%)
Rückfall sexuell »hands on«	9 (3,3%)	5 (3,7%)	3 (2,3%)
Rückfall sexuell »hands on« plus Haft	9 (3,3%)	5 (3,7%)	3 (2,3%)
Rückfall sexuell »hands off«	8 (2,9%)	7 (5,2%)	0
Rückfall sexuell »hands off« plus Haft	7 (2,6%)	7 (5,2%)	0
Rückfall »violent« (sexuell »hands-on« + nicht sexuelle Gewalt)	40 (14,6%)	12 (8,8%)	26 (19,7%)
Rückfall »violent« plus Haft	23 (8,4%)	6 (4,4%)	16 (12,1%)

Anm. zu Tab. 3: Sieben Täter der Gesamtgruppe konnten keiner der beiden nachfolgenden Deliktategorien zugeordnet werden.

Statistik

Als Validitätsindizes bezüglich der prognostischen Güte der Instrumente wurden ROC-Analysen durchgeführt. Entsprechende AUC-Werte werden angegeben. AUC-Werte entsprechen sowohl konzeptuell als auch dem Wert nach der Effektstärke (HARRIS et al. 2003 b). Es handelt sich hierbei um ein spezielles Verfahren der logistischen Regression mit dem Vorteil, von Basisraten weitgehend unabhängig zu sein. Im Wesentlichen drückt der AUC-Wert die Wahrscheinlichkeit aus, dass eine zufällig ausgewählte rückfällige Person einen höheren Scorewert als eine zufällig ausgewählte nichtrückfällige Person aufweist. AUC-Werte über 0,71 gelten als hoch, solche über 0,63 als moderat, und signifikante Werte darunter werden als schwach bezeichnet (DAHLE et al. 2007).

Ergebnisse

Die einzelnen Ergebnisse in Form von AUC-Werten sind in Tab. 4 zusammengefasst. Es zeigte sich, dass beide Instrumente für fast alle Rückfallkategorien – im Hinblick auf die Gesamtpopulation – eine hohe prognostische Güte aufweisen. Die Werte lagen meist über jenen, die bislang in der Literatur beschrieben worden waren (vgl. Tab. 1 und 2). Lediglich in der Kategorie des sexuellen »hands-on«-Delikts fand sich für beide Instrumente eine nur moderate Prognosegüte, wobei die AUC-Werte für den SORAG von vornherein nicht signifikant waren, diejenigen für den Static-99 nach Alpha-Fehler-Korrektur ebenfalls nicht. Es fanden sich in keiner Rückfallkategorie signifikante Unterschiede zwischen der Vorhersageleistung des Static-99 und des SORAG (paarweiser Vergleich der AUC-Werte).

Für die Gruppe der Kindesmissbraucher finden sich für beide Instrumente insbesondere für den Rückfall mit einem Sexualdelikt sehr gute Vorhersagewerte (AUC > 0,8), der SORAG erreichte sogar hohe Werte für die Vorhersage des sexuellen »hands-on«-Delikts bei Kindesmissbrauchern. Auch die Deliktategorie der gewalttätigen (»violent«) Rückfälle konnte mittels beider Instrumente gut vorhergesagt werden. Nicht mit der erforderlichen Güte vorhergesagt werden konnte der Rückfall in ein Sexualdelikt in der Gruppe der Vergewaltigter. Tendenziell waren in der Gruppe der Kindesmissbraucher die Werte mit höherer Prognosegüte zu finden (vgl. Tab. 4, S. 84).

Die Tab. 5 bis 8 führen die tatsächlichen Rückfallereignisse je nach Risikokategorie des entsprechenden Instruments an. Angegeben sind die kriminologisch relevanten Rückfallkategorien (»Rückfall in Sexualdelikt«, »Rückfall in sexuelles hands-on-Delikt«, »Rückfall in »violent« Delikt«, »Erneute Inhaftierung wegen eines »violent« Delikts«).

Static-99-Risikokategorien und Rückfallereignisse bei Kindesmissbrauchern

In Tab. 5 sind die entsprechenden Werte des Static-99 für die Gruppe der Kindesmissbraucher dargestellt. In der höchsten Risikokategorie fanden sich immerhin 33 % Rückfälle mit einem weiteren Sexualdelikt, wohingegen in den ersten drei Risikokategorien prozentual deutlich weniger derartige Rückfälle vorkamen.

Mit einem sexuellen »hands-on«-Delikt wurden allerdings selbst in der höchsten Risikokategorie nur 14 % rückfällig. Rückfälle in die Kategorie »violent« nahmen in der Gruppe der Kindesmissbraucher mit Anstieg der Risikokategorie stetig zu, wobei auch hier selbst in der höchsten Risikokategorie »nur« 19 % wieder verurteilt bzw. »nur« etwa 10 % wieder inhaftiert worden waren.

Static-99-Risikokategorien und Rückfallereignisse bei Vergewaltigern

Rückfälle mit Sexualdelikten standen in der Gruppe der Vergewaltigter nicht in Zusammenhang mit der durch den Static-99 ausgewiesenen Risikokategorie. Hingegen zeigte sich bezüglich

Tab. 4: Validitätsindizes des Static-99 und des SORAG (AUC-Werte)

	AUC-Werte für die Gesamtgruppe		AUC-Werte für die Gruppe der Kindesmissbraucher		AUC-Werte für die Gruppe der Vergewaltiger	
	Static-99	SORAG	Static-99	SORAG	Static-99	SORAG
Rückfall allgemein	0,70***	0,77***	0,69**	0,78***	0,67**	0,71***
Rückfall allgemein plus Haft	0,73***	0,78***	0,75**	0,83***	0,69**	0,71**
Rückfall Gewalt	0,69**	0,75***	0,64	0,66	0,67**	0,71***
Rückfall Gewalt plus Haft	0,70**	0,78***	0,60	0,74	0,67*	0,72***
Rückfall sexuell	0,74***	0,71**	0,81***	0,82***	0,54	0,65
Rückfall sexuell plus Haft	0,72***	0,71*	0,81***	0,82***	0,54	0,65
Rückfall sexuell »hands on«	0,68*	0,69	0,74	0,80*	0,54	0,65
Rückfall sexuell »hands on« plus Haft	0,68*	0,69	0,74	0,80*	0,54	0,65
Rückfall »violent« (inkl. sex. »hands-on«)	0,71***	0,74***	0,73*	0,74**	0,65*	0,67**
Rückfall »violent« plus Haft	0,71**	0,76***	0,74	0,83*	0,66*	0,67*

AUC-Wert > 0,71 = hoch, AUC-Wert > 0,63 = moderat, AUC-Wert < 0,63 = schwach

Tab. 5: Rückfälle in der Gruppe der sexuellen Kindesmissbraucher (N = 133) nach Static-99-Kategorie in Prozent (von drei Tätern lag kein Static-99 vor)

Static-99-Wert	Rückfall sexuell	Rückfall sexuell »hands on«	Rückfall »violent« inkl. sex. »hands-on«	Rückfall »violent« plus Haft
static 0,1 (n=53)	0	0	1 (1,89%)	0
static 2,3 (n=43)	5 (11,63%)	2 (4,65%)	4 (9,30%)	3 (6,98%)
static 4,5 (n=16)	0	0	2 (12,5%)	0
static 6,6+ (n=21)	7 (33,33%)	3 (14,29%)	4 (19,05%)	2 (9,52%)

eines »violent« Delikts ein kontinuierlicher prozentualer Anstieg der Rückfallereignisse in Abhängigkeit zur Static-99-Risikokategorie. In der Gruppe der Vergewaltiger lag die Rückfallrate, die aufgrund eines gewalttätigen Delikts zu einer Haftstrafe führte, in der höchsten Static-99-Risikogruppe immerhin bei 23 % (vgl. Tab. 6).

Tab. 6: Rückfälle in der Gruppe der Vergewaltiger (N = 128) nach Static-99-Kategorie in Prozent (von vier Tätern lag kein Static-99 vor)

Static-99-Wert	Rückfall sexuell	Rückfall sexuell »hands on«	Rückfall »violent« inkl. sex. »hands-on«	Rückfall »violent« plus Haft
static 0,1 (n=12)	0	0	1 (8,33%)	0
static 2,3 (n=43)	1 (2,33%)	1 (2,33%)	5 (11,62%)	4 (9,3%)
static 4,5 (n=47)	2 (4,26%)	2 (4,26%)	12 (25,53%)	6 (12,77%)
static 6,6+ (n=26)	0	0	8 (30,77%)	6 (23,08%)

SORAG-Risikokategorien und Rückfallereignisse bei Kindesmissbrauchern

Der SORAG zeigte in der Gruppe der Kindesmissbraucher ähnliche Ergebnisse wie der Static-99. Bezüglich des sexuell motivierten Rückfalls zeigte sich ein kontinuierlicher Anstieg der Rückfallzahlen in Korrelation zur Risikogruppe. Die höchsten Risikogruppen des SORAG bestätigten immerhin eine 30–40%ige Rückfallwahrscheinlichkeit bezüglich eines Sexualdelikts. Dieser Zusammenhang war für die Gruppe der sexuellen »hands-on«-Delikte weniger ausgeprägt. Auch hier war selbst in der zweithöchsten Risikokategorie nur ein Täter von fünf tatsächlich rückfällig geworden. Kein einziger hatte die Kriterien der höchsten Risikogruppe erfüllt. Auch der Zusammenhang zwischen Risikokategorie und Rückfall in ein gewalttätiges Delikt verhielt sich ähnlich wie beim Static-99, das heißt, es fand sich eine stetige Zunahme der Rückfallereignisse in Korrelation zur Höhe der Risikokategorie. Dennoch waren auch hier in den höchsten Risikokategorien weit mehr Täter nicht rückfällig als umgekehrt.

SORAG-Risikokategorien und Rückfallereignisse bei Vergewaltigern

In der Gruppe der Vergewaltiger zeigten sich bezüglich des SORAG ähnliche Ergebnisse wie beim Static-99. Es fand sich kein Zusammenhang zwischen der Risikokategorie und dem tatsächlichen sexuell motivierten Rückfallereignis, hingegen zeigte sich ein klarer Zusammenhang zwischen der Risikokategorie und der gewalttätigen Tat. Rückfallereignisse hinsichtlich der Deliktategorie »violent« zeigten sich in den beiden höchsten Risikokategorien des SORAG in über 40 %. Zu erneuten Haftstrafen wurden in den genannten Risikokategorien fast 30 % verurteilt.

Tab. 7: Rückfälle in der Gruppe der sexuellen Kindesmissbraucher (N=130) nach SORAG-Kategorie in Prozent (von sechs Tätern lag kein SORAG vor)

	Rückfall sexuell	Rückfall sexuell »hands on«	Rückfall »violent« inkl. sex. »hands-on«	Rückfall »violent« plus Haft
SORAG 1 (n=22)	0	0	1 (4,55%)	0
SORAG 2 (n=33)	1 (3,03%)	0	1 (3,03%)	0
SORAG 3 (n=22)	1 (4,55%)	1 (4,55%)	0	0
SORAG 4 (n=15)	1 (6,67%)	1 (6,67%)	1 (6,67%)	1 (6,67%)
SORAG 5 (n=13)	2 (15,38%)	1 (7,69%)	2 (15,38%)	1 (7,69%)
SORAG 6 (n=11)	2 (18,18%)	0	1 (9,09%)	1 (9,09%)
SORAG 7 (n=9)	3 (33,33%)	1 (11,11%)	3 (33,33%)	1 (11,11%)
SORAG 8 (n=5)	2 (40%)	1 (20%)	2 (40%)	1 (20%)
SORAG 9 (n=0)	–	–	–	–

Tab. 8: Rückfälle in der Gruppe der Vergewaltiger (N=123) nach SORAG-Kategorie in Prozent (von neun Tätern lag kein SORAG vor)

	Rückfall sexuell	Rückfall sexuell »hands on«	Rückfall »violent« inkl. sex. »hands-on«	Rückfall »violent« plus Haft
SORAG 1 (n=9)	0	0	0	0
SORAG 2 (n=13)	0	0	0	0
SORAG 3 (n=20)	0	0	3 (15%)	2 (10%)
SORAG 4 (n=16)	1 (6,25%)	1 (6,25%)	4 (25%)	2 (12,50%)
SORAG 5 (n=15)	2 (13,33%)	1 (13,33%)	5 (33,33%)	3 (20,0%)
SORAG 6 (n=16)	0	0	1 (6,25%)	1 (6,25%)
SORAG 7 (n=16)	0	0	3 (18,75%)	3 (18,75%)
SORAG 8 (n=11)	0	0	5 (45,45%)	3 (27,27%)
SORAG 9 (n=7)	0	0	3 (42,86%)	2 (28,57%)

Diskussion

Das wichtigste Ergebnis der Untersuchung ist die deutliche Diskrepanz zwischen den hohen Validitätsindizes von Static-99 und SORAG einerseits, die per definitionem dafür sprechen, dass mit diesen Instrumenten Rückfälle gut vorhergesagt werden können, und deren beschränktem praktischen Nutzen andererseits, wenn sie doch »Rückfällige« nicht hinreichend sicher identifizieren können, was daran ablesbar ist, dass selbst in den höchsten Risikokategorien oftmals weit unter 50 % der früheren Straftäter erneut verurteilt werden.

Wenngleich sich die Validität – also die statistisch nachgewiesene Gültigkeit – aktuarischer Prognoseinstrumente bei der Vorhersage von Sexual- und Gewalttaten klarer denn je abzeichnet und mittlerweile auch mittels Metaanalysen als erwiesen gilt (HANSON et al. 2007), so ist dennoch der praktische Wert der Information durch diese Instrumente für den Anwender kritisch zu bewerten. Allein in der Konstruktionsstichprobe des Static-99 findet sich zum Beispiel selbst in der Hochrisikogruppe (6, 6+) ein nur knapp über 50%-iges Rückfallrisiko für ein weiteres Sexualdelikt – und dies auch erst nach 15 Jahren »at risk«. Diesbezüglich besser vorhersagbar scheinen auf den ersten Blick die in der Konstruktionsstichprobe des SORAG angegebenen Werte für den »gewalttätigen« (sexuelle »hands-on« und nicht sexuell motivierte Gewaltdelikte) Rückfall bei Sexualstraftätern – diese sprechen von 100 % Rückfälligkeit in der höchsten Risikokategorie nach sieben und zehn Jahren. Allerdings ist zu berücksichtigen, dass die Konstruktionsstichprobe des SORAG relativ klein war, und insbesondere, dass in der höchsten Risikokategorie gerade einmal 1 % der Gesamtstichprobe – also nur drei Täter – vertreten waren. In einer Replikationsstudie mit 258 Tätern (NUNES et al. 2002) fanden sich ebenso wie in unserer Studie in der höchsten Risikokategorie gar keine Täter. Selbst in der zweithöchsten Risikostufe des SORAG fanden sich in dieser Studie lediglich 0,8 % der Gesamtstichprobe (zwei

Täter), wovon einer (50 %) mit einem gewalttätigen Rückfall in Erscheinung trat. Die Zuverlässigkeit, mittels hoher und höchster Werte von aktuarischen Prognoseinstrumenten auch wirklich später Rückfällige identifizieren zu können, ist also nicht sehr hoch – ganz im Gegensatz zur angenommenen Fähigkeit angesichts bester Validitätsindizes.

Prognoseinstrumente produzieren je nach Festlegung des Cut-Offs bzw. Schwellenwertes, oberhalb dessen jemand als gefährlich eingestuft wird, viele »falsch positive« Einschätzungen. Das heißt, es kommt durch eine unkritische Anwendung derartiger Instrumente zu einer systematischen Überschätzung des wirklichen (nachweisbaren) Risikos. Nimmt man zum Beispiel beim Static-99 den Gesamtsummenwert von 5 als Cut-Off, so werden nach einer durchschnittlichen Beobachtungszeit von etwa acht Jahren aus dieser Gruppe (Gesamtsummenwert 5, 5+) 46 % mit einem (sexuell und nicht sexuell motivierten) Gewaltdelikt rückfällig, was gleichzeitig bedeutet, dass 54 % nicht rückfällig werden (STADTLAND et al. 2006). Auch in der Originalstichprobe (HARRIS et al. 2003 a) waren selbst Täter aus der höchsten Risikokategorie (Werte von 6, 6+) nach fünf Jahren in 61 %, nach zehn Jahren in 55 %, und nach 15 Jahren in 48 % eben *nicht* mit einem neuerlichen Sexualdelikt rückfällig geworden. In einer anderen Studie (NUNES et al. 2002) lag die Rate an »Falsch Positiven« nach etwa sieben Jahren Nachbeobachtung beim gleichem Cut-Off (6, 6+) für die gleiche Rückfalldeliktkategorie gar bei 70 %. Ähnliche Befunde liegen für den SORAG vor. Die gleiche Arbeitsgruppe berichtet über eine 50%ige Rückfallrate bezüglich Sexualdelikten für die zweithöchste Risikostufe des SORAG und gar nur eine 25 % Rückfallrate für die dritthöchste Risikostufe.

In der von uns durchgeführten Studie wurden – insbesondere für bestimmte Untergruppen und Rückfallkategorien – sehr zufriedenstellende Validitätsindizes nachgewiesen. Diese lagen – insbesondere was die Vorhersage von sexuellen Rückfäl-

len bei Kindesmissbrauchern betraf – teilweise deutlich höher als die bislang publizierten. Betrachtet man die Validitätsindizes der Gesamtstichprobe, so finden sich für fast alle Rückfallkategorien sehr gute Werte. Ausgerechnet allerdings für die Rückfallkategorie der »hands-on«-Sexualdelikte sind die Werte weitaus weniger beeindruckend. Was die Gesamtgruppe betrifft, war die Vorhersageleistung des Static-99 hinsichtlich dieses Rückfallkriteriums mit einem AUC-Wert von 0,68 nur moderat; mittels des SORAG konnte diese Rückfallkategorie überhaupt nicht signifikant vorhergesagt werden. In der Gruppe der Kindesmissbraucher allerdings zeigte der SORAG für die Vorhersage der »hands-on«-Sexualdelikte mit einer AUC = 0,79 vorerst einen zufriedenstellenden Wert, der allerdings nach Typ-I-Fehler Korrektur ebenfalls nicht signifikant war.

Die schlechte bis fehlende Validität der Instrumente für die Kategorie der »hands-on«-Sexualdelikte ist wohl einerseits auf die in unserer Studie besonders niedrige Rückfallrate von nur 3,3 % zurückzuführen. Andererseits muss man aber aus den Daten wohl den Schluss ziehen, dass die Treffsicherheit der Instrumente gerade bei dieser so relevanten Rückfallkategorie gering ist (zumal Deliktkategorien vorliegen, die zwar noch weniger Rückfälle aufweisen, aber dennoch bessere Indizes erreichen). Offenbar wird diese Rückfallkategorie trotz aller Relevanz gerne »vergessen«. In vielen Validierungsstudien über Prognoseinstrumente wird nicht entsprechend differenziert (vgl. Tab. 1). Dabei liegt auf der Hand, dass es sowohl aus viktimologischer als auch aus allgemein gesellschaftspolitischer Sicht einen bedeutenden Unterschied macht, ob der Täter unmittelbar Missbrauchshandlungen an Opfern vornimmt oder z. B. »nur« Kinderpornografie konsumiert. Analysiert man die Daten genauer, ist die Schlussfolgerung unausweichlich, dass diese Instrumente, insbesondere in relevanten Täter-Untergruppen, für die Vorhersage des sexuell motivierten »hands-on«-Deliktes nicht valide sind oder zumindest der Validitätsnachweis noch aussteht: Beide Studien, die die Rückfallkategorie der sexuellen »hands-on«-Delikte untersuchten (BARBAREE et al. 2001; HARRIS et al. 2003 b), kamen entweder zu nur moderaten AUCs in der Gesamtgruppe (0,70), wobei eine weitere Differenzierung in Untergruppen nicht vorgenommen wurde (BARBAREE et al. 2001), oder die Werte waren von vornherein für die Gesamtgruppe schlecht (0,62 für Static-99 und 0,66 für SORAG), und die weitere Differenzierung zeigte, dass sie für die Gruppe der Vergewaltiger gar nicht (0,59 für Static-99, und 0,62 für SORAG) und für die Gruppe der Kindesmissbraucher nur mäßig valide waren (0,65 für Static-99 und 0,70 für SORAG) (HARRIS et al. 2003 b).

Sowohl SORAG als auch Static-99 konnten in unserer Studie den sexuell motivierten Rückfall in der Subkategorie der Vergewaltiger nicht vorhersagen. Dies lag einerseits an der niedrigen Rückfallrate, andererseits finden sich auch in anderen Studien ähnliche Probleme beschrieben (HARRIS et al. 2003 b; DUCRO & PHAM 2006). In der Gruppe der Vergewaltiger konnte mittels dieser Instrumente grundsätzlich besser das Gewaltdelikt vorhergesagt werden, wobei hier der SORAG dem Static-99 offenbar überlegen war. Dies war insofern nicht verwunderlich, als der SORAG für die Vorhersage des gewalttätigen Rückfalls konzipiert worden war. Andererseits zeigte sich in der Gruppe der Vergewaltiger – bei aller Heterogenität – die gewalttätige Komponente als das im Vergleich zur sexuellen Komponente größere Problem (PRENTKY et al. 1995).

Die Kategorie des gewalttätigen Rückfalls (»violent«: sexuelle »hands-on«-Delikte und nicht sexuell motivierte Gewaltdelikte) konnte auch in unserer Studie von beiden Instrumenten in beiden Untergruppen mit zufriedenstellender (Vergewaltiger) bis hoher (Kindesmissbraucher) Güte vorhergesagt werden.

Analysiert man allerdings die tatsächlichen Rückfallereignisse in den verschiedenen Risikokategorien, so relativiert sich für den Anwender der Wert der Prognoseinstrumente nochmals: bezüglich der hochrelevanten Rückfallkategorien sexuell motiviertes »hands-on«-Delikt und gewalttätiges (»violent«) Delikt fanden sich in der höchsten Risikogruppe (6, 6+) des Static-99 in der Gruppe der Kindesmissbraucher nur 14 % bzw. 19 % tatsächliche Rückfälle (9,5 % Verurteilungen zu einer Haftstrafe), in der Gruppe der Vergewaltiger gar 0 % bzw. 31 % tatsächliche Rückfälle (23 % Verurteilungen zu einer Haftstrafe). In den beiden höchsten Risikokategorien des SORAG fanden sich in der Gruppe der Kindesmissbraucher für die gleichen Rückfallkategorien 20 % bzw. 40 % (20 % Verurteilungen zu einer Haftstrafe), in der Gruppe der Vergewaltiger 0 % bzw. 44 % Rückfälle (28 % Verurteilungen zu einer Haftstrafe).

Dies zeigt, dass selbst in der Gruppe der Täter mit höchstem Risiko nach etwa 3,6 Jahren Nachbetrachtung nur etwa jeder fünfte bis zehnte Täter mit einem sexuell motivierten »hands-on«-Delikt, und nur etwa jeder zweite bis zehnte Täter mit einem Gewaltdelikt wieder verurteilt worden war (je nach Indexdelikt und Prognoseinstrument). Die Vorhersage der Verhängung einer Haftstrafe aufgrund derartiger Rückfallereignisse war ebenso ungenau: Die höchste Risikokategorie identifizierte lediglich jeden vierten bis zehnten Rückfälligen diesbezüglich richtig.

Unter Annahme dieser Cut-Offs für die Vorhersage des Rückfalls finden sich fast für jede Delikt- und Rückfallkategorie – was die Absolutzahl betrifft – gleich viele »Falsch Negative« (also Täter, die bei Vorhersage einer günstigen Prognose dennoch rückfällig werden) wie »Richtig Positive« (deren Rückfall richtig vorhergesagt wird). Die Gesamtzahl der Täter also, denen ein niedriges Risiko vorhergesagt wird, die aber dennoch rückfällig werden, ist (wenngleich prozentual gering) ebenso groß wie die Gesamtzahl derer, die aus der Gruppe der als hoch Rückfallgefährdeten tatsächlich rückfällig werden. Wenngleich der Prozentsatz der »Falsch Negativen« aufgrund der höheren Gesamtzahl von als nicht rückfällig prognostizierten zwar deutlich geringer ist, so spielen gerade diese Täter insofern in der forensischen Praxis eine relevante Rolle, als man diesen ja gerade ein nicht erhöhtes Risiko vorhergesagt hat, obwohl sie dann dennoch rückfällig werden. Dies lässt beim Laien den Glauben und das Vertrauen in die Professionalität der Prognosegutachter schwinden.

Gesetze, die eine zivilrechtliche oder strafrechtliche Maßnahme für Sexualstraftäter aufgrund ihrer »Gefährlichkeit« ermöglichen, verlangen zumindest eine hohe Wahrscheinlichkeit des Eintritts derartiger Straftaten, wobei die Definition des genauen Wahrscheinlichkeitsgrades oftmals ebenso unklar bleibt wie die Definition des zu verhindernden Übels (ABRACEN & LOOMAN 2006). Analysiert man unsere Daten und auch die Ergebnisse der oben genannten Studien, so zeigt sich, dass man selbst in den mit höchstem Gefährlichkeitsgrad definierten Untergruppen oftmals weit von der 50%igen geforderten Wahrscheinlichkeit einer negativen Prognose entfernt ist.

Die Vorhersage des Eintritts von Rückfällen bzw. des Schweregrads dieser Delikte mittels standardisierter Prognosemethoden ist bislang noch in vielerlei Hinsicht mangelhaft. Bei Sexualstraftätern gibt es in den bisherigen Publikationen zwar regelmäßig die Deliktategorien »sexuell« und »gewalttätig«, diese sind allerdings für den Rechtsanwender und Kliniker zu ungenau. Wie durch unserer Studie bestätigt, sind sexuelle Rückfalldelikte nicht selten »hands-off«-Delikte, die offenbar dennoch meist mit Haftstrafen geahndet werden. Aus dem Blickwinkel des Schadens für die Gesellschaft betrachtet, handelt es sich allerdings dabei nicht selten um minder schwere Delikte. Auch in der Kategorie der »gewalttätigen Rückfälle« finden sich oftmals Tathandlungen, die zwar offenbar »gewalttätig« sind, von den Gerichten allerdings – jedenfalls in unserer Studie –, nicht mit Haftstrafen geahndet wurden, was darauf schließen lässt, dass sie nicht so sehr schwerwiegend gewesen sein können.

Es ist demnach festzustellen, dass die genannten Prognoseinstrumente in den bisherigen Studien auf ihre Vorhersagekraft im Hinblick auf die wohl relevantesten Rückfallkategorien – nämlich sexuell motivierte »hands-on«-Delikte und schwere Gewaltdelikte, die mit Haftstrafen geahndet werden – bislang nicht ausreichend validiert waren.

Des Weiteren ist auf die Gefahr hinzuweisen, die Prognostizierbarkeit der einschlägigen Rückfälle bei Sexualstraftätern durch aktuarische Prognoseinstrumente zu überschätzen. Das unkritische Propagieren von Validitätsindizes kann eine derartige Überschätzung vorantreiben. Validitätsindizes alleine sind keine ausreichende Beschreibung der tatsächlichen Brauchbarkeit einer bestimmten Methode, und schon gar nicht erlaubt der Rückgriff auf noch so gute Validitätsindizes verlässliche Aussagen über das zukünftige Verhalten einzelner Täter. Dass bisher ausreichende Validitätsnachweise für die relevantesten Rückfallkategorien (sexuelle »hands-on«-Delikte und Gewalttaten, die zu Haftstrafen führen) fehlen, gibt Anlass für zusätzliche Kritik. Nun darf man das Kind allerdings nicht mit dem Bade ausschütten und aus den vorgelegten Daten den Schluss ziehen, die Anwendung aktuarischer Instrumente zur Gefährlichkeitseinschätzung sei damit etwa vollkommen nutzlos oder gar obsolet. Im Gegenteil, sie ist zur Einschätzung des Basisrisikos unterschiedlicher Tätertypen und Rückfallkategorien von großem Nutzen. Denn darauf können und sollen die weiteren individuellen Beurteilungen und Vollzugsentscheidungen Bezug nehmen.

Literatur

- ABRACEN J, LOOMAN J (2006) Evaluation of civil commitment criteria in a high risk sample of sexual offenders. *Journal of Sexual Offender Civil Commitment: Science and the Law*, 1, 124–140
- BARBAREE HE, SETO MC, LANTON CM, PEACOCK EJ (2001) Evaluating the predictive accuracy of six risk assessment instruments for adult sex offenders. *Criminal Justice and Behavior*, 28, 490–521
- BARTOSH DL, GARBY T, LEWIS D, GARY S (2003) Differences in the predictive validity of actuarial risk assessments in relation to sex offender type. *International Journal of Offender Therapy and Comparative Criminology*, 47, 422–438
- CRAIG LA, BEECH A, BROWNE KD (2006) Cross-validation of the risk matrix 2000 sexual and violent scales. *Journal of Interpersonal Violence*, 21, 612–633
- CRAIG LA, BROWNE KD, STRINGER I (2003) Risk scales and actors predictive of sexual offenses recidivism. *Trauma, Violence & Abuse: A Review Journal*, 4, 45–68
- DAHLE KP, SCHNEIDER V, ZIETHEN F (2007) Standardisierte Instrumente zur Kriminalprognose. *Forensische Psychiatrie, Psychologie, Kriminologie*, 1, 15–26
- DUCRO C, PHAM T (2006) Evaluation of the SORAG and the Static-99 on Belgian sex offenders committed to a forensic facility. *Sex Abuse*, 18, 15–26
- EHHER R, FRÜHWALD S, FROTTIER P (2006) Die standardisierte Begutachtung von Sexualstraftätern im Österreichischen Strafvollzug – erste Ergebnisse und Ausblicke. *Neuropsychiatrie*, 20, 50–55
- GRUBIN D (1998) Sex offending against children: Understanding the risk. *Police Research Series, Paper 99*. London: Home Office
- HANSON RK (1997) The Development of a Brief Actuarial Risk Scale for Sexual Offense Recidivism. User Report 1997-04. Ottawa: Department of the Solicitor General of Canada
- HANSON R K (2002) Recidivism and age follow-up data from 4,673 sexual offenders. *Journal of Interpersonal Violence*, 17, 1046–1062
- HANSON RK, BUSSIÈRE MT (1998) Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, 66, 348–362
- HANSON RK, MORTON-BOURGON KE (2007) The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis. (Rep. No. Corrections User Report No 2007–01.). Ottawa: Public Safety and Emergency Preparedness Canada
- HANSON RK, THORNTON D (2000) Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior*, 24, 119–136
- HARRIS AJ, PHENIX A, HANSON RK, THORNTON D (2003 a) Static-99 Coding Rules Revised – 2003. Correctional Service of Canada.
- HARRIS GT, RICE ME (2003) Actuarial assessment of risk among sex offenders. *Annals of the New York Academy of Science*, 989, 198–210
- HARRIS GT, RICE ME, QUINSEY VL, LALUMIÈRE ML, BOER D, LANG C (2003 b). A multisite comparison of actuarial risk instruments for sex offenders. *Psychological Assessment*, 15, 413–425
- HART SD (1998) The role of psychopathy in assessing risk for violence: Conceptual and methodological issues. *Legal and Criminological Psychology*, 3, 121–137
- HELMUS LMD, HANSON RK (2007) Predictive validity of the Static-99 and Static-2002 for sex offenders on community supervision. *Sexual Offender Treatment*, 2, 1–14 (www.sexual-offender-treatment.org/60.0.html)
- HOOD R, SHUTE S, FEILZER M, WILCOX A (2002) Sex offenders emerging from long-term imprisonment: A study of their long-term reconviction rates and of parole board members' judgements of their risk. *British Journal of Criminology*, 42, 371–394
- NEDOPIL N (2005) Prognosen in der Forensischen Psychiatrie. Ein Handbuch für die Praxis. Lengerich: Pabst
- NUNES KL, FIRESTONE P, BRADFORD JM, GREENBERG DM, BROOM I (2002) A comparison of modified versions of the Static-99 and the Sex Offender Risk Appraisal Guide. *Sex Abuse*, 14, 253–269
- PRENTKY RA, KNIGHT RA, LEE AS, CERCE DD (1995) Predic-

Failure of Static-99 and SORAG to predict relevant reoffense categories in relevant sexual offender subtypes: A prospective study

Reinhard Eher^{1,2}, Martin Rettenberger¹, Frank Schilling¹ & Friedemann Pfäfflin²

¹Evaluation Centre for Violent and Sexual Offenders, Vienna, Austria; ²Forensic Psychotherapy, Ulm University, Ulm, Germany

[Sexual Offender Treatment, Volume 3 (2008), Issue 1]

Abstract

An Austrian sample of 275 sexual offenders was assessed clinically as well as with actuarial tools including the Static-99 and the Sexual Offender Risk Appraisal Guide (SORAG) before being released from prison. After an average time of 3.6 years at risk, their relapse rates were correlated with their previous scores on these two instruments.

The results demonstrate that the Static-99 and the SORAG, although predicting certain types of relapses quite well, are much better in predicting the non-relapse than the relapse: Less than 50% of the high scorers on both instruments relapsed during the follow-up time of 3.6 years at risk and were again convicted for a new sexual or violent offence. For practical purposes, e. g. for the clinician and the judge being in charge of deciding about the release of an individual offender, these overall results as well as more specific results for certain types of relapses with low base rates are discussed.

Key words: Static-99, SORAG, risk assessment, sex offenders, relapse, validity

Introduction

The prediction of relapses may be primarily based on (1) individual clinical assessments or on (2) group-statistics evaluating actuarial data drawn from random samples of re-offenders, or, as is usually the case, a combination of both (Dahle et al. 2007). Standardized actuarial risk assessment tools like the Static-99 (Hanson & Thornton 2000) and the SORAG (Sexual Offender Risk Appraisal Guide; Quinsey et al. 2006) reflect mostly static, unchangeable factors and do not take into account dynamic, potentially changing factors. Their advantage is, however, easy handling and producing a total score allowing calculating the risk of certain types of re-offenses. Presently, the Static-99 is the most widely used actuarial risk assessment tool for sex offenders. In spite of high validity, their predictions rates are, however, moderate (Helmus & Hanson 2007).

In their meta-analysis, Hanson & Morton-Bourgon (2007) demonstrated the superiority of actuarial risk assessments to clinical assessments in predicting sex offenders' relapses with a new sex offense (Cohen's d 0.85 vs. 0.40). The same holds true for the prediction of a violent relapse (including sexually motivated violent offenses) (Cohen's d 0.85 vs. 0.31).

The Static-99 and the SORAG are empirically validated actuarial prognostic instruments (Hanson & Morton-Bourgon 2007). Their items were selected based on statistically validated correlations between individual items and relapse. Scoring is standardized. High reliability, constant observation periods, and few missing data increase the validity of such instruments (Harris & Rice 2003).

There is, however, also some controversy about actuarial risk assessments (Craig et al. 2006). Understanding and managing an offender by operating with just one total score may be of limited value (Hart 1998). Various subgroups of offenders might not have the same risk rates. Neither the Static-99 nor the SORAG predicted the sexual motivated relapse in a group of rapists (Harris et al. 2003b). Craig et al. (2003) examined 12 actuarial prognostic instruments and found that 10 out of 12 were better in predicting general than sexual recidivism.

Base rates play an outstanding role, and they vary between subgroups of sexual offenders especially so between the subgroups of rapists and the extra- and intra-familial child abuser (Hanson 2002). Also, the longer the follow-up time, the higher are the base rates. On the other hand, the lower the base rates are, the higher is the risk to overestimate the relapse rates and to produce false positives. With a base rate of 4%, the Static-99 would overestimate the risk in 49 out of 50 cases (Hood et al. 2002).

Recent results

Validation studies of the Static-99 (Harris et al. 2003a) and the SORAG (Quinsey et al. 1998) speak for the predictive validity of these instruments. Using ROC-analysis (Receiver Operating Characteristics), the results are quantified in AUC-values (area under the curve) (Dahle et al. 2007, Nedopil 2005). AUC-values $> .71$ speak for a high predictive validity, AUC-values between $.63$ and $.70$ for a moderate predictive validity and AUC-values $< .63$ for a weak predictive validity. As measured by AUC-values, the predictive validity for the prediction of subsequent sexual (Table 1) and violent (Table 2) relapses was found to be moderate up to high. Referring to our own data the predictive validity, as judged from a clinical perspective, often is weak. Only the minority of studies distinguishes between contact/hands-on offenses and non-contact/hands-off offenses.

In a previous retrospective study based on 1208 follow-ups of sexual offenders from England and Canada the Static-99 reached an AUC of $.71$ for sexually motivated relapse (there was no differentiation between contact and non-contact offenses). For violent relapse (including sexual hands-on offenses) an AUC of $.69$ was found (Hanson & Thornton 2000). After 15 years follow-up even in the highest risk group of the Static-99 (score of 6 or 6+ = 12% of the sample) just about 50% of the offenders had relapsed. The authors, nevertheless, concluded that the Static-99 allowed the identification of a significant subgroup of sexual offenders with a relapse risk of more than 50% (Tables 1 and 2).

Barbaree et al. (2001) compared a number of prognostic instruments by evaluating 215 sexual offenders with a time at risk of 4.5 years. They concluded that the SORAG came close to the Static-99 in predicting a sexual relapse (AUC $.70$ for both instruments) as well as in predicting other serious offenses (violent offense including sexual hands-on offense) (SORAG: AUC $.73$; Static-99: AUC $.70$). The study did not differentiate between sexual child abuse and rape (Tables 1 and 2).

Bartosh et al. (2003) examined re-imprisonment rates among 186 sexual offenders with a follow-up time of up to 5.5 years. They determined the validity indices for four prognostic instruments including the Static-99 and a modified form of the SORAG. They distinguished between rapists, intra- and extra-familial child abusers and non-contact offenders. Taking the total sample, the Static-99 predicted re-imprisonment as a result of a sexual offense only moderately (AUC $.64$), the SORAG not at all (AUC $.58$). The predictive validity of the Static-99 for re-imprisonment due to a violent offense (sexual and non-sexual) was high (AUC $.73$); using the SORAG it was just moderate (AUC $.64$). In the group of the extra-familial child abusers the Static-99 (AUC $.65$) and the SORAG (AUC $.69$) showed moderate validity. In predicting re-imprisonment due to a violent offense (sexual

and non-sexual), high AUC-values were found (AUC .94 and .93). For intra-familial offenders findings were similar. For rapists there could not be shown any significant predictions (Tables 1 and 2).

Similarly, Nunes et al. (2002) did not find differences in the predictive validity between the Static-99 and the SORAG, but they did not distinguish between the various sexual offender subgroups. The AUC-values for a sexual relapse (no differentiation between contact and non-contact offenses) was reported with .70 (Static-99) and .65 (SORAG). As regards the prediction of a violent offenses (including a sexual offense) there was no difference between the two instruments (AUC .69) (Tables 1 and 2).

In a retrospective study Harris et al. (2003b) compared the Static-99, the SORAG and other risk assessment tools by using a sample of 396 male sexual offenders. The follow-up time at risk averaged approximately 5 years. In the total sample the Static-99 showed an AUC of .62 (low) for sexual relapse (only contact-offenses, sexual relapse was included as a subgroup of violent offenses). The values for the subgroups of rapists (AUC .59) and child abusers (AUC .65) also ranked low. Violent recidivism could be predicted only moderately or not at all (total sample AUC .63; subgroup of rapists AUC .58; subgroup of child abusers AUC .64). In terms of both relapse criterions (sexual and violent) the SORAG scored better than the Static-99. As regards the total sample, predictive validity for a sexual relapse was moderate (AUC .66), for a violent offense it was high (AUC .73). The SORAG was better than the Static-99 in predicting relapses in the subgroups of rapists and sexual child abusers, and in predicting non-specific violent relapses as compared to sexual relapses (sexual relapse: rapist AUC .62, child abusers AUC .70; violent relapse: rapists AUC .70, sexual child abusers AUC .72) (Tables 1 and 2).

In a Belgian comparative study with an average follow-up time of 4.2 years (Ducro & Pham 2006) better values were found for the Static-99 (AUC .70 and AUC .71) than for the SORAG (AUC .65 and AUC .64) regarding sexually motivated relapses in the groups of sexual child abusers and rapists. In the subgroup of rapists, the SORAG predicted the violent relapse better than the Static-99 (AUC .77 and AUC .65) (Tables 1 and 2).

Table 1: Validity studies for the Static-99 and the SORAG for sexual relapses (AUC-values)

	Sexual relapse					
	Static-99			SORAG		
	Total N	Sexual child abusers	Rapists	Total N	Sexual child abusers	Rapists
HANSON & THORNTON, 2000+	.71	.72	.71	-	-	-
BARBAREE et al., 2001**	.70	-	-	.70	-	-
NUNES et al., 2002+	.70	-	-	.65	-	-
BARTOSH et al., 2003+	.64	.65/.74*	.71	.58	.70/.72*	.71
HARRIS et al., 2003b**	.62	.65	.59	.66	.70	.62
DUCRO & PHAM, 2006+	.66	.70	.71	.64	.65	.64

AUC-value > .71 = high, AUC-value > .63 = moderate, AUC-value < .63 = weak

* The authors differentiate between intra- and extra-familial sexual child abuse

** hands-on offenses only

+ hands-on as well as hands-off

Table 2: Validity studies for the Static-99 and the SORAG for violent relapses (AUC-values)

	Violent relapse					
	Static-99			SORAG		
	Total N	Sexual child abusers	Rapists	Total N	Sexual child abusers	Rapists
HANSON & THORNTON, 2000	.69	.71	.69	-	-	-
BARBAREE et al., 2001	.70	-	-	.73	-	-
NUNES et al., 2002	.69	-	-	.69	-	-
BARTOSH et al., 2003	.73	.94/.86*	.53	.64	.93/.82*	.57
HARRIS et al., 2003b	.63	.64	.58	.73	.72	.70
DUCRO & PHAM, 2006	.68	.72	.65	.72	.70	.77

AUC-value > .71 = high, AUC-value > .63 = moderate, AUC-value < .63 = weak

* The authors differentiate between intra- and extra-familial sexual child abuse

There are also some validation studies of the Static-99 from German-speaking countries based on samples of sexual offenders. Stadtland et al. (2006) found an AUC of .72 for the violent relapse; Rettenberger & Eher (2006b) an AUC of .74 for the sexually motivated and an AUC of .76 for the violent relapse. The validation of the German translation of the SORAG also resulted in AUC-values of .71 or higher for the sexual and violent relapses (Rettenberger & Eher 2007a,b), except for the sexually motivated relapse of rapists.

Aims of the Study

Once validations of the German translations of the Static-99 and the SORAG had been established (Rettenberger & Eher 2006 a,b) one could proceed to the next step: the quantitative analysis of actual relapses. Even though the statistical analysis is very important for the scientist, and even though the statistical values of the Static-99 and the SORAG are moderate or even high, the judge and the clinician want to know how to apply them to the individual case and how well they predict relapses individually. Against the background of limited resources the answer to these questions is of vital importance for the allocation of resources within the risk management of prisons. One does not want to produce false positives (Hood et al. 2002), and one wants to apply therapeutic activities of the staff as effectively as possible.

Methods

The prognostic tools Static-99 and SORAG

The Static-99 is a synthesis of two former prognostic instruments, the Rapid Risk Assessment for Sex Offender Recidivism (RRASOR, Hanson 1997) and the Structured Anchored Clinical Judgement Minimum (SACJ Min; Grubin 1998). All criteria were extracted from earlier empirical relapse studies (Hanson & Bussiere 1998). The short form of the SACJ and the RRASOR were combined into the 10 items of the Static-99. A complex revision was published in 2003 (Harris et al. 2003a). The SORAG is a modification of the Violence Risk Appraisal Guide (VRAG; Quinsey et al. 1998) developed for the purpose to predict violent (including sexually motivated violent) relapses of released offenders. The SORAG had been validated on a small sample of only 178 released sexual offenders. 14 of the most predictive variables were included into the SORAG in a weighted manner.

In the years 2002–2007, German versions of Static-99 and the SORAG were used in the Austrian prison system in the context of expert opinions at the Evaluation Centre for Violent and Sexual Offenders (BEST) an institution of the Federal Department of Justice for the assessment of violent and sexual offenders. During this period, 714 convicted sexual offenders were assessed, 629 completed the Static-99 and 614 the SORAG. Good inter-rater reliability and predictive validity had been found for both instruments in earlier studies (Rettenberger & Eher 2006a,b, 2007a,b).

Sample

Since 2002, all Austrian sexual offenders with a prison sentence have to be reported to the BEST. From the beginning of 2002 until September 2007 about 60% of all sex offenders imprisoned in Austria were thoroughly assessed clinically and under criminological aspects (Eher et al. 2006). For 275 of the 714 sex offenders information was collected about relapses in terms of criminal record advice of the Federal Department of the Interior. The average follow-up time at risk was 3.6 years (2.3–5.9 years). The average age at the time of release was 40.9 years (16.1–71.4). 136 offenders showed an index offense committed on minors and 132 an index offense committed on adults. 7 offenders had not committed a hands-on offense. The overall relapse rate was 30.2%. 17.5% were re-incarcerated. 6.2% had relapsed with a sexual offense, but only 3.3% with a hands-on offense. New violent offenses including sexual hands-on offenses were found in 14.6% of the cases, incarceration as a result of suchlike offenses was imposed in 8.36% (see Table 3).

Table 3: Relapse rates of total N and differentiated according to the subgroups of offenses against minors and adults after 3.6 years at risk

	Total N (N=275)	Index offense with minors (n=136)	Index offense with adults (n=132)
Relapse in general	83 (30.2%)	30 (22.1%)	48 (36.4%)
Relapse in general and incarceration	48 (17.5%)	18 (13.2%)	28 (21.2%)
Relapse “violent” (not sexually motivated)	40 (14.6%)	8 (5.9%)	31 (23.5%)
Relapse “violent” (not sexually motivated) and incarceration	24 (8.7%)	3 (2.2%)	20 (15.2%)

Relapse "sexual"	17 (6.2%)	12 (8.8%)	3 (2.3%)
Relapse "sexual" and incarceration	16 (5.8)	12 (8.8%)	3 (2.3%)
Relapse "sexual" with hands-on	9 (3.3%)	5 (3.7%)	3 (2.3%)
Relapse "sexual" with hands-on and incarceration	9 (3.3%)	5 (3.7%)	3 (2.3%)
Relapse "sexual" with hands-off	8 (2.9%)	7 (5.2%)	0
Relapse "sexual" with hands-off and incarceration	7 (2.6%)	7 (5.2%)	0
Relapse „violent“ ("sexual" with hands-on + not sexually motivated violence)	40 (14.6%)	12 (8.8%)	26 (19.7%)
Relapse „violent“ and incarceration	23 (8.4%)	6 (4.4%)	16 (12.1%)

7 of the total of N = 275 could not be assigned to one of the subgroups

Statistics

ROC-analyses were performed as validity indices regarding the prognostic value of the instruments. ROC-analysis is a procedure of logistic regression with the advantage of being largely independent of base rates. AUC-values correspond both, conceptually as well as numerically, with effect sizes (Harris et al. 2003b). Basically, the AUC-value indicates the probability to which a re-offender shows a higher score as compared to someone not re-offending. AUC-values above .71 are considered as high, values above .63 as moderate, and those below .63 as low (Dahle et al. 2007).

Results

The individual results in terms of AUC-values are summarized in Table 4. Taking the sample as a whole, both instruments demonstrated a high prognostic validity for almost all relapse categories. Values were mostly above those previously described in the literature (see Tables 1 and 2). Only in the category of sexual hands-on offenses both instruments showed only a moderate prognostic validity: for the SORAG, AUC-values were not significant from the very beginning, and neither were they significant for the Static-99 (after correction of type-I-error). In none of the relapse categories significant differences were found between the Static-99 and the SORAG (paired comparison of AUC-values).

In the group of sexual child abusers high prediction values were found, especially so as regards sexual relapses (AUC > 0.8). The SORAG even reached high values in predicting sexual hands-on offenses of sexual child abusers. Both instruments proved to be good in predicting violent relapses. Not satisfactorily predictable was the sexual relapse for rapists. Basically higher prognostic values were found in the group of child abusers (see Table 4).

Table 4: Validity indices of the Static-99 and the SORAG (AUC-values)

	AUC-values for the total N = 275	AUC-values for the subgroup of	AUC-values for the subgroup of
--	----------------------------------	--------------------------------	--------------------------------

			child abusers		rapists	
	Static-99	SORAG	Static-99	SORAG	Static-99	SORAG
Relapse in general	.70***	.77***	.69**	.78***	.67**	.71***
Relapse in general plus incarceration	.73***	.78***	.75**	.83***	.69**	.71**
Relapse violent	.69**	.75***	.64	.66	.67**	.71***
Relapse violent plus incarceration	.70**	.78***	.60	.74	.67*	.72***
Relapse sexual	.74***	.71**	.81***	.82***	.54	.65
Relapse sexual plus incarceration	.72***	.71*	.81***	.82***	.54	.65
Relapse sexual with hands-on	.68*	.69	.74	.80*	.54	.65
Relapse sexual with hands-on plus incarceration	.68*	.69	.74	.80*	.54	.65
Relapse violent (incl. sexual with hands-on)	.71***	.74***	.73*	.74**	.65*	.67**
Relapse violent plus incarceration	.71**	.76***	.74	.83*	.66*	.67*

AUC-value > .71 = high, AUC-value >.63 = moderate, AUC-value > .63 = weak

Tables 5 – 8 list the relapses according to the risk categories of the instruments. Indicated are the criminologically relevant relapse categories (relapse with sex offence , relapse with sexual hands-on offense , relapse with violent offense , re-imprisonment because of a violent offense).

Static-99 risk categories and relapse incidents in sexual child abusers

Table 5 shows the respective Static-99 values for the group of child abusers. The highest risk category shows 33% relapses with new sexual offenses whereas less relapses were found in the first three risk categories. Only 14% in the highest risk group relapsed with a sexual hands-on offense. Relapses in the category violent increased steadily with Static-99 risk categories, whereas even in the highest risk group only 19% were re-convicted and only about 10% were re-incarcerated.

Table 5: Relapses in the subgroup of sexual child abusers (N = 133) according to the Static-99-categories in absolute and relative numbers (missing data of 3 individuals)

Static-99-value	Relapse sexual	Relapse sexual, hands-on	Relapse violent (incl. sexual, hands-on)	Relapse violent and incarceration
static 0,1 (n=53)	0	0	1 (1.89%)	0

static 2,3 (n=43)	5 (11.63%)	2 (4.65%)	4 (9.30%)	3 (6.98%)
static 4,5 (n=16)	0	0	2 (12.5%)	0
static 6,6+ (n=21)	7 (33.33%)	3 (14.29%)	4 (19.05%)	2 (9.52%)

Static-99-risk categories and relapse incidents in rapists

There was no correlation between the risk category according to the Static-99 and sexual relapses. However, in terms of a violent offenses, a continuous rise in relapse incidents parallel to higher Static-99-risk categories was revealed. In the highest Static-99 risk group the relapse rate leading to incarceration was 23% (see Table 6).

Table 6: Relapses in the subgroup of rapists (N = 128) according to the Static-99-categories in absolute and relative numbers (missing data of 4 individuals)

Static-99-value	Relapse sexual	Relapse sexual, hands-on	Relapse violent (incl. sexual, hands-on)	Relapse violent and incarceration
static 0,1 (n=12)	0	0	1 (8.33%)	0
static 2,3 (n=43)	1 (2.33%)	1 (2.33%)	5 (11.62%)	4 (9.3%)
static 4,5 (n=47)	2 (4.26%)	2 (4.26%)	12 (25.53%)	6 (12.77%)
static 6,6+ (n=26)	0	0	8 (30.77%)	6 (23.08%)

SORAG-risk categories and relapse incidents in sexual child abusers

The SORAG revealed similar results as the Static-99 in the group of child abusers (see Table 7), i.e. an increase in the number of relapses in correlation to the risk group. The highest risk group in the SORAG confirmed after all a probability of 30-40% regarding a sexual relapse. This correlation was less distinct in the group of sexual hands-on offenses. In the second highest risk category only one out of five offenders actually re-offended. No single one met the criteria of the highest risk group. The correlation between risk category and relapse with a violent offense was again similar to the findings with the Static-99, i.e. a steady relapse incident increase in correlation to the risk category. In the highest risk category, however, far more offenders did not re-offend than vice versa.

Table 7: Relapses in the subgroup of child abusers (N = 130) according to SORAG-categories in absolute and relative numbers (missing data of 6 individuals)

	Relapse sexual	Relapse sexual, hands-on	Relapse violent (incl. sexual, hands-on)	Relapse violent and incarceration
SORAG 1 (n=22)	0	0	1 (4.55%)	0
SORAG 2 (n=33)	1 (3.03%)	0	1 (3.03%)	0
SORAG 3 (n=22)	1 (4.55%)	1 (4.55%)	0	0
SORAG 4 (n=15)	1 (6.67%)	1 (6.67%)	1 (6.67%)	1 (6.67%)
SORAG 5 (n=13)	2 (15,38%)	1 (7.69%)	2 (15.38%)	1 (7.69%)
SORAG 6 (n=11)	2 (18,18%)	0	1 (9.09%)	1 (9.09%)
SORAG 7 (n=9)	3 (33,33%)	1 (11.11%)	3 (33.33%)	1 (11.11%)
SORAG 8 (n=5)	2 (40%)	1 (20%)	2 (40%)	1 (20%)
SORAG 9 (n=0)	-	-	-	-

SORAG-risk categories and relapse incidents in rapists

In the group of rapists the SORAG showed similar results as the Static-99. No correlation was found between the risk category and the actual sexually motivated relapse, however a clear correlation was shown between risk category and violent relapse: more than 40% in the two highest risk categories of the SORAG relapsed with a violent offence, and in both categories nearly 30% of the offenders were again incarcerated (see Table 8).

Table 8: Relapses in the subgroup of rapists (N = 123) according to SORAG-category in absolute and relative numbers (missing data of 9 individuals)

	Relapse sexual	Relapse sexual, hands-on	Relapse violent (incl. sexual, hands-on)	Relapse violent and incarceration
SORAG 1 (n=9)	0	0	0	0
SORAG 2 (n=13)	0	0	0	0
SORAG 3 (n=20)	0	0	3 (15%)	2 (10%)
SORAG 4 (n=16)	1 (6.25%)	1 (6.25%)	4 (25%)	2 (12.50%)
SORAG 5 (n=15)	2 (13.33%)	1 (13.33%)	5 (33.33%)	3 (20.0%)
SORAG 6 (n=16)	0	0	1 (6.25%)	1 (6.25%)
SORAG 7 (n=16)	0	0	3 (18.75%)	3 (18.75%)
SORAG 8 (n=11)	0	0	5 (45.45%)	3 (27.27%)
SORAG 9 (n=7)	0	0	3 (42.86%)	2 (28.57%)

Discussion

The most important conclusion is the obvious discrepancy between the high validity indices of the Static-99 and the SORAG on the one hand which, by definition, suggest that relapses can be predicted very well with these instruments, and, on the other hand, their limited practical benefit when it comes to identifying recidivists with high enough certainty. This can be concluded from the fact that in the highest risk category often far below 50% of the offenders finally were re-incarcerated.

Although the (statistically verified) validity of actuarial prognostic instruments has been demonstrated by meta-analysis (Hanson et al. 2007), their practical value has to be discussed when the clinician or the judge have to decide in individual cases. In the primary sample of the Static-99 even in the high risk group (6,6+) a relapse risk for another sexual offense of only just above 50% was found and this only after 15 years at risk. Taking the primary sample of the SORAG for comparison, the results of this instrument seem to allow better predictions at first glance, at least for violent relapses (sexual hands-on and not sexually motivated violent offenses) of sexual offenders: in the highest risk group 100% relapsed after 7–10 years. It has, however, to be taken into account that the primary sample of the SORAG was very small, and only just 1%–3 offenders was represented in its highest risk category. In a replication study including 258 offenders (Nunes et al. 2002) as well as in our study not a single offender was found in the highest risk category. Even in the second highest risk category of the SORAG only 8%–2 offenders were found, and one (50%) of them relapsed with a violent offence (Nunes et al. 2002). Taking this into account, one has to conclude that even high and highest validity scores of such instruments do not predict who will actually re-offend and who will not.

Depending on the definition of cut-offs for dangerousness, prognostic instruments produce many false positives. Uncritical use of these instruments causes a systematic overestimation of the real (verifiable) risk. For example, taking 5 as the cut-off in the Static-99 after an average follow-up time of about 8 years at risk, 46% will re-offend in the 5, 5+ group with a sexual and/or non-sexual violent offense, but 54% will not (Stadtland et al. 2006). In the original sample (Harris et al. 2003a) 61% out of the highest risk group (6,6+) had not relapsed after 5 years; after 10 years 55%, and after 15 years 48% had not relapsed with a new sexual offense (Harris et al. 2003a). In the study of Nunes et al. (2002) the rate of false positives for the same cut-off (6,6+) and the same relapse category was 70% after approximately 7 years at risk. Comparably differentiated data do not exist for the SORAG, but Nunes et al. (2002) reported a 50% relapse rate for sexual offenses in the second highest risk category and only a 25% relapse rate in the third highest risk category.

In our study satisfactory validity indices were found, especially so for certain subgroups and relapse categories. Some of them were higher than previously published ones, e.g. for sexual relapses of child abusers. The values in the category hands-on sexual offenses were, however, much less impressive. The predictive performance of the Static-99 for this category was only moderate with an AUC of .68; and with the SORAG it could not be predicted at all in a significant manner. The poor or missing validity of the instruments for hands-on sexual offenses is due to the exceptional low relapse rate of 3,3% in our sample. On the other hand, these data lead to the conclusion that the accuracy of the instruments in this specifically relevant relapse category is of low level even more so as there are other offense categories with much less relapses but still better indices. Apparently this relapse category is likely to be forgotten in spite of all its relevance. In a number of validation studies the respective differentiation is missing (see Table 1). Yet it is obvious from a victimological as well as from a more general sociopolitical point of view that it makes a difference if the offender directly commits victim abuse with physical harm or for example only consumes child pornography. If one analyses the data accurately one must conclude that these

instruments are not valid for relevant subgroups, e.g. sexual-motivated hands-on relapses, or at least, that the proof of validity is still missing. The two studies, that examined the relapse category of sexual hands-on offenses either found only moderate AUCs in the total group (.70) a further differentiation in subgroups was not performed (Barbaree et al. 2001), or the values proved to be poor in the first place for the total group (.62 for the Static-99 and .66 for the SORAG) and further differentiation showed no validity for rapists (.59 for the Static-99, and .62 for the SORAG) and moderate validity for child abusers (.65 for the Static-99 and .70 for the SORAG) (Harris et al. 2003b).

In our study the SORAG as well as the Static-99 could not predict the sexually motivated relapse in the subgroup of rapists. This was certainly also due to the low relapse rate. On the other hand, Harris et al. (2003b) as well as Ducro & Pham (2006) had similar problems in predicting rapists sexually motivated relapses, whereas non sexual violent offenses could be predicted better with the SORAG as compared to the Static-99. This is no surprise since the SORAG had been designed for the prediction of violent relapse. On the other hand, in the group of rapists with all its heterogeneity the violent element is the more important problem in comparison to the sexual element (Prentky et al. 1995).

In our study the violent relapse (sexual hands-on offense and not sexually motivated violent offense) could be predicted with satisfactory (rapists) and high (child abusers) performance using both instruments in both subgroups.

If analyzing the actual relapse incidents in the different risk categories the value of the prognostic instruments is being put into perspective once more for the user: Regarding the highly relevant relapse categories sexually motivated hands-on offense and violent offense we found in the highest risk group (6,6+) of the Static-99 in the sample of child abusers only 14% and 19%, respectively, actual relapses (and 9.5% re-incarcerations). In the group of rapists there were only 0%, respectively 31% actual relapses (and 23% re-incarcerations). The respective numbers for the two highest risk categories of the SORAG in the group of child abusers are 20% and 40% (with 20% re-incarcerations), in the group of rapists 0% and 44%, respectively (and 28% re-incarcerations).

In other words: At the average follow-up time of 3.6 years at risk, the highest risk category correctly identified only every 5th to every 10th recidivist with a sexually motivated hands-on relapse and only every 2nd to every 10th violent recidivist (with the range depending on risk category and prognostic instrument). The prediction of a new sentence for imprisonment due to such relapse incidents was just as imprecise: the highest risk category identified only every 4th to every 10th recidivist correctly.

When relying on these cut-offs for relapse prediction, the absolute numbers of true positives and false negatives are the same. Although the percentage of false negatives is lower due to the absolute number of those who are predicted not to re-offend, it is exactly those false negatives that play an outstanding role in the public discussion, nourishing doubts as to the qualification and the professionalism of the forensic prognostic expert.

Laws enforcing civil commitment or prison sentences for sexual offenders due to their dangerousness demand at least a high probability as regards prospective offences, although in most of these laws neither high probability nor the evil that should be prevented are properly defined and quantified (Abracen & Looman 2006). Analyzing our data and the results of the studies quoted it is obvious that even in those subgroups with the highest risk levels we are often far from being able to establish at least a probability of 50% for a negative prognostic outcome.

The prediction of relapses and their severity with standardized prognostic instruments is still

insufficient in many respects. For sexual offenders there are the two major categories sexual and violent but these are much too imprecise for the legal practitioner or the clinician. As shown in our study as well as in other studies, sexual relapses frequently are hands-off offenses. Nevertheless, the legal reaction is mostly a prison sentence. Under the perspective of the damage to society caused by such offenses, many of them are minor offenses rarely justifying severe sanctions. On the other hand, the category violent in our study included relapses not resulting in a prison sentence, which indicates that the courts did not evaluate them as severe enough to justify such a sanction.

In conclusion we think that the Static-99 and the SORAG are not yet sufficiently validated as regards their predictive power for the most important relapse categories, i.e. sexually motivated hands-on offenses and those violent offenses resulting in re-incarcerations. Although both tools have high validity indices, their practical usefulness is restricted and they certainly do not allow the prediction of the behavior of individuals. This, however, has never been the claim, and one should not throw out the baby with the bathwater. They are most useful in establishing base rates of recidivism for various categories of offenders and offences on which the clinician and the legal practitioner may build their individual assessments and decisions.


References

1. Abracen, J., & Looman, J. (2006). Evaluation of civil commitment criteria in a high risk sample of sexual offenders. *Journal of Sexual Offender Civil Commitment: Science and the Law*, 1, 124-140.
2. Barbaree, H.E., Seto, M.C., Lanton, C.M., & Peacock, E.J. (2001). Evaluating the predictive accuracy of six risk assessment instruments for adult sex offenders. *Criminal Justice and Behavior*, 28, 490-521.
3. Bartosh, D.L., Garby, T., Lewis, D., & Gary, S. (2003). Differences in the predictive validity of actuarial risk assessments in relation to sex offender type. *International Journal of Offender Therapy and Comparative Criminology*, 47, 422-438.
4. Craig, L.A., Beech, A., & Browne, K.D. (2006). Cross-validation of the risk matrix 2000 sexual and violent scales. *Journal of Interpersonal Violence*, 21, 612-633.
5. Craig, L.A., Browne, K.D., & Stringer, I. (2003). Risk scales and actors predictive of sexual offenses recidivism. *Tauma, Violence & Abuse: A Review Journal*, 4, 45-68.
6. Dahle, K.P., Schneider, V., & Ziethen, F. (2007). Standardisierte Instrumente zur Kriminalprognose. *Forensische Psychiatrie, Psychologie, Kriminologie*, 1, 15-26.
7. Ducro, C., & Pham, T. (2006). Evaluation of the SORAG and the Static-99 on Belgian sex offenders committed to a forensic facility. *Sex Abuse*, 18, 15-26.
8. Eher, R., Fruehwald, S., & Frottier, P. (2006). Die standardisierte Begutachtung von Sexualstraftätern im Österreichischen Strafvollzug - erste Ergebnisse und Ausblicke. *Neuropsychiatrie*, 20, 50-55.
9. Grubin, D. (1998). Sex offending against children: Understanding the risk. *Police Research Series, Paper 99*. London: Home Office.
10. Hanson, R.K. (1997). The Development of a Brief Actuarial Risk Scale for Sexual Offense Recidivism. User Report 1997-04. Ottawa: Department of the Solicitor General of Canada.
11. Hanson, R.K. (2002). Recidivism and age follow-up data from 4,673 sexual offenders. *Journal of Interpersonal Violence*, 17, 1046-1062.
12. Hanson, R.K., & Bussiere, M.T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, 66, 348-362.
13. Hanson, R.K., & Morton-Bourgon, K.E. (2007). The accuracy of recidivism risk assessments for sexual offenders : A meta-analysis. (Rep. No. Corrections User Report No 2007-01.). Ottawa: Public Safety and Emergency Preparedness Canada.

14. Hanson, R.K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior*, 24, 119-136.
15. Harris, A.J., Phenix, A., Hanson, R.K., & Thornton, D. (2003a). *Static-99 Coding Rules Revised - 2003*. Correctional Service of Canada.
16. Harris, G.T., & Rice, M.E. (2003). Actuarial assessment of risk among sex offenders. *Annals of the New York Academy of Science*, 989, 198-210.
17. Harris, G.T., Rice, M.E., Quinsey, V.L., Lalumiere, M.L., Boer, D., & Lang, C. (2003b). A multisite comparison of actuarial risk instruments for sex offenders. *Psychological Assessment*, 15, 413-425.
18. Hart, S.D. (1998). The role of psychopathy in assessing risk for violence: Conceptual and methodological issues. *Legal and Criminological Psychology*, 3, 121-137.
19. Helmus, L.M.D., & Hanson, R.K. (2007). Predictive validity of the Static-99 and Static-2002 for sex offenders on community supervision. *Sexual Offender Treatment*, 2, 1-14 (www.sexual-offender-treatment.org/60.0.html)
20. Hood, R., Shute, S., Feilzer, M., & Wilcox, A. (2002). Sex offenders emerging from long-term imprisonment: A study of their long-term reconviction rates and of parole board members' judgements of their risk. *British Journal of Criminology*, 42, 371-394.
21. Nedopil, N. (2005). *Prognosen in der Forensischen Psychiatrie. Ein Handbuch für die Praxis*. Lengerich: Pabst.
22. Nunes, K.L., Firestone, P., Bradford, J.M., Greenberg, D.M., & Broom, I. (2002). A comparison of modified versions of the Static-99 and the Sex Offender Risk Appraisal Guide. *Sex Abuse*, 14, 253-269.
23. Prentky, R.A., Knight, R.A., Lee, A.S., Cerce, D.D. (1995). Predictive validity of lifestyle impulsivity for rapists. *Criminal Justice and Behavior*, 22, 106-128.
24. Quinsey, V.L., Harris, G.T., Rice, M.E., & Cormier, C. (1998). *Violent Offenders - Appraising and Managing Risk*. (1st ed.) Washington DC: American Psychological Association.
25. Quinsey, V.L., Harris, A.J., Rice, M.E., & Cormier, C. (2006). *Violent Offenders: Appraising and Managing Risk*. (2nd ed.) Washington DC: American Psychological Association.
26. Rettenberger, M., & Eher, R. (2006a). Actuarial assessment of sex offender recidivism risk: A validation of the German version of the Static-99. *Sexual Offender Treatment*, 1[3], 1-11 (www.sexual-offender-treatment.org/51.0.html).
27. Rettenberger, M., & Eher, R. (2006b). Die deutsche Übersetzung und Adaptierung des Static-99 zur aktuarischen Kriminalprognose verurteilter Sexualstraftäter. *Monatsschrift für Kriminologie und Strafrechtsreform*, 89, 352-365.
28. Rettenberger, M., & Eher, R. (2007a). Aktuarische Kriminalprognosemethoden und Sexualdelinquenz: Die deutsche Version des SORAG. *Monatsschrift für Kriminologie und Strafrechtsreform*, 90, 484-497.
29. Rettenberger, M., & Eher, R. (2007b). Predicting reoffence in sexual offender subtypes: A prospective validation study of the German version of the Sexual Offender Risk Appraisal Guide (SORAG). *Sexual Offender Treatment*, 2, 1-12 (www.sexual-offender-treatment.org/60.1.html).
30. Stadtland, C., Hollweg, M., Kleindienst, N., Dietl, J., Reich, U., & Nedopil, N. (2006). Rückfallprognosen bei Sexualstraftätern - Vergleich der prädiktiven Validität von Prognoseinstrumenten. *Nervenarzt*, 77, 587-595.

Note: A similar version of this paper was first published in German in the *Journal Recht & Psychiatrie*, 26: 79-88, 2008. The authors appreciate the permission of Psychiatrie Verlag, Bonn for republishing it in English

Author address

Reinhard Eher, MD, Associate Professor
Evaluation Centre for Violent and Sexual Offenders
Gerichtsgasse 6
A-1210 Wien
Austria
 reinhard.eher@justiz.gv.at

Aktuarische Prognose bei Sexualstraftätern

Ergebnisse einer prospektiven Studie mit 785 Tätern unter besonderer Berücksichtigung von relevanten Tätergruppen und Rückfallkategorie

von Reinhard Eher, Martin Rettenberger und Anna Matthes

Zusammenfassung

Statistisch aktuarische Prognosemethoden gelten heute bei der Vorhersage von Rückfallwahrscheinlichkeiten als unverzichtbar und stellen einen festen Bestandteil von Prognosegutachten dar. Unabhängig von der Tatsache ihrer grundsätzlichen und immanenten Beschränktheit, künftiges Verhalten aufgrund von bereits früher gezeigtem vorherzusagen, muss eine Differenzierung ihrer Anwendbarkeit auch vom Vorliegen relevanter Untergruppen bei Sexualstraftätern gefordert werden. Beispielhaft anhand des Static-99 – einem statistisch-aktuarischen Prognoseinstrument zur Vorhersage der sexuell motivierten Rückfälligkeit mit allgemein akzeptierter guter Vorhersageleistung – wird gezeigt, dass die Prognosegüte je nach Tätergruppe und je nach vorhergesagter Rückfallkategorie stark variiert. Der kritische Umgang mit dergleichen Instrumenten durch den gut ausgebildeten Forensiker muss einmal mehr gefordert werden.

Schlüsselwörter: Sexualstraftäter, Rückfall, Basisrate, Prognoseinstrumente, Static-99

1. Einleitung

Sexualstraftäter geben unabhängig von der tatsächlichen Gefahr, die von ihnen ausgeht, nicht selten in besonderer Weise Anlass für emotional geführte Diskussionen. Hingegen belegen empirische Studien oftmals niedrige Rückfallraten (*Hanson & Bussière* 1998). Durch adäquate Behandlungsstrategien kann die Rückfallgefahr außerdem reduziert werden (*Loesel & Schmucker* 2005; *Hanson et al.* 2002). So beschrieben *Hanson & Bussière* (1998) eine mittlere Rückfallrate mit Sexualdelikten bei entlassenen Sexualstraftätern von 13,4 % nach vier bis fünf Jahren, wobei sie in dieser Metaanalyse 61 Studien mit insgesamt über 23.000 Tätern einschlossen. *Hood* und Kollegen untersuchten im Auftrag des Home Office die Rückfallraten entlassener langzeithaftierter Sexualstraftäter (*Hood, Shute, Feilzer & Wilcox* 2002). Nach vier Jahren waren 4,3 %, nach sechs Jahren 8,5 % aufgrund eines Sexualdeliktes wieder inhaftiert. Rückfalldaten aus dem US-Bundesstaat Washington anhand einer Kohorte von über 4.000 verurteilten Sexualstraftätern zeigten eine Rückfallrate von 2,7 % mit einem neuerlichen Sexualdelikt (*Barnoski* 2005). Eingeschlossen waren allerdings nicht nur aus einer Strafhafte entlassene Täter, sondern auch solche, die unter so genannter »Community Supervision« standen. Eine Studie des US Department of Justice aus dem Jahr 2003 führt u.a. die Rückfallraten von insgesamt 9.691 entlassenen Sexualstraftätern aus 15 US-Bundesstaaten im Jahr 1994 an (*Langan, Schmitt & Durose* 2003). 5,3 % waren nach einem Beobachtungszeitraum von drei Jahren wiederum aufgrund eines Sexualdeliktes in Haft. Bemerkenswert war, dass diese Täter im Durchschnitt bereits nach einem Drittel ihrer Haftstrafe entlassen worden waren (nach durchschnittlich dreieinhalb Jahren von ursprünglich verurteilten acht Jahren). 4.295 dieser Täter waren aufgrund von sexuellem Kindesmissbrauch in Haft. Auch diese Gruppe wurde nach durchschnittlich 43 % der Dauer ihrer Haftstrafe entlassen. Die Rückfallrate lag nach drei Jahren bei 3,3 % Wiederverurteilung aufgrund eines erneuten Kindesmissbrauchsdeliktes.

Niedrige Wiederverurteilungsbasisraten – wengleich natürlich gewollt und auch aktiv durch Therapie und Betreuung angestrebt – erschweren aus methodischer Sicht allerdings die Vorhersage des Rückfalls. Nichtsdestotrotz sind aber gerade die Vorhersagen dieser selte-

nen Ereignisse von großer Bedeutung. Sie sind notwendig bei Entscheidungen über die Verhängung und die Aufhebung einer freiheitsentziehenden Maßregel, sie sind wichtig bei Ausgangsentscheidungen, und sie spielen eine Rolle bei allgemeinen Fragen nach Gefährlichkeit und Wiederholungsgefahr. Gefährlichkeits- und Rückfallprognosen allerdings haben natürliche Grenzen, die oftmals zu wenig beachtet werden oder gar nicht bekannt sind. Zum einen zeigt sich, dass die den Prognosen zugrunde liegenden Diagnosen oftmals nur mit weit weniger Genauigkeit gestellt werden können als angenommen (*Packard & Levenson* 2006). Aber auch die Prognosegüte selbst zeigt sich oftmals deutlich schlechter als angesichts der im Hinblick auf den Schweregrad der Konsequenz zu fordernden Treffsicherheit (*Abracen & Looman* 2006). So konnte eben in dieser Studie selbst für die höchste messbare – im Sinne von besonders rückfallgefährdete – Risikokategorie letztendlich eine nicht annähernd 50 %ige Rückfallrate festgestellt werden. Auch von anderer Seite (*Hood et al.* 2002) wird angemerkt, dass 9 von 10 als gefährlich und »hoch risikoreich« eingeschätzten Sexualtätern innerhalb eines Vier-Jahres-Zeitraums nicht mit einem weiteren Sexualdelikt rückfällig wurden.

Andererseits gibt es Untersuchungen über richterliche Entscheidungen und deren implizite Definition von »Gefährlichkeit«, wonach Richter bereits bei einer 26 %igen Wahrscheinlichkeit einer neuerlichen Begehung eines Gewaltdeliktes ihre Zustimmung für eine – jedenfalls kurzfristige – zivilrechtliche freiheitsentziehende Maßnahme gaben (*Monahan & Silver* 2003).

Wenngleich die kriminologische Prognoseleistung allgemein und die Prognosegüte von bestimmten Rückfallvorhersageinstrumenten, insbesondere für gewalttätige und sexuell motivierte Delikte während der letzten Jahre aufgrund der verstärkten Forschung zugenommen hat (*Hanson & Morton-Bourgon* 2007), und wenngleich entsprechende Instrumente nunmehr auch im deutschsprachigen Raum aufgrund ihrer akzeptablen Gütekriterien berechnete Anwendung finden (*Rettenberger & Eher* 2007; *Rettenberger & Eher* 2006; *Stadtland et al.* 2006; *Eher, Rettenberger, Schilling & Pfafflin* 2008), so sind die immanenten Schwächen und Grenzen des Prognostizierens vielfach zu wenig bekannt. Zu wenig bekannt ist zum Beispiel das Dilemma der empirisch-aktuarischen Prognose, dass aufgrund der – aus statistischer Sicht – Seltenheit des Rückfallgeschehens eine weitere Verbesserung der Treffsicherheit nicht selten mit dem Phänomen der Produktion hoher »Falsch-Positiver«-Raten einhergeht (was bedeutet, dass die Wahrscheinlichkeit, auch solche Täter als »gefährlich« einzuschätzen, die in Zukunft keine gravierenden Straftaten begehen, größer wird). Die Abbildung konkreter Wahrscheinlichkeiten lässt Entscheidungsträger im Zweifel zunächst eher das Negative annehmen. Die Gefahr, einen der relevantesten dynamischen Rückfallprädiktoren zu vernachlässigen – nämlich die professionelle Behandlung – ist dadurch zumindest gegeben (*Bonta & Andrews* 2008).

Im vorliegenden Artikel wird das Verhältnis zwischen Prognosestellung mittels aktuarischer Verfahren und der tatsächlichen Legalbewährung anhand einer prospektiven Studie mit 785 entlassenen Sexualstraftätern dargestellt. Insbesondere werden die Täter auch in kriminologisch relevante Untergruppen unterteilt und auch in diesen Gruppen das jeweilige Verhältnis zwischen ursprünglicher Prognose und tatsächlicher Legalbewährung dargestellt. Ziel dieser Darstellung ist es, einen realistischen Zugang zu den Möglichkeiten der aktuarisch-statistischen Prognose bei Sexualstraftätern zu bekommen und ungerechtfertigte Erwartungen an die empirische Kriminalprognose zu korrigieren.

2. Methode

Mit dem 1.1.2002 besteht von Seiten der österreichischen Justizanstalten eine Meldepflicht an die Begutachtungs- und Evaluationsstelle für Gewalt- und Sexualstraftäter (BEST) für alle mit Rechtskraft zu einer Gefängnisstrafe verurteilten Sexualstraftäter (mit Ausnahme von Tätern

mit ausschließlich Delikten aus dem Bereich Menschenhandel und Prostitution). Dies führte dazu, dass mit dem 1.1.2002 alle zu diesem Zeitpunkt Inhaftierten und in der Folge jeder neu in Straftat genommene Täter gemeldet wurden. An der Begutachtungs- und Evaluationsstelle kommt es nach einem standardisierten Screening-Verfahren zur Einberufung der problematischsten Tätergruppe zur intensiven klinisch-forensischen Begutachtung. Bei allen gemeldeten Tätern werden Basisdaten erhoben, z.B. über Deliktart, -schwere und Opfertyp. Ebenso wird mittels des Static-99 (Rettenberger & Eher 2006), des SORAG (Quinsey, Harris, Rice & Cormier 2006; Rettenberger & Eher 2007) und des SSPI (Seto, Harris, Rice & Barbaree 2004) eine standardisierte und strukturierte Einschätzung des Basisrückfallrisikos vorgenommen.

Mit dem Stichtag 17.9.2007 waren insgesamt 1216 Täter gemeldet, wobei Straftäter, die im Zuge der Straftat abgeschoben wurden oder während der Haft verstorben waren, in der Folge nicht berücksichtigt werden. Letztendlich waren 785 Täter zum Stichtag bereits seit mindestens zwei Jahren entlassen. Diese wurden mit Hilfe von Strafregisterauszügen (Stichtag 17.9.2007) des Bundesministeriums für Inneres auf neuerliche Einträge hin überprüft. Rückfallereignisse wurden dann vermerkt, wenn es wiederum zu neuerlichen Eintragungen von rechtskräftigen Verurteilungen gekommen war. Als »Rückfallzeitpunkt« (besser: »Wiederverurteilungszeitpunkt«) wurde das Datum der Rechtskraft des Urteils bestimmt.

Ebenso wurde einerseits eine Unterteilung in verschiedene – forensisch relevante – Deliktkategorien vorgenommen (Hanson, Morton & Harris 2003; Greenberg, Bradford, Firestone & Curry 2000), andererseits auch die Anzahl der Vorstrafen getrennt nach relevanten Deliktkategorien erhoben. Alle Täter wurden jeweils nach ihrem Indexdelikt den Gruppen der »innerfamiliären Kindesmissbraucher«, der »außerfamiliären Kindesmissbraucher« und »Vergewaltiger« oder »anderen« (solche, die nicht in diese drei Kategorien passten) zugeordnet. Die entsprechenden Differenzierungskriterien waren einerseits der verurteilte Straftatbestand (Delikte an Opfern unter/über 14 Jahre), und andererseits das Verhältnis des Opfers zum Täter. Bei allen Tätern wurde standardmäßig ein kriminologisches Screening-Instrument (Static-99) für Sexualstraftäter angewandt, das in der Lage ist, differenzierte Basisraten für die jeweilige Risikokategorie auszugeben (Harris, Phenix, Hanson & Thornton 2003). Dieses Instrument wurde von unserer Arbeitsgruppe ins Deutsche übersetzt. In einer Reihe von Validierungsstudien konnten sehr gute prädiktive Validitätsmaße nachgewiesen werden (Rettenberger & Eher 2006).

3. Ergebnisse

3.1 Rückfallraten der Gesamtstichprobe und relevanter Untergruppen

Die Rückfallraten der Gesamtstichprobe (und die relevanter Untergruppen, wie außerfamiliäre Kindesmissbraucher, innerfamiliäre Kindesmissbraucher und Vergewaltiger) durchschnittlich vier Jahre nach der Entlassung sind in *Tabelle 1* dargestellt. Die allgemeine Wiederverurteilungsraten der Gesamtstichprobe lag nach vier Jahren »on risk« bei 27,5 %. 3,8 % waren mit einem neuerlichen Sexualdelikt rückfällig geworden, wobei die Rückfallrate bei »hands-on«-Sexualdelikten bei 2,3 % lag. Verurteilungsraten aufgrund von sexuell oder nicht sexuell motivierten Gewalttaten (»violent«) lagen in der Gesamtgruppe bei 12,7 %, aufgrund dieser Deliktkategorie (»violent«) wurden 8,3 % zu einer Haftstrafe verurteilt.

Die Gruppe der außerfamiliären Kindesmissbraucher wurde allgemein mit 26,3 % wieder straffällig und lag damit knapp unter dem Durchschnitt der Gesamtpopulation. Diese Gruppe wies mit einem Wert von 8,0 % die höchste Rückfallrate bei sexuell motivierten Straftaten auf. »Hands-on«-Delikte wurden allerdings nur in 4,0 % der Fälle verurteilt. Gewalttätige (»violent«) Straftaten wurden in 9,4 % der Fälle verurteilt und führten in knapp 6,3 % der Fälle zu Haftstrafen.

Wesentlich geringer war die Wiederverurteilungsrate (in allen Rückfallkategorien) bei den innerfamiliären Missbrauchern. Sie wurden in 14,7 % der Fälle allgemein, und nur in 1,5 % der Fälle sexuell motiviert rückfällig. Sexuelle »hands-on«-Delikte wurden in etwa 0,7 % der Fälle wiederverurteilt, Gewalttaten in 4,4 %, die in knapp 1,5 % zu Haftstrafen führten.

Die Gruppe der Vergewaltiger wurde im Nachbeobachtungszeitraum in 30,9 % der Fälle wiederverurteilt. Die Rate an sexuell motivierten Rückfallereignissen lag allerdings nur bei 1,6 % (alles »hands-on«-Sexualdelikte). Gewalttätige (»violent«) Delikte wurden in knapp 16,7 % der Fälle wiederverurteilt, in 11,0 % wurden sie mit Haftstrafen geahndet.

Die kriminelle Vorbelastung war in den Gruppen der Vergewaltiger und außerfamiliären Kindesmissbraucher erwartungsgemäß vergleichbar und lag deutlich über derjenigen der innerfamiliären Missbraucher. Entsprechend lag der Durchschnittswert des Static-99 in den Gruppen der außerfamiliären Kindesmissbraucher bzw. Vergewaltiger bei 3,0 bzw. 3,16, hingegen in der Gruppe der innerfamiliären Kindesmissbraucher mit 0,88 deutlich niedriger (vgl. *Tabelle 1*).

Tabelle 1 Rückfälle, Validitätsindizes des Static-99 und kriminelle Vorbelastung

	Gesamt- gruppe (N = 785)	außer- familiäre Kindesmiss- braucher (N = 224)	inner- familiäre Kindesmiss- braucher (N = 136)	Vergewal- tiger (N = 372)
Rückfälle				
Rückfall allgemein	27,52 (0,68***)	26,34 (0,63**)	14,71 (0,58*)	30,92 (0,67***)
Rückfall sexuell	3,82 (0,75***)	8,04 (0,72***)	1,47 (0,89)	1,61 (0,65)
Rückfall sexuell »hands-on« ¹	2,29 (0,70**)	4,02 (0,62)	0,74 (0,84)	1,61 (0,65)
Rückfall »violent« ²	12,74 (0,70***)	9,38 (0,67*)	4,41 (0,77*)	16,67 (0,67***)
Rückfall »violent« – Haft ²	8,28 (0,74***)	6,25 (0,62)	1,47 (0,94*)	11,02 (0,71***)
kriminelle Vorbelastung				
Vordelikte allgemein gesamt ³	3,09	3,27	1,23	3,16
Vordelikte sexuell ³	0,30	0,49	0,06	0,15
Static-99 ³	2,69	3,00	0,88	3,16

1 »Hands-on«-Delikte sind Sexualdelikte mit körperlichem Übergriff, im Gegensatz zu »hands-off«-Delikten wie z.B. Kinderpornographie, Exhibitionismus.

2 Rückfallkategorie »violent«: Hier sind sexuell motivierte »hands-on«-Delikte und nicht sexuell motivierte Gewaltdelikte zusammengefasst.

3 Werte sind Mittelwerte.

* p < 0,05 ** p < 0,01 *** p < 0,001

Dargestellt sind Gesamtrückfallraten in Prozent nach durchschnittlich 1450 Tagen »at risk« (3,97 Jahre) und AUC-Werte (in Klammern) sowie Mittelwerte (bei krimineller Vorbelastung). AUC-Werte entsprechen einem Genauigkeitsmaß, mit dem das Instrument in der Lage ist, die angegebene Rückfallkategorie vorherzusagen (AUC-Wert > 0,71 = hoch, AUC-Wert > 0,63 = moderat, AUC-Wert < 0,63 = schwach; Signifikanz vorausgesetzt (nach *Dable u.a.* 2007)).

3.2 Zusammenhang zwischen aktuariischem Risiko (ermittelt durch Static-99) und Rückfallraten

Table 2 zeigt die Rückfallraten der Gesamtgruppe in Abhängigkeit vom Static-99 Wert. In allen angeführten Rückfallkategorien zeigt sich ein kontinuierlicher Anstieg der Rückfallereignisse mit Zunahme des Static-99-Wertes. Sexuell motivierte Rückfälle waren in der höchsten Risikokategorie nach Static-99 (Static 6, 6+) in über 16 % vorzufinden, während hingegen in der Niedrigrisikokategorie (Static 0–1) kaum entsprechende Rückfälle zu verzeichnen waren (< 1 %). Gewalttätige (»violent«) Rückfälle lagen in der Hochrisikokategorie bereits bei jedem 4. Täter vor, während hingegen in der niedrigsten Risikokategorie nach Static-99 nur knapp über 5 % in dieser Deliktategorie rückfällig wurden. Immerhin war jeder 5. Täter der Hochrisikokategorie aufgrund eines »violent« Rückfalls wieder in Haft, während in der niedrigsten Risikokategorie nur < 2 % wieder inhaftiert waren.

Table 2 Rückfallraten der Gesamtgruppe in Abhängigkeit vom Static-99 Wert in Prozent

	Static 0–1 N = 234	Static 2–3 N = 277	Static 4–5 N = 144	Static 6, 6+ N = 67
Rückfall allgemein	13,7	23,83	45,14	44,78
Rückfall sexuell	0,43	3,97	4,17	16,42
Rückfall sexuell – Haft	0,43	3,97	4,17	13,43
Rückfall sexuell »hands-on«	0,43	2,53	4,17	5,97
Rückfall »violent«	5,56	9,75	23,561	25,37
Rückfall »violent« – Haft	1,71	6,86	15,97	19,4

In der Gruppe der außerfamiliären Kindesmissbraucher (bei 224 Tätern lagen alle Informationen, die für den Static-99 notwendig sind, vor) konnte – insbesondere was die allgemeine, die sexuell motivierte und die gewalttätige (»violent«) Rückfälligkeit betraf – ein sehr deutlicher Zusammenhang zwischen Static-99-Werten und Rückfallereignissen nachgewiesen werden. Insbesondere hinsichtlich der Rückfälle mit Sexualdelikten hatte die Hochrisikogruppe (Static 6, 6+) mit einem Prozentsatz von 31 eine deutlich erhöhte Wahrscheinlichkeit gegenüber den anderen Risikogruppen. Während beinahe alle sexuell motivierten Rückfallereignisse – unabhängig davon, ob ein physischer Kontakt zum Opfer bestanden hatte oder nicht – in allen Risikogruppen mit Gefängnisstrafen sanktioniert wurden, zeigte sich z.B. in der höchsten Risikogruppe, dass nur etwa ein Drittel aller verurteilten sexuell motivierten Rückfallereignisse auch »hands-on«-Delikte waren (10,34 %). Die hohe Wiederverurteilungsrates aufgrund von Sexualdelikten ging also in erster Linie auf das Konto von »non-contact«-Ereignissen. Gewalttätige Rückfälle wurden in der höchsten Risikokategorie in 17,2 % der Fälle verurteilt, in 10,3 % wurden entsprechende Rückfälle mit Haftstrafen sanktioniert (siehe *Table 3*).

Tabelle 3 Rückfallraten in Prozent in Abhängigkeit vom Static-99 Wert – außer-familiäre Kindesmissbraucher

	Static 0–1 N = 59	Static 2–3 N = 96	Static 4–5 N = 40	Static 6, 6+ N = 29
Rückfall allgemein	16,95	23,96	37,5	37,93
Rückfall sexuell	1,69	7,29	2,5	31,03
Rückfall sexuell – Haft	1,69	7,29	2,5	27,59
Rückfall sexuell »hands-on«	1,69	5,21	2,5	10,34
Rückfall »violent«	5,08	8,33	15	17,24
Rückfall »violent« – Haft	5,08	5,21	10	10,34

Die Gruppe der innerfamiliären Kindesmissbraucher war bezüglich der Verteilung der Static-99 Werte nicht homogen (insgesamt lagen von 136 Tätern Static-99 Werte vor) (siehe *Tabelle 4*). 107 Täter hatten Werte von 0 oder 1, 26 Täter von 2 oder 3, und nur drei Täter erreichten einen Gesamtwert von 4 bis 5. Die höchste Risikostufe (6, 6+) wurde überhaupt nicht erreicht. Allgemeine Wiederverurteilungsraten lagen in der ersten und zweiten Risikokategorie bei knapp 12,2 % und 15,4 %, in der dritten bei 100 % (allerdings bei insgesamt nur drei Tätern). Sexuell motivierte Rückfälle kamen nur in der 2. Risikokategorie vor (7,7 % gesamt, 3,9 % für »hands-on«-Delikte). Rückfälle aus der Kategorie »violent« kamen insgesamt selten vor und wurden kaum mit Haftstrafen sanktioniert.

Tabelle 4 Rückfallraten in Prozent in Abhängigkeit vom Static-99 Wert – inner-familiäre Kindesmissbraucher

	Static 0–1 N = 107	Static 2–3 N = 26	Static 4–5 N = 3	Static 6, 6+ N = 0
Rückfall allgemein	12,15	15,38	100	
Rückfall sexuell	0	7,69	0	
Rückfall sexuell – Haft	0	7,69	0	
Rückfall sexuell »hands-on«	0	3,85	0	
Rückfall »violent«	1,87	11,54	33,3	
Rückfall »violent« – Haft	0	7,69	0	

In der Gruppe der Vergewaltiger (bei 344 Tätern lagen Informationen über den Static-99 vor) zeigte sich – abgesehen von sexuell motivierten Rückfallereignissen, die kaum vorkamen – ein eindeutiger Zusammenhang zwischen den Static-99 Werten und den Rückfallereignissen. In der höchsten Risikogruppe kam es zwar zu keinem einzigen Rückfall in ein Sexualdelikt, allerdings waren in dieser Gruppe nach vier Jahren 50 % wiederverurteilt, davon knapp 26,5 % aufgrund von gewalttätigen Handlungen zu einer Haftstrafe. Auch in der dritten Risikogruppe (Static-99-Werte 4, 5) kam es in fast 46 % der Fälle zu Wiederverurteilungen, etwa 18,4 % waren aufgrund eines Gewaltdelikttes wiederum inhaftiert worden. In dieser Gruppe lag die Rückfallrate aufgrund von sexuell motivierten Straftaten bei 4 % und war damit am höchsten (siehe *Tabelle 5*).

Tabelle 5 Rückfallraten in Prozent in Abhängigkeit zum Static-99 Wert – Vergewaltiger

	Static 0–1 N = 60	Static 2–3 N = 152	Static 4–5 N = 98	Static 6, 6+ N = 34
Rückfall allgemein	13,33	25	45,92	50
Rückfall sexuell	0	1,32	4,08	0
Rückfall sexuell – Haft	0	1,32	4,08	0
Rückfall sexuell »hands-on«	0	1,32	4,08	0
Rückfall »violent«	11,67	11,18	26,53	32,35
Rückfall »violent« – Haft	1,67	8,55	18,36	26,47

4. Diskussion

Die vorliegende Untersuchung stellt das Verhältnis von erhobenen aktuarischen Prognosevariablen und tatsächlicher Rückfälligkeit (genauer: Wiederverurteilung) etwa vier Jahre nach Entlassung einer Gruppe von Sexualstraftätern dar. Die untersuchte Stichprobe ist mit 785 Tätern wesentlich größer als in vielen vergleichbaren Studien, und sie ist vor allem repräsentativ für eine inhaftierte Sexualstraftäterpopulation, da sie alle derartigen Täter im österreichischen Strafvollzug umfasst.

Die Vier-Jahres-Rückfallraten (Wiederverurteilung) – insbesondere in weitere Sexualdelikte – sind mit knapp 4 % auf den ersten Blick niedrig, entsprechen allerdings bei genauerer Analyse vergleichbaren Studien an vergleichbaren Populationen (*Birkbauer & Hirtenlehner 2005; Hood et al. 2002; Barnoski 2005; Langan et al. 2003; Bartosh, Garby, Lewis & Gary 2003*).

Sexuell motivierte Rückfallraten waren in der Gruppe der außerfamiliären Kindesmissbraucher mit etwa 8 % am höchsten. Sowohl innerfamiliäre Täter als auch Vergewaltiger wurden jeweils nur in knapp 2 % (1,5 % und 1,6 %) mit Sexualdelikten rückfällig. Die Rückfallzahlen insbesondere für die Gruppen der Vergewaltiger und der innerfamiliären Kindesmissbraucher waren somit sehr niedrig, auch gegenüber vergleichbaren Studien. So finden *Greenberg et al. (2000)* nach etwas über sieben Jahren Nachuntersuchung 5 % der biologischen und Stiefväter mit einem Sexualdelikt rückfällig und etwa 11 % der Täter aus dem erweiterten Familienkreis. Innerfamiliäre Täter sind diejenige Gruppe mit der niedrigsten statistischen Rückfallwahrscheinlichkeit (*Hanson, Morton & Harris 2003; Hanson & Bussière 1998*), wengleich auch Hinweise darüber vorliegen, dass diese – durch Wiederverurteilungsraten ermittelten – Rückfallzahlen insbesondere in dieser Gruppe nicht die ganze Wahrheit abbilden und gerade hier möglicherweise die Dunkelziffer höher ist als in den anderen Untergruppen (*Eher & Ross 2006; Studer, Clelland, Aylwin, Reddon & Monro 2000; Rice & Harris 2002*). Vergewaltiger allerdings werden in den meisten Studien ähnlich oder sogar mehr als rückfallgefährdet angesehen als Kindesmissbraucher (*Hanson & Bussière 1998; Hanson, Morton & Harris 2003; Harris et al. 2003*), was jedenfalls den Ergebnissen unserer Studie nicht entspricht. Dieser Unterschied liegt vermutlich darin begründet, dass hier eine repräsentative Population entlassener Strafgefangener vorliegt, die offenbar ein deutlich niedrigeres Basisrisiko aufweist gegenüber den Populationen der oben zitierten Studien, in denen vorwiegend Hochrisikotäter untersucht wurden. Geht es um vergleichbare Populationen (repräsentative Straftäterpopulationen), so finden sich ähnlich niedrige Werte – wie z.B. in einer Studie über entlassene Täter aus dem US-Bundesstaat Washington (*Barnoski 2005*), in der die Fünf-Jahres-Wiederverurteilungsraten bei Vergewaltigern bei

3,9 % lagen, oder eine Gesamt-US-amerikanische Studie über entlassene Sexualstraftäter, die von 3,2 % Wiederverurteilungen aufgrund sexuell motivierter Straftaten bei Vergewaltigern nach drei Jahren berichtet (*Langan et al.* 2003).

Selten erhoben oder gar differenziert werden Rückfallraten zwischen sexuell motivierten »hands-on«- (also mit physischem Kontakt) und »hands-off«-Delikten (ohne entsprechenden physischen Kontakt mit dem Opfer). In unserer Studie war die Rückfallrate mit sexuell motivierten »hands-on«-Delikten mit 2,3 % nochmals deutlich niedriger als der Prozentsatz derer, die überhaupt aufgrund eines Sexualdelikts wieder verurteilt wurden (3,8 %). Wesentlich höher waren die Prozentsätze betreffend allgemeiner Rückfälligkeit, die nach vier Jahren für die Gesamtgruppe bei knapp 28 % lag. Mit der Deliktkategorie »violent« wurden insgesamt fast 13 % rückfällig, wengleich auch nur in 8,3 % der Fälle eine Haftstrafe verhängt wurde.

Relevante Kategorien im Sinne von sexuell motivierten »hands-on«-Delikten oder gewalttätigen (»violent«) Delikten – insbesondere dann, wenn sie aufgrund ihres Schweregrades auch mit einer Haftstrafe geahndet werden – sind offenbar seltene Ereignisse, die die Prognose allein schon aus statistischen Gründen erschweren. Noch schwieriger wird die Prognose solcher Rückfallereignisse für verschiedene relevante Sexualtäter-Untergruppen, da die Wahrscheinlichkeit ihres Auftretens in manchen Gruppen noch geringer ist. Dies wirft zu Recht die Frage auf, ob in bestimmten Untergruppen das sexuell motivierte, insbesondere »hands-on«-Rückfallereignis überhaupt vorhergesagt werden kann (*Eber u.a.* 2008). Tatsächlich ist die Vorhersage des sexuell motivierten Rückfallereignisses mittels aktuuarischer Prognoseinstrumente bei aus der Strafhaft entlassenen Vergewaltigern kaum möglich. Die Basisraten sind so niedrig, dass jedenfalls anhand einer Vier-Jahres-Rückfallstudie, wie sie hier vorliegt, trotz Vorhandenseins einer vergleichsweise sehr großen Stichprobe derartige Aussagen nicht gemacht werden können. AUC-Werte legen den Schluss nahe, dass insbesondere das sexuell motivierte Rückfallereignis bei einer aus der Strafhaft entlassenen Vergewaltigerpopulation nicht vorhergesagt werden kann, ganz im Unterschied zum gewalttätigen Delikt. Bei der Gruppe der Kindesmissbraucher verhält es sich hingegen anders. Für die Gruppe der außerfamiliären Kindesmissbraucher zeigen sich nicht nur besonders gute AUC-Werte, sondern es lässt sich auch der Anstieg der Wiederverurteilungen in Abhängigkeit von Static-99-Werten gut nachverfolgen, wengleich selbst in der höchsten Risikokategorie lediglich 30 % rückfällig sind, und – zieht man die »hands-off«-Delikte ab – selbst hier nur eine 10 %ige Rückfallwahrscheinlichkeit mit einem »hands-on«-Delikt vorhergesagt werden kann (also auch hier die Güte der Vorhersage eines relevanten »hands-on«-Delikts sehr bescheiden ist). Auch die Analyse von innerfamiliären Kindesmissbrauchern ergab zwar hohe AUC-Werte bei der Vorhersage des sexuell motivierten Rückfalls, diese waren allerdings – möglicherweise bedingt durch die niedrigen Rückfallbasisraten – nicht signifikant. Wengleich also in unserer Studie (noch) nicht nachweisbar, gibt es dennoch Hinweise, dass auch diese Vorhersage mittels aktuuarischer Instrumente möglich ist (*Bartosh et al.* 2003). Sehr gut möglich jedenfalls war bei innerfamiliären Tätern – ebenso wie in oben angeführter Studie – die Vorhersage des gewalttätigen Delikts.

Vorliegende Daten legen nahe, die manchmal überzogenen Erwartungen an das Prognostizieren etwas bescheidener ausfallen zu lassen. Geht man davon aus, dass die aktuuarische Prognose derzeit der klinischen gegenüber als überlegen gilt (*Hanson et al.* 2007) – wengleich auch dazu Kritik anzumerken ist, da die »klinische« Prognose oftmals schlecht oder sogar unfair definiert ist (*Dable* 2007) –, so kann angesichts des Vorliegens dieser repräsentativen Daten zu Recht geschlossen werden, dass die Prognose nach gruppenstatistischen Regeln zwar richtig ist und ihre Gültigkeit hat, dass sie allerdings als alleinige Einschätzung für den Einzelfall unzureichend ist. Selbst in den höchsten, durch validierte Instrumente

erhobenen Risikostufen sind in der Regel weniger als 50 % der Probanden tatsächlich innerhalb von vier Jahren rückfällig. Dies ist insbesondere von Relevanz, wenn es um die Verhängung oder Aufhebung von freiheitsentziehenden Maßnahmen geht, muss man sich doch eingestehen, dass die Treffsicherheit derartiger Entscheidungen momentan leider noch nicht sehr hoch sein kann (Douard 2007; Abracen et al. 2006). Die entsprechende kritische Würdigung solcher Ergebnisse und vor allem die Anwendung derartiger Instrumente ausschließlich durch den gut ausgebildeten Forensiker sind zu fordern.

Actuarial prognoses of sex offenders

Results from a study involving 785 offenders with an emphasis on relevant offender groups and recidivism categories

Summary

Methods of actuarial risk assessment are indispensable when analyzing recidivism base rates for sexual offenders. As such, an actuarial risk assessment should be performed at the beginning of each risk evaluation. However, in spite of the fact that there are good validity indices for the sexual offender population in general, actuarial statistics have not yet proven to be of consistent predictive value when applied to some particular subgroups of sexual offenders and relapse categories. Using a validated German version of the Static-99, this study demonstrates that predictive validity differs quite dramatically in respect of offender subgroup and relapse category. Therefore, the critical use and importance of actuarial risk assessment methods conducted by experienced and trained professionals must once again be emphasized.

Keywords: Sexual offenders, relapse, base rate, risk assessment, Static-99

Literatur

Abracen, J. & Looman, J. (2006). Evaluation of civil commitment criteria in a high risk sample of sexual offenders. *Journal of Sexual Offender Civil Commitment: Science and the Law* 1, 124-140. – Barnoski, R. (2005). Sex Offender Sentencing in Washington State: Recidivism Rates. Washington State Institute for Public Policy; <http://www.wsipp.wa.gov/rptfiles/05-08-1203.pdf>. – Bartosh, D.L., Garby, T., Lewis, D. & Gary, S. (2003). Differences in the predictive validity of actuarial risk assessments in relation to sex offender type. *International Journal of Offender Therapy and Comparative Criminology* 47, 422-438. – Birkelbauer, A. & Hirtenlehner, H. (2005). Bewährung nach bedingter Entlassung aus dem Strafvollzug. *Österreichische Juristen Zeitung* 55, 938-948. – Bonta, J. & Andrews, D.A. (2008). Risk-Need-Responsivity Model for Offender Assessment and Rehabilitation. Public Safety Canada; http://www.ps-sp.gc.ca/res/cor/rep/fl/Risk_Need_2007-06_e.pdf. – Dable, K.P., Schneider, V. & Zietzen, F. (2007). Standardisierte Instrumente zur Kriminalprognose: Forensische Psychiatrie, Psychologie, Kriminologie 1, 15-26. – Douard, J. (2007). Loathing the sinner, medicalizing the sin: Why sexually violent predator statutes are unjust. *International Journal of Law and Psychiatry* 30, 36-48. – Eber, R., Rettenberger, M., Schilling, F. & Pfäfflin, F. (2008). Validität oder praktischer Nutzen? Rückfallvorhersagen mittels Static-99 und SORAG. Eine prospektive Rückfallstudie an 275 Sexualstraftätern. *Recht und Psychiatrie* 26, 79-88. – Eber, R. & Ross, T. (2006). Reconsidering risk for reoffense in interfamilial child molesters: New aspects on clinical and criminological issues. *Sexual Offender Treatment*; <http://www.sexual-offender-treatment.org/39.0.html>. – Greenberg, D., Bradford, J., Firestone, P. & Curry, S. (2000). Recidivism of child molesters: A study of victim relationship with the perpetrator. *Child Abuse and Neglect* 24, 1485-1494. – Hanson, R.K. & Bussière, M.T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology* 66, 348-362. – Hanson, R.K., Gordon, A., Harris, A.J., Marques, J.K., Murphy, W., Quinsey, V.L. et al. (2002). First report of the collaborative outcome data project on the effectiveness of psychological treatment for sex offenders. *Sexual Abuse* 14, 169-194. – Hanson, R.K., Morton, K.E. & Harris, A.J. (2003). Sexual offender recidivism risk: What we know and what we need to know. *Annals of the New York Academy of Science* 989, 154-166. – Hanson, R.K. & Morton-Bourgon, K.E. (2007). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis. Public Safety Canada; <http://www.publicsafety.gc.ca/>

res/cor/rep/_fl/crp2007-01-en.pdf. – Harris, A.J., Phenix, A., Hanson, R.K. & Thornton, D. (2003). Static-99 Coding Rules Revised - 2003. Public Safety Canada; http://ww2.psepc-sppcc.gc.ca/publications/corrections/pdf/Static-99-coding-Rules_e.pdf. – Harris, G.T., Rice, M.E., Quinsey, V.L., Lalumière, M.L., Boer, D. & Lang, C. (2003). A multisite comparison of actuarial risk instruments for sex offenders. *Psychological Assessment* 15, 413-425. – Hood, R., Shute, S., Feilzer, M. & Wilcox, A. (2002). Sex offenders emerging from long-term imprisonment: A study of their long-term reconviction rates and of parole board members' judgments of their risk. *British Journal of Criminology* 42, 371-394. – Langan, P.A., Schmitt, E.L. & Durose, M.R. (2003). Recidivism of sex offenders released from prison in 1994. U.S. Department of Justice; <http://www.ojp.usdoj.gov/bjs/abstract/rsorp94.htm>. – Loesel, F. & Schmucker, M. (2005). The effectiveness of treatment for sexual offenders: A comprehensive meta-analysis. *Journal of Experimental Criminology* 1, 117-146. – Monahan, J. & Silver, E. (2003). Judicial decision thresholds for violence risk management. *International Journal of Forensic Mental Health* 2, 1-6. – Packard, R.L. & Levenson, J.S. (2006). Revisiting the reliability of diagnostic decisions in sex offender civil commitment. *Sexual offender treatment*; <http://www.sexual-offender-treatment.org/50.0.html>. – Quinsey, V.L., Harris, A.J., Rice, M.E. & Cormier, B.M. (2006). *Violent Offenders: Appraising and Managing Risk*. 2nd ed. Washington DC: American Psychological Association. – Rettenberger, M. & Eher, R. (2006). Die deutsche Übersetzung und Adaptierung des Static-99 zur aktuarischen Kriminalprognose verurteilter Sexualstraftäter. *MschKrim* 89, 352-365. – Rettenberger, M. & Eher, R. (2007). Aktuarische Kriminalprognosemethoden und Sexualdelinquenz: Die deutsche Version des SORAG. *MschKrim* 90, 484-497. – Rice, M.E. & Harris, G.T. (2002). Men who molest their sexually immature daughters: Is a special explanation required? *Journal of Abnormal Psychology* 111, 329-339. – Seto, M.C., Harris, G.T., Rice, M.E. & Barbaree, H.E. (2004). The screening scale for pedophilic interests predicts recidivism among adult sex offenders with child victims. *Archives of Sexual Behavior* 33, 455-466. – Stadland, C., Hollweg, M., Kleindienst, N., Dietl, J., Reich, U. & Nedopil, N. (2006). Rückfallprognosen bei Sexualstraftätern. Vergleich der prädiktiven Validität von Prognoseinstrumenten. *Nervenarzt* 77, 587-595. – Studer, L.H., Clelland, S.R., Aylwin, A.S., Reddon, J.R. & Monro, A. (2000). Rethinking risk assessment for incest offenders. *International Journal of Law and Psychiatry* 23, 15-22. – Ziegler, A., Lange, S. & Bender, R. (2004). Überlebenszeitanalyse: Die Cox-Regression. *Deutsche Medizinische Wochenschrift* 129, T1-T3.

(Anshr. d. Verf.: PD Dr. Reinhard Eher und Dipl.-Psych. Martin Rettenberger, Begutachtungs- und Evaluationsstelle für Gewalt- und Sexualstraftäter, Vollzugsdirektion, Gerichtsgasse 6, 1210 Wien/Austria; reinhard.eher@justiz.gv.at; martin.retttenberger@justiz.gv.at; Dipl.-Psych. Anna Matthes, Institut für Psychologie, Christian-Albrechts-Universität zu Kiel, Olshausenstr. 75, 24118 Kiel/Germany; matthes@psychologie.uni-kiel.de)

Die deutsche Übersetzung und Adaptierung des Static-99 zur aktuarischen Kriminalprognose verurteilter Sexualstraftäter

Erste Validierungsdaten¹

von Martin Rettenberger und Reinhard Eher

Zusammenfassung

Das Static-99 ist ein aktuarisches Prognoseinstrument, das für die Vorhersage sexuell motivierter und gewalttätiger Rückfälligkeit von Sexualstraftätern konzipiert wurde. Es wurde im Jahre 1999 von *Hanson* (Kanada) und *Thornton* (Großbritannien) entwickelt und wird in Nordamerika und manchen europäischen Ländern mittlerweile regelmäßig bei der kriminalprognostischen Begutachtung von Sexualstraftätern eingesetzt. In der vorliegenden Arbeit werden die Interraterreliabilität sowie die konvergente und die prädiktive Validität einer deutschsprachigen Adaption der im Jahre 2003 von *Harris*, *Phenix*, *Hanson* und *Thornton* revidierten Version des Static-99 berichtet. Die Daten wurden retrospektiv anhand von Akteninformationen von Sexualstraftätern aus Österreich, die zwischen 1968 und 2002 verurteilt wurden, erhoben. Dabei wurden für die Interraterreliabilität und die konvergente Validität zufriedenstellende Ergebnisse ermittelt. Die Überprüfung der prädiktiven Validität ergab gute Vorhersageleistungen für die hier verwendete deutschsprachige Adaption des Instruments (allgemeine Rückfälligkeit: $r = .41$, $AUC = .74$; sexuell motivierte Rückfälligkeit: $r = .35$, $AUC = .74$; gewalttätige Rückfälligkeit: $r = .41$, $AUC = .76$). Die Ergebnisse der vorliegenden Untersuchung können als Beleg dafür interpretiert werden, dass die hier verwendete deutschsprachige Adaption des revidierten Static-99 zur Erhöhung der Validität kriminalprognostischer Einschätzungen bei Sexualstraftätern beitragen kann.

Schlüsselwörter: Static-99, Kriminalprognose, Sexualstraftäter, Validität, Rückfälligkeit

1. Einleitung

Kriminalprognostischen Einschätzungen² kommt im deutschen und österreichischen Strafrecht große Bedeutung zu. In der Bundesrepublik Deutschland beispielsweise beeinflussen sie neben der Auswahl und Bemessung von Sanktionen auch den Vollzug und besitzen entscheidenden Einfluss auf die Beendigung vor allem freiheitsentziehender Sanktionen (vgl. z.B. *Dable* 2005; *Nedopil* 2000). In den letzten Jahren und Jahrzehnten konnten auf dem Gebiet der forensisch-kriminologischen Prognoseforschung beachtliche Fortschritte erzielt

- 1 Dieser Artikel basiert auf der Diplomarbeit von *Martin Rettenberger*, die im Studiengang Psychologie am Institut für Forensische Psychiatrie der Freien Universität Berlin unter der Betreuung von *Klaus-Peter Dable* geschrieben wurde. Die wissenschaftliche Begleitung vor Ort (Zentrale Dokumentations- und Koordinationsstelle für Sexualtäter, Wien, und Forensisch Therapeutisches Zentrum Wien) erfolgte durch *Reinhard Eher*. Die deutschsprachige Adaption der revidierten Version des Static-99 sowie die Diplomarbeit sind bei bestehender IATSO-Mitgliedschaft verfügbar unter www.iatso.org/iatso_user_tool/publications.
- 2 In Anlehnung an *Dable* (2000; 2005) ist in dieser Untersuchung der Begriff Kriminalprognose im Sinne einer individuellen Kriminalrückfallprognose zu verstehen, d.h. Gegenstand der individuellen Prognose sind Personen, die bereits einschlägig in Erscheinung getreten sind. Im Gegensatz dazu definiert *Nedopil* individuelle Kriminalrückfallprognosen als Rückfallprognosen, während sich der Begriff Kriminalprognose ausschließlich darauf bezieht, »ob ein Mensch, eine bestimmte Gruppe von Menschen oder bestimmte Bevölkerungsanteile in Zukunft überhaupt kriminell werden – und zwar unabhängig von der Frage, ob sie bislang kriminell waren« (*Nedopil* 2005, 17).

werden, so dass aktuell nicht mehr diskutiert wird, ob psychiatrische oder psychologische Sachverständige überhaupt in der Lage sind, einen sinnvollen Beitrag zur Kriminalprognose zu leisten, sondern vielmehr, welche Methoden dabei zur Anwendung kommen sollten (vgl. Noll, Endrass, Rossegger & Urbaniok 2006). In den meisten Darstellungen kriminalprognostischer Methoden (z.B. Dable 1997; 2000; 2005; Rehder 2004; Foerster 2000) werden die folgenden drei methodischen Grundformen genannt: die statistische, die klinische und die intuitive Kriminalprognose. Dable (1997) zufolge handelt es sich bei dieser schematischen Einteilung um eine Idealisierung, die in der Praxis kaum vorkommt; in der Regel findet man vielfältige Überschneidungen der drei Methoden (vgl. König, Schnoor, Auer, Rebernick, Schläfke & Fegert 2005).

Nach der Definition von Dable (1997; 2000; 2005) versteht man unter der intuitiven Methode³ ein individuumszentriertes Vorgehen, bei dem sich der Prognostiker ohne Bezugnahme auf vorgegebene Regeln oder allgemeine theoretische oder empirische Konzepte ausschließlich von den spezifischen individuellen Gegebenheiten der zu beurteilenden Person leiten lässt. Als Konsequenz daraus folgt, dass eine intuitive Prognose zwangsläufig immer frei von Prognosefehlern ist – allerdings nicht von Prognoseirrtümern –, da es keine methodischen Regeln gibt, gegen die sie verstoßen könnte⁴. In rechtlicher Hinsicht verstößt ein rein intuitives Vorgehen damit gegen das Transparenzgebot, das bei Kriminalprognosen von Sachverständigen vor Gericht die Nachvollziehbarkeit und Überprüfbarkeit der kriminalprognostischen Einschätzung gewährleistet (Dable 2005).

Dable (1997; 2000; 2005) definiert die statistisch-nomothetische («aktuarische») Prognose als den methodischen Idealtypus einer vollständig regelgeleiteten Vorgehensweise. Sowohl die Auswahl der für die Prognose benötigten Informationen, als auch die dafür erforderlichen Erhebungsmethoden und die Art und Weise der Verknüpfung der gewonnenen Daten zu einem prognostischen Urteil werden von den methodischen Vorgaben geleitet. Das Grundprinzip dieser Vorgehensweise besteht darin, personen- und/oder tatbezogene Merkmale zu identifizieren, für die in empirischen Untersuchungen ein Zusammenhang mit Rückfälligkeit festgestellt wurde. Es wird angenommen, dass diese Merkmale nicht nur dazu geeignet sind, bereits rückfällig gewordene Straftäter zu beschreiben, sondern auch für die Vorhersage zukünftiger Rückfälle verwendet werden können. Zur Untermauerung dieser Annahme werden Kreuzvalidierungen bei einzelnen Instrumenten vorgenommen. Mit dem Begriff «aktuarisch» wird eine Weiterentwicklung der statistischen Prognoseinstrumente in den letzten Jahren beschrieben (vgl. Nedopil 2005).

Die klinische Prognosemethode versucht die Vorteile der Individuumszentrierung der intuitiven Methode mit einer regelgeleiteten und wissenschaftlichen Standards genügenden Vorgehensweise zu verbinden (vgl. Dable 2005)⁵. Ähnlich wie bei den oben genannten Definitionen handelt es sich dabei um einen methodischen Idealtypus, der sich von den statisti-

3 Dable (2005) zufolge handelt es sich beim intuitiven Vorgehen nicht um eine Methode im engeren Sinn, da die Person des Beurteilers die Methodik gewissermaßen ersetzt.

4 Nach Dable (2005) versteht man unter einem Prognosefehler eine fehlerhafte Anwendung einer Prognosemethode, d.h. der Prognostiker verstößt gegen die der jeweiligen Methode zugrundeliegenden Standards und Regeln. Als Prognoseirrtum wird hingegen das Nichteintreffen der Verhaltensvorhersage bezeichnet, obwohl die Prognosemethode korrekt angewandt wurde. Fehlerfreie Prognosen in diesem Sinne schließen somit Irrtümer nicht aus. Es ist aber zu erwarten, dass eine methodisch fehlerfreie Prognose das Risiko eines Prognoseirrtums auf das Ausmaß des nach aktuellem wissenschaftlichen Kenntnisstand Möglichen reduziert.

5 In der wissenschaftlichen Literatur werden in Bezug auf die klinische Prognosemethode teilweise verschiedene definitorische Schwerpunkte gesetzt, wengleich die Kernaussagen große Parallelen aufweisen (vgl. z.B. Nedopil 2000; 2005; Rehder 2004; Foerster 2000; Nowara 2001).

schen Prognosemethoden im Hinblick auf das Ausmaß an methodischen Vorgaben und Reglementierungen unterscheidet. Bei der klinischen Prognose sind die vorgegebenen Regeln eher als allgemeine Leitlinien oder Prinzipien zu verstehen, die grundsätzlich erforderliche Denkschritte reglementieren. Dem Prognostiker wird seine Vorgehensweise jedoch nicht in allen Einzelheiten vorgeschrieben, wie dies etwa bei statistischen Methoden in der Regel der Fall ist. Im Gegensatz zur intuitiven Methode können bei der klinischen Prognose mögliche Fehler im prognostischen Prozess identifiziert werden. Die Beurteilung der Prognosefehler fällt im Vergleich zu den statistischen Methoden allerdings schwerer, da der Spielraum bei der Prognosestellung größer ist (*Dable* 2005).

Wie bereits erwähnt, ist eine Trennung der einzelnen Prognosemethoden lediglich künstlicher Natur, da es sich bei den genannten methodischen Vorgehensweisen um Idealtypen handelt. Vor allem in den letzten Jahren wurde verstärkt auf eine Integration der verschiedenen Methoden gesetzt, wobei im Mittelpunkt derartiger Bemühungen eine Verbindung der statistischen und der klinischen Vorgehensweise steht (vgl. *König u.a.* 2005; *Nedopil* 2005). Ein Beispiel hierfür ist der Vorschlag von *Endres* (2000), der anhand einer zweistufigen statistisch-klinischen Prognosemethode versucht, die Vorteile beider Methoden zu nutzen. Zunächst sollen statistische Prognosemethoden angewendet werden, um allgemein zu klären, ob die in Frage stehende Person zu einer Gruppe mit hohem Rückfallrisiko gehört. Anschließend wird unter Verwendung klinischer Methoden nach den individuellen Ursachen gesucht und geklärt, inwieweit diese fortbestehen. *König u.a.* (2005) sehen in der Verwendung aktuarischer Instrumente zu Beginn des Prognoseprozesses eine Möglichkeit, allgemeine Anhaltspunkte für die Einschätzung der Basiswahrscheinlichkeit (s.u.) zu gewinnen. Darauf aufbauend soll dann eine klinisch-intuitive Beurteilung des individuellen Einzelfalles erfolgen. Neben diesen integrativen Ansätzen wurden in den letzten Jahren diverse Weiterentwicklungen der genannten Idealtypen der kriminalprognostischen Methodik vorgestellt: Hier sind vor allem die so genannten dimensionalen Konzepte von *Rasch* (1999) und *Nedopil* (1995; 2000) und das darauf basierende allgemeine Prozessmodell der klinisch-idiographischen Prognosebeurteilung von *Dable* (1997; 2000; 2005) zu nennen.

Das *Static-99* (*Hanson & Thornton* 1999) ist den statistisch-nomothetischen (»aktuarischen«) Prognoseinstrumenten zuzuordnen und kann gemäß *Dable* (2005) zu den bekannteren Prognoseinstrumenten dieser Art gezählt werden (vgl. dazu auch *Noll u.a.* 2006). Das *Static-99* gewährleistet den Autoren zufolge eine valide Einschätzung der Basiswahrscheinlichkeit von sexuell motivierten und gewalttätigen Rückfällen bei bereits mindestens einmal einschlägig in Erscheinung getretenen erwachsenen männlichen Sexualstraftätern. Das Instrument ist eine Synthese aus dem *Rapid Risk Assessment for Sex Offender Recidivism* (RRASOR; *Hanson* 1997) und dem *Structured Anchored Clinical Judgement-Minimum* (SACJ-Min; *Grubin* 1998). Das RRASOR besteht aus vier Items, die sich alle auf einfach zu erhebende, statische Merkmale beziehen (vgl. *Hanson* 1997). Die vier Items wurden von *Hanson & Bussière* (1996) anhand einer Meta-Analyse von Verlaufsuntersuchungen aus dem angloamerikanischen Raum extrahiert. Für diese vier Merkmale wurden stabile Zusammenhänge mit einem sexuell motivierten Rückfallereignis ermittelt. Basierend auf dieser Meta-Analyse führte *Hanson* (1997) eine empirische Untersuchung über Rückfallprädiktoren durch und kam zu dem Ergebnis, dass diese vier Items über verschiedene Studien hinweg bei geringer Redundanz zwischen rückfälligen und nicht-rückfälligen Sexualstraftätern diskriminierten. Anschließend lieferte eine Kreuzvalidierungsstudie, die anhand einer Stichprobe aus Großbritannien durchgeführt wurde, einen ersten Beleg für die prädiktive Validität des Instruments (*Hanson* 1997). Etwa zur gleichen Zeit entwickelte *Thornton* anhand mehrerer Stichproben aus Großbritannien das *Structured Anchored Clinical Judgement* (SACJ; *Grubin* 1998), für das ebenfalls Hinweise bezüglich der prädiktiven Validität vorlie-

gen. Das SACJ wird aktuell regelmäßig in Großbritannien und Wales angewandt und enthält neben Items, die das Ausmaß der sexuellen Devianz erfassen, auch Items, die sich auf die Vordelinquenz des Straftäters beziehen. Das SACJ-Min ist eine Kurzform des SACJ und besteht aus insgesamt neun Items, von denen sieben auch im Static-99 enthalten sind.

Die Kurzform des SACJ und das RRASOR wurden zum Static-99 zusammengefasst mit dem Ziel, ein Instrument zu entwickeln, das einfach zu erhebende, statische Risikofaktoren erfasst (siehe *Tabelle 1*) und in seiner prädiktiven Validität seinen Vorgängern überlegen ist. Neun der insgesamt zehn Items werden dichotom mit »Ja« vs. »Nein« bzw. »Vorhanden« vs. »Nicht vorhanden« bewertet, während ein Item auf einer Skala von »0« bis »3« Punkten bewertet wird, so dass sich daraus ein möglicher Gesamtscorebereich zwischen »0« und »12« Punkten ergibt. Diese 12 Items können drei thematischen Kategorien zugeordnet werden: Zwei Items beziehen sich auf demographische Informationen (*Demographic Information*), fünf Items beziehen sich auf Informationen bezüglich der kriminellen Vorgeschichte (*Official Criminal Record*) und drei Items werden anhand der Charakteristika des bzw. der Opfer (*Victim Information*) bewertet. Das Static-99 kann ausschließlich anhand von Akteninformationen bewertet werden. Gemäß den Instruktionen des Manuals der revidierten Version des Instruments (*Harris, Phenix, Hanson & Thornton 2003*) dürfen für die Datenerhebung hinsichtlich der Vordelinquenz des Straftäters Explorationsinformationen bei der Bewertung nicht berücksichtigt werden. Liegen diesbezüglich keine offiziellen Akteninformationen vor, kann das Static-99 nicht verwendet werden. Bei den restlichen fünf Items können unter Umständen Explorationsdaten in den Entscheidungsprozess integriert werden, allerdings sollten diese durch zusätzliche Hinweise aus offiziellen Informationsquellen abgesichert werden. Anschließend an die Bewertung aller Items werden die einzelnen Itemscores zu einem Gesamtscore addiert. Der Straftäter kann nun anhand dieses Gesamtscores einer von vier möglichen Risikokategorien (»niedrig«, »niedrig bis durchschnittlich«, »durchschnittlich bis hoch« und »hoch«) zugeordnet werden, welche wiederum eine bestimmte, empirisch ermittelte Basisrate für ein erneutes sexuell motiviertes bzw. ein erneutes Gewaltdelikt nach 5, 10 oder 15 Jahren ergibt (vgl. *Hanson & Thornton 1999; Harris, Phenix, Hanson & Thornton 2003*). Die prognostische Bedeutung dieser zehn Items konnte anhand der bereits genannten Meta-Analyse von *Hanson & Bussière (1998)* bestätigt werden. Wie auch das RRASOR (vgl. *Hanson 1997*) besitzt das Static-99 nicht den Anspruch, eine umfassende kriminalprognostische Einschätzung zu leisten, die den rechtlichen und methodischen Anforderungen gerecht wird. Es dient vielmehr der Einschätzung der Basiswahrscheinlichkeit⁶, die am Anfang eines umfassenderen Prognoseprozesses stehen kann, und anhand derer eine erste Einstufung in eine von vier Risikokategorien möglich ist (vgl. *Hanson & Thornton 1999; Harris, Phenix, Hanson & Thornton 2003*).

6 Unter der Basisrate versteht man das Vorhandensein eines Merkmals in einer Population (vgl. *Nedopil 2005*). Im forensischen Kontext ist dieses Merkmal eine Straftat, die innerhalb eines Zeitraums in einer Population – beispielsweise in der Bevölkerung eines bestimmten Landes – begangen wurde. Die daraus resultierende Deliktrate wird auch als Basisrate bezeichnet. Rückfallraten (oder Rezidivraten) wiederum beziehen sich auf das wiederholte Auftreten eines Merkmals in einer Population von Merkmalsträgern, d.h. sie stellen eine Maßzahl für den Anteil an Personen dar, die bei Entlassung mit erneuten (gravierenden) Straftaten rückfällig werden würden.

Table 1 Die zehn Items des revidierten Static-99 (vgl. Harris, Phenix, Hanson & Thornton 2003) und in der deutschsprachigen Adaption (Rettenberger & Eher 2006)

Item-Nr.	Original	Adaption	Erläuterung
1	Young	Alter des Straftäters (zum Zeitpunkt der Prognose)	Ist der Straftäter zu dem Zeitpunkt, für den die kriminalprognostische Einschätzung gelten soll, zwischen 18 und 25 Jahre alt, wird er bei diesem Item mit »1« bewertet. Ist er hingegen 25 Jahre alt oder älter, wird dieses Item mit »0« bewertet.
2	Ever lived with an intimate partner – 2 years	Beziehungsstatus – partnerschaftliche Beziehungen, die mindestens zwei Jahre andauerten	Wenn der Straftäter in der Vergangenheit eine partnerschaftliche Beziehung führte, die mindestens zwei Jahre anhielt, wird dieses Item mit »0« bewertet. War dies nicht der Fall, resultiert daraus der Itemscore »1«.
3	Index non-sexual violence (NSV) – any convictions	Verurteilungen beim Index-Delikt aufgrund nicht-sexueller Gewalt	Falls der Strafregisterauszug bzw. der Bundeszentralregisterauszug des Straftäters zusätzlich eine Verurteilung aufgrund nicht-sexueller Gewaltanwendung aufweist, die zusammen mit der Verurteilung des Index-Delikts ausgesprochen wurde, wird dieses Item mit »1« bewertet. Ist kein zusätzlich verurteiltes Gewaltdelikt enthalten, wird dieses Item mit »0« bewertet.
4	Prior non-sexual violence – any convictions	frühere Verurteilungen aufgrund nicht-sexueller Gewalt	Falls der Strafregisterauszug bzw. der Bundeszentralregisterauszug des Straftäters Verurteilungen aufgrund nicht-sexuell motivierter Gewaltdelikte aufweist, die vor dem Index-Delikt begangen wurden, wird dieses Item mit »1« bewertet. Existieren keine derartigen Verurteilungen wird dieses Item mit »0« bewertet.
5	Prior sex offences	frühere Anklagen und/oder Verurteilungen aufgrund sexuell motivierter Straftaten	Dies ist das einzige Item des Static-99, das nicht dichotom bewertet wird. Je nach Anzahl früherer Anklagen und/oder Verurteilungen wird dieses Item zwischen »0« und »3« bewertet.
6	Prior sentencing dates	Vorstrafen	Enthält der Strafregisterauszug bzw. der Bundeszentralregisterauszug des Delinquenten vier oder mehr Einträge aufgrund von Straftaten, die vor dem Index-Delikt begangen wurden, wird dieses Item mit »1« bewertet. Wenn der Straftäter drei oder weniger Delikte vor dem Index-Delikt beging, wird dieses Item mit »0« bewertet.

Item-Nr.	Original	Adaption	Erläuterung
7	Any convictions for non-contact sex offences	Verurteilungen aufgrund von Sexualstraftaten ohne Opfer bzw. ohne körperlichen Kontakt	Enthält der Strafregisterauszug bzw. der Bundeszentralregisterauszug des Straftäters separate – d.h. unabhängig vom Index-Delikt – Einträge aufgrund von Verurteilungen wegen Sexualstraftaten ohne Opfer bzw. ohne körperlichen Kontakt, wird dieses Item mit »1« bewertet. Sind keine derartigen Eintragungen vorhanden, wird dieses Item mit »0« bewertet.
8	Any unrelated victims	verwandtschaftliches Verhältnis zwischen Täter und Opfer	Falls mindestens ein Opfer der vom Täter begangenen Sexualdelikte nicht mit diesem verwandt ist, wird dieses Item mit »1« bewertet. Handelt es sich hingegen ausschließlich um verwandte Opfer, erhält dieses Item den Score »0«.
9	Any stranger victims	Bekanntheitsgrad zwischen Täter und Opfer	Ist dem Täter mindestens ein Opfer der von ihm begangenen Sexualstraftaten zum Zeitpunkt der im Index-Delikt enthaltenen sexuellen Handlungen noch nicht mindestens 24 Stunden bekannt, wird dieses Item mit »1« kodiert. Kannte der Täter hingegen sämtliche Opfer der von ihm begangenen Sexualstraftaten vor der Tat bereits seit mindestens 24 Stunden, wird dieses Item mit »0« bewertet.
10	Any male victims	Geschlecht des Opfers	Ist mindestens ein Opfer der vom Straftäter begangenen Sexualdelikte männlich, wird dieses Item mit »1« bewertet. Falls sämtliche Opfer weiblich sind, wird dieses Item mit »0« bewertet.

Neben dem *Static-2002*⁷ (Hanson & Thornton 2003), das sich deutlich vom Static-99 unterscheidet, wurde von Harris, Phenix, Hanson & Thornton (2003) eine weitere revidierte Version des Static-99 veröffentlicht, die der ursprünglichen Version zwar sehr ähnlich ist, zusätzlich aber einige Schwachpunkte des Vorgängers behebt. Durch die unterschiedliche Herkunft der beiden Autoren (Hanson aus Kanada und Thornton aus Großbritannien) und des Datenmaterials waren die Operationalisierungen und Kodierungsregeln für die Anwendung relativ kompliziert geworden (vgl. Nedopil 2005). Da dies im Widerspruch zum anfangs formulierten Ziel der einfachen Handhabung stand, wurde das Original von den Autoren überarbeitet (vgl. Harris, Phenix, Hanson & Thornton 2003). Diese revidierte Version der Originalversion ist Gegenstand der vorliegenden Arbeit.

Um die Vorhersagegenauigkeit des Instruments zu überprüfen, wurden in den letzten Jahren und Jahrzehnten zahlreiche Studien durchgeführt (vgl. z.B. Noll u.a. 2006). Dabei konnte die konvergente wie auch die prädiktive Validität sowohl in Untersuchungen aus

7 Das Static-2002 besteht aus 12 Items und ist – ebenso wie die Revision des Static-99 – als eine Weiterentwicklung des ursprünglichen Static-99 konzipiert. Das Hauptanliegen war, die Interraterreliabilität und prädiktive Validität des Instruments zu erhöhen. Erste Ergebnisse geben Hinweise darauf, dass das Vorhaben der Autoren gelungen ist (Hanson & Thornton 2003), wobei weitere Validierungsstudien notwendig sind (vgl. Nedopil 2005).

dem angloamerikanischen Sprachraum (z.B. *Hanson & Thornton* 1999; *Barbaree, Seto, Langton & Peacock* 2001; *Nunes, Firestone, Bradford, Greenberg & Broom* 2002; *Harris, Quinsey, Lalumière, Boer & Lang* 2003) als auch aus Europa (z.B. *Sjöstedt & Långstrom* 2001; *Vogel, Ruitter, Beek & Mead* 2004; *Stadtland, Hollweg, Kleindienst, Dietl, Reich & Nedopil* 2005) weitgehend bestätigt werden. Obwohl das Static-99 mittlerweile auch außerhalb der USA und Kanada in immer mehr Ländern angewendet wird⁸, handelt es sich beim Großteil der publizierten Validierungsstudien um Untersuchungen aus dem angloamerikanischen Raum, wodurch die Generalisierung dieser vielversprechenden Ergebnisse für andere Sprach- und Kulturräume erschwert wird. Deshalb ist bei der Anwendung des Static-99 im deutschsprachigen Raum zu beachten, dass nach wie vor ein Mangel an Validierungsstudien mit deutschen, österreichischen oder schweizerischen Probanden festzustellen ist, weshalb das Instrument aktuell nur mit Einschränkungen zu empfehlen ist (vgl. *Noll u.a.* 2006).

Mit der vorliegenden Arbeit soll ein weiterer Beitrag zur Beantwortung der Frage geliefert werden, ob bzw. inwieweit das Static-99 auch im deutschsprachigen Raum für die kriminalprognostische Einschätzung bei Sexualstraftätern geeignet ist. Um dies zu überprüfen, wurde zunächst die von *Harris, Phenix, Hanson & Thornton* (2003) revidierte Version des Instruments ins Deutsche übersetzt und an die rechtlichen Rahmenbedingungen in Deutschland und Österreich angepasst (vgl. *Rettenberger & Eher* 2006). Anschließend wurde die konvergente und die prädiktive Validität dieser adaptierten Version des Static-99 anhand von Stichproben aus Österreich überprüft. Des Weiteren wurde der Frage nach der Beurteilerübereinstimmung (Interraterreliabilität) nachgegangen, um die Qualität der Operationalisierungen und Kodierungsregeln des Manuals bzw. dessen deutschsprachiger Adaption zu überprüfen. Um den Nachweis der konvergenten Validität zu erbringen, wurden die Ergebnisse verschiedener Instrumente, die vorgeben, das gleiche oder ähnliche Konstrukt zu messen, miteinander korreliert. In der vorliegenden Studie wurden die Ergebnisse der adaptierten Version des revidierten Static-99 mit den Ergebnissen von drei anderen etablierten Prognoseinstrumenten korreliert. Dabei handelt es sich um das *Sexual Violence Risk-20* Schema (SVR-20; *Boer, Hart, Kropp & Webster* 1997), der *Sex Offender Risk Appraisal Guide* (SORAG; *Quinsey, Harris, Rice & Cormier* 1998) und die *Psychopathy Checklist-Revised* (PCL-R; *Hare* 1991)⁹. Die Auswahl der Vergleichsinstrumente basierte auf der Überlegung, dass es sich dabei um international anerkannte Verfahren handelt (z.B. *Müller-Isberner, Cabeza & Eucker* 2000; *Habermeyer & Herpertz* 2005), für die bereits eine Vielzahl von Validierungsstudien existieren (z.B. *Barbaree et al.* 2001; *Stadtland et al.* 2005), deren Ergebnisse ihre Anwendung rechtfertigen. Es wird davon ausgegangen, dass das Static-99 hoch mit diesen drei Instrumenten korreliert und die Ergebnisse insgesamt dem internationalen Erfahrungsstand bezüglich der konvergenten Validität dieser Instrumente entsprechen.

Im Mittelpunkt der vorliegenden Untersuchung stand die Überprüfung der prädiktiven Validität, d.h. es wurde der Frage nachgegangen, ob anhand des Gesamtscores des Static-99 weitere strafrechtsrelevante Handlungen von bereits mindestens einmal wegen sexuell motivierter Straftaten verurteilter Personen prognostiziert werden können. Das Konstrukt

8 *Giovannangeli, Cornet & Mormont* (2000) kamen in ihrer Studie zu dem Ergebnis, dass das Static-99 in Europa neben Großbritannien regelmäßig in Finnland, Irland und Schweden angewendet wird; des Weiteren findet es im asiatischen Raum in Taiwan seine Anwendung (*Hanson*, persönliche Mitteilung, 19. 2. 2001).

9 Auf eine ausführliche Darstellung der Vergleichsinstrumente wurde im Rahmen dieses Beitrags verzichtet (ausführlich z.B. bei *Nedopil* 2005).

Rückfälligkeit wurde anhand der folgenden Differenzierung erfasst: (a) Die *allgemeine Rückfälligkeit* bezieht sich auf sämtliche neuerliche Eintragungen im Strafregisterauszug der Republik Österreich, ohne Einschränkung auf bestimmte Delikttypen. (b) Bei der *einschlägigen Rückfälligkeit* sind ausschließlich sexuell motivierte Straftaten von Interesse. (c) Anhand des letzten Rückfallkriteriums, der *gewalttätigen Rückfälligkeit*, wurden Straftaten erfasst, die u.a. eine gewalttätige Komponente besaßen. Dabei diente die im Manual der revidierten Version des Static-99 verwendete Definition von Gewalt als Orientierung: Sämtliche sexuell motivierten Straftaten (»Hands-on«-Delikte) wurden gleichzeitig auch als Gewaltdelikte interpretiert (vgl. *Harris, Phenix, Hanson & Thornton* 2003).

2. Methode

Zunächst wurde für alle an der Untersuchung teilnehmenden Probanden retrospektiv auf der Basis eines detaillierten und ausführlichen Aktenstudiums das Static-99 (*Hanson & Thornton* 1999) kodiert, wobei die deutschsprachige Adaption (*Rettenberger & Eher* 2006) der revidierten Version des Manuals (*Harris, Phenix, Hanson & Thornton* 2003) die theoretische Grundlage hierfür darstellte. Die Bewertung der drei Vergleichsinstrumente für die Überprüfung der konvergenten Validität – PCL-R, SVR-20 und SORAG – wurde von forensisch erfahrenen und im Umgang mit diesen Prognoseinstrumenten theoretisch und praktisch geschulten Mitarbeitern der Zentralen Dokumentations- und Koordinationsstelle für Sexualstraftäter (Justizanstalt Wien-Mittersteig, Außenstelle Floridsdorf) durchgeführt. Im Hinblick auf die Frage nach der prädiktiven Validität erfolgte die Auswertung der Strafregisterauszüge vom Zweitautor. In beiden Fällen waren die an der Datenerhebung beteiligten Personen blind in Bezug auf andere relevante Ergebnisse. Bei der Datenauswertung kamen – neben Produkt-Moment-Korrelationen – die aus den ROC-Analysen (Analysen der *Receiver Operation Characteristics*) resultierenden AUC-Werte (Maßzahlen für die *Area Under Curve*) zur Anwendung. Die ROC-Analysen gehen auf die Signalentdeckungstheorie zurück (z.B. *Egan* 1975) und weisen gegenüber anderen Verfahren vor allem deshalb Vorteile auf, weil sie weitgehend unabhängig von der Basisrate sind (*Hanley & McNeil* 1982). Die ROC-Kurven stellen für jeden möglichen Ausprägungsgrad des Prognoseinstruments den Anteil der jeweils korrekt positiv klassifizierten Probanden (in der vorliegenden Arbeit folglich die richtig identifizierten rückfälligen Sexualstraftäter) und die zugehörige Rate falsch-positiver Probanden (in diesem Fall diejenigen Straftäter, die trotz schlechter Prognose nicht rückfällig wurden) graphisch dar. Die AUC-Maße, die auf der dabei ermittelten und quantifizierten Fläche unter der Kurve basieren, ist ein Validierungsmaß für das Instrument, wobei eine AUC von 0,5 genau der Zufallswahrscheinlichkeit entspricht, während eine AUC von 1,0 eine in jedem Untersuchungsfall korrekte Prognosestellung bedeuten würde. *Dable* (2005) gibt unter Bezugnahme auf die von *Cohen* (1992) formulierten Kriterien für die Interpretation der AUC-Maße an, dass nach allgemeinen Standards Werte von AUC > .70 für Prognoseinstrumente als gut, Werte zwischen .65 und .70 als moderat bezeichnet werden können.

Um die Frage nach der Interraterreliabilität des Static-99 zu beantworten, wurde ein circa dreistündiges Rater-Training durchgeführt. Insgesamt gingen 27 Fälle, die jeweils von vier Ratern unabhängig voneinander bewertet wurden, mit in die Überprüfung der Beurteilerübereinstimmung ein. Die zur Überprüfung der Beurteilerübereinstimmung erhobenen Daten wurden anhand von Intraklassen-Korrelationen (ICC) und Produkt-Moment-Korrelationen ausgewertet.

Die genannten Fragestellungen wurden anhand zweier unterschiedlicher Stichproben überprüft, wobei es sich bei den Probanden ausschließlich um österreichische Staatsbürger handelte, die in den Jahren zwischen 1968 und 2002 in Österreich aufgrund der Begehung

sexuell motivierter Straftaten rechtskräftig verurteilt wurden. Bezüglich der Fragestellung nach der konvergenten Validität der deutschsprachigen Adaption der revidierten Version des Static-99 wurden $N = 202$ Probanden (Stichprobe I) aus der Datenbank der Zentralen Dokumentations- und Koordinationsstelle für Sexualstraftäter ausgewählt, die in den Jahren 2002 bzw. 2003 dort begutachtet wurden. Bei den 27 Probanden, deren Daten in die Überprüfung der Interraterreliabilität einfließen, handelte es sich um eine Zufallsstichprobe aus den 202 Probanden der Stichprobe I. Das durchschnittliche Alter der 202 Probanden der Stichprobe I betrug $M = 40,41$ Jahre ($SD = 11,73$ Jahre), wobei der jüngste Proband 18, der älteste 69 Jahre alt war. Die Index-Delikte dieser Probanden sind der *Tabelle 2* zu entnehmen. Die aus dem Index-Delikt resultierende Freiheitsstrafe betrug im Durchschnitt 72,44 Monate ($SD = 153,27$ Monate). 66,3 % ($n = 134$) der Probanden waren vorbestraft, wobei 28,7 % ($n = 58$) einschlägige Vorstrafen aufwiesen, d.h. bei 71,3 % ($n = 144$) handelte es sich beim Index-Delikt um die erste sexuell motivierte Straftat.

Tabelle 2 Absolute und relative Häufigkeiten des Index-Delikts der Stichprobe I (konvergente Validität)

Deliktkategorie	absolute Häufigkeit	relative Häufigkeit in %
Vergewaltigungsdelikte	91	45,1
pädosexuelle Delikte	92	45,5
»Hands-off«-Delikte	7	3,5
sexuell motivierte Tötungsdelikte	12	5,9
Gesamt	202	100,0

Bezüglich der Fragestellung nach der prädiktiven Validität wurden alle rechtskräftig verurteilten Sexualstraftäter ($N = 81$) aus dem Archiv des Forensisch Therapeutischen Zentrums Wien (FTZW; Nachbetreuungsumbulanz) ausgewählt, die zum Zeitpunkt der Datenerhebung mindestens drei Jahre aus der Haft entlassen (»on risk«) waren (Stichprobe II). Bei den 81 Probanden der Stichprobe II betrug das durchschnittliche Alter zum Zeitpunkt der Entlassung $M = 36,43$ Jahre ($SD = 10,78$ Jahre), wobei der jüngste Proband 18, der älteste 71 Jahre alt war. Die Index-Delikte dieser Probanden sind der *Tabelle 3* zu entnehmen. Die aus dem Index-Delikt resultierende Freiheitsstrafe betrug im Durchschnitt 52,01 Monate ($SD = 144,48$ Monate). 71,6 % ($n = 58$) der Probanden waren vorbestraft, wobei 44,4 % ($n = 36$) einschlägige Vorstrafen aufwiesen, d.h. bei 55,6 % ($n = 45$) handelte es sich beim Index-Delikt um die erste sexuell motivierte Straftat. Des Weiteren wurde bei 38,3 % ($n = 31$) der Probanden beim Index-Delikt eine Maßregel der Besserung, Behandlung und Sicherung nach § 21 StGB Österreichisches Strafgesetzbuch angeordnet. Durchschnittlich lag der Entlassungszeitpunkt der Untersuchungsteilnehmer $M = 7,72$ Jahre ($SD = 3,42$ Jahre) zurück.

Tabelle 3 Absolute und relative Häufigkeiten des Index-Delikts der Stichprobe II (prädiktive Validität)

Deliktkategorie	absolute Häufigkeit	relative Häufigkeit in %
Vergewaltigungsdelikte	15	18,5
pädosexuelle Delikte	55	67,9
»Hands-off«-Delikte	4	4,9
sexuell motivierte Tötungsdelikte	7	8,7
Gesamt	81	100,0

3. Ergebnisse

3.1 Interraterreliabilität

Der Intraclass-Korrelationskoeffizient (ICC) betrug .90 (Single Measure; 95 %-KI = .83 – .95) bzw. .97 (Average Measure; 95 %-KI: .95 – .99). Um die Vergleichbarkeit mit anderen Studien zu gewährleisten, wurden zusätzlich Produkt-Moment-Korrelationen zwischen den vier Ratern berechnet und anschließend der Mittelwert daraus gebildet. Die einzelnen Korrelationen zwischen jeweils zwei der an der Untersuchung beteiligten Beurteiler betragen zwischen .83 und .97 und waren alle hoch signifikant ($p < .001$). Der Mittelwert dieser Korrelationen betrug .92.

3.2 Konvergente Validität

Der Mittelwert der Gesamtscores des Static-99 betrug $M = 3,66$ ($SD = 2,41$). Bei 191 Probanden waren Informationen über den Gesamtscore des SVR-20 verfügbar, wobei der Mittelwert bei $M = 19,40$ ($SD = 7,34$) lag. Das SORAG wurde bei 193 Probanden angewendet ($M = 7,83$; $SD = 14,12$). Der Mittelwert der Gesamtscores der an der Bewertung beteiligten 198 Probanden betrug bei der PCL-R $M = 18,88$ ($SD = 9,21$). Die Korrelationen zwischen Static-99 und den drei Vergleichsinstrumenten betrugen für die PCL-R .46, für das SVR-20 .57 und für das SORAG .65 und waren allesamt hoch signifikant ($p < .001$).

3.3 Prädiktive Validität

Bei 51,9 % ($n = 42$) der entlassenen Straftäter wurden neuerliche Eintragungen im Strafregisterauszug festgestellt. 25,9 % ($n = 21$) der Probanden wurden einschlägig rückfällig, d.h. sie begingen nach ihrer Entlassung erneut eine sexuell motivierte Straftat. 8,6 % ($n = 7$) der Probanden begingen ein nicht-sexuell motiviertes Gewaltdelikt und 13,6 % ($n = 11$) verwirklichten einen anderen Straftatbestand nach StGB, d.h. sie begingen eine Straftat, die nicht sexuell und/oder gewalttätig motiviert war. Die Rückfallraten entsprechen damit dem nationalen und internationalen Erfahrungsstand (vgl. z.B. *Hanson & Thornton* 1999; *Pritt-witz* 2000). Gemäß dem von *Harris, Phenix, Hanson & Thornton* (2003) revidierten Manual werden im Sinne des Static-99 auch sexuell motivierte »Hands-on«-Delikte als gewalttätige Rückfälle definiert. Demnach müssen bei diesem Rückfallkriterium einschlägige und gewalttätige Rückfälle zusammengefasst werden, um die gewalttätige Rückfälligkeit zu quantifizieren. Dieser Definition zufolge wurden 34,5 % ($n = 28$) der Probanden gewalttätig rückfällig. Bei diesen 81 Probanden betrug der Mittelwert des Static-99 $M = 4,31$ ($SD = 2,45$). Die Produkt-Moment-Korrelation zwischen dem Gesamtscore des Static-99 und dem allgemeinen Rückfallkriterium betrug $r = .41$ und war hoch signifikant ($p < .001$). Die Korrelation zwischen dem Gesamtscore des Static-99 und der sexuell motivierten bzw. einschlägigen Rückfälligkeit betrug .35 ($p < .01$), während sie zwischen dem Gesamtscore des Static-99 und der gewalttätigen Rückfälligkeit bei .41 ($p < .001$) lag. Die AUC-Werte für die drei Rückfallkriterien ergaben für die allgemeine Rückfälligkeit $AUC = .74$ ($p < .001$), für die einschlägige Rückfälligkeit $AUC = .74$ ($p < .001$) und für die gewalttätige Rückfälligkeit $AUC = .76$ ($p < .001$).

4. Diskussion

Aus den Ergebnissen wird ersichtlich, dass mit der vorliegenden deutschsprachigen Adaption des Static-99 eine zufriedenstellende Interraterreliabilität erreicht werden kann, die mit der Originalversion vergleichbar ist (vgl. z.B. *Harris, Quinsey, Lalumière, Boer &*

Lang 2003). Des Weiteren können die dargestellten Ergebnisse als Beleg für die konvergente Validität des Instruments interpretiert werden. Das Static-99 korreliert hoch signifikant (jeweils $p < .001$) mit den anderen hier verwendeten Prognoseinstrumenten PCL-R, SVR-20 und SORAG, wobei die hier berichteten Resultate vergleichbar sind mit denen angloamerikanischer Studien (vgl. z.B. *Barbaree et al.* 2001; *Harris, Quinsey, Lalumière, Boer & Lang* 2003).

Entscheidend für die Vorhersagegüte eines Prognoseinstrumentes ist allerdings der Nachweis der prädiktiven Validität (vgl. z.B. *Witt* 2000). In dieser Untersuchung betrug der durchschnittliche Beobachtungszeitraum, d.h. die durchschnittliche Zeitspanne möglicher Rückfallstätigkeit (Zeit zwischen Entlassung nach dem Index-Delikt und dem Untersuchungszeitpunkt) $M = 7,72$ Jahre ($SD = 3,42$ Jahre) und lag damit zum Teil deutlich über dem Beobachtungszeitraum anderer Static-99-Validierungsstudien (vgl. z.B. *Sjöstedt & Långström* 2001), wobei manche Untersuchungen vergleichbare (vgl. z.B. *Stadtland et al.* 2005) oder längere Beobachtungszeiträume (vgl. z.B. *Vogel et al.* 2004) angaben.

Anhand des Kriteriums der *allgemeinen Rückfälligkeit* wurde erfasst, inwieweit die Anwendung der vorliegenden Adaption des revidierten Static-99 Aussagen darüber erlaubt, ob ein Sexualstraftäter wegen der Verwirklichung eines beliebigen Straftatbestandes des StGB erneut verurteilt wird. Es konnte ein hoch signifikanter Zusammenhang zwischen dem Gesamtscore des Static-99 und allgemeiner Rückfälligkeit ermittelt werden ($r = .41$; $p < .001$). Die aus der Analyse der ROC-Kurven hervorgehenden AUC-Werte, die in der vorliegenden Arbeit ermittelt wurden, können *Dahle* (2005) zufolge als Beleg für die gute Vorhersageleistung des Instruments in Bezug auf die prädiktive Validität bezeichnet werden ($AUC = .74$; $p < .001$). Neben der vorliegenden Untersuchung überprüften lediglich zwei weitere Validierungsstudien die Vorhersageleistung des Static-99 bei der allgemeinen Rückfälligkeit von Sexualstraftätern: Sowohl *Barbaree et al.* (2001) als auch *Vogel et al.* (2004) ermittelten jedoch zum Teil deutlich niedrigere Validitätskoeffizienten bei diesem Rückfallkriterium.

Der Zusammenhang zwischen dem Gesamtscore des Static-99 und *einschlägiger Rückfälligkeit* war ebenfalls hoch signifikant ($r = .35$; $p < .001$). Des Weiteren betrug der AUC-Wert $AUC = .74$ ($p < .001$), was gemäß *Dahle* (2005) als ein gutes Ergebnis für die Vorhersageleistung des Instruments bezeichnet werden kann. Die hier berichteten Ergebnisse sind vergleichbar mit den Resultaten angloamerikanischer (z.B. *Hanson & Thornton* 1999; *Barbaree et al.* 2001; *Nunes et al.* 2002) und anderer europäischer (z.B. *Sjöstedt & Långström* 2001; *Vogel et al.* 2004; *Stadtland et al.* 2005) Studien. Mit Ausnahme der Arbeit von *Harris, Quinsey, Lalumière, Boer & Lang* (2003) konnten für die einschlägige Rückfälligkeit in vergleichbaren Untersuchungen durchweg moderate bis gute Werte für die prädiktive Validität ermittelt werden, die in der vorliegenden Studie bestätigt wurden.

Der dem Manual entnommenen Gewaltdefinition zufolge werden alle sexuell motivierten »Hands-on«-Delikte im Sinne des Static-99 auch als Gewaltdelikte interpretiert (vgl. *Hanson & Thornton* 1999; *Harris, Phenix, Hanson & Thornton* 2003). Auch bei der *gewalttätigen Rückfälligkeit* kann die Vorhersageleistung des Instruments gemäß *Dahle* (2005) als gut bezeichnet werden ($r = .41$, $p < .001$; $AUC = .76$, $p < .001$). Die bereits genannten Vergleichsstudien ermittelten ebenfalls durchweg moderate bis gute Vorhersageleistungen für die Kategorie »gewalttätige Rückfälligkeit«, so dass die in der vorliegenden Arbeit ermittelten Ergebnisse weitgehend mit dem internationalen Erfahrungsstand übereinstimmen.

Nachdem die moderaten bis guten Vorhersageleistungen, die in angloamerikanischen Untersuchungen ermittelt wurden, auch anhand von Stichproben aus Schweden (vgl. *Sjöstedt & Långström* 2001), aus den Niederlanden (vgl. *Vogel et al.* 2004) und aus Deutschland (vgl. *Stadtland et al.* 2005) bestätigt wurden, lieferte die vorliegende Untersuchung einen Beleg für die prädiktive Validität einer publizierten deutschsprachigen Version des Static-99.

Wenngleich nach wie vor weitere Untersuchungen notwendig sind, um die positiven Resultate der vorliegenden Validierungsstudie zu untermauern, erscheint die Anwendung der hier vorliegenden deutschsprachigen Version des Static-99 – in Anbetracht der Forderung nach validen und reliablen Prognosemethoden bei der kriminalprognostischen Begutachtung von Sexualstraf Tätern – als gerechtfertigt und sinnvoll.

Wozu die vorliegende Untersuchung nicht in der Lage war, ist die von *Noll u.a.* (2006) angeregte Validierung an unterschiedlichen Subgruppen von Sexualstraf Tätern. Nach *Bartosh, Garby, Lewis & Gray* (2003) zeigt sich die englischsprachige Originalversion für die Vorhersage von sexueller und gewalttätiger Rückfälligkeit bei den unterschiedlichen Gruppen von Indexdelikt Tätern (Vergewaltigter, extra- und innerfamiliäre Kindesmissbraucher) in ähnlicher Weise bedeutsam. Kritisch sollte auch – insbesondere an die Adresse nicht professioneller Entscheidungsträger gerichtet – angemerkt werden, dass das Static-99 keinesfalls eine fachlich fundamentierte Prognosebegutachtung ersetzen kann, sondern – wie oben schon ausführlich dargestellt – eine gruppenspezifische Basisrate der Rückfallwahrscheinlichkeit abgeben kann. Das Instrument – aufgrund seiner vordergründigen einfachen Handhabung – lädt dazu ein, auch Laien glauben zu lassen, sie könnten Sexualstraf Täter prognostisch einschätzen. Vor einer Handhabung des Instruments durch Laien wird aus dem Grund ausdrücklich gewarnt.

Nicht zuletzt muss auch vor einer kritiklosen Anwendung des Instruments durch erfahrene Forensiker gewarnt werden, leitet sich doch unser ganzes Wissen über die Vorhersagekraft dieses Instruments von *Wiederverurteilungsraten*, nicht aber von *Delikt-Wiederbegehungsraten* ab. Das Static-99 verfügt über mehrere Items, die den heutigen Wissensstand über die Täter-Opfer-Beziehung und dessen glaublichen Zusammenhang mit »Rückfälligkeit« (in Wahrheit aber: *Wiederverurteilung*) abbilden. Demzufolge verfügen innerfamiliäre Täter oftmals über ein sehr geringes Rückfallrisiko, was nicht selten der klinischen Einschätzung und der allgemeinen forensischen Erfahrung widerspricht (*Studer, Clelland, Aylwin, Reddon & Monro* 2000; *Eher* 2006). Hält man sich dann vor Augen, dass das Static-99 eben an dem Kriterium »Wiederverurteilung« validiert ist, und ist man sich andererseits bewusst, wie leicht es innerfamiliären Tätern gelingen mag, ihre Taten über viele Jahre geheimzuhalten (also trotz eines Rückfalls nicht »verurteilt« zu werden), muss man insbesondere vor der kritiklosen Anwendung des Instruments bei innerfamiliären Tätern warnen.

Summary

The Static-99 is an actuarial risk assessment tool for evaluating the risk of sexual and violent recidivism in sexual offenders. It was developed in 1999 by *Hanson* (Canada) and *Thornton* (Great Britain) and is in the mean time regularly used for risk assessment in North America and some countries in Europe. In this retrospective study, interrater reliability, concurrent validity and predictive validity of a German adaptation of the by *Harris, Phenix, Hanson & Thornton* in 2003 revised version of the Static-99 are presented. The instrument was validated from file information of Austrian sexual offenders, who were convicted between 1968 and 2002. Interrater reliability and concurrent validity were good. The instrument showed good predictive validity for general ($r = .41$, $AUC = .74$), sexual ($r = .35$, $AUC = .74$) and violent ($r = .41$, $AUC = .76$) recidivism. These results support the utility of this German version of the revised version of the Static-99 in improving risk assessment of sexual offenders.

Keywords: Static-99, risk assessment, sexual offender, validity, recidivism

Literatur

Barbaree, H.E., Seto, M.C., Langton, C.M. & Peacock, E.J. (2001). Evaluating the predictive accuracy of six risk assessment instruments for adult sex offenders. *Criminal Justice and Behavior* 28, 490–521. – *Bartosh, D.L., Garby, T., Lewis, D. & Gray, S.* (2003). Differences in the predictive validity of actuarial risk assessments in relation to sex offender type. *International Journal of Offender Therapy and*

- Comparative Criminology 47, 422–438. – *Boer, D.P., Hart, S.D., Kropp, P.R. & Webster, C.D.* (1997). Manual for the Sexual Violence Risk-20: Professional Guidelines for Assessing Risk of Sexual Violence. Vancouver/CA: The Mental Health, Law & Policy Institute. – *Cohen, J.* (1992). A power primer. Psychological Bulletin 112, 155–159. – *Dable, K.P.* (1997). Kriminalprognosen im Strafrecht. Psychologische Aspekte individueller Verhaltensvorhersagen, in: M. Steller & R. Volbert (Hrsg.), Psychologie im Strafverfahren, 119–140. Bern. – *Dable, K.P.* (2000). Psychologische Begutachtung zur Kriminalprognose, in: H.-L. Kröber & M. Steller (Hrsg.), Psychologische Begutachtung im Strafverfahren, 77–111. Darmstadt. – *Dable, K.P.* (2005). Psychologische Kriminalprognose. Herbolzheim. – *Egan, J.P.* (1975). Signal Detection Theory and ROC Analysis. New York. – *Eher, R.* (2006). Reconsidering risk for reoffense in intrafamilial child molesters: new aspects on clinical and criminological issues. Sexual Offender Treatment 1/2; www.sexual-offender-treatment.org. – *Endres, J.* (2000). Die Kriminalprognose im Strafvollzug: Grundlagen, Methoden und Probleme der Vorhersage von Straftaten. ZfStrVo 49, 67–83. – *Foerster, K.* (2000). Gefährlichkeitsprognose gemäß § 454 StPO, in: R. Egg (Hrsg.), Behandlung von Sexualstraf Tätern im Justizvollzug, 293–305. Wiesbaden: KrimZ. – *Giovannangeli, D., Cornet, J.P. & Mormont, C.* (2000). Étude comparative dans les quinze pays de l'Union européenne: les méthodes et les techniques d'évaluation de la dangerosité et du risque de récidive des personnes présumées ou avérées délinquants sexuels. Rapport de recherche n°99/STOPII/028. Commission Européenne: unveröffentlicht. – *Grubin, D.* (1998). Sex Offending Against Children: Understanding the Risk. Police Research Series, Paper 99. London: Home Office. – *Habermeyer, E. & Herpertz, S.* (2005). Bedeutung des »psychopathy«-Konzepts von Hare für Sexualstrafaten, in: D. Schläfke, F. Häßler & J.M. Fegert (Hrsg.), Sexualstrafaten, 131–141. Stuttgart. – *Hanley, J.A. & McNeil, B.J.* (1982). The meaning and use of the area under a Receiver Operating Characteristic (ROC) Curve. Radiology 143, 29–36. – *Hanson, R.K. & Bussière, M.T.* (1996). Predictors of Sexual Offender Recidivism: A Meta-Analysis. User Report 1996–04. Ottawa: Department of the Solicitor General of Canada. – *Hanson, R.K.* (1997). The Development of a Brief Actuarial Risk Scale for Sexual Offense Recidivism. User Report 1997–04. Ottawa: Department of the Solicitor General of Canada. – *Hanson, R.K. & Bussière, M.T.* (1998). Predicting relapse: a meta-analysis of sex offender recidivism studies. Journal of Consulting and Clinical Psychology 66, 348–362. – *Hanson, R.K. & Thornton, D.* (1999). Static-99: Improving Actuarial Risk Assessments for Sex Offenders. User Report 1999–02. Ottawa: Department of the Solicitor General of Canada. – *Hanson, R.K. & Thornton, D.* (2003). Notes on the Development of Static-2002. User Report 2003–01. Ottawa: Department of the Solicitor General of Canada. – *Hare, R.D.* (1991). Manual for the Psychopathy-Checklist-Revised. Toronto: Multi Health Systems. – *Harris, A., Phenix, A., Hanson, R.K. & Thornton, D.* (2003). Static-99 Coding Rules Revised-2003. Ottawa: Department of the Solicitor General of Canada. – *Harris, G.T., Quinsey, V., Lalumière, M.L., Boer, D. & Lang, C.* (2003). A multisite comparison of actuarial risk instruments for sex offenders. Psychological Assessment 15, 413–425. – *König, C., Schnoor, K., Auer, U., Rebernik, E., Schläfke, D. & Fegert, J.M.* (2005). Modellprojekt forensisch-psychiatrischer Gutachtertätigkeit in Mecklenburg-Vorpommern bei Sexualdelikten, in: D. Schläfke, F. Häßler & J.M. Fegert (Hrsg.), Sexualstrafaten, 57–76. Stuttgart. – *Müller-Isberner, R., Cabeza, S.G. & Eucker, S.* (2000). Die Vorhersage sexueller Gewalttaten mit dem SVR-20. Haina: Institut für Forensische Psychiatrie. – *Nedopil, N.* (1995). Neues zur Kriminalprognose. Gibt es das?, in: D. Dölling (Hrsg.), Die Täter-Individualprognose, 83–95. Heidelberg. – *Nedopil, N.* (2000). Forensische Psychiatrie. Stuttgart. – *Nedopil, N.* (2005). Prognosen in der Forensischen Psychiatrie. Ein Handbuch für die Praxis. Lengerich. – *Noll, T., Endrass, J., Rossegger, A. & Urbaniok, F.* (2006). Die Risikokalkulation für die Begehung von Gewalt- und Sexualstrafaten mit Hilfe des Static-99. MschrKrim 89, 24–33. – *Nowara, S.* (2001). Die Beurteilung der Gefährlichkeit von Straftätern, in: G. Rehn, B. Wischka, F. Lösel & M. Walter (Hrsg.), Behandlung »gefährlicher« Straftäter, 104–110. Herbolzheim. – *Nunes, K.L., Firestone, P., Bradford, J.M., Greenberg, D.M. & Broom, I.* (2002). A comparison of modified versions of the Static-99 and the Sex Offender Risk Appraisal Guide (SORAG). Sexual Abuse: A Journal of Research and Treatment 14, 253–269. – *Prittowitz, C.* (2000). Reform des Sexualstrafrechts - Verantwortbare Risiken?, in: J.M. Fegert & F. Häßler (Hrsg.), Qualität forensischer Begutachtung, insbesondere bei Jugenddelinquenz und Sexualstrafaten, 107–135. Herbolzheim. – *Quinsey, V.L., Harris, G.T., Rice, M.E. & Cormier, C.* (1998). Violent Offenders: Appraising and Managing Risk. Washington/DC: American Psychological Association. – *Rasch, W.* (1999). Forensische Psychiatrie. Stuttgart. – *Rehder, U.* (2004). Legalprognose bei Tätern, die wegen sexuellen Missbrauchs verurteilt wurden, in: W. Körner & A. Lenz (Hrsg.), Sexueller Missbrauch. Band I: Grundlagen und Konzepte, 581–599. Göttingen. – *Ret-*

tenberger, M. & Eher, R. (2006). Die revidierten Kodierungsrichtlinien des Static-99 [bei bestehender IATSO-Mitgliedschaft verfügbar unter www.iatso.org/iatso_user_tool/publications]. – *Sjöstedt, G. & Långstrom, N.* (2001). Actuarial assessment of sex offender recidivism risk: a cross-validation of the RRASOR and the Static-99 in Sweden. *Law and Human Behavior* 25, 629–645. – *Stadtland, C., Hollweg, M., Kleindienst, N., Dietl, J., Reich, U. & Nedopil, N.* (2005). Risk assessment and prediction of violent and sexual recidivism in sex offenders: long-term predictive validity of four risk assessment instruments. *Journal of Forensic Psychiatry and Psychology* 16, 92–108. – *Studer, L.H., Clelland, S.R., Aylwin, A.S., Reddon, J.R. & Monro, A.* (2000). Rethinking risk assessment for incest offenders. *International Journal of Law and Psychiatry* 23, 15–22. – *Vogel, V. d., Ruiter, C. d., Beek, D. v. & Mead, G.* (2004). Predictive validity of the SVR-20 and Static-99 in a Dutch sample of treated sex offenders. *Law and Human Behavior* 28, 235–251. – *Witt, P.H.* (2000). A practitioner's view of risk assessment: the HCR-20 and SVR-20. *Behavioral Sciences and the Law* 18, 791–798.

(Anshr. d. Verf.: Cand. Psych. *Martin Rettenberger*, PD Dr. med. *Reinhard Eher*, Justizanstalt Wien-Mittersteig, Außenstelle Floridsdorf, Gerichtsgasse 6, A-1210 Wien und Universitätsklinikum Ulm, Abteilung für Psychosomatische Medizin und Psychotherapie, Sektion Forensische Psychotherapie; reinhard.eher@justiz.gv.at)

Actuarial Assessment of Sex Offender Recidivism Risk: A Validation of the German version of the Static–99¹

Martin Rettenberger¹, Reinhard Eher²

¹Free University of Berlin, Germany

²Federal Documentation Centre for Sex Offenders, Vienna, Austria

[Sexual Offender Treatment, Volume 1 (2006), Issue 3]

Abstract

The Static–99 and the RRASOR are actuarial risk assessment tools for evaluating the risk of sexual and violent recidivism in sexual offenders. The Static–99 was developed in 1999 by Karl R. Hanson (Canada) and David Thornton (Great Britain) and is in the mean time regularly used for risk assessment in North America and some countries in Europe. The RRASOR can be described as a predecessor of the Static–99 and was published by Hanson in 1997. At first we translated the revised version of the Static–99 (Harris, Phenix, Hanson & Thornton, 2003) and adapted the instrument and the manual to the forensic context in Germany and Austria (Rettenberger & Eher, 2006). In this retrospective study, interrater reliability and concurrent validity of the RRASOR and of the German adaption of the Static–99 is presented. Furthermore we evaluated the predictive accuracy of the Static–99 and the RRASOR and compared their results. The instruments were validated from file information of Austrian sexual offenders, who were convicted between 1968 and 2002. Both the Static–99 and the RRASOR had good interrater reliability and concurrent validity. The Static–99 showed good predictive validity for general ($r = .41$, $AUC = .74$), sexual ($r = .35$, $AUC = .74$) and violent ($r = .41$, $AUC = .76$) recidivism, whereas the predictive accuracy of the RRASOR was moderate for general ($r = .29$, $AUC = .66$), sexual ($r = .30$, $AUC = .68$) and violent ($r = .28$, $AUC = .67$) recidivism. The Static–99 exhibited a higher accuracy for the prediction of sexual offender recidivism. Although further validation studies on German–speaking populations of sex offenders are necessary, these results support the utility of the German version of the revised version of the Static–99 in improving risk assessment of sexual offenders.

Key words: Static–99, RRASOR, risk assessment, sexual offenders, actuarial risk assessment, validity, recidivism

Introduction

In recent years there has been a growing interest in validated risk–assessment tools for sex offenders in the German–speaking part of Europe (Noll, Endrass, Rossegger, & Urbaniok, 2006). Compared to North America – where actuarial assessment instruments are regarded as “state of the art” – there are only very few studies in Germany or Austria that have dealt with these instruments. Although the results of a number of studies from North America support the utility of the actuarial approach to risk assessment of sexual offenders (e.g. Barbaree, Seto, Langton, & Peacock, 2001), validation studies on German–speaking populations are necessary before the use of these instruments can be recommended in Germany and Austria. In this study we translated the revised version of the *Static–99* (Harris, Phenix, Hanson, & Thornton, 2003) and adapted the instrument and the manual to the forensic context in Germany and Austria (Rettenberger & Eher, 2006). The *Static–99* is a brief actuarial instrument for the assessment of risk for sexual and violent recidivism of adult sexual offenders. It was developed in 1999 by Karl R. Hanson (Canada) and David Thornton (Great Britain) and is in the meantime regularly used for risk assessment in North America and

SEXUAL OFFENDER TREATMENT: Martin Rettenberger, Reinhard Eher

some countries in Europe. The instrument is derived from a fusion of two previously developed risk assessment instruments, the *Rapid Risk Assessment of Sexual Offender Recidivism* (RRASOR; Hanson, 1997) and a shorter version of the *Structured Anchored Clinical Judgement* (SACJ–Min; Grubin, 1998) composed of ten mainly static risk factors (see Table 1). The individual risk factors of a sexual offender add up to a maximum total score of 12 that is subsequently translated into four risk categories: low (0–1), medium low (2–3), medium high (4–5), and high (6 or more).

Table 1: Items and coding form of the Static–99

<i>Item number</i>	<i>Risk Factor</i>	<i>Codes</i>		<i>Score</i>
1	Young	Aged 25 or older		0
		Aged 18–24.99		1
2	Ever lived with an intimate partner for at least two years?	Yes		0
		No		1
3	Index non–sexual violence Any Convictions?	No		0
		Yes		1
4	Prior non–sexual violence Any Convictions?	No		0
		Yes		1
5	Prior sex offenses	Charges	Conventions	
		None	None	0
		1–2	1	1
		3–5	2–3	2
		6+	4+	3
6	Prior sentencing dates	3 or less		0
		4 or more		1
7	Any convictions for non–contact sex offenses	No		0
		Yes		1
8	Any unrelated victims	No		0
		Yes		1
9	Any stranger victims	No		0
		Yes		1
10	Any male victims	No		0
		Yes		1

SEXUAL OFFENDER TREATMENT: Martin Rettenberger, Reinhard Eher

The RRASOR was published in 1997 by Hanson and consists of four items: Prior sex offenses, age at release (current age), victim gender, and relationship to victim. In a meta-analytic review Hanson and Bussière (1996) found that indicators of deviant sexual interests (e.g. the number of prior sex offenses) consistently predicted sexual recidivism. Drawing from these results, Hanson (1997) selected variables with a minimum correlation of .10 with sexual recidivism and developed the RRASOR. Across seven development samples, comprising a total of 2,592 sexual offenders, Hanson found that RRASOR scores had an average correlation of .27 and an average *Area Under the Curve of the Receiver Operating Characteristic* (AUC of the ROC) of .71 with sexual recidivism.

Hanson and Thornton (1999) tested the predictive accuracy of the RRASOR and the Static-99 using four different, mixed data sets of 1,301 sexual offenders (including child molesters and rapists from prison as well as secure forensic psychiatric settings) collected in Canada and the United Kingdom. Average follow-up times ranged from 4 to 23 years and sexual recidivism base rates were 15–35%. The RRASOR exhibited a predictive accuracy significantly higher than chance with an AUC of .68 for sexual and .64 for any violent (including sexual) recidivism. The same analyses for the Static-99 rendered an AUC of .71 for sexual and .69 for any violent (including sexual) recidivism. The authors emphasize the importance of replication, because if similar results are found across samples, then the instrument would appear robust and can be adopted in other sociocultural and legal contexts. Meanwhile results from studies with Canadian, Dutch, German, and Swedish samples of sex offenders were presented, whereas in the Dutch and German studies only the Static-99 was used. In the Canadian study published by Barbaree, Seto, Langton, and Peacock (2001) in a mixed sample of 215 child molesters and rapists assessed for a prison-based treatment program, the base-rate for registered sexual offenses involving physical contact was 9% after an average follow-up time of 4.5 years. The RRASOR and the Static-99 exhibited moderate and significant predictive accuracy (AUC of .77 for the RRASOR and of .70 for the Static-99) concerning sexual recidivism. In a large cross-validation sample of 1,400 sexual offenders followed for an average of 3.7 years in Sweden, Sjöstedt and Långström (2001) found that the RRASOR had a correlation of .22 and an AUC of .72 with sexual recidivism. For the Static-99 they reported an AUC of .76 for sexual recidivism and of .74 for nonsexually violent recidivism. In the retrospective study of de Vogel, de Ruiter, van Beek, and Mead (2004) with a Dutch sample of 122 sex offenders the base rate for sexual recidivism was 39%, for nonsexual violent offenses 46%, and for general offenses 74%. The average follow-up period was 11.7 years. The authors reported a moderate and significant predictive accuracy for their Dutch adaptation of the Static-99 for sexual recidivism: For Static-99 total score they found a correlation of .38 and an AUC of .71. However, for nonsexual violent and for general recidivism they did not find significant results: For violent reoffending the correlation was .11 and the AUC .54, for general reoffending the correlation was .13 and the AUC .57. In a German validation study of Stadtland, Hollweg, Kleindienst, Dietl, Reich, and Nedopil (2005) with 134 sex offenders and a mean follow-up time of 9 years the Static-99 showed an AUC of .71 (including treatment dropouts) and .72 (excluding treatment dropouts), respectively. In this study the base rate for sexual reoffending was 37% and for violent (but not sexual) reoffending 5%.

The primary aim of this retrospective study was the verification of the application of the revised version of the Static-99 to the forensic context in Austria. Therefore we present interrater reliability, concurrent validity, and predictive validity of this German adaptation of the Static-99 that was evaluated with an Austrian sample. Furthermore we calculated the interrater reliability, the concurrent validity, and the predictive accuracy of the RRASOR and compared the results of both assessment tools; we expect that our findings are in accordance with previous studies from North America and Europe, and the predictive accuracy of the Static-99 will be higher than the performance of the RRASOR.

Method

Subjects

In order to examine the predictive accuracy of the German version of the Static–99, we cross–validated the instrument using file information of 81 sexual offenders released either from prison (61.7%, $n = 50$) or from secure forensic psychiatric settings (38.3%, $n = 31$) in Austria (mean age = 36.43 years, $SD = 10.78$, range 18–71). The average follow–up time was 7.7 years ($SD = 3.42$, range 3–16). The group consisted of 58 child molesters, 15 rapists, 7 sexual murderers and one exhibitionist, all male. The mean sentence length for all convicted sex offenders was 52.01 months ($SD = 144.48$, range 2–240). 44.4% ($n = 36$) had a previous conviction for a sexual offense and 71.6% ($n = 58$) had any prior conviction. In order to examine the interrater reliability and the concurrent validity of the German version of the Static–99, we gathered information about 202 sex offenders from the database of the *Federal Documentation Centre of Sex Offender* (Justizanstalt Wien–Mittersteig, Außenstelle Floridsdorf) in Austria, who were still in prison at the time of data collection². The mean age of these 202 sexual offenders was 40.41 years ($SD = 11.73$, range 18–69). The group consisted of 98 child molesters, 91 rapists, 12 sexual murderers and one exhibitionist, all male. The mean sentence length for all convicted sex offenders was 72.44 months ($SD = 153.27$, range 2–240). 28.7% ($n = 58$) had a previous conviction for a sexual offense and 66.3% ($n = 134$) had any prior conviction.

Statistical Analysis

The interrater reliability of the Static–99 and the RRASOR was examined by Intraclass–Correlations (ICC) and average Pearson correlations between the raters. The predictive ability of the Static–99 and RRASOR was analyzed using Pearson correlations and the AUC of the ROC. To calculate the difference between the AUC values of the Static–99 and the RRASOR we used the program ROCKit version 0.9.1 (Metz, 1998). In order to establish the concurrent validity of the instruments, we calculated Pearson correlations between the Static–99 and the RRASOR on the one hand and the *Sexual Violence Risk–20* (SVR–20; Boer, Hart, Kropp, & Webster, 1997), the *Sex Offender Risk Appraisal Guide* (SORAG; Quinsey, Harris, Rice, & Cormier, 1998), and the *Psychopathy Checklist–Revised* (PCL–R; Hare, 1991) on the other hand. We chose the SVR–20, SORAG, and PCL–R, because there are already a number of validation studies (e.g. Quinsey, Rice, & Harris, 1995; Dempster & Hart, 2002; Barbaree et al., 2001; Harris, Rice, Quinsey, Lalumière, Boer, & Lang, 2003), so these instruments can be called internationally recognized risk–assessment tools for sex offenders (e.g. Müller–Isberner, Cabeza, & Eucker, 2000; Habermeyer & Herpertz, 2005). The SVR–20 is a structured clinical guideline designed for the assessment of risk for sexual violence in adult sex offenders. The instrument was developed from a thorough consideration of the empirical literature and the clinical expertise of a number of clinicians. The SVR–20 consists of 20 items, divided into three domains: Psychosocial adjustment, Sexual offenses, and Future plans, that have to be coded by an experienced forensic clinician. The SORAG is a modification of the *Violence Risk Appraisal Guide* (VRAG; Harris, Rice, & Quinsey, 1993), developed specifically to predict violent recidivism (which includes sexual offenses involving physical contact with the victim) among sex offenders. 10 of its 14 items are the same items as in the VRAG, so both instruments are highly correlated with each other, and both significantly predicted violent and sexual recidivism (Harris, Rice, Quinsey, Lalumière, Boer, & Lang, 2003). In contemporary research and clinical practice, Hare’s *Psychopathy Checklist–Revised* (PCL–R) is the psycho–diagnostic tool most commonly used to assess Psychopathy. The PCL–R is based on semistructured interviews and review of file information. Participants are assigned ratings of “0” (absent), “1” (some indication), or “2” (present) on each of the 20 PCL–R items, tapping characteristics such as impulsivity, irresponsibility, and callousness. Scale scores are obtained by summing the ratings, for a total possible score of 40. The conventional cutoff for making a diagnosis of Psychopathy in North America is 30, whereas in Europe the cutoff is mostly 25 (Hart & Hare, 1997; Hartmann, Hollweg, & Nedopil, 2001). In the development of the PCL–R, Hare was specifically interested in the construction of an instrument to quantify psychopathic personality traits, not an instrument to predict reoffenses. Nevertheless, research shows that the PCL–R does a reasonable job in predicting recidivism (e.g. Quinsey, Rice, & Harris, 1995).

Procedure

In order to establish the interrater reliability, four raters independently rated 27 cases. For the concurrent validity we gathered information from the files of 202 sex offenders listed in the database of the *Federal Documentation Centre of Sex Offender* (Justizanstalt Wien–Mittersteig, Außenstelle Floridsdorf) in Austria using the SVR–20, SORAG, and PCL–R, and we compared these results with our ratings of the Static–99 and RRASOR. Regarding the predictive validity we used the Static–99 and the RRASOR retrospectively on the basis of database files of a care facility for ex–prisoners in Vienna (*Forensisch Therapeutisches Zentrum Wien* [FTZW]). Data on recidivism were retrieved from the official Federal Central Register of the Austrian Department of Justice, whereas each new conviction, that was listed on the official criminal record, was counted as a reoffense. Sexual recidivism was defined as a new conviction for a sex offense in accordance with Austrian criminal law, comprising both hands–on and hands–off offenses³. In agreement with Hanson and Thornton (1999) we included sexual reoffenses in violent recidivism. Furthermore, we explored general recidivism by counting all new entries of offenses.

Results

Reconviction Rates

The reconviction rates after the average follow–up time of 7.72 years were 52% (n = 42) for general, 26% (n = 21) for sexual, and 35% (n = 28) for any violent recidivism (including sexual offenses).

Interrater Reliability

Using the critical values for ICC (single measure) of Fleiss (1986; ICC $\geq .75$ = excellent; $.60 \leq$ ICC $< .75$ = good; $.40 \leq$ ICC $< .60$ = moderate; ICC $< .40$ = poor), the interrater reliability of the Static–99 and the RRASOR were excellent: For the Static–99 we calculated an ICC = .90 (single measure, 95%–Confidence Interval [CI] = .83–.95) and an average correlation of $r = .92$. The interrater reliability of the RRASOR was lower but still very high: The ICC was .78 (single measure, 95%–CI = .64–.88) and the average correlation was $r = .81$.

Concurrent Validity

Both the Static–99 and the RRASOR showed good concurrent validity. The Static–99 scores correlated significantly ($p < .001$ each) with the scores of the SVR–20, SORAG and PCL–R: The correlation between the Static–99 and the SVR–20 was $r = .57$, between the Static–99 and the SORAG $r = .65$ and between the Static–99 and the PCL–R $r = .46$. The RRASOR scores correlated lightly lower but significant with the three instruments of comparison: The correlation between the RRASOR and the SVR–20 was $r = .33$ ($p < .001$), between the RRASOR and the SORAG $r = .43$ ($p < .001$), and between the RRASOR and the PCL–R $r = .22$ ($p < .05$).

Predictive Validity

The mean total scores of the Static–99 and the RRASOR for the 81 sex offender of the follow–up sample were 4.31 (SD = 2.45) and 2.10 (SD = 1.50), respectively. Regarding the Static–99, 16.1% (n = 13) of all sex offender were categorized into the low risk group (0–1 risk points), 19.8% (n = 16) medium low (2–3), 33.3% (n = 27) medium high (4–5), and 30.8% (n = 25) high. For the RRASOR, 16.0% (n = 13) had a score of 0, 23.5% (n = 19) of 1, 21.0% (n = 17) of 2, 22.2% (n = 18) of 3 and 8.6% (n = 7) each of 4 and 5⁴.

Figure 1 to 3 present the ROC Curves of the Static–99 and the RRASOR for general, sexual, and violent

SEXUAL OFFENDER TREATMENT: Martin Rettenberger, Reinhard Eher

(including sexual) reoffending. Regarding the critical values for AUC of Cohen (1992; $AUC > .70$ = good; $.65 < AUC \leq .70$ = moderate; $AUC \leq .65$ = insufficient), the predictive validity of the Static-99 for general, sexual, and violent (including sexual) recidivism was good, whereas the predictive validity of the RRASOR for each kind of recidivism was moderate⁵ (see Table 2). The difference of the AUC values was significant for general ($z = 2,27$; $p < .05$) and violent ($z = 2,23$; $p < .05$), but not for sexual ($z = 1,45$; $p = .15$) recidivism.

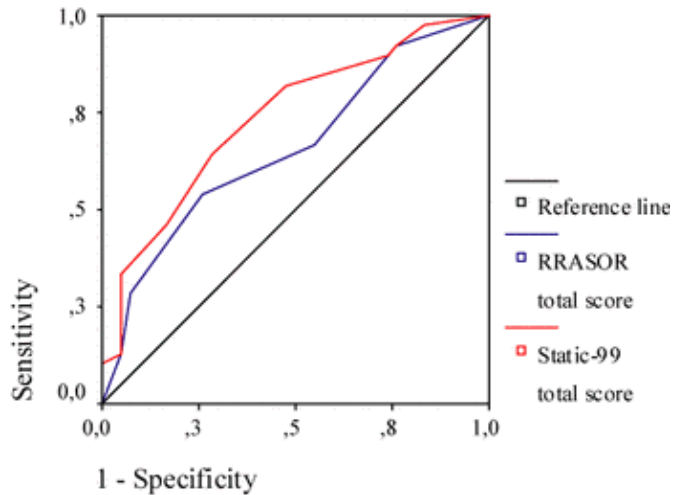


Figure 1: ROC curve of the Static-99 and the RRASOR for general recidivism.

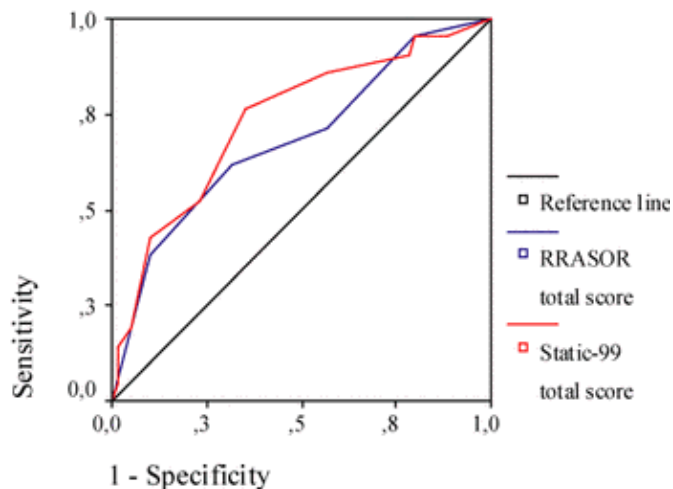


Figure 2: ROC curve of the Static-99 and the RRASOR for sexual recidivism.

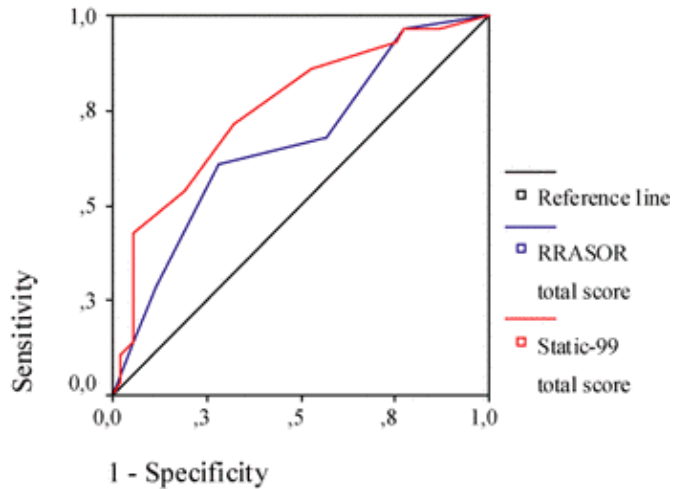


Figure 3: ROC curve of the Static–99 and the RRASOR for violent (including sexual) recidivism.

Table 2: Predictive validity of the Static–99 and the RRASOR

	<i>General recidivism</i>	<i>Sexual recidivism</i>	<i>Violent (including sexual) recidivism</i>
Static–99 Pearson correlation	.41**	.35**	.41**
Static–99 AUC	.74**	.74**	.76**
RRASOR Pearson correlation	.29*	.30**	.28*
RRASOR AUC	.66*	.68*	.67*
* p < .05 ** p < .001			

Discussion

The present study is the first one in Austria to assess the predictive validity of the Static–99 and the RRASOR. The base rates for general, sexual, and violent (including sexual) reoffending of this study are comparable to those found in international studies about sex offender recidivism (e.g. Hanson & Thornton, 1999). In order to calculate the predictive validity of the instrument, we developed a German version of the Static–99 (Rettenberger & Eher, 2006). This German version showed an excellent interrater reliability, and the concurrent validity of the instrument is comparable to results from other studies and can be interpreted as satisfactory. The results for interrater reliability for the RRASOR was lower but still very good. Furthermore even the RRASOR showed for concurrent validity results that are comparable to studies from North America (e.g. Barbaree et al., 2001); therefore these results can be interpreted as satisfactory, too.

The present study yielded good predictive validity for the Static-99 for each recidivism criterion. Although the instrument was originally designed to predict sexual and violent (including sexual) reoffenses, it turned out that the Static-99 also had significant predictive accuracy for general reoffending. With reference to the critical AUC values expressed by Cohen (1992) the predictive validity of the RRASOR can be interpreted as moderate. Regarding general and violent (including sexual) recidivism the results show that the predictive accuracy of the Static-99 is better than that one of the RRASOR. Although the AUC values of the Static-99 are higher than the AUC values of the RRASOR, the not significant difference regarding sexual recidivism indicates that they are equally accurate at predicting sexual reoffenses⁶. These results can be interpreted as a support of the opinion of Harris, Phenix, Hanson, and Thornton (2003): "On the whole, if the information is available to score the Static-99 it is preferable to use the Static-99 over the RRASOR as estimates based on the Static-99 utilize more information than those based upon RRASOR score" (p. 4). However, the performance of the RRASOR is remarkable: The predictive accuracy of the instrument depended on only four simply scored items, so the RRASOR is by far the easiest actuarial risk assessment tool to score (Barbaree et al., 2001). Even if actuarial risk assessment instruments like the Static-99 and the RRASOR cannot be recommended as stand-alone devices for risk assessment in clinical or legal practice (Sjöstedt & Långström, 2001)⁷, our data support previous claims that these instrument could serve as an initial baseline tool to guide decisions concerning sex offenders' need of further clinical evaluation and intervention. Therefore the results of our cross-validation study with an Austrian sample of sex offenders confirmed the results of previous studies from North America and Europe. Further research is still required on the performance of actuarial tools in different subpopulations of sex offenders. Regarding the recidivism risk of intrafamilial child molesters, some recent studies indicate that their risk of reoffense is actually higher than generally assumed, whereas actuarial risk assessment instruments like the Static-99 and the RRASOR categorize intrafamilial child molesters mostly as low-risk-offender (Studer, Clelland, Aylwin, Reddon, & Monro, 2000; Eher & Ross, 2006). Our sample was too small to answer this question. Furthermore the data showed that the majority of the sex offenders in our study are a high-risk group⁸, so this sample is not representative for all Austrian sex offenders. Future research will have to include more representative samples of sex offenders and will have to focus on the predictive validity of these instruments in different subgroups of sex offenders. Regardless of the discussion which sex offender risk assessment approach (actuarial vs. clinical) is the better way to evaluate the recidivism risk (e.g. Berlin, Galbreath, Geary, & McClone, 2003; Boer, 2006), our results support the utility of the RRASOR and particularly the Static-99 in Austria. Because of the higher predictive accuracy of the Static-99 compared with the RRASOR, we recommend to prefer the Static-99 in evaluating sex offender recidivism risk.

References

1. Barbaree, H. E., Seto, M. C., Langton, C. M., & Peacock, E. J. (2001). Evaluating the predictive accuracy of six risk assessment instruments for adult sex offenders. *Criminal Justice and Behavior*, 28, 490–521.
2. Berlin, F. S., Galbreath, N. W., Geary, B., & McClone, G. (2003). The use of actuarials at civil commitment hearings to predict the likelihood of future sexual violence. *Sexual Abuse: A Journal of Research and Treatment*, 15, 377–382.
3. Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1997). *Manual for the Sexual Violence Risk – 20: Professional guidelines for assessing risk of sexual violence*. Vancouver/CA: The Mental Health, Law, & Policy Institute.
4. Boer, D. P. (2006). Sexual offender risk assessment strategies: Is there a convergence of opinion yet? *Sexual Offender Treatment*, 1, 1–4. <http://www.iatso.org/ejournal/>
5. Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159.
6. Dahle, K. P. (2005). *Psychologische Kriminalprognose*. Herbolzheim: Centaurus.
7. Dempster, R. J. & Hart, S. D. (2002). The relative utility of fixed and variable risk factors in discriminating sexual recidivists and nonrecidivists. *Sexual Abuse: A Journal of Research and*

SEXUAL OFFENDER TREATMENT: Martin Rettenberger, Reinhard Eher

- Treatment, 14, 121–138.
7. Douglas, K. S. (2001). HCR–20 violence risk assessment scheme: Overview and annotated bibliography. <http://www.sfu.ca/psychology/groups/faculty/hart/violink.htm>
 8. Eher, R. & Ross, T. (2006). Reconsidering risk for reoffense in intrafamilial child molesters: New aspects on clinical and criminological issues. *Sexual Offender Treatment*, 1, 1–9. <http://www.iatso.org/ejournal/>
 9. Fleiss, J. L. (1986). *The design and analysis of clinical experiments*. New York: Wiley.
 10. Grubin, D. (1998). *Sex offending against children: Understanding the risk*. Police Research Series, Paper 99. London: Home Office.
 11. Habermeyer, E. & Herpertz, S. (2005). Bedeutung des "psychopathy"-Konzepts von Hare für Sexualstraftaten. In D. Schläfke, F. Häßler, & J. M. Fegert (Eds.), *Sexualstraftaten* (pp. 131–141). Stuttgart: Schattauer.
 12. Hanson, R. K. (1997). *The development of a brief actuarial risk scale for sexual offense recidivism*. User Report 1997–04. Ottawa: Department of the Solicitor General of Canada.
 13. Hanson, R. K. & Bussière, M. T. (1996). *Predictors of sexual offender recidivism: A meta-analysis*. User Report 1996–04. Ottawa: Department of the Solicitor General of Canada.
 14. Hanson, R. K. & Thornton, D. (1999). *Static–99: Improving actuarial risk assessments for sex offenders*. User Report 99–02. Ottawa: Department of the Solicitor General of Canada. verfügbar unter: www.sgc.gc.ca
 15. Hare, R. D. (1991). *Manual for the Psychopathy–Checklist–Revised*. Toronto: Multi Health Systems.
 16. Harris, A., Phenix, A., Hanson, R. K., & Thornton, D. (2003). *Static–99 coding rules revised – 2003*. Ottawa: Department of the Solicitor General of Canada.
 17. 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.
 18. Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumière, M. L., Boer, D., & Lang, C. (2003). A multisite comparison of actuarial risk instruments for sex offenders. *Psychological Assessment*, 15, 413–425.
 19. Hart, S. D. & Hare, R. D. (1997). Psychopathy: Assessment and association with criminal conduct. In D. Stoff, J. Breiling, & J. D. Maser (Eds.), *Handbook of antisocial behavior* (pp. 22–35). New York: John Wiley & Sons.
 20. Hartmann, J., Hollweg, M., & Nedopil, N. (2001). Quantitative Erfassung dissozialer und psychopathischer Persönlichkeiten bei der strafrechtlichen Begutachtung. *Nervenarzt*, 72, 365–370.
 21. Metz, C. E. (1998). *IBM compatible ROCKIT user`s guide*. Chicago: The University of Chicago/Department of Radiology.
 22. Müller–Isberner, R., Cabeza, S. G., & Eucker, S. (2000). Die Vorhersage sexueller Gewalttaten mit dem SVR–20. Haina: Institut für Forensische Psychiatrie.
 23. Noll, T., Endrass, J., Rossegger, A., & Urbaniok, F. (2006). Die Risikokalkulation für die Begehung von Gewalt– und Sexualstraftaten mit Hilfe des Static–99. *Monatsschrift für Kriminologie und Strafrechtsreform*, 89, 24–33.
 24. Quinsey, V. L., Rice, M. E., & Harris, G. T. (1995). Actuarial prediction of sexual recidivism. *Journal of Interpersonal Violence*, 10, 85–105.
 25. Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. (1998). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.
 26. Rettenberger, M. & Eher, R. (2006). Die revidierten Kodierungsrichtlinien des STATIC–99. http://www.iatso.org/iatso_user_tool/publications
 27. Sjöstedt, G. & Långström, N. (2001). Actuarial assessment of sex offender recidivism risk: A cross-validation of the RRASOR and the Static–99 in Sweden. *Law and Human Behavior*, 25, 629–645.
 28. Stadtland, C., Hollweg, M., Kleindienst, N., Dietl, J., Reich, U., & Nedopil, N. (2005). Risk assessment and prediction of violent and sexual recidivism in sex offenders: Long-term predictive validity of four risk assessment instruments. *Journal of Forensic Psychiatry and Psychology*, 16,

92–108.

29. Studer, L. H., Clelland, S. R., Aylwin, A. S., Reddon, J. R., & Monro, A. (2000). Rethinking risk assessment for incest offenders. *International Journal of Law and Psychiatry*, 23, 15–22.
30. Vogel, V. d., Ruiter, C. d., Beek, D. v., & ead, G. (2004). Predictive validity of the SVR–20 and Static–99 in a Dutch sample of treated sex offenders. *Law and Human Behavior*, 28, 235–251.

Footnotes

¹ This article is short version of the diploma thesis written by Martin Rettenberger at the Institute for Forensic Psychiatry of the Free University of Berlin supervised by PD Dr. Klaus–Peter Dahle (Institute for Forensic Psychiatry, Free University Berlin) and PD Dr. Reinhard Eher (Federal Documentation Centre for Sexual Offenders, Austrian Prison System). The whole version of the diploma thesis in German is for IATSO–members available from: http://www.iatso.org/iatso_user_tool/publications

² Therefore it was not possible in this study to examine the predictive accuracy of the SVR–20, SORAG, and PCL–R; the quality of the available files of our follow–up sample was not sufficient for scoring more elaborate instruments like SVR–20, SORAG or PCL–R.

³ For the purposes of a Static–99 assessment a sexual offense is an officially recorded sexual misbehaviour or criminal behaviour with sexual intent (Harris, Phenix, Hanson, & Thornton, 2003). You can find exact definitions of the different kind of sexual offenses and a list of paragraphs of the Austrian and German criminal codes in the German version of the manual of the revised version of the Static–99 (for IATSO–members available from: http://www.iatso.org/iatso_user_tool/publications)

⁴ These distributions of scores differ from validation studies from North America (e.g. Hanson & Thornton, 1999; Barbaree et al., 2001), but are comparable with other European studies (e.g. de Vogel et al., 2004; Stadtland et al., 2005).

⁵ De Vogel, de Ruiter, van Beek, and Mead (2004) use with reference to Douglas (2001) the following interpretation of critical AUC values: AUC of .70 and above are considered moderate, and above .75 good.

⁶ However, the results of the Static–99 and the RRASOR regarding sexual recidivism can be interpreted as an advice of a better predictive accuracy of the Static–99; it has to be assumed that one reason for the not significant result is the small sample size.

⁷ Sjöstedt and Långstrom (2001) explain that the validity of instruments like the Static–99 or the RRASOR is still too low to recommend either model to be used as stand–alone devices for risk assessment in clinical or legal practice: “We maintain that they should be used only with extreme caution outside of the research context, and only if properly supplemented with other (nonactuarial) approaches” (p. 640). One reason for that is the clinical need for dynamik risk factors, another one are legal requirements of risk assessment in the German–speaking part of Europe regarding actuarial methods in common (e.g. Dahle, 2005).

⁸ It can be speculated that the persons in charge of aftercare operations programs for sex offenders selected mainly high risk offenders for their treatment programs. The subjects of this study attended treatment programs of an Austrian aftercare operations institution, so it is possible that these sex offenders are also part of a high risk group.

Author address

*Reinhard Eher, M.D., Associate Professor
Federal Documentation Centre for Sexual Offenders
Austrian Prison System
Gerichtsgasse 6
A'210 Vienna
Austria
Phone: ++43 1 27 85 300 31
Fax: ++43 1 27 85 300 32
E mail: reinhard.eher@justiz.gv.at*

Predicting Reoffence in Sexual Offender Subtypes: A Prospective Validation Study of the German Version of the Sexual Offender Risk Appraisal Guide (SORAG)

Martin Rettenberger, Reinhard Eher

Federal Documentation Centre for Sex Offenders, Austrian Prison System, Gerichtsgasse 6, A–1210 Wien

[Sexual Offender Treatment, Volume 2 (2007), Issue 2]

Abstract

This study is part of a prospective, longitudinal research project to evaluate the reliability and validity of different recidivism risk assessment methods for sexual offenders under community supervision for scientific and practical use in the German-speaking part of Europe. In this paper we present the German adaptation of the Sexual Offender Risk Appraisal Guide (SORAG), a specific risk assessment tool for sexual offenders that was developed and published in 1998 in Canada. We examined interrater reliability, concurrent validity and predictive validity of the German version of the instrument with a sample of 254 male sexual offenders arrested in the Austrian penal system. The German SORAG showed good results for interrater reliability and concurrent validity. The predictive validity was determined using subgroups of the sample based on each offender's index offence. The results were predominantly good, but the ability of the instrument to predict sexual and violent recidivism varied depending on offender type. The best results could be obtained for the prediction of re-imprisonment. The conclusion of this study is that the German version of the SORAG performs equally well and is most useful in predicting serious reoffenses.

Key words: SORAG, risk assessment, sexual offender, validity, recidivism

Introduction

In the last years there has been growing interest in validated risk assessment tools for sex offenders in the German speaking part of Europe (Noll, Endrass, Rossegger, & Urbaniok, 2006). Compared to North America – where actuarial risk assessment instruments are regarded as “state of the art” – there exist only a few studies in Germany, Austria or Switzerland that have dealt with these instruments. Although the results of a number of studies from North America support the utility of the actuarial approach for the risk assessment for sexual offenders (e.g. Barbaree, Seto, Langton, & Peacock, 2001), validation studies of these instruments also on German speaking populations are required before giving a general recommendation for the use of these instruments in Germany and Austria.

In this study we present the results of a cross-validation of the German adaptation of the *Sexual Offender Risk Appraisal Guide* (SORAG; Quinsey, Harris, Rice, & Cormier, 2006; Rettenberger & Eher, 2007) and its utility to predict relapse in a prospective longitudinal design of sexual offenders

under community supervision. The SORAG is an actuarial risk assessment tool for sexual offenders and was developed by the Canadian forensic researcher Vernon L. Quinsey and his colleagues. This instrument is a modification of the *Violence Risk Appraisal Guide* (VRAG) that was developed to predict violent and sexual recidivism among male offenders; 10 of the 14 items of the SORAG are the same items as in the VRAG. The SORAG is conceptualized for the risk assessment for sexual offenders to assess the likelihood of violent recidivism which includes sexual offences involving physical contact with the victim.

The instrument consists of 14 weighted items (see Table 1). After scoring these the evaluator adds up the item scores and gets the total score of the SORAG. Based on the total score the evaluator can allocate the offender to one of nine risk categories. By means of these risk categories it is possible to infer to empirically calculated probabilities of violent (including sexual) recidivism after seven and ten years, respectively.

Table1: Items and basic coding rules of the SORAG (Quinsey et al., 2006)

Item Number	Risk Factor	Coding Rule
1	Lived with both biological parents to age 16 (except for death of parent) Score <i>no</i> if offender did not live continuously with both biological parents until age 16, except if one or both parents died. In case of parent death, score as for yes	Yes = -2 No = +3
2	Elementary school maladjustment (up to and including Grade 8)	No problems = -1 Slight or moderate problems = +2 Severe problems = +5
3	History of alcohol problems Allot one point for each of the following: alcohol abuse in biological parent, teenage alcohol problem, adult alcohol problem, alcohol involved in a prior offense, alcohol involved in the index offense	0 = -1 1 or 2 = 0 3 = +1 4 or 5 = +2
4	Marital status (or lived common law in the same home for at least 6 months) At time of index offense	Ever married = -2 Never married = +1
5	Criminal history score for convictions and charges for nonviolent offenses prior to the index offense (from the Cormier–Lang system)	Score 0 = -2 Score 1 or 2 = 0 Score of 3 or above = +3
6	Criminal history score for convictions and charges for violent offenses prior to the index offense (from the Cormier–Lang system)	Score 0 = -1 Score 2 = 0 Score of 3 or above = +6

7	Number of convictions for previous sexual offenses (pertains to convictions for sexual offenses that occurred prior to the index offense) Count any offenses known to be sexual, including, for example, indecent exposure	0 = -1 1 or 2 = +1 e 3 = +5
8	History of sex offenses against girls under age 14 <i>only</i> (includes index offense; if offender was less than 5 years older than victim, always score +4)	Yes = 0 No = +4
9	Failure on prior conditional releases (includes parole violation or revocation; breach of or failure to comply with recognizance or probation; bail violation; and any new charges, including the index offense, while on a conditional release)	No = 0 Yes = +3
10	Age at index offense (at most recent birthday)	e 39 = -5 34–38 = -2 28–33 = -1 27 = 0 d 26 = +2
11	Meets DSM–III criteria for any personality disorder	No = -2 Yes = +3
12	Meets DSM–III criteria for schizophrenia	Yes = -3 No = +1
13	Phallometric test results	<i>All</i> indicate nondeviant sexual preferences = -1 <i>Any</i> test indicates deviant sexual preferences = +1
14	Hare Psychopathy Checklist Revised score (PCL–R; Hare, 1991)	d 4 = -5 5–9 = -3 10–14 = -1 15–24 = 0 25–34 = +4 e 35 = +12

In order to use the SORAG in Austria, we had to adapt the original coding guidelines to the forensic context in this country (Rettenberger & Eher, 2007). This concerns especially the items number 11 and 12, because of changes of the diagnostic criteria in the new version of the DSM (American Psychiatric Association [2003], APA). A further important modification concerns the item number 13: In the German-speaking part of Europe phallometric evaluations are not conducted routinely, so this information is simply not available¹. Therefore, a diagnosis of sexual deviance according to DSM or ICD criteria replaced the phallometric evidence. If a sexual offender met the DSM or ICD criteria for any paraphilia the evaluator had to score +1, otherwise he had to score -1.

There are no published validation studies of the SORAG in the German-speaking part of Europe until now². Most of the validation studies derive from Anglo-American countries. Harris, Rice, Quinsey, Lalumère, Boer, and Lang (2003) reviewed the results of these studies and concluded that several studies (e.g., Barbaree, Seto, Langton, & Peacock, 2001; Nunes, Firestone, Bradford,

Greenberg, & Broom, 2002; Rice & Harris, 2002) have shown high accuracy of the SORAG in the prediction of violent (including sexual) recidivism and moderate accuracy in predicting sexual offenses. For the prediction of violent (including sexual) recidivism they calculated a median AUC value of AUC = .75.

Although the results of the predictive accuracy of the SORAG has been reasonably consistent across studies, Bartosh, Garby, Lewis, and Gray (2003) suggested that the predictive validity of the instrument varied depending on the type of sexual offender. According to these authors, the SORAG could significantly predict sexual, violent, and overall recidivism for extrafamilial child molesters (AUC values ranged from .70 to .93) and for incest offenders (AUC ranged from .72 to .91). As regards to rapists and hands-off offenders, however, the SORAG showed much lower predictive power (AUC ranged from .46 to .71). Ducro and Pham (2006), retrospectively, evaluated the predictive accuracy of the SORAG on Belgian sexual offenders committed to a forensic facility. For the total sample the instrument showed strong predictive validity for general (AUC = .70) and violent (AUC = .72) recidivism and moderate predictive validity for sexual recidivism (AUC = .64). Depending on offender subgroup and recidivism criterion the AUC values ranged from .64 to .77. The results of Bartosh et al. (2003) and Ducro & Pham (2006) support the evidence that the SORAG shows good predictive validity, whereas the results varied depending on sex offender type.

The primary aim of this prospective, longitudinal study was to investigate the usefulness and the predictive validity of the German version of the SORAG in the forensic context of a German speaking country. Therefore, we present data on the interrater reliability, the concurrent validity, and the predictive validity of this German version. Furthermore, the predictive validity was also investigated in subgroups of the sample based on offender's index offence. On the one hand, we expected the SORAG to show high predictive accuracy and that our findings would be in accordance with previous studies from North America and Europe. On the other hand, we expected differences in the predictive accuracy of the SORAG depending on sexual offender subgroup and recidivism criterion.

Method

Subjects

Two hundred and fifty-four male sex offenders were investigated in 2002 and 2003 at the *Federal Documentation Centre of Sex Offender of the Austrian Prison System* and followed up after prison release until December, 31, 2005 (mean age at time of release = 40.92 years, SD = 12.53, range 18–72). The follow-up time periods ranged from 24.27 to 66.67 months (M = 38.82 months; SD = 9.51; 24 months was defined to be a minimal follow-up period). The group consisted of 132 child molesters (58 extrafamilial molest offenders and 74 incest offenders), and 120 rapists; one offender was convicted because of sexual burglary and one because of child pornography. The mean duration of imprisonment for the index offense was M = 32.37 months (SD = 22.14), 20.9% (n = 53) had at least committed one previous sexual offense (M = 0.54; SD = 1.80, range 0–16), and 58.7% (n = 149) had at least one prior conviction for any offense (M = 3.88, SD = 6.14, range 0–37).

Of the 254 offenders, 10.2% (n = 26) failed compulsory school or attended school for mentally or physically handicapped children, 34.3% (n = 87) completed school, but had no professional education, 49.6% (n = 126) completed technical school or professional education, 3.1% (n = 8) had a university-entrance diploma or completed technical college, and 0.8% (n = 2) had a university degree. Regarding the level of education, there were no information on 2.0% (n = 5) of offenders.

Procedure and Statistical Analysis

The interrater reliability of the SORAG was examined by the Intraclass Correlation Coefficient (ICC). The predictive ability of the instrument was analyzed using Pearson correlations and the area under the curve (AUC) of the ROC. Referring to Cohen (1992), Dahle, Schneider, and Ziethen (2007) formulated the following criteria for the classification of the predictive accuracy of risk assessment tools: AUC values of .72 or above (r e .37) are classified as “good” and AUC values between .64 and .71 (r e .24) are classified as “moderate”. Significant AUC values that are below the value of .64 (r < .24) are classified as “small”³.

In order to establish the concurrent validity of the instrument, we calculated Pearson correlations between the SORAG on the one hand and the *Sexual Violence Risk–20* (SVR–20; Boer, Hart, Kropp, & Webster, 1997), the *Static–99* (Hanson & Thornton, 1999), and the *Psychopathy Checklist–Revised* (PCL–R; Hare, 1991) on the other hand⁴. We chose the SVR–20, SORAG, and PCL–R, because there exists already a number of validation studies (e.g. Rettenberger & Eher, 2006a; Barbaree et al., 2001; Harris, Rice, et al., 2003), so these instruments represent internationally recognized risk assessment tools for sex offenders (e.g. Müller–Isberner, Cabeza, & Eucker, 2000; Habermeyer & Herpertz, 2005). For Pearson correlations and ICC values we use the raw scores of each instrument.

The SVR–20 is a structured clinical guideline designed for the assessment of risk for sexual violence in adult sex offenders. The instrument was developed from a thorough consideration of the empirical literature and the clinical expertise of a number of clinicians. The SVR–20 consists of 20 items, divided into three domains: Psychosocial adjustment, Sexual offenses, and Future plans, which have to be coded by an experienced forensic clinician. Although originally designed as a structured clinical guideline, it is not uncommon for research purposes to add up the items. In this case, the instrument becomes a conceptual actuarial measure (Hanson & Morton–Bourgon, 2007).

The Static–99 is a brief actuarial instrument for the assessment of risk for sexual and violent recidivism of adult sexual offenders. This instrument is derived from the fusion of two previously developed risk assessment instruments, the *Rapid Risk Assessment of Sexual Offender Recidivism* (RRASOR; Hanson, 1997) and a shorter version of the *Structured Anchored Clinical Judgement* (SACJ–Min; Grubin, 1998) composed of ten mainly static risk factors. In terms of the Static–99, there exists already a German adaptation (Rettenberger & Eher, 2006b) of the revised version of this instrument (Harris, Phenix, Hanson, & Thornton, 2003). The German version of the Static–99 showed high predictive accuracy and good results for interrater reliability and concurrent validity (Rettenberger & Eher, 2006a).

In contemporary research and clinical practice, Hare’s Psychopathy Checklist–Revised (PCL–R) is the psycho–diagnostic tool most commonly used to assess psychopathy. The PCL–R is based on semistructured interviews and review of file information. Participants are assigned to ratings of “0” (absent), “1” (some indication), or “2” (present) on each of the 20 PCL–R items, tapping characteristics such as impulsivity, irresponsibility, and callousness. Scale scores are obtained by summing the ratings, for a total possible score of 40. The conventional cutoff for making a diagnosis of psychopathy in North America is 30, whereas in Europe the cutoff is mostly 25 (Hart & Hare, 1997; Hartmann, Hollweg, & Nedopil, 2001). In the development of the PCL–R, Hare was specifically interested in the construction of an instrument to quantify psychopathic personality traits, not an instrument to predict reoffenses. Nevertheless, research shows that the PCL–R does a reasonable job in predicting recidivism (e.g. Quinsey, Rice, & Harris, 1995).

In order to establish the interrater reliability, three randomly selected colleagues, that have to be

theoretically and practically skilled in the application of actuarial risk assessment methods, rated independently ten cases. These ten cases were randomly selected within the database of the *Federal Documentation Centre of Sex Offender of the Austrian Prison System*. The second author gave the raters official file information of each case to code the actuarial instruments (e.g. charges, convictions, official criminal records, etc.).

Data on recidivism were retrieved from the official Federal Central Register of the Austrian Ministry of Internal Affairs, whereas each new conviction (with or without imprisonment), that was listed on the official criminal record, was counted as a reoffense. Furthermore, we counted separately each new imprisonment. We defined the following recidivism criteria: Sexual recidivism (with or without imprisonment), nonsexual violent recidivism (with or without imprisonment), general recidivism (with or without imprisonment), and violent recidivims (including sexual and non sexual violent offenses, with and without imprisonment)⁵.

Results

Reconviction Rates

Offenders (N = 254) were categorized as rapists (47.2%; n = 120), extrafamilial child molesters (22.8%; n = 58), and incest offenders (29.1%; n = 74). Two offenders (0.8%) could not be allocated to either of these groups and therefore were excluded. The reconviction rates for the total sample and the three subgroups are shown in Table 2.

Table 2: Reconviction rates of the total sample and each category of index offense type

<i>Recidivism Criterion</i>	<i>Total Sample (N = 254)</i>	<i>Rapists (n = 120)</i>	<i>Extrafamilial Molest Offenders (n = 58)</i>	<i>Incest Offenders (n = 74)</i>
Any Recidivism (Conviction)	25.2% (n = 64)	33.3% (n = 40)	20.7% (n = 12)	16.2% (n = 12)
Any Recidivism (Imprisonment)	15.0% (n = 38)	22.5% (n = 27)	15.5% (n = 9)	2.7% (n = 2)
Sexual Recidivism (Conviction)	3.5% (n = 9)	1.7% (n = 2)	12.1% (n = 7)	0% (n = 0)
Sexual Recidivism (Imprisonment)	3.1% (n = 8)	1.7% (n = 2)	10.3% (n = 6)	0% (n = 0)
Violent non–sexual Recidivism (Conviction)	12.2% (n = 31)	20.8% (n = 25)	1.7% (n = 1)	6.8% (n = 5)
Violent non–sexual Recidivism (Imprisonment)	6.7% (n = 17)	12.5% (n = 15)	1.7% (n = 1)	1.4% (n = 1)

Sexual and/or Violent Recidivism (Conviction)	15.4% (n = 39)	21.7% (n = 26)	13.8% (n = 8)	6.8% (n = 5)
Sexual and/or Violent Recidivism (Imprisonment)	9.4% (n = 24)	13.3% (n = 16)	12.1% (n = 7)	1.4% (n = 1)

Interrater Reliability

Using the critical values for ICC (single measure) of Fleiss (1986; ICC $\geq .75$ = excellent; $.60 \leq$ ICC $< .75$ = good; $.40 \leq$ ICC $< .60$ = moderate; ICC $< .40$ = poor), the interrater reliability of the German version of the SORAG was excellent: We calculated an ICC = .93 (single measure, $p < .001$). Furthermore, we also evaluated the interrater reliability of the instruments used for the concurrent validity analyses. All instruments showed excellent interrater reliability: for the Static–99 we calculated an ICC = .98, for the SVR–20 an ICC = .84, and for the PCL–R = .93 (all $p < .001$).

Concurrent Validity

The SORAG showed good concurrent validity: The instrument correlated significantly ($p < .001$) with the Static–99, SVR–20, and PCL–R: The correlation coefficients were $r = .71$, $r = .78$ and $r = .79$, respectively.

Predictive Validity

The mean total score of the SORAG for the 254 sex offenders was 5.59 (SD = 13.49). Regarding the SORAG risk categories, 12.6% (n = 32) were categorized into the first risk group (SORAG total score of ≤ -10), 17.3% (n = 44) into the second risk group (-9 to -4), 17.0% (n = 43) into the third risk group (-3 to $+2$), 12.2% (n = 31) into the fourth risk group ($+3$ to $+8$), 11.4% (n = 29) into the fifth risk group ($+9$ to $+14$), 10.6% (n = 27) into the sixth risk group ($+15$ to $+19$), 9.8% (n = 25) into the seventh risk group ($+20$ to $+24$), 6.3% (n = 16) into the eighth risk group ($+25$ to $+30$), and 2.8% (n = 7) into the ninth risk group (≥ 31).

The values of the predictive validity (AUC values of the ROC and Pearson correlations) of the SORAG for the total sample and the three subgroups are shown in Table 3. For the total sample the AUC values are at least .73 and for all recidivism criteria of the total group the AUC and correlation values reach statistical significance. For the rapist subgroup the SORAG shows also significant predictive accuracy, comparable to the values of the total sample. However, the results for sexual recidivism (where just two relapses occurred) for the rapist subgroup reach no statistical significance because of too few cases of relapse within this category. With exception of the results for violent nonsexual recidivism (where just one relapse occurred) the results for the extrafamilial child molester group are at least AUC = .70, all $p < .001$. For the incest offender subgroup the recidivism base rates are altogether very low, so the results reach statistical significance only for the “any recidivism”–criteria. Overall, the AUC values for the “imprisonment”–criteria are consistently higher than for the “conviction”–criteria.

Table 3: Predictive validity of the SORAG for the total sample and each category of index offense type

<i>Recidivism Criterion</i>	<i>Total Sample (N = 254)</i>	<i>Rapists (n = 120)</i>	<i>Extrafamilial Molest Offenders (n = 58)</i>	<i>Incest Offenders (n = 74)</i>
Any Recidivism (Conviction)	AUC = .73** (r = .35**)	AUC = .73** (r = .37**)	AUC = .70* (r = .26)	AUC = .62 (r = .26*)
Any Recidivism (Imprisonment)	AUC = .83** (r = .41**)	AUC = .80** (r = .41**)	AUC = .76* (r = .30*)	AUC = .95* (r = .31**)
Sexual Recidivism (Conviction)	AUC = .73* (r = .14*)	AUC = .71 (r = .09)	AUC = .76* (r = .28*)	--- ^a
Sexual Recidivism (Imprisonment)	AUC = .73* (r = .14*)	AUC = .71 (r = .09)	AUC = .76* (r = .26)	--- ^a
Violent non–sexual Recidivism (Conviction)	AUC = .75** (r = .29**)	AUC = .72** (r = .31**)	AUC = .56 (r = .01)	AUC = .63 (r = .21)
Violent non–sexual Recidivism (Imprisonment)	AUC = .83** (r = .30**)	AUC = .80** (r = .34**)	AUC = .56 (r = .01)	AUC = .92 (r = .18)
Sexual and/or Violent Recidivism (Conviction)	AUC = .76** (r = .33**)	AUC = .73** (r = .33**)	AUC = .74* (r = .27*)	AUC = .63 (r = .21)
Sexual and/or Violent Recidivism (Imprisonment)	AUC = .82** (r = .33**)	AUC = .81** (r = .36**)	AUC = .74* (r = .24)	AUC = .92 (r = .18)

* p < .05 ** p < .001 ^a

Discussion

The present study is the first assessing the validity of a German version of the SORAG in a prospective design. The results show a normal distribution of the total scores and exhibited the mean score of the SORAG to be comparable to results of previous studies (e.g. Bartosh et al., 2003; Barbaree et al., 2001). This indicates that our sample is representative for a general sexual offender population. Regarding the critical values of ICC (single measure) of Fleiss (1986), interrater reliability of the German version of the instrument is excellent. These results are also comparable to previous studies (e.g. Ducro & Pham, 2006; Bartosh et al., 2003). Furthermore, the values of the concurrent validity analyses can be interpreted as satisfactory (Harris, Rice, et al., 2003). Although the PCL–R is not designed as an actuarial measure for sex offender risk assessment, the high correlation between SORAG and PCL–R is not surprising: The PCL–R is included in the SORAG, so the PCL–R total score has a direct influence on the SORAG total score. The high correlations between the SORAG on the one hand and Static–99 and SVR–20 on the other hand reveal the fact that that all 3 instruments are designed to predict sexual recidivism partially using the same items (e.g. criminal history). But there are, however, less than perfect associations between these instruments, so it has to be one aim of further research on this topic to clarify similarities and differences between actuarial methods⁷.

Referring to the critical value descriptors for AUC values by Dahle, Schneider, and Ziethen (2007), the predictive validity of the German SORAG for the total sample can be classified as “good”. The AUC values for the “imprisonment”–criteria were consistently higher than for the “conviction”–criteria. This result is an interesting finding with importance for forensic practitioners, since it suggests the instrument is better at predicting more serious relapses (which are punished by a prison sentence).

Originally, the SORAG was developed to predict nonsexual violent and/or sexual recidivism (Quinsey et al., 2006). Therefore, the recidivism criterion is of particular interest and the present study showed good results for this criterion (AUC = .76 for new convictions and AUC = .82 for new imprisonments). Even more importantly, these AUC values are slightly higher than those from previous results of Anglo–American studies (e.g. Barbaree et al., 2001; Harris, Rice, et al., 2003)⁶.

The German version of the SORAG tends to yield more accurate results for rapists than for both pedosexual subgroups. With the exception of the results for sexual recidivism, the instrument shows good predictive accuracy for the rapist subgroup. Although the AUC value for sexual recidivism can be classified as moderate, there are too few cases of relapse within this category in order to interpret these results reasonably (which is also reflected by the fact that this ROC is not significant). For the extrafamilial child molester group the results are similar: With exception of the results for violent nonsexual recidivism – where just one relapse occurred – the predictive validity of the SORAG can be interpreted as satisfactory.

As mentioned before, it is important to take a more precise look at the recidivism base rates and the follow–up time periods when interpreting the results of the sexual offender subgroup analyses: During the follow–up time period in our study there are subgroups and recidivism criteria with no (incest offender subgroup with no sexually motivated relapses) or very few relapses (only one nonsexually motivated violent relapse in the extrafamilial child molester group). In these recidivism categories the SORAG shows the lowest predictive accuracy. In the innerfamilial child sexual abuse group no sexual relapses occurred. Therefore, no AUC values could be calculated. When taking into account the low recidivism rates in some categories of some offender subgroups, one has to address one major limitation of this study: the small sample sizes, even though AUC measures usually are seen to be less sensitive to low base rates than other statistical procedures. Given to our results one can expect that predictive accuracy of the instrument will be demonstrated even for these subgroups and recidivism criteria after a longer at–risk period.

To conclude, our results support the utility of actuarial risk assessment methods like the SORAG even in non English speaking countries. The results of the German version to predict relapse in sexual offenders are predominantly good. Nevertheless, the ability to predict recidivism varied depending on offender type and relapse category with a consistently better performance on more serious reoffenses. The results of the present study again indicate the helpfulness of actuarial instruments like the SORAG for assessing the recidivism risk in sexual offenders.

References

1. American Psychiatric Association (2003). Diagnostisches und statistisches Manual psychischer Störungen: Textrevision DSM–IV–TR. Göttingen: Hogrefe.
2. Barbaree, H. E., Seto, M. C., Langton, C. M., & Peacock, E. J. (2001). Evaluating the Predictive Accuracy of Six Risk Assessment Instruments for Adult Sex Offender. *Criminal*

- Justice and Behavior, 28, 490–521.
3. Bartosh, D. L., Garby, T., Lewis, D., & Gray, S. (2003). Differences in the Predictive Validity of Actuarial Risk Assessments in Relation to Sex Offender Type. *International Journal of Offender Therapy and Comparative Criminology*, 47, 422–438.
 4. Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1997). *Manual for the Sexual Violence Risk – 20: Professional Guidelines for Assessing Risk of Sexual Violence*. Vancouver/CA: The Mental Health, Law, & Policy Institute.
 5. Cohen, J. (1992). A Power Primer. *Psychological Bulletin*, 112, 155–159.
 6. Dahle, K. P., Schneider, V., & Ziethen, F. (2007). Standardisierte Instrumente zur Kriminalprognose. *Forensische Psychiatrie, Psychologie, Kriminologie*, 1, 15–26.
 7. Douglas, K. S. (2001). HCR–20 Violence Risk Assessment Scheme: Overview and Annotated Bibliography. <http://www.sfu.ca/psychology/groups/faculty/hart/violink.htm>.
 8. Ducro, C. & Pham, T. (2006). Evaluation of the SORAG and the Static–99 on Belgian Sex Offenders Committed to a Forensic Facility. *Sexual Abuse: A Journal of Research and Treatment*, 18, 15–26.
 9. Fleiss, J. L. (1986). *The Design and Analysis of Clinical Experiments*. New York: Wiley.
 10. Grubin, D. (1998). *Sex Offending Against Children: Understanding the Risk*. Police Research Series, Paper 99. London: Home Office.
 11. Habermeyer, E. & Herpertz, S. (2005). Bedeutung des "psychopathy"-Konzeptes von Hare für Sexualstraftaten. In D. Schläfke, F. Häßler, & J. M. Fegert (Eds.), *Sexualstraftaten* (pp. 131–141). Stuttgart: Schattauer.
 12. Hanson, R. K. (1997). *The Development of a Brief Actuarial Risk Scale for Sexual Offense Recidivism*. User Report 1997–04. Ottawa: Department of the Solicitor General of Canada.
 13. Hanson, R. K. & Thornton, D. (1999). *Static–99: Improving Actuarial Risk Assessments for Sex Offenders*. User Report 99–02. Ottawa: Department of the Solicitor General of Canada.
 14. Hanson, R. K. & Morton–Bourgon, K. E. (2007). *The Accuracy of Recidivism Risk Assessment for Sexual Offenders: A Meta–Analysis*. User Report 07–01. Ottawa: Department of the Solicitor General of Canada.
 15. Hare, R. D. (1991). *Manual for the Psychopathy–Checklist–Revised*. Toronto: Multi Health Systems.
 16. Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumière, M. L., Boer, D., & Lang, C. (2003). A Multisite Comparison of Actuarial Risk Instruments for Sex Offenders. *Psychological Assessment*, 15, 413–425.
 17. Harris, G. T., Phenix, A., Hanson, R. K., & Thornton, D. (2003). *Static–99 Coding Rules Revised – 2003*. Ottawa: Department of the Solicitor General of Canada.
 18. Hart, S. D. & Hare, R. D. (1997). Psychopathy: Assessment and Association with Criminal Conduct. In D. Stoff, J. Breiling, & J. D. Maser (Eds.), *Handbook of Antisocial Behavior* (pp. 22–35). New York: Wiley.
 19. Hartmann, J., Hollweg, M., & Nedopil, N. (2001). Quantitative Erfassung dissozialer und psychopathischer Persönlichkeiten bei der strafrechtlichen Begutachtung. *Nervenarzt*, 72, 365–370.
 20. Marshall, W. L. (2005). Grenzen der Phallometrie. *Recht & Psychiatrie*, 23, 11–23.
 21. Müller–Isberner, R., Cabeza, S. G., & Eucker, S. (2000). Die Vorhersage sexueller Gewalttaten mit dem SVR–20. Haina: Institut für Forensische Psychiatrie.
 22. Noll, T., Endrass, J., Rossegger, A., & Urbaniok, F. (2006). Die Risikokalkulation für die Begehung von Gewalt– und Sexualstraftaten mit Hilfe des Static–99. *Monatsschrift für Kriminologie und Strafrechtsreform*, 89, 24–33.
 23. Nunes, K. L., Firestone, P., Bradford, J. M., Greenberg, D. M., & Broom, I. (2002). A Comparison of Modified Versions of the Static–99 and the Sex Offender Risk Appraisal Guide (SORAG). *Sexual Abuse: A Journal of Research and Treatment*, 14, 253–269.
 24. Quenzer, C. (2005). Prädiktive Validität ausgewählter aktuarischer Verfahren bei

- Sexualdelinquenten. Berlin: Unpublished Diploma Thesis.
25. Quinsey, V. L., Rice, M. E., & Harris, G. T. (1995). Actuarial Prediction of Sexual Recidivism. *Journal of Interpersonal Violence*, 10, 85–105.
 26. Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. (1998). *Violent Offenders: Appraising and Managing Risk*. Washington, DC: American Psychological Association.
 27. Rettenberger, M. & Eher, R. (2006a). Actuarial Assessment of Sex Offender Recidivism Risk: A Validation of the German Version of the Static–99. *Sexual Offender Treatment*, 1, 1–11.
 28. Rettenberger, M. & Eher, R. (2006b). Die revidierten Kodierungsrichtlinien des Static–99. <http://www.vffw.org>.
 29. Rettenberger, M. & Eher, R. (2007). Aktuarische Kriminalprognose bei Sexualstraftätern: Die deutsche Version des Sex Offender Risk Appraisal Guide (SORAG). <http://www.vffw.org>.
 30. Rice, M. E. & Harris, G. T. (2002). Men who Molest their Sexually Immature Daughters: Is a Special Explanation Required? *Journal of Abnormal Psychology*, 111, 329–339.
 31. Vogel, V. d., Ruiter, C. d., Beek, D. v., & Mead, G. (2004). Predictive Validity of the SVR–20 and Static–99 in a Dutch Sample of Treated Sex Offenders. *Law and Human Behavior*, 28, 251.

Footnotes

¹ Furthermore, the evidence of the reliability and criterion validity of phallometric testing in common seems to be weak, although research has suggested a limited value in predicting subsequent recidivism. Therefore Marshall (2005) concluded that the routine use of phallometric assessments as part of the evaluation of sexual offenders can not be recommended.

² However, there are unpublished validation studies in Germany that indicate good predictive validity results (e.g. Quenzer, 2005).

³ De Vogel, de Ruiter, van Beek, and Mead (2004) use with reference to Douglas (2001) the following interpretation of critical AUC values: AUC values of .70 and above are classified as „moderate“, and values above .75 are classified as „good“.

⁴ The upcoming task of the risk assessment research project of the Federal Documentation Centre of Sex Offender of the Austrian Prison System is the advanced analysis of other risk assessment tools (e.g. the SVR–20, Static–99, and PCL–R) and the comparison of the predictive accuracy of these instruments.

⁵ The SORAG was developed to predict violent or/and sexual reoffenses (Quinsey et al., 2006).

⁶ In this context, it should be noted that the small modifications of the German coding rules could be one reason for differences in the results of the predictive accuracy. On the other hand these results show that the German adaptation of the SORAG works at least as good as the original version: This indicates, for example, that a clinical diagnoses of paraphilia represents a good substitution for phallometric assessment results.

⁷ For example, it is possible that the instruments measure different dimensions of recidivism risk. In this study we were not able to focus on this issue, but future research may provide more precise findings about this topic.

Author address

*Reinhard Eher, M.D., Associate Professor
Federal Documentation and Evaluation Centre for Violent and Sexual Offenders
Gerichtsgasse 6
A 210 Vienna
Austria
E Mail:  reinhard.eher@justiz.gv.at
Phone: ++43 1 27 85 300 31
Fax: ++43 1 27 85 300 32*

Aktuarische Kriminalprognosemethoden und Sexualdelinquenz: Die deutsche Version des SORAG

von Martin Rettenberger und Reinhard Eher

Zusammenfassung

Die vorliegende Studie ist Teil eines prospektiv-längsschnittlichen Forschungsprojekts, dessen Gegenstand die Evaluierung unterschiedlicher Kriminalprognosemethoden bei Sexualstraftätern ist. Im vorliegenden Artikel wird die deutsche Adaption des Sexual Offender Risk Appraisal Guide (SORAG) und dessen Vorhersageleistung sexuell motivierter und gewalttätiger Rückfälle dargestellt. Dabei handelt es sich um ein speziell für Sexualstraftäter entwickeltes und erstmals 1998 in Kanada veröffentlichtes aktuarisches Prognoseinstrument. In der vorliegenden Untersuchung werden die Interraterreliabilität sowie die konvergente und prädiktive Validität des Instruments dargestellt, die anhand von 178 in Österreich verurteilten und inhaftierten männlichen Sexualstraftätern erhoben wurden. Dabei erzielte die deutsche Version des SORAG zufriedenstellende Ergebnisse für die Interraterreliabilität und die konvergente Validität. Zur Überprüfung der prädiktiven Validität wurde neben der Gesamtgruppe auch die Vorhersageleistung bei verschiedenen Subgruppen von Sexualstraftätern untersucht. Zusammenfassend belegen die dabei erzielten Ergebnisse eine gute Vorhersageleistung des Instruments für das sexuell motivierte und gewalttätige Rückfallereignis.

Schlüsselwörter: SORAG, Kriminalprognose, Sexualstraftäter, Validität, Rückfälligkeit

1. Einleitung

Aktuarische Kriminalprognosemethoden¹ erfreuen sich im deutschsprachigen Raum mittlerweile einer wachsenden Beliebtheit und werden zunehmend bei kriminalprognostischen Begutachtungen vor allem im Straf- und Maßregelvollzug eingesetzt (*Dable, Schneider & Zietzen* 2007). Auch die jüngst veröffentlichten »Mindestanforderungen für Prognosegutachten« (*Boetticher u.a.* 2006) empfehlen explizit die Verwendung aktuarischer Prognoseinstrumente, wobei einschränkend festgehalten wird, dass diese Instrumente die hermeneutische und hypothesengeleitete Individualprognose, wie sie vom Rechtsanwender gefordert wird, nicht ersetzen kann und darf. Der Nutzen standardisierter aktuarischer Prognosemethoden besteht vielmehr in der Bereitstellung empirischen Wissens über rückfallrelevante Risikovariablen und in der Einhaltung internationaler Prognosestandards, so dass sie als wertvolle Hilfsmittel bei der Erstellung einer auf den Einzelfall bezogenen ideographischen Prognosebegutachtung bezeichnet werden können. In Anbetracht der von juristischer Seite an den psychologischen oder psychiatrischen Sachverständigen formulierten Anforderungen an ein wissenschaftlich fundiertes Prognosegutachten diskutiert *Dable* (2007) die Vorzüge und Begrenzungen verschiedener methodischer Zugänge einer kriminalprognostischen Stellungnahme. Vor allem die in den genannten Mindeststandards formulierte Forderung der

1 In der Literatur gelegentlich auch als »statistische« oder »nomothetische« Prognosemethoden bezeichnet, basiert eine aktuarische Kriminalprognose auf empirisch gesicherten Erkenntnissen, die man bis dato mit Rückfälligkeit von Straftätern gewonnen hat. Ausgangspunkt sind gruppenstatistisch gewonnene Ergebnisse über Variablen, die die Rückfallwahrscheinlichkeit beeinflussen. Die Prognosemethode besteht darin, die zu beurteilende Person aufgrund der individuell vorliegenden Variablenkonfiguration einer empirisch konstruierten Tätergruppe mit einer ähnlichen Konfiguration zuzuordnen, um so von der bekannten Rückfallwahrscheinlichkeit der Tätergruppe auf die individuelle Rückfallwahrscheinlichkeit des zu begutachtenden Einzelfalls zu schließen (ausführlich z.B. bei *Dable* 2005).

wissenschaftlichen Fundierung zählt zu den besonderen Stärken aktuarischer Prognoseinstrumente, da das durch diese Instrumente vorgegebene Vorgehen transparent ist, einer nachvollziehbaren Logik folgt, systematisch empirisch gut gesicherte Kenntnisse nutzt und somit eine objektive und nachprüfbare Fallbeurteilung erlaubt. Zu Recht wird jedoch gleichzeitig der Einsatz dieser Instrumente nur unter Vorbehalt empfohlen, da diese »die Anforderung einer auf den Einzelfall zugeschnittenen Individualprognose schon von ihrer Anlage her nicht bewältigen« können (Dable 2007, 105). Gleichzeitig überzeugen diese Instrumente jedoch mit ihrer Vorhersagefähigkeit zukünftiger strafrechtlich relevanter Handlungen, die zum Teil erheblich über der Vorhersageleistung anderer Prognosemethoden² liegt (Hanson & Morton-Bourgon 2007).

Um die Vorhersagegenauigkeit aktuarischer Prognoseinstrumente weiter zu verbessern, wurden für zahlreiche Subgruppen von Straftätern spezifische Instrumente entwickelt (Dable 2005). So wurden auch für die Subgruppe der Sexualstraftäter eine Reihe spezifischer Instrumente konzipiert und in den letzten Jahren einer Vielzahl von Validierungsstudien unterzogen, wobei sowohl die meisten Instrumente als auch die überwiegende Mehrheit dieser Studien aus dem angloamerikanischen Sprachraum stammen (einen Überblick liefert z.B. Hanson & Morton-Bourgon 2007). Aufgrund der vorhandenen sozialen, kulturellen und juristischen Unterschiede zwischen den USA und den deutschsprachigen Ländern ist eine ungeprüfte Übertragung dieser Methoden auf den deutschsprachigen Raum allerdings nicht zulässig. Die Forderung nach wissenschaftlich fundierten Beurteilungsmethoden schließt den Nachweis der Validität und Objektivität an einer hinreichend großen Stichprobe aus der jeweiligen Jurisdiktion ein. Die vorliegende Studie soll einen Beitrag zur Adaption bewährter angloamerikanischer Kriminalprognosemethoden zur Prognosebegutachtung von Sexualstraftätern im deutschsprachigen Raum leisten und gleichzeitig die Reliabilität und Validität dieser deutschen Version überprüfen.

1.1 Der Sex Offender Risk Appraisal Guide

Der *Sex Offender Risk Appraisal Guide* (SORAG; Quinsey et al. 2006) zählt neben dem *Static-99* (Hanson & Thornton 1999) und dem *Sexual Violence Risk-20* Schema (SVR-20; Boer et al. 1997) zu den bekanntesten und meist verwendeten Prognoseverfahren für Sexualstraftäter (Dable 2005; Hanson & Morton-Bourgon 2004). Sowohl für das *Static-99* (Rettenberger & Eher 2006a) als auch für den SVR-20 (Müller-Isberner, Cabeza & Eucker 2000) liegen deutschsprachige Adaptionen vor, die anschließend Gegenstand von Kreuzvalidierungsstudien an deutschsprachigen Stichproben waren (Rettenberger & Eher 2006b; Stadtland et al. 2005).

Der SORAG ist eine Modifikation des *Violence Risk Appraisal Guide* (VRAG; Quinsey et al. 2006)³ und wurde konzipiert, um gewalttätige (inklusive sexuell motivierte Gewalttaten) Rückfälle entlassener Sexualstraftäter vorherzusagen. Die zur Konstruktion des Instruments verwendete Stichprobe umfasste 178 Sexualstraftäter, die aus einer Hochsicherheitseinrichtung in der kanadischen Provinz Ontario entlassen worden waren. Anhand einer ausführlichen Analyse der während der Haft erhobenen Daten wurde der Zusammenhang zwischen relevanten Merkmalen und der Wahrscheinlichkeit eines erneuten (sexuell motivierten) Gewaltdelikts berechnet. Die 14 vorhersagestärksten Variablen wurden zum

2 Zur Beschreibung alternativer Prognosemethoden sei auf die einschlägigen Lehrbücher oder auf die aktuellen Beiträge von Dable u.a. (2007) und Hanson & Morton-Bourgon (2007) verwiesen.

3 Das VRAG wurde zur Vorhersage gewalttätiger Rückfälle entlassener Täter aus Straf- und Maßregelvollzug entwickelt; neben der Originalpublikation von Quinsey et al. (2006) findet sich bei Dable u.a. (2007) eine ausführliche Darstellung des Instruments.

SORAG zusammengefasst und – je nach der empirisch ermittelten Rückfallrelevanz – unterschiedlich gewichtet (siehe *Tabelle 1*).

Tabelle 1 Die 14 Items des SORAG im englischen Original von Quinsey et al. (2006)

Item-Nummer	Beschreibung	Kodierung
1	Lived with both biological parents to age 16 (except for death of parent) – Score <i>no</i> if offender did not live continuously with both biological parents until age 16, except if one or both parents died. In case of parent death, score as for <i>yes</i>	Yes = -2 No = +3
2	Elementary school maladjustment (up to and including Grade 8)	No problems = -1 Slight or moderate problems = +2 Severe problems = +5
3	History of alcohol problems – Allot one point for each of the following: alcohol abuse in biological parent, teenage alcohol problem, adult alcohol problem, alcohol involved in a prior offense, alcohol involved in the index offense	0 = -1 1 or 2 = 0 3 = +1 4 or 5 = +2
4	Marital status (or lived common law in the same home for at least 6 months) – At time of index offense	Ever married = -2 Never married = +1
5	Criminal history score for convictions and charges for nonviolent offenses prior to the index offense (from the Cormier-Lang system)	Score 0 = -2 Score 1 or 2 = 0 Score 3 or over = +3
6	Criminal history score for convictions and charges for violent offenses prior to the index offense (from the Cormier-Lang system)	Score 0 = -1 Score 2 = 0 Score 3 or above = +6
7	Number of previous convictions for sexual offenses (pertains to convictions for sexual offenses that occurred prior to the index offense) – Count any offenses known to be sexual, including, for example, indecent exposure	0 = -1 1 or 2 = +1 ≥ 3 = +5
8	History of sex offenses against girls under age 14 <i>only</i> (includes index offense; if offender was less than 5 years older than victim, always score +4)	Yes = 0 No = +4
9	Failure on prior conditional releases (includes parole violation or revocation; breach of or failure to comply with recognizance or probation; bail violation; and any new charges, including the index offense, while on conditional release)	No = 0 Yes = +3
10	Age at index offense (at most recent birthday)	≥ 39 = -5 34–38 = -2 28–33 = -1 27 = 0 ≤ 26 = +2
11	Meets DSM-III criteria for any personality disorder	No = -2 Yes = +3
12	Meets DSM-III criteria for schizophrenia	Yes = -3 No = +1
13	Phallometric test results	<i>All</i> indicate nondeviant sexual preferences = -1 <i>Any</i> test indicates deviant sexual preferences = +1
14	Hare Psychopathy Checklist-Revised score (PCL-R; Hare 1991)	≤ 4 = -5 5–9 = -3 10–14 = -1 15–24 = 0 25–34 = +4 ≥ 35 = +12

Die einzelnen Itemscores der 14 Items werden anschließend zum Gesamtscore addiert, von dem aus wiederum auf eine der insgesamt neun normierten Risikokategorien geschlossen werden kann. Des Weiteren kann auf prozentuale Schätzwerte für die Rückfallwahrscheinlichkeit nach sieben bzw. zehn Jahren zurückgegriffen werden, die anhand der oben genannten Konstruktionsstichprobe ermittelt wurden. Zusätzlich stellen *Quinsey et al.* (2006) eine Tabelle mit Perzentilwerten in Bezug auf die Gesamtscoreverteilung zur Verfügung. Wie aus der dargestellten Tabelle ersichtlich wird, erfordert die Anwendung des SORAG in seiner Originalversion neben einer ausführlichen psychopathologischen Diagnostik die Ergebnisse phallometrischer Untersuchungen und die Anwendung der PCL-R. Zu Letzterem sei erwähnt, dass von den Autoren eine Alternative zur PCL-R angeboten und ausführlich beschrieben wird: die sog. *Child and Adolescent Taxon Scale* (CATS), auf die an dieser Stelle nicht näher eingegangen werden soll, wobei hinzugefügt werden muss, dass deren Anwendung anstelle der PCL-R von *Quinsey* und Mitarbeitern (2006) explizit empfohlen wird und die für Forschungszwecke bereits verwendet wurde (*Bartosh et al.* 2003).

Sowohl VRAG als auch SORAG wurden vor allem im nordamerikanischen Raum einer Reihe von Validierungsstudien unterzogen und lieferten fast ausschließlich zufriedenstellende Vorhersageergebnisse, die in der bereits zitierten aktuellen Meta-Analyse von *Hanson & Morton-Bourgon* (2007) bestätigt wurden: Die berichtete Vorhersageleistung des Instruments reichte dabei von $AUC = .64$ (*Bartosh et al.* 2003) über $AUC = .69$ (*Nunes et al.* 2002) bis zu $AUC = .72$ (*Ducro & Pham* 2006) bzw. $AUC = .73$ (*Harris et al.* 2003; *Barbaree et al.* 2001)⁴. Abhängig von Rückfallkriterium und Subgruppe sexualdelinquenten Täter wurden Vorhersageleistungen bis zu einem Wert von $AUC = .93$ (innerfamiliäre pädosexuelle Täter) berichtet (*Bartosh et al.* 2003). Bis auf wenige Ausnahmen zeigte der SORAG nach den gängigen Bewertungskriterien für Prognoseinstrumente demnach moderate bis hohe Validitätsindizes.

1.2 Die deutsche Version des SORAG

Der SORAG war auch im deutschsprachigen Raum bereits Gegenstand wissenschaftlicher Untersuchungen und erreichte hierbei zufriedenstellende Prognosewerte (*Quenzer* 2005), allerdings fehlt bislang eine einheitliche und öffentlich zugängliche Übersetzung des Instruments. Ein Anliegen der vorliegenden Arbeit besteht deshalb darin, eine deutschsprachige Adaption des Instruments bereitzustellen, in dem die einzelnen Items auf den deutschen Sprachraum übertragen und mit eindeutig formulierten Operationalisierungen versehen werden (*Rettenberger & Eber* 2007)⁵. Die deutschsprachige Adaption des SORAG basiert auf einer für die interne Verwendung an der *Zentralen Dokumentations- und Koordinationsstelle für Sexualstraftäter (Begutachtungsstation)* in Wien übersetzte Version des Instruments. Diese Übersetzung wurde überarbeitet (ohne die inhaltliche Ausarbeitung der einzelnen Items zu verändern) und die intern verwendeten Operationalisierungen manualisiert, um das Instrument weiteren potentiellen Nutzern zugänglich und seine Anwendung praktikabel zu machen.

Bedingt durch die Unterschiede in den forensischen Kontexten im nordamerikanischen und deutschen Sprachraum, aber auch durch die Weiterentwicklung des Diagnosemanuals

4 Die genannten AUC-Werte beziehen sich auf gewalttätige (inklusive sexuell motivierte Gewalttaten) Rückfälle, sprich auf das Kriterium, für das das SORAG laut *Quinsey et al.* (2006) konzipiert wurde. Für weitere Validitätsindizes und Rückfallkriterien sei auf die Originalarbeiten verwiesen.

5 Neben anderen deutschsprachigen Prognosemanualen und -instrumenten kann das deutschsprachige Manual des SORAG über den *Verein für Forensische Forschung und Weiterbildung (VFFW)* unter www.vffw.org bezogen werden.

DSM bestand die Notwendigkeit, Veränderungen an einzelnen Definitionen und Operationalisierungen vorzunehmen. Explizit erwähnt werden soll an dieser Stelle das Item 13 (Original: *Phallometric test results*): Da im deutschsprachigen Raum keine phallometrischen Untersuchungen vorgenommen werden⁶, wurde als alternative Operationalisierung der sexuellen Devianz eines Straftäters die Diagnosekriterien einer nach den offiziellen psychiatrischen Diagnosesystemen ICD-10 (*Dilling, Mombour & Schmidt* 2004) bzw. DSM-IV (*American Psychiatric Association* 2000) diagnostizierten Störung der Sexualpräferenz (ICD-10) bzw. Paraphilie (DSM-IV) verwendet. Auch andere, ursprünglich nach DSM-III diagnostizierte Störungen wurden durch Diagnosekriterien des DSM-IV oder ICD-10 ersetzt.

1.3 Fragestellungen

Neben der Darstellung der Interraterreliabilität und der konvergenten Validität ist Hauptgegenstand der vorliegenden Arbeit die Überprüfung der prädiktiven Validität der vorgestellten deutschsprachigen Adaption des SORAG. Die prädiktive Validität eines Prognoseinstruments wird naturgemäß durch eine Gegenüberstellung vorhergesagter und tatsächlicher Rückfallereignisse gemessen. In der vorliegenden Studie wurde das Konstrukt »Rückfälligkeit« anhand unterschiedlicher Kriterien erfasst (ausführlich dazu siehe bei *Nedopil* 2005): Zunächst wurde nach den üblichen Deliktcharakteristika zwischen allgemeiner (jede neuerliche Eintragung im Strafregisterauszug, ohne Einschränkung auf bestimmte Delikttypen), einschlägiger (sexuell motivierter) und gewalttätiger (inklusive bzw. exklusive sexuell motivierter) Rückfälligkeit unterschieden. Des Weiteren wurde im Hinblick auf die Sanktionierung zwischen den beiden Kriterien »Verurteilung allgemein« und »Verurteilung zu einer Haftstrafe« differenziert, um auch die Schwere der Rückfalltat zu erfassen. Es sei an dieser Stelle nochmals darauf hingewiesen, dass der SORAG ursprünglich ausschließlich für die Vorhersage von gewalttätigen (inklusive sexuell motivierten, sog. »Hands-on«-Delikten) Rückfällen konzipiert wurde. Aus diesem Grund wurde außerdem das Rückfallkriterium »Schwere gewalttätige Rückfälligkeit« definiert, das sowohl gewalttätige als auch sämtliche sexuell motivierten »Hands-on«-Delikte berücksichtigt mit der zusätzlichen Forderung, dass diese zur Verhängung einer Freiheitsstrafe führten.

Die forensische Prognoseforschung leidet oftmals unter einer mangelhaften Berücksichtigung der Heterogenität der Gesamtgruppe (*Dahle* 2007). Aus diesem Grund wurde die prädiktive Validität auch für forensisch relevante Untergruppen bestimmt. Dieses Vorgehen wird als Überprüfung der differentiellen Validität bezeichnet, deren Hauptgegenstand darin besteht, relevante Moderatorvariablen zu identifizieren, anhand derer Validitätsunterschiede in der Gesamtpopulation feststellbar sind.

2. Methode

2.1 Datenerhebung und Datenauswertung

Die notwendigen Daten wurden im Sinne eines prospektiv-längsschnittlichen Studiendesigns im Zuge einer umfassenden kriminalprognostischen Begutachtung, die in den Jahren 2002 bis 2004 an der *Zentralen Dokumentations- und Koordinationsstelle für Sexualstraftäter des Österreichischen Strafvollzugs* in Wien vorgenommen wurde, von forensisch erfah-

6 Unabhängig davon muss die Interpretation und diagnostische Verwendung phallometrischer Testergebnisse allgemein mit Skepsis betrachtet werden (ausführlich bei *Marshall* 2005).

renen und im Umgang mit aktuarischen Prognoseinstrumenten theoretisch und praktisch geschulten Mitarbeitern erhoben.

Die Indizes der Interraterreliabilität und der konvergenten Validität wurden anhand von Intraklassenkorrelationen (ICC) bzw. von Produkt-Moment-Korrelationen quantifiziert.

Um die Frage der prädiktiven Validität des SORAG zu beantworten, erfolgte eine aktuelle Auswertung der Strafregistrauszüge (Stand: Mai 2007), die vom Zweitautor vorgenommen wurde. In jedem Stadium der Datenerhebung waren alle beteiligten Personen blind bzgl. anderer relevanter Ergebnisse. Es wurden neben Produkt-Moment-Korrelationen auch die aus den ROC-Analysen (Analysen der *Receiver Operation Characteristics*) resultierenden AUC-Werte (Maßzahlen für die *Area Under Curve*) berechnet. Die ROC-Analysen gehen auf die Signaldeckungstheorie zurück (z.B. Egan 1975) und weisen gegenüber anderen Verfahren vor allem deshalb Vorteile auf, weil sie weitgehend unabhängig von der Basisrate sind (Hanley & McNeil 1982). Die ROC-Kurven stellen für jeden möglichen Ausprägungsgrad des Prognoseinstruments den Anteil der jeweils korrekt positiv klassifizierten Täter (in der vorliegenden Arbeit folglich die richtig identifizierten rückfälligen Sexualstraftäter) und die zugehörige Rate falsch-positiver Probanden (in diesem Fall diejenigen Straftäter, die trotz schlechter Prognose nicht rückfällig wurden) graphisch dar. Die AUC-Maße, die auf der dabei ermittelten und quantifizierten Fläche unter der Kurve basieren, können als Validitätsmaß für ein Prognoseinstrument verwendet werden, wobei ein AUC-Wert von 0,5 genau der Zufallswahrscheinlichkeit entspricht, während ein AUC-Wert von 1,0 eine in jedem Untersuchungsfall korrekte Prognosebeurteilung bedeuten würde. Dable (2007) gibt unter Bezugnahme auf die von Cohen (1992) formulierten Kriterien für die Interpretation von AUC-Werten und Korrelationskoeffizienten an, dass nach allgemeinen Standards die Vorhersagegüte von Prognoseinstrumenten bei Werten von $r \geq .37$ ($AUC \geq .72$) als »hoch«, von $r \geq .24$ ($AUC \geq .64$) als »moderat« und bei darunter liegenden, signifikanten Werten als »schwach« einzustufen ist.

2.2 Stichprobe

Die 178 Sexualstraftäter, deren Daten und Prognosegutachten bei der vorliegenden Studie verwendet werden konnten, wurden zwischen April 1992 und März 2004 in Österreich aufgrund einer sexuell motivierten Straftat rechtskräftig zur Verbüßung einer un- bzw. teilbedingten Freiheitsstrafe im Strafvollzug verurteilt und aufgrund dieser Verurteilung in den Jahren 2002 bis 2004 an der *Zentralen Dokumentations- und Koordinationsstelle für Sexualstraftäter des Österreichischen Strafvollzugs* in Wien u.a. kriminalprognostisch begutachtet. Das durchschnittliche Alter der Probanden betrug zum Zeitpunkt der Auswertung der Strafregistrauszüge $M = 43,63$ Jahre (Standardabweichung $SD = 12,75$ Jahre), wobei der jüngste 22 und der älteste 74 Jahre alt war. 9,6 % ($n = 17$) der Straftäter besuchten die Sonderschule und/oder wiesen keinen Pflichtschulabschluss auf, wohingegen 38,2 % ($n = 68$) die Pflichtschule erfolgreich absolvierten. 48,9 % ($n = 87$) verfügten über eine abgeschlossene Berufsausbildung oder hatten eine Fachschule besucht. 2,8 % ($n = 5$) der Verurteilten besaßen die allgemeine Hochschulreife und einer (0,6 %) einen universitären Abschluss. Die den aktuellen Verurteilungen zugrundeliegenden Anlasstaten (»Index-Delikte«) verteilten sich wie folgt: 48,9 % ($n = 87$) wurden aufgrund von Vergewaltigungsdelikten, 48,3 % ($n = 86$) aufgrund von pädosexuellen Straftaten verurteilt, wobei bei Letztgenannten bezogen auf die Gesamtgruppe 20,2 % ($n = 36$) wegen extrafamiliärer und 28,1 % ($n = 50$) wegen innerfamiliärer Taten verurteilt und inhaftiert wurden. Von den verbleibenden fünf Untersuchten wurden 2,8 % ($n = 4$) aufgrund eines sog. »Hands-off-«-Delikts und 0,6 % ($n = 1$) aufgrund eines sexuell assoziierten Tötungsdelikts zu einer unbedingten Freiheitsstrafe verurteilt. Die aufgrund des Index-Delikts ausgesprochene Strafe betrug laut Urteil im Durchschnitt

M = 31,75 Monate (SD = 23,27 Monate) mit einem Minimum von 4 und einem Maximum von 157 Monaten.

Die kriminelle Vorbelastung der Täter wurde anhand der zum Zeitpunkt der Begutachtung angeforderten Strafregisterauszüge ermittelt (siehe *Tabelle 2*). Des Weiteren wurde die kriminelle Vorbelastung subgruppenspezifisch erhoben.

Tabelle 2 *Kriminelle Vorbelastung der Gesamtstichprobe (N = 178) und für die Subgruppen der Vergewaltigungstäter (n = 87) und der extrafamiliären (n = 36) und innerfamiliären (n = 50) pädosexuellen Täter*

Vordelinquenzkriterium	Gesamtstichprobe (N = 178)	Vergewaltigungstäter (n = 87)	extrafamiliäre pädosexuelle Täter (n = 36)	innerfamiliäre pädosexuelle Täter (n = 50)
Anteil vorbestrafter Probanden	60,7 % (n = 108)	72,4 % (n = 63)	55,6 % (n = 20)	46,0 % (n = 23)
Gesamtanzahl der Vordelikte	M = 4,03 (SD = 6,52)	M = 4,21 (SD = 5,97)	M = 5,86 (SD = 9,59)	M = 2,02 (SD = 4,10)
gewalttätige (nicht sexuell motivierte) Vordelinquenz	M = 1,16 (SD = 2,39)	M = 1,61 (SD = 2,77)	M = 0,86 (SD = 2,10)	M = 0,60 (SD = 1,78)
sexuell motivierte Vordelinquenz	M = 0,66 (SD = 2,09)	M = 0,29 (SD = 0,65)	M = 2,06 (SD = 4,11)	M = 0,14 (SD = 0,50)
Strafmaß Vordelinquenz (Summe verurteilter Freiheitsstrafen in Monaten)	M = 35 (SD = 43,96)	M = 28,42 (SD = 35,96)	M = 64,13 (SD = 55,15)	M = 18,89 (SD = 25,74)

3. Ergebnisse

3.1 Interraterreliabilität

Der Intraklassenkorrelationskoeffizient (Single Measure) betrug für den SORAG ICC = .93. Die Reliabilitätskoeffizienten der Beurteilerübereinstimmung wurden auch für die zur Überprüfung der konvergenten Validität herangezogenen Prognoseinstrumente erhoben und lagen für das Static-99 bei ICC = .98, für den SVR-20 bei ICC = .84 und für die PCL-R ebenfalls bei ICC = .93. Alle Korrelationen waren hoch signifikant ($p < .001$).

3.2 Konvergente Validität

Zur Überprüfung der konvergenten Validität wurden Produkt-Moment-Korrelationen zwischen den Gesamtscores des SORAG auf der einen Seite und den Gesamtscores des Static-99, des SVR-20 und der PCL-R auf der anderen Seite berechnet. Die Korrelation des SORAG mit dem Static-99 betrug $r = .72$, mit dem SVR-20 $r = .78$ und mit der PCL-R $r = .79$, wobei alle Korrelationskoeffizienten hoch signifikant waren ($p < .001$).

3.3 Rezidivraten und Katamnesezeitraum

Zum Zeitpunkt der Auswertung der aktuellen Strafregisterauszüge lag der durchschnittliche Follow-up-Zeitraum bei M = 42,58 Monaten (SD = 7,74), wobei der Entlassungszeitpunkt der Probanden mindestens 30 Monate zurückliegen musste, um in die Gesamtstichprobe aufgenommen zu werden. Das Minimum lag bei 31,10 Monaten, das Maximum bei 65,75 Monaten.

Die Rezidivraten für die genannten Rückfallkriterien sind in *Tabelle 3* dargestellt, wobei auch hier – neben der Rückfälligkeit für die Gesamtstichprobe – die relevanten Subgruppen der Vergewaltigungstäter und der extra- und innerfamiliären pädosexuellen Täter angeführt wurden. Bei insgesamt fünf Probanden der Gesamtstichprobe erfolgte keine Zuteilung zu einer der drei genannten Subgruppen, da es sich beim Index-Delikt zwar um eine sexuell motivierte Straftat handelte, eine eindeutige Zuordnung zu den Subkategorien aber nicht möglich war.

Tabelle 3 Darstellung der Rezidivraten nach unterschiedlichen Rückfallkriterien für die Gesamtstichprobe (N = 178) und die Subgruppen der Vergewaltigungstäter (n = 87) und der extra- (n = 36) und innerfamiliären (n = 50) pädosexuellen Täter

Rückfallkriterium	Gesamtstichprobe (N = 178)	Vergewaltigungstäter (n = 87)	extrafamiliäre pädosexuelle Täter (n = 36)	innerfamiliäre pädosexuelle Täter (n = 50)
allgemeine Rückfälligkeit (Verurteilung)	29,2 % (n = 52)	37,9 % (n = 33)	25,0 % (n = 9)	14,0 % (n = 7)
allgemeine Rückfälligkeit (Haft)	18,0 % (n = 32)	25,3 % (n = 22)	22,2 % (n = 8)	2,0 % (n = 1)
sexuell motivierte Rückfälligkeit (Verurteilung)	5,1 % (n = 9)	1,1 % (n = 1)	16,7 % (n = 6)	0 % (n = 0)
sexuell motivierte Rückfälligkeit (Haft)	4,5 % (n = 8)	1,1 % (n = 1)	16,7 % (n = 6)	0 % (n = 0)
gewalttätige Rückfälligkeit (Verurteilung)	14,6 % (n = 26)	26,4 % (n = 23)	2,8 % (n = 1)	4,0 % (n = 2)
gewalttätige Rückfälligkeit (Haft)	9,6 % (n = 17)	18,4 % (n = 16)	2,8 % (n = 1)	0 % (n = 0)
gewalttätige (inkl. sexuell motivierte) Rückfälligkeit (Verurteilung)	19,1 % (n = 34)	26,4 % (n = 23)	19,4 % (n = 7)	4,0 % (n = 2)
gewalttätige (inkl. sexuell motivierte) Rückfälligkeit (Haft)	13,5 % (n = 24)	18,4 % (n = 16)	19,4 % (n = 7)	0 % (n = 0)
schwere gewalttätige Rückfälligkeit (inkl. sexuell motivierte »Hands-on«-Delikte [Haft])	11,2 % (n = 20)	18,4 % (n = 16)	8,3 % (n = 3)	0 % (n = 0)

3.4 Prädiktive Validität

In *Tabelle 4* sind sämtliche Indizes der prädiktiven Validität für die Gesamtstichprobe und für die drei Subgruppen der Vergewaltigungstäter und der extra- bzw. innerfamiliären pädosexuellen Täter dargestellt. Für die Rückfallkriterien »sexuell motivierte Rückfälligkeit (Verurteilung)«, »sexuell motivierte Rückfälligkeit (Haft)«, »gewalttätige Rückfälligkeit (Haft)«, »gewalttätige (inkl. sexuell motivierte) Rückfälligkeit (Haft)« und »schwere gewalttätige Rückfälligkeit (inkl. sexuell motivierte »Hands-on«-Delikte [Haft])« war eine Darstellung der Validitätsindizes bei den innerfamiliären Tätern nicht möglich, da keine Rückfallereignisse vorlagen.

Tabelle 4 *Indizes der prädiktiven Validität des SORAG im Hinblick auf unterschiedliche Rückfallkriterien für die Gesamtstichprobe (N = 178) und die Subgruppen der Vergewaltigungstäter (n = 87) und der extra- (n = 36) und innerfamiliären (n = 50) pädosexuellen Täter*

Rückfallkriterium	Gesamtstichprobe (N = 178)	Vergewaltigungstäter (n = 87)	extrafamiliäre pädosexuelle Täter (n = 36)	innerfamiliäre pädosexuelle Täter (n = 50)
allgemeine Rückfälligkeit (Verurteilung)	AUC = .72** (r = .35**)	AUC = .72** (r = .37**)	AUC = .72 n.s.: p = .06 (r = .28*)	AUC = .48 n.s. (r = .07 n.s.)
allgemeine Rückfälligkeit (Haft)	AUC = .82** (r = .42**)	AUC = .76** (r = .39**)	AUC = .81** (r = .41**)	AUC = .98 n.s. (r = .33**)
sexuell motivierte Rückfälligkeit (Verurteilung)	AUC = .73** (r = .17*)	AUC = .56 n.s. (r = .03 n.s.)	AUC = .79* (r = .34*)	– ^a
sexuell motivierte Rückfälligkeit (Haft)	AUC = .73** (r = .16*)	AUC = .56 n.s. (r = .03 n.s.)	AUC = .79* (r = .34*)	– ^a
gewalttätige Rückfälligkeit (Verurteilung)	AUC = .72** (r = .27**)	AUC = .72** (r = .33**)	AUC = .58 n.s. (r = .02 n.s.)	AUC = .11 n.s. (r = –.21 n.s.)
gewalttätige Rückfälligkeit (Haft)	AUC = .80** (r = .31**)	AUC = .76** (r = .35**)	AUC = .58 n.s. (r = .02 n.s.)	– ^a
gewalttätige (inkl. sexuell motivierte) Rückfälligkeit (Verurteilung)	AUC = .74** (r = .32**)	AUC = .72** (r = .33**)	AUC = .76* (r = .33*)	AUC = .11 n.s. (r = –.21 n.s.)
gewalttätige (inkl. sexuell motivierte) Rückfälligkeit (Haft)	AUC = .80** (r = .36**)	AUC = .76** (r = .35**)	AUC = .76* (r = .33*)	– ^a
schwere gewalttätige Rückfälligkeit (inkl. sexuell motivierte »Hands-on«-Delikte [Haft])	AUC = .80** (r = .33**)	AUC = .76** (r = .35**)	AUC = .65 n.s. (r = .11 n.s.)	– ^a

* Die Korrelation bzw. der AUC-Wert ist auf dem Niveau von 0,01 (1-seitig) signifikant.

** Die Korrelation bzw. der AUC-Wert ist auf dem Niveau von 0,05 (1-seitig) signifikant.

n.s. = nicht signifikant

^a = kann nicht berechnet werden, da mindestens eine der Variablen konstant ist bzw. im Sinne der ROC-Kurven die Zustandsgruppen keine gültigen Beobachtungen aufweisen.

4. Diskussion

4.1 Interraterreliabilität

Der berichtete Intraklassenkorrelationskoeffizient belegt eine zufriedenstellende Beurteilerübereinstimmung der adaptierten deutschsprachigen Version des SORAG. Die Höhe des Interraterreliabilitätsindex ist mit ICC = .93 vergleichbar mit den Ergebnissen angloamerikanischer (Harris et al. 2003; Bartosh et al. 2003) bzw. europäischer (Ducro & Pham 2006) Studien.

Die Indizes der Beurteilerübereinstimmung der zur Überprüfung der konvergenten Validität herangezogenen Prognoseinstrumente sind ebenfalls als zufriedenstellend zu interpretieren (Greve & Wentura 1997) und in ihrer Höhe mit dem ICC-Wert des SORAG vergleichbar. Der mit ICC = .84 niedrigere Wert des SVR-20 liegt in der Konzeption des Instruments als strukturierte Checkliste für die klinisch-idiographische Prognosebegutachtung begründet. Im Vergleich zum SVR-20 liegt ein entscheidender Vorteil statistisch-nomothetisch

tischer und aktuarischer Instrumente in einer völlig regelgeleiteten Vorgehensweise (Dahle 2005), die sich in entsprechend höheren Interraterreliabilitätsindizes ausdrückt.

4.2 Konvergente Validität

Die im Ergebnisteil dargestellten Validitätskennwerte belegen die konvergente Validität des SORAG: Die Produkt-Moment-Korrelationen zwischen dem SORAG einerseits und dem Static-99, dem SVR-20 und der PCL-R andererseits sind mit Werten zwischen $r = .72$ und $r = .79$ als hoch einzustufen und sämtlich hochsignifikant ($p < .001$). Diese Validitätsindizes sind teilweise identisch mit den Werten, die in angloamerikanischen Vergleichsstudien berichtet werden (Nunes et al. 2002), liegen jedoch zum Teil deutlich über vorangegangenen Untersuchungen (z.B. Harris et al. 2003; Barbaree et al. 2001; Ducro & Pham 2006). Der hohe Korrelationskoeffizient zwischen dem SORAG und der PCL-R ist nicht überraschend, ist doch der PCL-R-Wert einerseits in der Kodierung des SORAG integriert (vgl. Item 14), und ist die PCL-R andererseits dem SORAG auch konstruktnahe (Erfassung der Wahrscheinlichkeit späterer Gewalttaten). Die hier berichteten Werte sind Beweise für die Validität der deutschen SORAG-Version, handelt es sich doch bei den Vergleichsinstrumenten um in Deutschland und Österreich validierte Prognoseinstrumente (Rettenberger & Eher 2006a; Stadland et al. 2006).

4.3 Rezidivraten und Katamnesezeitraum

Der durchschnittliche Katamnesezeitraum ist mit $M = 42,58$ ($SD = 7,74$ Monate) Monaten vergleichbar mit einer Reihe vorangegangener, inhaltlich verwandter Studien, wobei diesbezüglich eine große Heterogenität zwischen den Studien besteht. Wir wählten in unserer Studie einen Mindest-Follow-up-Zeitraum von 30 Monaten, das heißt jeder der beteiligten Sexualstraftäter musste mindestens einen möglichen Nachuntersuchungszeitraum (und damit auch die Möglichkeit zum Rückfall) von 2,5 Jahren aufweisen. Obwohl vor allem die erste Phase nach der Entlassung für die Rückfälligkeit von Straftätern als besonders relevant angesehen wird (Schmucker 2004), sind Kriminalprognosestudien mit kürzeren als von uns verwendeten Katamnesezeiträumen in ihrer Aussagekraft sehr beschränkt. Die im Vergleich zu anderen Untersuchungen niedrige Standardabweichung des Katamnesezeitraums zeigt, dass in unserer Studie homogene Follow-up-Zeiträume vorlagen, die sich auf die Interpretierbarkeit der Ergebnisse positiv auswirken⁷. Auffallend ist, dass die in der vorliegenden Arbeit berichteten Rückfallraten teilweise deutlich niedriger sind als die in früheren angloamerikanischen (z.B. Hanson & Bussière 1998) oder deutschsprachigen Untersuchungen (z.B. Egg 2004). Dies mag zum einen daran liegen, dass der hier berichtete Katamnesezeitraum relativ kurz ist, zum anderen könnte die niedrigere Rückfälligkeit auch auf verbesserte Betreuungsangebote und Behandlungskonzepte zurückzuführen sein, die im österreichischen Strafvollzug und in Form von diversen Nachbetreuungsinstitutionen in den letzten Jahren in Österreich implementiert wurden (Eher 2007; Eher u.a. 2006). Die besonders niedrigen Rezidivraten der innerfamiliären pädosexuellen Täter müssen mit Vorsicht interpretiert werden, da neuere Untersuchungen die Hypothese stützen, dass im Vergleich zu anderen Subgruppen bei jeweils gleichen Katamnesezeiträumen nicht zwangsläufig die Rückfallraten bei den innerfamiliären pädosexuellen Tätern niedriger sind, sondern lediglich die Wiederverurteilungsraten, da bei innerfamiliären Delikten der Zeitraum zwischen Delikt

7 Die niedrige Standardabweichung ist ein Beleg für die geringe Variabilität der Katamnesezeiträume innerhalb der Stichprobe. Damit wird gewährleistet, dass die Basiswahrscheinlichkeiten für die Rückfälligkeit so homogen wie möglich für alle Untersuchten sind.

und polizeilicher Anzeige, Anklage und Verurteilung durchschnittlich bedeutend länger ist (Eher & Ross 2006).

4.4 Prädiktive Validität

Auf den ersten Blick scheinen die Ergebnisse der prädiktiven Validität der vorliegenden Arbeit die Resultate vorangegangener Untersuchungen zunächst zu bestätigen und eine uneingeschränkte Empfehlung der Anwendung des SORAG für kriminalprognostische Einschätzungen zu rechtfertigen. Bei Verwendung der AUC-Werte als Referenzsystem sind die Validitätsindizes für die Gesamtstichprobe bezogen auf die aktuell gültigen Bewertungskriterien für Prognoseinstrumente (Dable 2007) ausschließlich als hoch (und damit sehr gut) einzustufen. Sämtliche Vorhersagewerte liegen bei $AUC = .72$ oder höher und dokumentieren damit die gute Prognoseleistung des Instruments. Von besonderer Bedeutung sind dabei die Rückfallkriterien »gewalttätige (inklusive sexuell motivierte) Rückfälligkeit (Verurteilung)« und »gewalttätige (inklusive sexuell motivierte) Rückfälligkeit (Haft)«, da das Instrument ursprünglich für diese Rückfallereignisse konzipiert wurde (Quinsey et al. 2006). Anhand der als hoch einzustufenden AUC-Werte von $.74$ bzw. $.80$ konnten bisherige Befunde aus angloamerikanischen Studien repliziert werden (z.B. Barbaree et al. 2001; Harris et al. 2003). In der Gesamtgruppe zeigte das Instrument bei der wichtigen Zielvariablen der sexuell motivierten Rückfälligkeit ebenfalls eine hohe Vorhersageleistung, die zum Teil deutlich über den Werten vorangegangener Studien lag (z.B. Harris et al. 2003; Nunes et al. 2002; Bartosh et al. 2003). Um aus viktimologischer Sicht auf besonders gravierende Rückfälle zu fokussieren, wurde in der vorliegenden Studie ein weiteres, besonders restriktives Rückfallkriterium kreiert: Ein besonders gravierendes Rückfalldelikt lag dann vor, wenn der Täter zu einer erneuten Haftstrafe aufgrund eines gewalttätigen und/oder sexuell motivierten »Hands-on«-Delikts verurteilt wurde (siehe Tabelle 5: »schwere gewalttätige Rückfälligkeit [inklusive sexuell motivierte »Hands-on«-Delikte (Haft)]«). Mit einem AUC-Wert von $.80$ kann auch dieser Validitätsindex als sehr zufriedenstellend bezeichnet werden. Auch wenn sich die Rezidivraten zwischen den einzelnen Rückfallkriterien oft nicht sonderlich unterscheiden, soll mit Hilfe des zusätzlich aufgenommenen letztgenannten Kriteriums dem Umstand Rechnung getragen werden, dass es vor allem die gravierenden »Hands-on«-Rückfalldelikte sind, die die öffentliche und politische Diskussion lenken. Zu Recht werden von verschiedenen Seiten kriminalprognostische Methoden gefordert, die im Stande sind, gerade diese Straftaten valide vorherzusagen. Darüber hinaus muss es als ein zusätzliches Qualitätsmerkmal des SORAG angesehen werden, dass bei höherschweligen Rückfallkriterien tendenziell bessere Prognoseergebnisse erzielt wurden. Die Vorhersageleistung des SORAG für das Kriterium »allgemeine Rückfälligkeit« war mit einem AUC-Wert von $.72$ bzw. $.82$ ebenfalls als hoch einzustufen.

Die ausschließlich positiven Prognoseleistungen für die Gesamtgruppe werden zum Teil durch die Analyse der einzelnen Subgruppen relativiert. Anhand vorangegangener Prognosestudien (Bartosh et al. 2003) und aufgrund des gesicherten empirischen Wissens über die Heterogenität der Gesamtpopulation sexuell motivierter Straftäter (z.B. Fiedler 2004) wurde in der vorliegenden Studie die Gesamtstichprobe anhand des Index-Delikts in folgende drei Subgruppen unterteilt: Vergewaltigungstäter und extra- bzw. innerfamiliäre pädosexuelle Täter.

Für die Subgruppe der Vergewaltigungstäter konnten die hohen Validitätswerte für die meisten Rückfallkriterien bestätigt werden: Mit Ausnahme der sexuell motivierten Rückfälligkeit lagen sämtliche AUC-Werte bei $.72$ oder darüber und gelten somit als Beleg für die hohe prädiktive Validität des SORAG bei diesen Straftätern. Für die sexuell motivierte Rückfälligkeit wurde mit $AUC = .56$ ein (nicht signifikanter) Wert ermittelt, der nahe an der

Zufallswahrscheinlichkeit liegt. Allerdings muss hierbei berücksichtigt werden, dass lediglich ein Täter von den insgesamt 87 dieser Subgruppe auch tatsächlich eine erneute sexuell motivierte Straftat beging. Auch wenn ROC-Analysen prinzipiell unabhängig von der jeweiligen Basisrate sind, können nur solche Ereignisse prognostiziert werden, die auch tatsächlich auftreten. Der einzige Rückfalltäter dieser Subgruppe wurde mit einem Gesamtscore von 14 Punkten in die fünfte (mit einer aus der Originalversion geschätzten Rückfallwahrscheinlichkeit nach sieben Jahren Follow-up-Zeitraum von 45 % bzw. nach zehn Jahren mit einer Rückfallwahrscheinlichkeit von 76 %) von insgesamt neun Risikokategorien zugeteilt, womit eine durchschnittliche bis hohe Rückfallwahrscheinlichkeit prognostiziert werden würde. Demgegenüber ist in dieser Subgruppe auf den hohen Anteil Falsch-Positiver hinzuweisen, d.h. auf den Straftäteranteil, bei dem Rückfälle zwar mit hoher Wahrscheinlichkeit prognostiziert wurden, die aber tatsächlich innerhalb des untersuchten Katamnesezeitraums nicht rückfällig wurden. Anhand der Normtabellen der Originalversion würde bei 14 Begutachteten von den insgesamt 87 Vergewaltigungstätern nach sieben Jahren eine Rückfallwahrscheinlichkeit von 75 %, bei sechs Untersuchten gar eine von 100 % prognostiziert werden. Zumindest nach durchschnittlich ca. der Hälfte des fraglichen Beobachtungszeitraumes beging keiner dieser Begutachteten erneut eine sexuell motivierte Straftat. Freilich sind bei dieser Gegenüberstellung einige beträchtliche Einschränkungen zu machen: Zum einen handelt es sich streng genommen nicht um Rückfallraten im Sinne von Delikt-Wiederbegehungsraten, sondern um Wiederverurteilungs- bzw. Wiederinhaftierungsraten mit zusätzlich relativ kurzen Follow-up-Zeiträumen. Zum anderen ist das beispielhaft dargestellte Rückfallkriterium, bei dem der hohe Anteil Falsch-Positiver auffällt, eben nicht jenes Kriterium, für das der SORAG konzipiert wurde (das Instrument wurde konstruiert, um gewalttätige *und/oder* sexuell motivierte Rückfälle zu prognostizieren). Bei genau diesem Kriterium zeigt das Instrument jedoch hohe Prognoseleistungen und in Relation zum Anteil der Validen-Positiven einen niedrigeren Anteil Falsch-Positiver.

Bei den extrafamiliären pädosexuellen Tätern zeigt sich im Vergleich zur Subgruppe der Vergewaltigungstäter ein ähnliches Ergebnismuster – allerdings mit unterschiedlichen Rückfallkriterien. Für die allgemeine, die sexuell motivierte und die gewalttätige (inklusive sexuell motivierte) Rückfälligkeit (jeweils sowohl erneute Verurteilung als auch Inhaftierung) liegen die Indizes der prädiktiven Validität des SORAG zum Teil deutlich über dem kritischen Wert von $AUC = .72$ und belegen somit die hohe Prognoseleistung des Instruments. Für Kategorien mit reiner nicht sexuell motivierter Gewalttätigkeit lagen die Werte jedoch nahe der Zufallswahrscheinlichkeit von $AUC = .50$. Allerdings sind in diesen Kriterien auch die Rezidivraten äußerst niedrig.

Die vorhandenen Validitätsindizes der Subgruppe der innerfamiliären pädosexuellen Täter sind aufgrund der zu geringen Rezidivraten nicht sinnvoll interpretierbar. Darüber hinaus wurde bei fünf der insgesamt neun Kriterien laut Strafregisterauszug überhaupt kein Rückfall verzeichnet. Es gibt allerdings Studien, die eine Unterteilung von pädosexuellen Sexualdelinquenten in inner- bzw. extrafamiliäre Täter insgesamt in Frage stellen, da laut aktuellen empirischen Erkenntnissen der Überschneidungsbereich zwischen beiden Gruppen viel größer als ursprünglich angenommen ist (Studer & Aylwin 2006). So gesehen ist es nicht unwahrscheinlich, dass nach einem längeren Katamnesezeitraum ähnlich gute Vorhersagewerte wie bei der Gruppe der extrafamiliären Täter vorliegen.

Zusammenfassend dokumentieren die Resultate den Wert der deutschsprachigen Version des SORAG. Einerseits zeigen sich durchwegs gute Validitätswerte für die Gesamtgruppe. Andererseits zeigt sich in der Vergewaltigergruppe eine gute Vorhersageleistung für die forensisch relevante Kategorie der schweren Gewaltdelikte und in der Gruppe der extra-

familiären pädosexuellen Täter eine hervorragende Vorhersageleistung für die Kategorie »sexuell motivierte Rückfälligkeit«.

Neben der hier vorgenommenen Aufteilung sind weitergehende Untersuchungen mit unterschiedlichen Subgruppen notwendig, um die Möglichkeiten und Grenzen von Rückfallprognoseinstrumenten wie dem SORAG besser abschätzen zu können. Auch wenn die forensische Prognoseforschung in den letzten Jahren bedeutende Fortschritte vorweisen kann, zeigen die vorliegenden Ergebnisse, dass man noch weit davon entfernt ist, konkret formulierte und empirisch abgesicherte Handlungsanweisungen zur Erstellung von Kriminalprognosegutachten bereitstellen zu können, die eindeutige Antworten darauf geben können, bei welchem Sexualdelinquenten welche Prognosemethode aus welchen Gründen wie gut vorhersagen wird.

Summary

This study is part of a prospective, longitudinal research project to evaluate the reliability and validity of different recidivism risk assessment methods for sexual offenders for scientific and practical use in the German-speaking part of Europe. In this paper we present the German adaptation of the Sexual Offender Risk Appraisal Guide (SORAG), a specific risk assessment tool for sexual offenders that was developed and published in 1998 in Canada. We examined the interrater reliability, concurrent validity and predictive validity of the German version of the instrument with a sample of 178 male sexual offenders arrested in Austria. The German SORAG showed good results for interrater reliability and concurrent validity. The predictive validity was determined utilizing subgroups of the sample based on each offender's index offense. The results were predominantly good, but the effectiveness of the instrument varied depending on offender type. The use of actuarial risk assessment methods is helpful for the assessment process, but further research is necessary to know the exact drawbacks and opportunities of these instruments.

Keywords: SORAG, risk assessment, sexual offender, validity, recidivism

Literatur

American Psychiatric Association (2000). Diagnostisches und statistisches Manual psychischer Störungen: DSM-IV. Göttingen. – *Barbaree, H.E., Seto, M.C., Langton, C.M. & Peacock, E.J.* (2001). Evaluating the predictive accuracy of six risk assessment instruments for adult sex offenders. *Criminal Justice and Behavior* 28, 490–521. – *Bartosh, D.L., Garby, T., Lewis, D. & Gray, S.* (2003). Differences in the predictive validity of actuarial risk assessments in relation to sex offender type. *International Journal of Offender Therapy and Comparative Criminology* 47, 422–438. – *Boer, D.P., Hart, S.D., Kropp, P.R. & Webster, C.D.* (1997). Manual for the Sexual Violence Risk-20: Professional Guidelines for Assessing Risk of Sexual Violence. Vancouver. – *Boetticher, A., Kröber, H.-L., Müller-Isberner, R., Böhm, K.M., Müller-Metz, R. & Wolf, T.* (2006). Mindestanforderungen für Prognosegutachten. *NStZ* 26, 537–544. – *Cohen, J.* (1992). A power primer. *Psychological Bulletin* 112, 155–159. – *Dable, K.P.* (2007). Methodische Grundlagen der Kriminalprognose. *Forensische Psychiatrie, Psychologie, Kriminologie* 1, 101–110. – *Dable, K.P.* (2005). Psychologische Kriminalprognose. Herbolzheim. – *Dable, K.P., Schneider, V. & Zietzen, F.* (2007). Standardisierte Instrumente zur Kriminalprognose. *Forensische Psychiatrie, Psychologie, Kriminologie* 1, 15–26. – *Dilling, H., Mombour, W. & Schmidt, M.H.* (2004). Internationale Klassifikation psychischer Störungen. ICD-10 Kapitel V (F). Klinisch-diagnostische Leitlinien. Bern. – *Ducro, C. & Pham, T.* (2006). Evaluation of the SORAG and the Static-99 on Belgian sex offenders committed to a forensic facility. *Sexual Abuse: A Journal of Research and Treatment* 18, 15–26. – *Egan, J.P.* (1975). Signal Detection Theory and ROC Analysis. New York. – *Eher, R.* (2007). Die Zentrale Dokumentations- und Koordinationsstelle für Sexualstraftäter im Österreichischen Strafvollzug. Maßnahmenpakete zur Rückfallvermeidung inhaftierter Sexualstraftäter. *Journal für Strafrecht* 2, 51–55. – *Eher, R., Schilling, F., Graf, T., Frühwald, S. & Frottier, P.* (2006). Die standardisierte Begutachtung von Sexualstraftätern im Österreichischen Strafvollzug - erste Ergebnisse und Ausblicke. *Neuropsychiatrie* 20, 50–56. – *Eher, R. & Ross, T.* (2006). Reconsidering risk for reoffense in intrafamilial child molesters: New aspects on clinical and criminological issues. *Sexual Offender Treatment* 1, 1–9. – *Fiedler, P.* (2004).

Sexuelle Orientierung und sexuelle Abweichung. Weinheim. – Greve, W. & Wentura, D. (1997). Wissenschaftliche Beobachtung in der Psychologie. München. – Hanley, J.A. & McNeil, B.J. (1982). The meaning and use of the area under the Receiver Operating Characteristic (ROC) Curve. *Radiology* 143, 29-36. – Hanson, R.K. & Thornton, D. (1999). *Static-99: Improving Actuarial Risk Assessments for Sex Offenders*. User Report 99-02. Ottawa: Department of the Solicitor General of Canada. – Hanson, R.K. & Morton-Bourgon, K. (2004). Predictors of Sexual Recidivism: An Updated Meta-Analysis. Ottawa: Public Safety and Emergency Preparedness Canada. – Hanson, R.K. & Morton-Bourgon, K. (2007). *The Accuracy of Recidivism Risk Assessments for Sexual Offenders: A meta-analysis*. Ottawa: Public Safety and Emergency Preparedness Canada. – Hare, R.D. (1991). *Manual for the Psychopathy-Checklist-Revised*. Toronto. – Harris, G.T., Rice, M.E., Quinsey, V., Lalumière, M.L., Boer, D. & Lang, C. (2003). A multisite comparison of actuarial risk instruments for sex offenders. *Psychological Assessment* 15, 413-425. – Marshall, W.L. (2005). Grenzen der Phallometrie. *Recht & Psychiatrie* 23, 11-23. – Müller-Isberner, R., Cabeza, S.G. & Eucker, S. (2000). Die Vorhersage sexueller Gewalttaten mit dem SVR-20. Haina. – Nedopil, N. (2005). Prognosen in der Forensischen Psychiatrie. Ein Handbuch für die Praxis. Lengerich. – Nunes, K.L., Firestone, P., Bradford, J.M., Greenberg, D.M. & Broom, I. (2002). A comparison of modified versions of the Static-99 and the Sex Offender Risk Appraisal Guide (SORAG). *Sexual Abuse: A Journal of Research and Treatment* 14, 253-269. – Quenzer, C. (2005). *Prädiktive Validität ausgewählter aktuarischer Verfahren bei Sexualdelinquenten*. Berlin: unveröff. Diplomarbeit. – Quinsey, V.L., Harris, G.T., Rice, M.E. & Cormier, C. (2006). *Violent Offenders: Appraising and Managing Risk*. Washington/DC. – Rettenberger, M. & Eher, R. (2006a). Die deutsche Übersetzung des Static-99 zur aktuarischen Kriminalprognose verurteilter Sexualstraftäter. *MschR/Krim* 89, 352-365. – Rettenberger, M. & Eher, R. (2006b). Die revidierten Kodierungsrichtlinien des Static-99; www.vffw.org. – Rettenberger, M. & Eher, R. (2007). *Aktuarische Kriminalprognose bei Sexualstraftätern: Die deutsche Version des Sex Offender Risk Appraisal Guide (SORAG)*; www.vffw.org. – Schmucker, M. (2004). *Kann Therapie Rückfälle verhindern?* Herbolzheim. – Stadtland, C., Hollweg, M., Kleindienst, N., Diel, J., Reich, U. & Nedopil, N. (2005). Risk assessment and prediction of violent and sexual recidivism in sex offenders: Long-term predictive validity of four risk assessment instruments. *Journal of Forensic Psychiatry and Psychology* 16, 92-108. – Studer, L.H. & Aylwin, A.S. (2006). Sexual offender subtyping: The incest offender question. *Sexual Offender Treatment* 1, 1-8.

(Ansch. d. Verf.: Dipl.-Psych. Martin Rettenberger, Institut für Gewaltforschung und Prävention (IGF), Jörgerstraße 22/3, A-1170 Wien; martin.retttenberger@igf.or.at; PD Dr. med. Reinhard Eher, Justizanstalt Wien-Mittersteig, Außenstelle Floridsdorf, Gerichtsgasse 6, A-1210 Wien; reinhard.eher@justiz.gv.at; Universitätsklinikum Ulm, Abteilung für Psychosomatische Medizin und Psychotherapie, Sektion Forensische Psychotherapie, Am Hochsträß 8, D-89081 Ulm)

Prospective Actuarial Risk Assessment

A Comparison of Five Risk Assessment Instruments in Different Sexual Offender Subtypes

Martin Rettenberger

Federal Evaluation Centre for Violent and Sexual Offenders, Vienna, Austria

Anna Matthes

Christian-Albrechts University of Kiel, Germany

Douglas P. Boer

University of Waikato, Hamilton, New Zealand

Reinhard Eher

University Hospital Ulm, Germany

Federal Evaluation Centre for Violent and Sexual Offenders, Vienna, Austria

This study examines the predictive validity of the most commonly used risk assessment instruments for sexual offenders: Static-99, Rapid Risk Assessment for Sexual Offense Recidivism, Sex Offender Risk Appraisal Guide, Sexual Violence Risk-20, and Psychopathy Checklist-Revised in a prospective research design. Although risk assessment is part of a regime leading to various efforts to reduce risk by treatment and aftercare, all instruments show good predictive validity. However, depending on the instrument, recidivism category, and subgroup, the predictive accuracy varies markedly. Furthermore, the authors fail to demonstrate predictive validity for sexual violent reoffences—for the whole sample and for all subgroups. The results, nevertheless, support the utility and predictive validity of actuarial risk assessment complementary to treatment efforts to reduce risk. On the other hand, forensic practitioners have to be aware of the limitations of actuarial risk assessment methods, in particular as regards to variable predictive accuracy for different sexual offender subgroups and reoffence categories.

Keywords: *sexual offending; sexual offenders; recidivism; risk assessment; actuarial instrument*

Authors' Note: The authors would like to thank Lea Studer and Scott Aylwin, Canada, for very helpful assistance. Please address correspondence to Reinhard Eher, Federal Evaluation Centre for Violent and Sexual Offenders, Gerichtsgasse 6, A-1210 Vienna, Austria; e-mail: reinhard.eher@justiz.gv.at.

Sexual offender recidivism risk assessment has made important advances during the past decades (Barbaree, Langton, & Peacock, 2006). However, in spite of many research studies about the reliability, validity, and predictive accuracy of actuarial risk assessment tools, there are still unresolved problems and inconsistent research results (e.g., Craig, Browne, Stringer, & Beech, 2004). For example, Seto (2005) investigated whether combining multiple actuarial risk tools increases accuracy in predicting sexual offender recidivism and concluded that such combinations would not provide a consistent advantage over the predictive accuracy of the single best actuarial scale. However, the author cited an unpublished survey of evaluators involved in sexual offender civil commitment proceedings, demonstrating that most evaluators use multiple actuarial risk scales. Barbaree et al. (2006) found that different actuarial risk scales for sexual offenders produce different risk rankings. Their study concerns the question of whether different risk tools predict the same outcomes or vary in their ability to predict different outcomes, so the use of multiple actuarial risk scales may provide more questions than answers.

Another issue is the predictive accuracy of instruments for different sexual offender subtypes. In the developmental study of the Static-99, Hanson and Thornton (2000) concluded that the predictive accuracy of the instrument proved satisfactory for both rapists and child molesters. Although they found no differences in these two subgroups, more recent studies about the differential validity of tests across different sexual offender subgroups indicate that the predictive accuracy of particular instruments varies depending on offender type (e.g., Bartosh, Garby, Lewis, & Gray, 2003; Rettenberger & Eher, 2007a). Furthermore, there are subgroups of sexual offenders—for example, hands-off offenders or sexual offenders with adult male victims—for which there is little or no empirical evidence of predictive validity of actuarial risk assessment tools (Barbaree et al., 2006). Other relevant sexual offender subgroups include juvenile sexual offenders (e.g., Parks & Bard, 2006) and intellectually disabled sexual offenders (e.g., Lambrick, 2003).

To emphasize the importance of this aspect we want to reflect results of the Sex Offender Risk Appraisal Guide (SORAG) in detail. Bartosh et al. (2003) for example showed that the SORAG significantly predicted sexual, violent, and overall recidivism for extrafamilial child molesters (area under the curve [AUC] values ranged from .70 to .93) and for incest offenders (AUCs ranged from .72 to .91). However, the SORAG showed much less predictive power (AUCs ranged from .46 to .71) for rapists and hands-off offenders. In a Belgian validation study of the SORAG, the AUC values ranged from .64 (sexual recidivism in the rapist subgroup) to .77 (violent recidivism in the rapist subgroup), demonstrating strong variations depending on offender subgroup and recidivism criterion (Ducro & Pham, 2006). In an Austrian replication study, the ability to predict recidivism also varied strongly depending on offender type and reoffence category. AUCs ranged from .56 (violent non-sexual recidivism in the rapist subgroup) to .95 (any recidivism in the incest offender

subgroup; Rettenberger & Eher, 2007a). These results on one hand provide evidence of good predictive validity of the SORAG and on the other emphasize the importance of looking at the different performance of risk prediction tools depending on sexual offender subgroups and recidivism criteria.

This study is part of a prospective longitudinal research project to evaluate the reliability and validity of different recidivism risk assessment methods for sexual offenders allocated to treatment and other interventions for risk reduction. First, we investigated whether commonly reported validity indices could also be replicated in a prospective design. In a second step, we checked validity indices for all risk assessment tools in different sexual offender subtypes and for different reoffence categories. We hypothesised that the instruments still would show predictive validity independently of any treatment. We further hypothesised that they would perform better with the recidivism criterion they were particularly developed for, so Static-99, Rapid Risk Assessment for Sexual Offense Recidivism (RRASOR), and Sexual Violence Risk-20 (SVR-20) would be better predicting sexual recidivism whereas SORAG would be better predicting violent behaviour. Although originally not designed for risk assessment, we also included the Psychopathy Checklist-Revised (PCL-R) because its underlying construct is in close relationship to criminal behaviour (Nedopil, 2005). We hypothesised that the PCL-R would be more accurate for general and violent recidivism.

Method

Risk Assessment Instruments

For our study we used the RRASOR (Hanson, 1997), the Static-99 (Hanson & Thornton, 1999, 2000), the SORAG (Quinsey, Harris, Rice, & Cormier, 2006), the SVR-20 (Boer, Hart, Kropp, & Webster, 1997), and the PCL-R (Hare, 1991).

The RRASOR was published in 1997 by R. Karl Hanson for the prediction of sexual recidivism of convicted sexual offenders and consists of four items (Hanson, 1997): prior sex offences, age when exposed to risk, victim gender, and relationship to victim. In a meta-analytic review, Hanson and Bussière (1996) found that indicators of deviant sexual interests (e.g., the number of prior sex offences) consistently predicted sexual recidivism. Drawing from these results, Hanson (1997) selected variables with a minimum correlation of .10 with sexual recidivism and developed the RRASOR. Across six development samples and one validation sample, comprising a total of 2,592 sexual offenders, Hanson found that RRASOR scores had an average correlation (r) of .27 and an average AUC of .71 with sexual recidivism. The predictive accuracy of the RRASOR in the independent validation sample ($n = 303$) was represented by an r of .25 and an AUC of .67. In a Canadian study published by Barbaree, Seto, Langton, and Peacock (2001) in a mixed sample of 215 child molesters

and rapists assessed for a prison-based treatment program, the RRASOR exhibited good predictive accuracy (AUC = .77) concerning sexual recidivism, moderate predictive accuracy (AUC = .65) for serious recidivism, and small but still significant predictive accuracy (AUC = .60) for any reoffences.¹ The authors evaluated the predictive accuracy of six different risk assessment tools, whereas the RRASOR reached the highest AUC value for sexual recidivism. In a cross-validation study with an Austrian sample of adult male sexual offenders ($N = 81$), the instrument showed moderate significant predictive accuracy for an average follow-up time of 7.7 years for general (AUC = .66), sexual (AUC = .68), and violent (AUC = .67) recidivism (Rettenberger & Eher, 2006a). In a large cross-validation sample of 1,400 sexual offenders followed for an average of 3.7 years in Sweden, Sjöstedt and Långström (2001) found that the RRASOR had a correlation r of .22 and an AUC of .72 with sexual recidivism.

The Static-99 is a brief actuarial instrument for the assessment of risk for sexual and violent recidivism of adult sexual offenders (Harris, Phenix, Hanson, & Thornton, 2003). It is probably the most commonly used and the best validated actuarial instrument. It was developed in 1999 by Karl R. Hanson (Canada) and David Thornton (Great Britain) and is regularly used for risk assessment in North America and some countries in Europe. The instrument has been derived from a fusion of two previously developed risk assessment instruments, the RRASOR mentioned above and a shorter version of the Structured Anchored Clinical Judgement (SACJ-Min; Grubin, 1998). The Static-99 consists of 10 static risk factors: age when exposed to risk, any live-in intimate relationship for 2 or more years, any index offence of non-sexual violence, prior offences of non-sexual violence, prior charges or convictions for sexual offences, prior sentencing dates, any convictions for non-contact sexual offences, any unrelated victims, any stranger victims, and any male victims. The individual risk factors of a sexual offender add up to a maximum total score of 12 that is subsequently translated into four risk categories: low (0-1), medium low (2-3), medium high (4-5), and high (6 or more). Hanson and Thornton (1999) tested the predictive accuracy of the Static-99 using four different, mixed data sets of 1,301 sexual offenders (including child molesters and rapists from prison as well as high-security forensic psychiatric settings) collected in Canada and the United Kingdom. The Static-99 rendered predictive accuracy values of AUC = .71 for sexual and AUC = .69 for any violent (including sexual) recidivism. Among others, further results from studies with Canadian (Barbaree et al., 2001), Dutch (de Vogel, de Ruiter, van Beek, & Mead, 2004), Belgian (Ducro & Pham, 2006), Austrian (Rettenberger & Eher, 2006a), and Swedish (Sjöstedt & Långström, 2001) samples of sex offenders were published. For sexual recidivism the predictive accuracy values ranged from moderate (AUC = .66; Ducro & Pham, 2006) to good (AUC = .76; Sjöstedt & Långström, 2001).

The SORAG is an actuarial risk assessment tool for sexual offenders that was developed by Canadian forensic researchers. This instrument is a modification of the Violence Risk Appraisal Guide (VRAG; Quinsey et al., 2006), which was developed to predict violent and sexual recidivism among adult male offenders; 10 of the 14

items of the SORAG are the same as in the VRAG. The SORAG is conceptualised for sexual offenders to assess violent recidivism risk, which includes sexual offences involving physical contact with the victim. The instrument consists of 14 weighted items: lived with biological parents up to age 16, elementary school maladjustment, history of alcohol problems, marital status, criminal history for nonviolent offences, criminal history for violent offences, previous convictions for sexual offences, sexual offences against girls under age 14 only, failure on prior conditional releases, age at index offence, *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed.) (American Psychiatric Association, 1980) criteria for any personality disorder, *DSM-III* criteria for schizophrenia, phallometric test results indicating pedophilia or sexual sadism, and PCL-R score. Based on the total score the evaluator can allocate the offender to one of nine risk categories. By means of these risk categories, it is possible to infer to empirically calculated probabilities of violent (including sexual) recidivism after 7 and 10 years at risk, respectively. Most of the validation studies were completed in Anglo-American countries. Harris et al. (2003) reviewed the results of these studies and concluded that several of them (e.g., Barbaree et al., 2001; Nunes, Firestone, Bradford, Greenberg, & Broom, 2002; Rice & Harris, 1995) have shown the SORAG to have a high accuracy in the prediction of violent (including sexual) recidivism and moderate accuracy in predicting sexual offences. For the prediction of violent (including sexual) recidivism they calculated a median AUC value of .75. With regard to European samples, Ducro and Pham (2006) evaluated the predictive accuracy of the SORAG on Belgian sexual offenders committed to a forensic facility. For the total sample ($N = 147$) the instrument showed strong predictive validity for general ($AUC = .70$) and violent ($AUC = .72$) recidivism and moderate predictive validity for sexual recidivism ($AUC = .64$). Rettenberger and Eher (2007a), however, also found good predictive validity values of a German version for general ($AUC = .73$), violent ($AUC = .76$), and sexual ($AUC = .73$) recidivism.

The SVR-20 is a structured clinical guideline designed for the assessment of risk for sexual violence in adult sex offenders. The instrument was developed from a thorough consideration of the empirical literature and the clinical expertise of a number of clinicians. The SVR-20 consists of 20 items, divided into three domains: psychosocial adjustment (11 items: sexual deviance, victim of child abuse, psychopathy, major mental illness, substance use problems, suicidal or homicidal ideation, relationship problems, employment problems, past non-sexual violent offences, past nonviolent offences, and past supervision failure), sexual offences (7 items: high density sex offences, multiple sex offence types, physical harm to victim(s) in sex offences, use of weapons or threats of death in sex offences, escalation in frequency of severity of sex offences, extreme minimisation or denial of sex offences, and attitudes that support or condone sex offences), and future plans (2 items: lacks realistic plans and negative attitude towards intervention), which have to be coded by an experienced forensic clinician. Although originally designed as a structured clinical guideline, it is not uncommon for research purposes to add up the items. In this case, the instrument becomes a conceptual actuarial measure (Hanson & Morton-Bourgon,

2007). In comparison to the above-mentioned risk assessment tools there is only limited knowledge about the psychometric properties of the SVR-20 (de Vogel et al., 2004). However, current studies provide first indications of the predictive validity of the SVR-20 and its cross-cultural transferability (Hanson & Morton-Bourgon, 2007; Stadtland et al., 2005; de Vogel et al., 2004). For example, in a retrospective study with a Dutch sample of treated sexual offenders, the SVR-20 showed good predictive validity indices, with AUC values up to .83 for sexual recidivism. Furthermore, the SVR-20 final risk judgement was a significantly better predictor of sexual recidivism than the Static-99 risk category (de Vogel et al., 2004).

Although the PCL-R was originally developed as a psychometric instrument that measures the construct of psychopathy, research shows that the PCL-R performs reasonably in predicting general, sexual, and violent recidivism in both prison and forensic psychiatric populations (e.g., Quinsey, Rice, & Harris, 1995; Rice & Harris, 1995; Salekin, Rogers, & Sewell, 1996). Because the PCL-R is often used as an assessment tool for dangerousness and recidivism risk (Hare, 1999), the PCL-R was also included in the present study. The PCL-R is based on semi-structured interviews and review of file information. It measures glibness or superficial charm, grandiose sense of self-worth, need for stimulation and proneness to boredom, pathological lying, conning and manipulative, lack of remorse or guilt, shallow affect, callous and lack of empathy, parasitic lifestyle, poor behavioural control, promiscuous sexual behaviour, early behaviour problems, lack of realistic and long-term goals, impulsivity, irresponsibility, failure to accept responsibility for own actions, many short-term marital relationships, juvenile delinquency, revocation of conditional release, and criminal versatility. Scale scores are obtained by summing the ratings, for a total possible score of 40. The conventional cut-off for the diagnosis of psychopathy in North America is 30, whereas in Europe the cut-off commonly used is 25 (Hart & Hare, 1997; Hartmann, Hollweg, & Nedopil, 2001). As mentioned above, the PCL-R does a reasonable job as a recidivism risk prediction instrument for different outcomes in various forensic populations (e.g., Quinsey et al., 1995; Rice & Harris, 1995; Salekin et al., 1996).

During the past few years German versions of all instruments used have become available (Noll, Endrass, Rossegger, & Urbaniok, 2006; Rettenberger & Eher, 2006a, 2006b, 2007a, 2007b; Stadtland et al., 2005). In general, they show good reliability and validity indices.

Interrater Reliability

The interrater reliability of the instruments was examined by the Intraclass-Correlation Coefficient (ICC). The results of interrater reliability have already been reported elsewhere (Rettenberger & Eher, 2006a, 2007a). Using the critical values for ICC (single measure) reported by Fleiss (1986; $ICC \geq .75$ = excellent; $.60 \leq ICC < .75$ = good; $.40 \leq ICC < .60$ = moderate; $ICC < .40$ = poor), the interrater reliability of each instrument was

excellent: for the Static-99 we calculated an ICC = .98, for the RRASOR an ICC = .90, for the SORAG an ICC = .93, for the SVR-20 an ICC = .84, and for the PCL-R = .93 (all $p < .001$).

Participants

Three hundred ninety-four male sex offenders were investigated between 2002 and 2005 at the Federal Evaluation Centre for Violent and Sexual Offenders in the Austrian Prison System and followed up after prison release until December 31, 2007 (mean age at time of release = 41.43 years, $SD = 12.59$, range 16-72). The follow-up time periods ranged from 18.41 to 66.64 months ($M = 36.79$ months, $SD = 11.68$; 18 months was defined to be a minimal follow-up period). Although follow-up time periods of this study correspond approximately with the conventional follow-up periods reported by Prentky, Lee, Knight, and Cerce (1997), many studies have had access to longer follow-up periods (e.g., Hanson & Thornton, 1999, 2000; Langton et al., 2007).

The group consisted of 200 child molesters (92 extrafamilial offenders and 108 incest offenders), 177 rapists, 12 hands-off offenders (6 exhibitionists, 5 offenders convicted of child pornography, 1 offender convicted of sexual burglary), and 5 sexual murderers. The mean duration of imprisonment for the index offence was 30.57 months ($SD = 26.61$, range 2.30-235.46). Of these, 55.8% ($n = 220$) had at least one prior conviction for any offence ($M = 3.76$, $SD = 6.01$, range 0-37), 39.3% ($n = 155$) had at least one prior conviction for a non-sexual violent offence ($M = 1.29$, $SD = 2.58$, range 0-18), and 17.3% ($n = 68$) had at least one prior conviction for a sexual offence ($M = 0.41$, $SD = 1.46$, range 0-16). All offenders had been allocated for at least one treatment regime during their prison sentence, afterwards, or both (no further treatment data were available at this point).

Procedure and Statistical Analysis

All 394 offenders were evaluated between 2002 and 2005 at the Federal Evaluation Centre for Violent and Sexual Offenders in the Austrian Prison System (Eher, Schilling, Graf, Fruehwald, & Frottier, 2006). For each of the offenders a comprehensive forensic psychological appraisal was provided to plan the correctional scheme. Risk assessment is an important part of these appraisals, and actuarial instruments are an integral part of the overall risk assessment process. Predictions were subsequently compared with official recidivism data (in terms of reconviction information) within a prospective longitudinal research design. Data on reconviction were retrieved from the Federal Central Register of the Austrian Ministry of Internal Affairs. Each new conviction listed on the official criminal record was counted as a reoffence. We used the following recidivism criteria: sexual reconviction (each new conviction because of both sexual hands-on and hands-off offences), sexual violent

reconviction (i.e., we counted each new conviction because of a sexual offence involving physical contact or, in other words, each new conviction because of a sexual hands-on reoffence),² general violent reconviction (each new conviction because of non-sexual violent and sexual hands-on reoffences), and general criminal reconviction (each new conviction for any kind of reoffence).

The concurrent validity of the instruments was analysed using Pearson correlations of the raw scores of each instrument.³ The predictive accuracy of risk assessment instruments was determined by calculating AUCs of the receiver operating characteristics (ROCs; Hanley & McNeil, 1982). Direct comparisons between AUC values of the five actuarial risk assessment tools have previously been conducted pairwise for dependent AUC curves (software package MedCalc). In the present study we only used raw scores rather than risk categories to determine predictive validity. The ROC curve is produced by plotting the hit and false-alarm rates across the possible cut-off values. This statistical procedure is commonly used to examine the predictive accuracy of binary decisions, such as “release” or “do not release” (Mossman, 1994). The AUC values lie between 0 and 1, with .5 indicating prediction at the chance level and 1 indicating perfect prediction. In the context of recidivism prediction, the AUC is commonly interpreted as the probability that a randomly selected recidivist will have a higher score on a risk variable than will a randomly selected non-recidivist (Seto, 2005). Because of its low sensitivity to base rates of recidivism and to users’ biases for or against Type I or Type II prediction error and the easy interpretation, the AUC is a standard measure of diagnostic and predictive accuracy in clinical and forensic research (Mossman, 1994). Furthermore, none of the measures commonly used in the past provides an efficient way to evaluate the usefulness of predictive tests (Rice & Harris, 1995). Referring to Cohen (1992), Dahle, Schneider, and Ziethen (2007) formulated the following criteria for classification of predictive accuracy of risk assessment tools: AUC values of .72 or above ($r \geq .37$) are classified as good and values between .64 and .71 ($r \geq .24$) are classified as moderate. Significant AUC values that are below .64 ($r < .24$) are classified as small. Douglas et al. (2001) report the following interpretation of critical AUC values: AUC values of .70 and above are classified as moderate and values above .75 are classified as good.

Results

Instruments

Ninety-seven percent ($n = 382$) of all offenders included in this study were scored on all 10 items of the Static-99. The average Static-99 raw score was 3.32 ($SD = 2.21$, range 0-10). Regarding RRASOR, 97.0% ($n = 382$) of the total sample were coded with an average raw score of 1.39 ($SD = 1.13$, range 0-5). PCL-R scores were available for 98.0% ($n = 386$) of the offenders; average PCL-R score was 19.69

Table 1
Reconviction Rates of the Total Sample and Subgroups

	Recidivism Criterion			
	Sexual Recidivism, % (n)	Sexual Violent Recidivism, % (n)	General Violent Recidivism, % (n)	General Criminal Recidivism, % (n)
Total sample ($N = 394$)	4.3 (17)	1.8 (7)	12.2 (48)	25.9 (102)
Rapists ($n = 177$)	1.7 (3)	1.7 (3)	16.9 (30)	32.8 (58)
Extrafamilial child sexual offenders ($n = 92$)	12.0 (11)	4.3 (4)	9.8 (9)	22.8 (21)
Incest offenders ($n = 108$)	0.9 (1)	0 (0)	5.6 (6)	14.8 (16)

($SD = 8.04$, range 3-40). Regarding SVR-20 and SORAG, we had information on 93.9% ($n = 366$) and 94.4% ($n = 372$) of the total sample, respectively. The average raw score of SVR-20 was 18.68 ($SD = 6.90$, range 2-36) and for SORAG it was 6.05 ($SD = 13.44$, range -21 to 39).

Reconviction Rates

The total sample of the present study ($N = 394$) consisted of rapists (44.9%, $n = 177$), extrafamilial child molesters (23.4%, $n = 92$), and incest offenders (27.4%, $n = 108$). Twelve hands-off offenders (3.1% of the total sample) and 5 sexual murderers (1.3%) were not allocated to either of these groups and therefore were excluded from further analysis. Reconviction rates for the total sample and the three subgroups are shown in Table 1. The average time-at-risk periods of groups did not differ significantly.

Low reconviction rates in some relapse categories impeded further analysis. Even though the AUC statistic is seen to be independent of low base rates, numbers of only one or two recidivists in one reoffence category would preclude meaningful conclusions of ROC analyses. For the present study, we defined a minimal reconviction rate of 1.5% per follow-up year as a threshold for meaningful analysis. This leads to the minimum required reconviction rate of 4.5% for AUC analyses (see also Harris & Hanson, 2004; Sjöstedt & Långström, 2001).

Concurrent Validity

Pearson correlation coefficients between raw scores of the risk assessment instruments are shown in Table 2. As expected given their shared items, the RRASOR and Static-99 show the highest intercorrelation. The high correlation between SORAG and

Table 2
Correlations Between Raw Scores of the Risk Assessment Instruments

Instrument	RRASOR	Static-99	SORAG	SVR-20	PCL-R
RRASOR	–				
Static-99	.82**	–			
SORAG	.45**	.71**	–		
SVR-20	.37**	.63**	.79**	–	
PCL-R	.24**	.51**	.78**	.77**	–

Note: RRASOR = Rapid Risk Assessment for Sexual Offense Recidivism; SORAG = Sex Offender Risk Appraisal Guide; SVR-20 = Sexual Violence Risk–20; PCL-R = Psychopathy Checklist–Revised.

** $p < .01$.

PCL-R is also not surprising, because both instruments focus strongly on antisociality and lifestyle impulsivity. Furthermore, SORAG (as well as the SVR-20) includes the PCL-R score in risk assessment. The high correlation between SORAG, Static-99, and SVR-20 scores reveals that all three instruments are designed to predict sexual recidivism partially using similar items (e.g., criminal history). Taken together, the results are comparable to previous studies from North America (e.g., Barbaree et al., 2001; Langton et al., 2007) and Europe (e.g., Rettenberger & Eher, 2007a) and could be therefore interpreted as evidence of the concurrent validity of the instruments.

Predictive Validity

AUC values of the five instruments are shown in Table 3 for the total sample ($n = 394$). Because of very low base rates, we did not calculate any AUC values for the sexual violent recidivism criterion. However, for sexual recidivism we conducted further statistical analyses, because this reconviction rate (4.3%) approximately reached the threshold of 4.5%. For the total sample, RRASOR, Static-99, SORAG, and SVR-20 (but not PCL-R) could be shown to significantly predict sexual recidivism. Concerning this recidivism criterion, there were no significant differences between the instruments. Also, for general violent and general criminal recidivism all five instruments showed significant predictive power. As for the prediction of general criminal recidivism, the SORAG showed significantly better results than RRASOR ($p = .014$), SVR-20 ($p = .002$), and PCL-R ($p = .001$). For general violent recidivism the SORAG could be found to predict significantly better than SVR-20 ($p = .001$) and PCL-R ($p = .012$). Furthermore, for general violent recidivism the difference of the AUC values between Static-99 and SVR-20 was significant ($p = .038$). All other pairwise comparisons failed to reach significance.

In Table 4 the five instruments' predictive validity data for the rapist subgroup ($n = 177$) are presented. Because of very low base rates, we did not calculate AUC values

Table 3
Predictive Validity of Five Risk Assessment Instruments
for the Total Sample ($N = 394$)

Instrument	Sexual Recidivism	Sexual Violent Recidivism ^a	General Violent Recidivism	General Criminal Recidivism
RRASOR	.74** (.61-.87)	–	.66** (.59-.74)	.66** (.60-.72)
Static-99	.74** (.61-.87)	–	.71** (.64-.76)	.70** (.65-.75)
SORAG	.69** (.60-.78)	–	.72** (.64-.80)	.75** (.69-.80)
SVR-20	.71** (.60-.81)	–	.61* (.52-.71)	.67** (.61-.74)
PCL-R	.59 (.49-.69)	–	.65** (.57-.73)	.67** (.61-.73)

Note: RRASOR = Rapid Risk Assessment for Sexual Offense Recidivism; SORAG = Sex Offender Risk Appraisal Guide; SVR-20 = Sexual Violence Risk–20; PCL-R = Psychopathy Checklist–Revised. Values in parentheses are 95% confidence intervals.

a. Reconviction rate too low.

* $p < .05$. ** $p < .01$.

for sexual and sexual violent recidivism. For general violent and general criminal recidivism, RRASOR, Static-99, and SORAG demonstrated significant predictive accuracy.

The AUC values of the five instruments' predictive validity for the extrafamilial child molester subgroup ($n = 92$) are shown in Table 5. With the exception of the RRASOR, all instruments showed significant predictive accuracy for sexual recidivism. In spite of very few recidivists for sexual violent recidivism we conducted further statistical analyses for this recidivism category, because the reconviction rate (4.3%) approximately reached the threshold of 4.5%. None of the instruments was able to significantly predict this relapse category. As regards the general violent recidivism criterion, SORAG, SVR-20, and PCL-R performed well, whereas RRASOR and Static-99 could not demonstrate significant predictive power. For the prediction of general criminal recidivism, only the RRASOR failed. Furthermore, SORAG and SVR-20 performed significantly better than Static-99 ($p = .044$ and $p = .036$, respectively).

In Table 6 the predictive validity indices of all five risk assessment instruments are shown for the incest offender subgroup ($n = 108$). As for the rapist subgroup, we did not calculate AUC values for sexual and violent sexual recidivism. Within these two recidivism criteria, no more than one offender relapsed. For general violent recidivism no instrument showed significant predictive power, whereas in terms of general criminal recidivism Static-99, SORAG, SVR-20, and PCL-R (but not RRASOR) performed equally well.

Discussion

The aim of this study was to test German versions of the five most commonly used actuarial sexual offender risk prediction instruments for predictive validity in a

Table 4
Predictive Validity of Five Risk Assessment
Instruments for the Rapist Subgroup (n = 177)

Instrument	Sexual Recidivism ^a	Sexual Violent Recidivism ^a	General Violent Recidivism	General Criminal Recidivism
RRASOR	–	–	.70** (.60-.81)	.69** (.61-.78)
Static-99	–	–	.70** (.60-.80)	.69** (.61-.77)
SORAG	–	–	.68** (.58-.79)	.72** (.63-.80)
SVR-20	–	–	.53 (.40-.66)	.58 (.49-.68)
PCL-R	–	–	.55 (.43-.67)	.55 (.46-.65)

Note: RRASOR = Rapid Risk Assessment for Sexual Offense Recidivism; SORAG = Sex Offender Risk Appraisal Guide; SVR-20 = Sexual Violence Risk–20; PCL-R = Psychopathy Checklist–Revised. Values in parentheses are 95% confidence intervals.

a. Reconviction rate too low.

***p* < .01.

Table 5
Predictive Validity of the Five Risk Assessment Instruments for the
Extrafamilial Molest Offender Subgroup (n = 92)

Instrument	Sexual Recidivism	Sexual Violent Recidivism	General Violent Recidivism	General Criminal Recidivism
RRASOR	.64 (.45-.83)	.53 (.11-.94)	.61 (.40-.81)	.60 (.46-.75)
Static-99	.68* (.50-.87)	.53 (.17-.88)	.69 (.53-.86)	.67* (.53-.81)
SORAG	.71* (.57-.85)	.62 (.45-.80)	.81** (.69-.93)	.77** (.66-.88)
SVR-20	.75** (.60-.91)	.51 (.25-.77)	.81** (.69-.92)	.77** (.65-.89)
PCL-R	.73* (.61-.84)	.70 (.46-.94)	.75* (.60-.89)	.80** (.70-.89)

Note: RRASOR = Rapid Risk Assessment for Sexual Offense Recidivism; SORAG = Sex Offender Risk Appraisal Guide; SVR-20 = Sexual Violence Risk–20; PCL-R = Psychopathy Checklist–Revised. Values in parentheses are 95% confidence intervals.

p* < .05. *p* < .01.

prospective research design. After a mean follow-up time of 3 years after prison release, the RRASOR, Static-99, SORAG, SVR-20, and PCL-R yielded good predictive accuracy in the total sample for most reoffence categories, with comparable or even better results than have been previously reported (Hanson & Morton-Bourgon, 2007). Therefore, we assume that those risk assessment instruments perform equally well in a German sexual offender population and in a prospective research design. Because of the prospective approach in our sample, the risk assessment itself had

Table 6
Predictive Validity of Five Risk Assessment Instruments for the Incest
Offender Subgroup ($n = 108$)

Instrument	Sexual Recidivism ^a	Sexual Violent Recidivism ^a	General Violent Recidivism	General Criminal Recidivism
RRASOR	–	–	.67 (.44-.90)	.65 (.50-.80)
Static-99	–	–	.72 (.51-.93)	.70* (.55-.84)
SORAG	–	–	.58 (.31-.86)	.69* (.53-.84)
SVR-20	–	–	.57 (.31-.82)	.73** (.59-.86)
PCL-R	–	–	.66 (.48-.87)	.77** (.66-.88)

Note: RRASOR = Rapid Risk Assessment for Sexual Offense Recidivism; SORAG = Sex Offender Risk Appraisal Guide; SVR-20 = Sexual Violence Risk-20; PCL-R = Psychopathy Checklist-Revised. Values in parentheses are 95% confidence intervals.

a. Reconviction rate too low.

* $p < .05$. ** $p < .01$.

triggered intervention regimes like treatment and aftercare leading to a possibly substantial change of the relative risk (Andrews & Bonta, 2003). In other words, within a prospective research design, forensic professionals are working against their own prediction of (high) risk. However, despite this fact no evidence was found for a deterioration of the general predictive accuracy of instruments. Nevertheless, low relapse base rates—possibly also due to treatment effects—reduced the utility and applicability of instruments for some offender types and reoffense categories.

Overall, we found substantial differences between instruments in predicting different relapse categories in different offender subgroups. Because of generally low base rates in our sample, we abstained from evaluating the predictive validity of instruments for some recidivism categories and offender subgroups. Regarding the total sample, most instruments showed moderate to good predictive accuracy for the recidivism criteria “general violent” and “general criminal.” For sexual recidivism—the outcome that they were designed for—the RRASOR, Static-99, and SVR-20 showed good predictive power and the SORAG continued to show moderate predictive power. Only the PCL-R failed to demonstrate satisfactory predictive accuracy for sexual reoffences in the total sample.

Because of very low reconviction rates for sexual reoffences in the rapist subgroup, after 3 years at risk we refrained from further analyses of predictive accuracy. However, both general violent and general criminal recidivism were predicted significantly by the RRASOR, Static-99, and SORAG, whereas the PCL-R and the SVR-20 failed to do so.

For the extrafamilial child molester subgroup the SORAG, SVR-20, and PCL-R showed good predictive accuracy for sexual, general violent, and general criminal

recidivism, exceeding AUC values previously reported (for an overview, see e.g., Eher, Rettenberger, Schilling, & Pfäfflin, 2008). By contrast, the RRASOR completely failed to predict general criminal recidivism, and the Static-99 was found to be significant for only sexual and general recidivism. Neither of the instruments could significantly predict sexual violent reoffences.

Intrafamilial child sexual abusers usually have a low risk for reoffence (Firestone et al., 1999; Harris & Hanson, 2004). Very low relapse base rates made it impossible to calculate AUCs for most reoffence categories. Only general criminal recidivism, yielding a base rate of about 15%, was predicted significantly by the Static-99, SORAG, SVR-20, and PCL-R (but not the RRASOR). Analysing offender populations with longer follow-up periods might provide higher reconviction rates and subsequently also show good predictive validity of instruments. This was previously shown in a retrospective research design by Bartosh et al. (2003), who demonstrated high predictive accuracy for the Static-99 and SORAG in predicting violent and overall relapse in incest offenders.

In general, our results show that there is a need to be aware of the variability in predictive validity depending on the instrument being used, the sexual offender subtype, and the recidivism category of interest. Whereas all instruments provided good predictive accuracy for each reoffence category once the whole sex offender sample was tested, substantial differences occurred when subgroups were analysed: The RRASOR completely failed to predict any kind of reoffence category in both types of child molesters. The Static-99 also failed to predict general violent reoffences in both types of child molesters. The SORAG failed to predict violent reoffences in intrafamilial child molesters, and the SVR-20 and PCL-R failed to predict general violent and general recidivism in rapists and violent recidivism in intrafamilial child molesters. No instrument in our study, however, could predict sexual violent reoffences.

In a recent study, Langton et al. (2007) presented meaningful predictive accuracy values for actuarial instruments (among others they also used RRASOR, Static-99, and SORAG) regarding sexual violent reoffences. However, AUC values for sexual violent reoffences were only reported between AUCs of .64 and .68 and therefore missed the criterion for “good predictive validity” (Langton et al., 2007). Low relapse base rates for the sexual violent category in our study—because of short time-at-risk periods and possible treatment effects—made AUC calculations for most offender types impossible. Only the extrafamilial child molester group reached relapse rates for sexual violent offences high enough for further AUC analyses. Nevertheless, no significant prediction of sexual violent reoffences could be gained. Langton et al. (2007), on the other hand, dealing with a high-risk sample with considerably more prior offences, had inevitably higher recidivism base rates. In spite of this fact they failed to demonstrate good predictive validity for the sexual violent category. Also, Harris et al. (2003) failed to demonstrate good predictive validity for sexual violent reoffences in spite of high base rates (sexual reoffence rates in child molesters: 21%, in rapists: 29%) and substantially longer time at risk (5 years) in their sample (AUC for rapists: 0.59, AUC for child molesters: 0.65, for the Static-99; and AUC for rapists: 0.62, AUC for child molesters: 0.70, for the SORAG).

From the results of these studies and of the present study, the actuarial prediction of some reoffence categories in at least some offender subtypes is less accurate than generally assumed. One major aim of most criminal justice systems is to calculate risk by predicting the probability of severe sexual crimes. This goal obviously is not yet achieved satisfactorily by actuarial risk assessment, because results are far from ideal, especially when time-at-risk periods are relatively short.

Another more theory-driven explanation of risk assessment for sexual offenders focuses on dynamic risk factors. The findings of Hanson and Harris (2001) indicated that stable and acute dynamic risk factors were able to distinguish between recidivists and non-recidivists even after controlling for common static risk factors such as Static-99 scores. The authors concluded that predictions provided by actuarial risk assessment instruments can be improved by considering a range of dynamic risk factors related to behaviour while on community supervision. For short follow-up periods, dynamic risk factors could especially play a substantial role in the prediction of reoffences for sexual offenders. It can be hypothesised that in cases of short follow-up periods, the prediction of rare incidences—such as sexual reoffences—could be considerably improved by using dynamic factors.

The results of the present study support the benefit of multidimensional models of risk assessment (Doren, 2004): The low correlation between RRASOR and PCL-R supports the idea of at least two independent dimensions, suggesting the RRASOR as a measure of the sexual deviance and the PCL-R as a measure of the psychopathy or general criminality dimension (Eher, Neuwirth, Fruehwald, & Frottier, 2003). The SORAG, on the other hand, was constructed for predicting violent recidivism. Consistently, it demonstrated the highest AUC values for general violent and general criminal recidivism, whereas the RRASOR showed best predictive power for sexual recidivism.

To conclude, our results support the utility of German versions of actuarial risk assessment instruments. However, depending on the instrument, recidivism criterion, and subgroup, the ability to predict recidivism varied strongly. Because of very low base rates the results of the present study were not able to provide evidence for the ability of any risk assessment tool to predict sexual violent reoffence. Nevertheless, although actuarial risk assessment methods are clearly not a panacea for the problems that have plagued the prediction of recidivism among sexual offenders, these instruments should be an inherent part of a broad clinical risk assessment process. Actuarials, in any case, have to be validated on offender subtypes and various reoffence categories in a prospective design.

Notes

1. For the interpretation of predictive validity indices see, for example, Dahle, Schneider, and Ziethen (2007) or Douglas, Webster, Hart, Eaves, and Ogloff (2001).

2. This definition of violence is derived from the discussion about the definition of relevant outcome variables for actuarial risk assessment instruments proposed by Quinsey, Harris, Rice, and Cormier

(2006): In a nutshell, they decided to count all sexual assaults involving physical contact as violent. By contrast, sexually motivated non-contact or hands-off offences such as exhibitionism and voyeurism did not count as violent.

3. Langton et al. (2007) used risk categories rather than raw scores to determine risk predictions. As already indicated by de Vogel, de Ruiter, van Beek, and Mead (2004), there was no significant difference between the AUC values calculated using either raw scores or risk categories in the present study.

References

- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.
- Andrews, D. A., & Bonta, J. (2003). *Psychology of criminal conduct*. Cincinnati, OH: Anderson.
- Barbaree, H. E., Langton, C. M., & Peacock, E. J. (2006). Different actuarial risk measures produce different risk rankings for sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 18*, 423-440.
- Barbaree, H. E., Seto, M. C., Langton, C. M., & Peacock, E. J. (2001). Evaluating the predictive accuracy of six risk assessment instruments for adult sex offenders. *Criminal Justice and Behavior, 28*, 490-521.
- Bartosh, D. L., Garby, T., Lewis, D., & Gray, S. (2003). Differences in the predictive validity of actuarial risk assessments in relation to sex offender type. *International Journal of Offender Therapy and Comparative Criminology, 47*, 422-438.
- Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1997). *Manual for the Sexual Violence Risk-20: Professional guidelines for assessing risk of sexual violence*. Vancouver: The Mental Health, Law, & Policy Institute.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.
- Craig, L. A., Browne, K. D., Stringer, I., & Beech, A. (2004). Limitations in actuarial risk assessment of sexual offenders: A methodological note. *British Journal of Forensic Practice, 6*, 16-32.
- Dahle, K.-P., Schneider, V., & Zietzen, F. (2007). Standardisierte Instrumente zur Kriminalprognose [Actuarial instruments for the prediction of criminal reoffences]. *Forensische Psychiatrie, Psychologie, Kriminologie, 1*, 15-26.
- de Vogel, V., de Ruiter, C., van Beek, D., & Mead, G. (2004). Predictive validity of the SVR-20 and Static-99 in a Dutch sample of treated sex offenders. *Law and Human Behavior, 28*, 235-251.
- Doren, D. M. (2004). Toward a multidimensional model for sexual recidivism risk. *Journal of Interpersonal Violence, 19*, 835-856.
- Douglas, K. S., Webster, C. D., Hart, S. D., Eaves, D., & Ogloff, J. R. P. (2001). *HCR-20 violence risk management companion guide*. Burnaby: Mental Health Law and Policy Institute, Simon Fraser University and Louis de la Parte Florida Mental Health Institute, University of South Florida.
- Ducro, C., & Pham, T. (2006). Evaluation of the SORAG and the Static-99 on Belgian sex offenders committed to a forensic facility. *Sexual Abuse: A Journal of Research and Treatment, 18*, 15-26.
- Eher, R., Neuwirth, W., Fruehwald, S., & Frottier, P. (2003). Sexualization and lifestyle impulsivity: Clinically valid discriminators in sexual offenders. *International Journal of Offender Therapy and Comparative Criminology, 47*, 452-467.
- Eher, R., Rettenberger, M., Schilling, F., & Pfäfflin, F. (2008). Failure of Static-99 and SORAG to predict relevant reoffense categories in relevant sexual offender subtypes. A prospective study. *Sexual Offender Treatment, 3*, 1-14.
- Eher, R., Schilling, F., Graf, T., Fruehwald, S., & Frottier, P. (2006). Die standardisierte Begutachtung von Sexualstraftätern im Österreichischen Strafvollzug—Erste Ergebnisse und Ausblicke [Comprehensive assessment of sexual offenders in Austria's prison system]. *Neuropsychiatrie, 20*, 50-55.
- Firestone, P., Bradford, J. M., McCoy, M., Greenberg, D. M., Larose, M. R., & Curry, S. (1999). Prediction of recidivism in incest offenders. *Journal of Interpersonal Violence, 14*, 511-531.

- Fleiss, J. L. (1986). *The design and analysis of clinical experiments*. New York: John Wiley.
- Grubin, D. (1998). *Sex offending against children: Understanding the risk. Police Research Series, Paper 99*. London: Home Office.
- Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*, 29-36.
- Hanson, R. K. (1997). *The development of a brief actuarial risk scale for sexual offense recidivism (User Report No. 1997-04)*. Ottawa: Department of the Solicitor General of Canada.
- Hanson, R. K., & Bussière, M. T. (1996). *Predictors of sexual offender recidivism: A meta-analysis (User Report 1996-04)*. Ottawa: Department of the Solicitor General of Canada.
- Hanson, R. K., & Harris, A. J. (2001). A structured approach to evaluating change among sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 13*, 105-122.
- Hanson, R. K., & Morton-Bourgon, K. (2007). *The accuracy of recidivism risk assessment for sexual offenders: A meta-analysis (User Report No. 2007-01)*. Canada: Public Safety and Emergency Preparedness.
- Hanson, R. K., & Thornton, D. (1999). *Static-99: Improving actuarial risk assessments for sex offenders (User Report No. 1999-02)*. Ottawa: Department of the Solicitor General of Canada.
- Hanson, R. K., & Thornton, D. (2000). Improving risk assessment for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior, 24*, 119-136.
- Hare, R. D. (1991). *Manual for the Psychopathy-Checklist-Revised*. Toronto: Multi Health Systems.
- Hare, R. D. (1999). *Without conscience: The disturbing world of the psychopaths among us*. New York: Guilford.
- Harris, A. J. R., & Hanson, R. K. (2004). *Sex offender recidivism: A simple question (User Report 2004-03)*. Ottawa: Public Safety and Emergency Preparedness.
- Harris, G. T., Phenix, A., Hanson, R. K., & Thornton, D. (2003). *Static-99 coding rules revised—2003*. Ottawa: Department of the Solicitor General of Canada.
- Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumière, M. L., Boer, D., & Lang, C. (2003). A multisite comparison of actuarial risk instruments for sex offenders. *Psychological Assessment, 15*, 413-425.
- Hart, S. D., & Hare, R. D. (1997). Psychopathy: Assessment and association with criminal conduct. In D. Stoff, J. Breiling, & J. D. Maser (Eds.), *Handbook of antisocial behavior* (pp. 22-35). New York: John Wiley.
- Hartmann, J., Hollweg, M., & Nedopil, N. (2001). Quantitative Erfassung dissozialer und psychopathischer Persönlichkeiten bei der strafrechtlichen Begutachtung [Quantitative detection of dyssocial and psychopathic personalities in criminal expert assessment]. *Nervenarzt, 72*, 365-370.
- Lambrick, F. (2003). Issues surrounding the risk assessment of sexual offenders with intellectual disability. *Psychiatry, Psychology, and Law, 10*, 353-358.
- Langton, C. M., Barbaree, H. E., Seto, M. C., Peacock, E. J., Harkins, L., & Hansen, K. T. (2007). Actuarial assessment of risk for reoffense among adult sex offenders. *Criminal Justice and Behavior, 34*, 37-59.
- Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62*, 783-792.
- Nedopil, N. (2005). *Prognosen in der Forensischen Psychiatrie. Ein Handbuch für die Praxis* [Risk prediction in forensic psychiatry. A handbook for forensic practice]. Lengerich: Pabst Science.
- Noll, T., Endrass, J., Rossegger, A., & Urbaniok, F. (2006). Die Risikokalkulation für die Begehung von Gewalt- und Sexualstraftaten mit Hilfe des Static-99 [Risk prediction of violent and sexual offenses with the Static-99]. *Monatsschrift für Kriminologie und Strafrechtsreform, 89*, 24-33.
- Nunes, K. L., Firestone, P., Bradford, J. M., Greenberg, D. M., & Broom, I. (2002). A comparison of modified versions of the Static-99 and the sex offender risk appraisal guide (SORAG). *Sexual Abuse: A Journal of Research and Treatment, 14*, 253-269.
- Parks, G. A., & Bard, D. E. (2006). Risk factors for adolescent sex offender recidivism: Evaluation of predictive factors and comparison of three groups based upon victim type. *Sexual Abuse: A Journal of Research and Treatment, 18*, 319-342.
- Prentky, R. A., Lee, A. F. S., Knight, R. A., & Cerce, D. (1997). Recidivism rates among child molesters and rapists: A methodological analysis. *Law and Human Behavior, 21*, 635-659.

- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. (2006). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.
- Quinsey, V. L., Rice, M. E., & Harris, G. T. (1995). Actuarial prediction of sexual recidivism. *Journal of Interpersonal Violence, 10*, 85-105.
- Rettenberger, M., & Eher, R. (2006a). Actuarial assessment of sex offender recidivism risk: A validation of the German version of the Static-99. *Sexual Offender Treatment, 1*, 1-11.
- Rettenberger, M., & Eher, R. (2006b). *Die revidierten Kodierungsrichtlinien des Static-99* [The revised coding rules of the Static-99]. Wien: Verein für forensische Forschung und Weiterbildung. Available from <http://www.vffw.org>
- Rettenberger, M., & Eher, R. (2007a). Predicting reoffense in sexual offender subtypes: A prospective validation study of the German version of the sexual offender risk appraisal guide (SORAG). *Sexual Offender Treatment, 2*, 1-12.
- Rettenberger, M., & Eher, R. (2007b). *Aktuarische Kriminalprognose bei Sexualstraftätern: Die deutsche Version des Sex Offender Risk Appraisal Guide (SORAG)* [Actuarial risk prediction for sexual offenders: The German version of the sex offender risk appraisal guide (SORAG)]. Wien: Verein für forensische Forschung und Weiterbildung. Available from <http://www.vffw.org>
- Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology, 63*, 737-748.
- Salekin, R. T., Rogers, R., & Sewell, K. W. (1996). A review and meta-analysis of the Psychopathy Checklist and Psychopathy Checklist-Revised. *Clinical Psychology: Science and Practice, 3*, 203-215.
- Seto, M. C. (2005). Is more better? Combining actuarial risk scales to predict recidivism among adult sex offenders. *Psychological Assessment, 17*, 156-167.
- Sjöstedt, G., & Långström, N. (2001). Actuarial assessment of sex offender recidivism risk: A cross-validation of the RRASOR and the Static-99 in Sweden. *Law and Human Behavior, 25*, 629-645.
- Stadtland, C., Hollweg, M., Kleindienst, N., Dietl, J., Reich, U., & Nedopil, N. (2005). Risk assessment and prediction of violent and sexual recidivism in sex offenders: Long-term predictive validity of four risk assessment instruments. *Journal of Forensic Psychiatry and Psychology, 16*, 92-108.

Danksagung

An erster Stelle möchte ich PD Dr. med. Reinhard Eher meinen tief empfundenen Dank für die Betreuung meiner Dissertation aussprechen. Ohne seinen Einsatz und sein mir entgegengebrachtes Vertrauen wäre mein bisheriger beruflicher und wissenschaftlicher Werdegang in dieser Form nicht möglich gewesen.

Darüber hinaus möchte ich allen ehemaligen und gegenwärtigen Kollegen an der *Begutachtungs- und Evaluationsstelle für Gewalt- und Sexualstraftäter (BEST)*, die durch ihre tägliche Arbeit und Engagement jegliche wissenschaftliche Auseinandersetzung erst ermöglichen, danken. Gleiches gilt für die Kollegen, die viel Zeit und persönliche Energie aufwendeten, um mit mir zusammen die dargestellten wissenschaftlichen Untersuchungen durchzuführen und zu verfassen: Prof. Dr. med. Friedemann Pfäfflin, Dipl.-Psych. Anna Matthes, Mag. Frank Schilling und Douglas P. Boer, Ph. D., Associate Professor, sei hierfür herzlich gedankt.

Mein ganz persönlicher Dank geht an meine Familie, meine Freundin und Freunde, die mich auch in den letzten Jahren in jedem Moment unterstützten und mir durch ihre Nähe und Zuneigung die Sicherheit gaben, auf dem richtigen Weg zu sein.

Zuletzt richtet sich mein anonymer Dank an alle inhaftierten Straftäter, deren Begutachtungsdaten zum Gelingen der vorliegenden Arbeit beitrugen.

Wien, im März 2009

Martin Rettenberger