

Kriminalprognosen in der Praxis

Die Ergebnisse des International Risk Surveys (IRiS) aus Deutschland

Martin Rettenberger, Reinhard Eher, Sarah L. Desmarais, Cristina Hurducas, Karin Arbach-Lucioni, Carolina Condemarin, Kimberlie Dean, Michael Doyle, Jorge O. Folino, Verónica Godoy-Cervera, Martin Grann, Robyn Mei Yee Ho, Matthew M. Large, Thierry H. Pham, Louise Hjort Nielsen, Maria Francisca Rebocho, Kim A. Reeves, Corine de Ruiter, Katharina Seewald und Jay P. Singh

Zusammenfassung. Einschätzungen über das Risiko zukünftiger Gewalttätigkeit sind ein fester Bestandteil der alltäglichen Arbeit von Psychologinnen und Psychologen, wobei bis heute wenig darüber bekannt ist, in welcher Form kriminalprognostische Einschätzungen in der alltäglichen Berufspraxis vorgenommen werden. Durch die vorliegende Forschungsarbeit wird erstmals ein Einblick in die kriminalprognostische Praxis in Deutschland gegeben. In der vorliegenden Studie werden die Ergebnisse des International Risk Survey (IRiS; Singh et al., 2014) ausgewertet, an der weltweit 2135 Personen aus 44 Ländern teilgenommen haben. Aus Deutschland wurden 97 Psychologinnen und Psychologen sowie Angehörige anderer Berufsgruppen über ihre kriminalprognostischen Tätigkeiten befragt. Die Daten zeigen, dass mittlerweile in der Praxis mehrheitlich auf standardisierte Prognoseinstrumente zurückgegriffen wird. Die Instrumente werden nicht nur für die prognostische Einschätzung über das zukünftige Gewaltrisiko als nützlich eingestuft, sondern auch im Hinblick auf die Therapieindikation sowie die verlaufdiagnostische Untersuchung von Behandlungs- und Betreuungsfällen als hilfreich beurteilt.

Schlüsselwörter: Kriminalprognose, Begutachtung, Prognoseinstrument, Maßregelvollzug, Strafvollzug

Risk Assessment in Clinical Practice – Results of the International Risk Survey (IRiS) From Germany

Abstract. Forensic psychologists routinely assess the violence risk of their clients and patients. Nevertheless, there is only little knowledge about methods of risk assessment in the field. The present study provides the first insight into the clinical practice of risk assessment in Germany. Data were collected from 2,135 respondents from 44 countries as part of the International Risk Survey (IRiS). In Germany, 97 psychologists and members of other professional disciplines were surveyed about their clinical risk assessment practices. The data indicated that a majority of the respondents regularly use [author: edit ok?] standardized risk assessment instruments. The instruments were not only perceived as useful tools for risk assessment purposes alone but also for developing and evaluating risk management plans.

Keywords: risk assessment, assessment, risk assessment instrument, forensic psychiatry, prison system

Im deutschen Strafrechtssystem kommt kriminalprognostischen Einschätzungen eine große Bedeutung zu (Dahle, 2005), weshalb sie einen integralen Bestandteil der forensischen Sachverständigentätigkeit und der diagnostischen und therapeutischen Arbeit im Maßregel- und Strafvollzug darstellen (Basdekis-Jozsa, Rettenberger & Briken, 2013; Gretenkord, 2013). Um Kriminalprognosen zu erstellen, stehen psychowissenschaftlichen Sachverständigen unterschiedliche Methoden zur Verfügung, die

in der einschlägigen Literatur meist den drei folgenden idealtypischen Gruppen zugeordnet werden: die intuitive, die statistisch-nomothetische oder aktuarische und die klinisch-idiographische Prognosemethodik. Das intuitive Prognostizieren beschreibt die gefühlsmäßige Erfassung des Täters, die überwiegend oder ausschließlich geprägt wird durch die subjektive Vorerfahrung des Prognostikers; diese Prognoseform stellt für die richterliche bzw. die allgemein juristische Praxis das häufigste Vorgehen

Die Autorinnen und Autoren der vorliegenden Studie möchten sich sehr herzlich bei Prof. Dr. Jürgen L. Müller (Georg-August-Universität Göttingen), Prof. Dr. Daniela Hosser (Technische Universität Braunschweig), Prof. Dr. Denis Köhler (Fachhochschule Düsseldorf), Werner Stuckmann (Klinik Nette-Gut für Forensische Psychiatrie an der Rhein-Mosel-Fachklinik Andernach) und Dr. Stefan Suhling (Kriminologischer Dienst des niedersächsischen Justizvollzugs) für die Unterstützung bei der Aussendung des Links zur Teilnahme bedanken.

dar (Dahle, 2005). Aufgrund der mangelnden Transparenz intuitiven Vorgehens und der damit verbundenen eingeschränkten Nachvollziehbarkeit würde ein rein intuitives Vorgehen für den psychowissenschaftlichen Sachverständigen allerdings weder den methodischen Mindeststandards noch den rechtlichen Anforderungen genügen, die an Prognosegutachten von Sachverständigen gestellt werden (Boetticher et al., 2007).

Darüber hinaus wurde mehrfach nachgewiesen, dass intuitive Kriminalprognosen nur eine mangelhafte Vorhersageleistung aufweisen, die in der Regel nahe am Zufallsniveau liegt (Grove, Zald, Lebow, Snitz & Nelson, 2000; Quinsey, Harris, Rice & Cormier, 2006). Bereits seit den 1920er Jahren wurden wissenschaftliche Untersuchungen darüber durchgeführt, wie anhand statistisch-quantitativer Methoden die Vorhersageleistungen in psychologischen und kriminologischen Anwendungsbereichen verbessert werden kann (Burgess, 1928). Die Ergebnisse dieser Untersuchungen führte Paul E. Meehl (1954/2013) in seiner bis heute grundlegenden Monographie erstmals systematisch zusammen und folgerte aus seinen Ergebnissen, dass aktuarische Prognosemethoden einem klinisch-intuitiven Vorgehen in der Regel überlegen sind. Diese Schlussfolgerung wurde in den darauffolgenden Jahrzehnten sowohl in der allgemeinen psychologischen Diagnostik und Prognoseforschung als auch speziell im forensisch-rechtspsychologischen Prognosebereich mehrfach bestätigt (Grove & Meehl, 1996; Grove et al., 2000; Quinsey et al., 2006). Hanson und Morton-Bourgon (2009) unterstrichen die Bedeutung, die statistisch-aktuarische Prognosemethoden insbesondere im rechtspsychologisch-forensischen Bereich haben, im Rahmen einer Meta-Analyse, in der sie unter Einbezug von 118 Einzelstudien unterschiedliche Prognosemethoden bei Personen, die aufgrund sexuell motivierter Straftaten verurteilt worden waren, verglichen. Auch in dieser Untersuchung erzielte der statistisch-aktuarische Prognoseansatz die höchste Effektstärke (Cohens $d = 0.67$) im Vergleich zu einem klinisch-intuitiven ($d = 0.46$) bzw. strukturiert-klinischen ($d = 0.42$) Vorgehen.

Bei all der methodischen Überlegenheit statistisch-aktuarischer Prognosemethoden haftet ihnen im Hinblick auf die gutachterliche Tätigkeit kriminalprognostischer Sachverständiger im Rechtswesen ein entscheidendes Problem an: Unabhängig von der Höhe der jeweiligen Effektstärken bleibt die aktuarische Prognose per definitionem immer eine auf gruppenstatistisch gewonnenen Erkenntnissen basierende Vorhersage. Der Rechtsanwender fordert jedoch vom kriminalprognostischen Gutachter eine auf den jeweiligen Einzelfall abzielende individuelle Diagnostik und Prognose, die durch aktuarische Prognoseinstrumente alleine (noch) nicht erstellt werden kann (Dahle, 2010). Aus diesem Grund wurde

vorgeschlagen, gruppenstatistisch gewonnene Erkenntnisse mit einem einzelfallorientierten klinisch-strukturierten Vorgehen zu kombinieren und die Befunde aus unterschiedlichen Erkenntnisquellen im Rahmen eines übergeordneten Prognosemodells zu integrieren (in diesem Zusammenhang ist beispielsweise das von Klaus-Peter Dahle vorgeschlagene Prozessmodell der klinisch-idiographischen Prognosebeurteilung zu nennen; vgl. Dahle, 2005, 2010; zur Auswahl und Integration von unterschiedlichen Prognoseinstrumenten und -modellen siehe auch Dahle, 2013). Es ist allerdings einschränkend darauf hinzuweisen, dass die bisherige Forschung zur prognostischen Qualität integrativer Modelle noch am Anfang steht und bisher nur wenige empirische Untersuchungen diesbezüglich durchgeführt wurden. Während eine deutschsprachige Studie hierzu vielversprechende Ergebnisse erzielen konnte (Dahle, 2005), legen internationale Untersuchungen nahe, dass eine klinisch-individuelle Korrektur aktuarischer Prognoseinstrumente tendenziell eher zu einer Verschlechterung als zu einer Verbesserung der Prognoseleistung führt (diese Korrektur wird in der internationalen Literatur meist unter dem Stichwort „Clinical Override“ diskutiert; Hanson & Morton-Bourgon, 2009; Wormith, Hogg & Guzzo, 2012). Zusammenfassend überwiegt deshalb international die Skepsis gegenüber klinisch-einzelfallorientierten Anpassungen oder Korrekturen von auf statistisch-aktuarischem Wege erhaltenen prognostischen Einschätzungen (Grove & Meehl, 1996), während aus gutachterlicher Sicht die oben dargestellte Notwendigkeit eines (unter anderem) einzelfallorientierten Vorgehens zumindest im deutschsprachigen Raum bestehen bleibt (Boetticher et al., 2007).

Anknüpfend an die Forschungstradition zu strukturierten und empirisch geprüften Prognoseinstrumenten und -modellen steht dem psychowissenschaftlichen Sachverständigen heute eine Fülle von standardisierten Instrumenten und klinisch-strukturierten Verfahren zur Verfügung, deren Entwicklung und Implementierung in den letzten drei Jahrzehnten zu einer substantiellen Verbesserung der potentiellen Prognosequalität führten (Grettenkord, 2013; Rettenberger & Eher, 2012). Diese Instrumente und Verfahren werden idealtypisch meist einer von zwei Methodengruppen zugeordnet: Den aktuarischen (oder statistisch-nomothetischen) Prognoseinstrumenten sowie den klinisch-idiographischen Verfahren und Modellen. Bei den aktuarischen Prognoseinstrumenten handelt es sich um streng standardisierte Verfahren, die aus einer festgelegten Anzahl an Risikofaktoren bestehen, deren Zusammenhang mit Rückfälligkeit empirisch abgesichert sein muss. Die einzelnen Risikofaktoren sind meist klar operationalisiert und anhand eines Manuals nach festen Regeln zu bewerten. Aus diesen Einzelbewertungen wird anschließend – meist durch einfaches

Aufaddieren der einzelnen Itemwerte – ein Gesamtwert gebildet, von dem aus wiederum auf absolute (z.B. in Form von Normwerten oder Rückfallraten) und relative (z.B. in Form von Prozenträngen bzw. Perzentilen) Risiko- maße geschlossen werden kann (Eher, Schilling, Haubner-MacLean, Jahn & Rettenberger, 2012).

Demgegenüber zeichnen sich strukturiert-klinische Prognoseverfahren wie das *Historical-Clinical-Risk Management-20* (HCR-20; Webster, Douglas, Eaves & Hart, 1997) dadurch aus, dass sie die Besonderheiten des Einzelfalls sowie therapeutische und kriminalpräventive Aspekte stärker berücksichtigen (von Franqué, 2013). Instrumente aus dieser Methodengruppe, die unter Bezugnahme auf die nordamerikanischen Wurzeln meist als *Structured Professional Judgment* (SPJ)-Instrumente bezeichnet werden, bestehen zwar ebenfalls aus einer bestimmten Anzahl an Risiko- und/oder Schutzfaktoren, bieten jedoch deutlich mehr Gestaltungs- und Interpretationsspielraum bei der Erstellung der kriminalprognostischen Einschätzung. Darüber hinaus besteht ein zentrales Merkmal von SPJ-Instrumenten darin, dass keine numerische Risikokommunikation vorgenommen wird (d.h. es erfolgt keine probabilistische Prognoseangabe in Form von absoluten oder relativen Rückfallmaßen wie bei den zuvor dargestellten aktuarischen Prognoseinstrumenten), sondern die kriminalprognostische Einschätzung wird in Form von sogenannten nominalen Risikokategorien angegeben (d.h. das Rückfallrisiko wird als „niedrig“, „moderat“ oder „hoch“ eingestuft). Diese Einschätzung basiert nicht wie bei den zuvor genannten aktuarischen Prognoseinstrumenten auf einer numerischen (additiven) Verrechnung der einzelnen Itemwerte, sondern auf einer qualitativen und idiographischen Bewertung des Risikofaktors, die für jeden Einzelfall neu vorgenommen wird.

Neben den genannten Prognoseinstrumenten ist ein weiteres Instrument innerhalb der klinisch-forensischen Diagnostik von besonderer Relevanz: die *Psychopathy Checklist-Revised* (PCL-R; Hare, 2003) sowie damit eng verwandte Verfahren wie die *Screeningversion* der PCL (PCL:SV; Hart, Cox & Hare, 1995). Die PCL-R ist ein Fremdeinschätzungsverfahren, dessen 20 Items durch entsprechend ausgebildete Anwenderinnen und Anwender anhand von Akten- und Explorationsinformationen unter Verwendung eines ausführlichen Manuals (Hare, 2003) bewertet werden. Auch wenn die PCL-R ursprünglich nicht als Prognoseinstrument entwickelt wurde, wird sie seit jeher auch für kriminalprognostische Zwecke eingesetzt (Mokros, 2013). Mittlerweile existiert eine Fülle an empirischen Untersuchungen, die die kriminalprognostische Relevanz Psychopathie-assoziiertes Persönlichkeitseigenschaften belegen (Hare, 2003; Mokros, 2013; Mokros, Vohs & Habermeyer, 2014).

Über alle der genannten Instrumente und Verfahren wurde in den letzten Jahrzehnten eine kaum mehr zu überblickende Anzahl an Forschungsarbeiten durchgeführt und veröffentlicht, die meist die psychometrischen Eigenschaften der einzelnen Instrumente – vorrangig die prädiktive Validität – zum Gegenstand hatten. Gleichzeitig ist bis heute relativ wenig darüber bekannt, wie die genannten Prognosemethoden in der Praxis aufgenommen und eingesetzt werden (Viljoen, McLachlan & Vincent, 2010). Beispielsweise gibt es bisher keine verlässlichen Daten darüber, welche Instrumente von forensischen Sachverständigen in der Praxis wie häufig angewandt werden und wie hoch die wahrgenommene Nützlichkeit ist (Hurducas, Singh, de Ruiter & Petrila, 2014). Die bisher verfügbare Datenlage liefert ausschließlich Erkenntnisse für den angloamerikanischen Sprachraum, wobei, von wenigen Ausnahmen abgesehen (Viljoen et al., 2010), auch diese Studien nur sehr eingeschränkte Aussagen über eine breite Anwendungspraxis zulassen (Archer, Buffington-Vollum, Stredny & Handel, 2006; Jackson & Hess, 2007). Aus diesem Grunde wurde das *International Risk Survey* (IRiS) konzipiert, ein aus 17 internationalen Kooperationspartnern bestehender Forschungsverbund, dessen Aufgabe darin bestand, weltweit Daten über die Anwendungspraxis kriminalprognostischer Methoden zu sammeln und zu vergleichen (Singh et al., 2014). Gegenstand des vorliegenden Artikels sind die IRiS-Ergebnisse, die auf den in Deutschland erhobenen Daten basieren. Hierdurch bestand die Möglichkeit, Erkenntnisse über die praktische Anwendung kriminalprognostischer Methoden durch psychowissenschaftliche Sachverständige zu gewinnen und zu diskutieren.

Methoden

Datenerhebung und Datenauswertung

Die in IRiS enthaltenen Fragen wurden im Rahmen einer Konzeptionsphase in Zusammenarbeit mit 17 Kooperationspartnern aus unterschiedlichen Ländern (Deutschland, USA, Schweiz, Norwegen, Spanien, Chile, Australien, Großbritannien, Argentinien, Mexiko, Schweden, Hongkong, Dänemark, Portugal, Belgien, Kanada, Niederlande) formuliert (Singh et al., 2014) und orientierten sich in Teilen an bereits bestehenden angloamerikanischen Untersuchungen (Viljoen et al., 2010). Anschließend wurde die IRiS-Umfrage sowie die Einladungs- und Informationsschreiben von professionellen Übersetzern in die jeweiligen Landessprachen übersetzt und an die Kooperationspartnerinnen und -partner zur Rücküberset-

zung versandt. Nachdem auf diesem Wege die endgültige Version der IRiS-Umfrage mittels der Software *Qualtrics* (www.qualtrics.com) erstellt worden war, wurde jeder Kooperationspartner gebeten, den Link zur Umfrage über die E-Mail-Verteiler von zumindest drei nationalen Berufsverbänden zu versenden, deren Mitglieder dem rechtspsychologischen, forensisch-psychiatrischen und/oder forensisch-pflegerischen Bereich zugeordnet werden können.¹ Weltweit wurden die E-Mail-Verteiler von 59 Verbänden genutzt, wobei aus Deutschland die Mitglieder der Fachgruppe Rechtspsychologie der *Deutschen Gesellschaft für Psychologie* (DGPs), der Sektion Rechtspsychologie des *Berufsverbands Deutscher Psychologinnen und Psychologen* (BDP), des Referats Forensische Psychiatrie der *Deutschen Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde* (DGPPN), des Netzwerks Forensik der *Bundesfachvereinigung Leitender Krankenpflegepersonen der Psychiatrie e. V.* und der Abteilung Justizvollzug und Kriminologischer Dienst im *Niedersächsischen Justizministerium* um ihre Teilnahme gebeten wurden.

Die Analyse der Daten erfolgte zunächst überwiegend deskriptiv. Anschließend wurde mittels Varianzanalysen Unterschiede zwischen Angehörigen unterschiedlicher Berufsgruppen untersucht. Für diesen Analysezweck wurden die Angaben von Psychologinnen und Psychologen mit den Angaben von Psychiaterinnen und Psychiatern verglichen. Als dritte Vergleichsgruppe wurden alle übrigen Berufe in einer weiteren Gruppe zusammengefasst. Im Falle signifikanter Unterschiede wurden detailliertere Gruppenunterschiede mittels des als konservativ geltenden *Scheffé-Tests* untersucht. Häufigkeitsunterschiede wurden anhand von χ^2 -Tests sowie – aufgrund teilweise kleiner Zellenbesetzung – anhand des *exakten Tests* nach Fisher auf Signifikanz geprüft.

Stichprobe

Insgesamt nahmen 2135 Personen aus 44 Ländern an der IRiS-Umfrage teil, wovon 97 Personen aus Deutschland stammten. Gefragt nach dem beruflichen Hintergrund, ordneten sich die Teilnehmerinnen und Teilnehmer aus Deutschland zu 52% ($n = 50$) der Psychologie, zu 19% ($n = 18$) der Psychiatrie, zu 24% ($n = 23$) dem Krankenpflegepersonal und zu 6% ($n = 6$) der Antwortkategorie „Anderes“ (z. B. Sozialarbeit) zu. Die beiden letztgenannten Berufsgruppen wurden für die folgenden Auswertun-

gen zu einer gemeinsamen Substichprobe zusammengefasst (30%; $n = 29$). Das durchschnittliche Alter betrug $M = 45.58$ Jahre ($SD = 10.33$; Spannweite = 23–65), 63% ($n = 61$) waren männlich, 37% ($n = 36$) weiblich. Die durchschnittliche Berufserfahrung lag bei $M = 16.15$ Jahre ($SD = 9.98$; Spannweite = 2–40).

Ergebnisse

Die Gesamtzahl aller bisher durchgeführten kriminalprognostischen Einschätzungen der 97 teilnehmenden Personen lag im Durchschnitt bei $M = 310.02$ ($SD = 451.96$; Spannweite = 3–3000). In 56% aller Fälle wurde dabei ein standardisiertes Prognoseinstrument verwendet. Die Anzahl der kriminalprognostischen Einschätzungen in den letzten 12 Monaten betrug durchschnittlich $M = 26.40$ ($SD = 44.65$; Spannweite = 1–400). Für diesen Zeitraum wurde in 65% aller kriminalprognostischen Einschätzungen ein Prognoseinstrument eingesetzt. Sowohl im Hinblick auf die Zahl aller bisher durchgeführter kriminalprognostischer Einschätzungen, $F(2, 92) = 9.41$, $p < .001$, als auch bezüglich der vergangenen 12 Monate, $F(2, 94) = 7.77$, $p = .001$, unterschieden sich die drei Berufsgruppen signifikant, wobei Angehörige der Psychiatriegruppe für beide Zeiträume signifikant mehr Kriminalprognosen erstellten als die anderen beiden Berufsgruppen (jeweils $p < .01$). Auch bezogen auf den Anteil, in dem in allen bisherigen kriminalprognostischen Einschätzungen bzw. bei denen der letzten 12 Monate ein standardisiertes Prognoseinstrument verwendet wurde, unterschieden sich die drei Gruppen signifikant, $F(2, 94) = 13.86$, $p < .001$ bzw. $F(2, 91) = 7.95$, $p = .001$. Im Hinblick auf die Gesamtzahl aller bisher durchgeführten Einschätzungen griff die Psychologiegruppe signifikant häufiger als die Psychiatriegruppe ($p < .05$) und die dritte Gruppe ($p < .001$), in der alle sonstigen Berufsfelder zusammengefasst wurden, auf Prognoseinstrumente zurück, wohingegen sich die beiden letztgenannten Gruppen (Psychiatrie vs. Sonstige) nicht unterschieden. Im Hinblick auf die vergangenen 12 Monate war lediglich der Unterschied zwischen Vertreterinnen und Vertretern der Psychologie und der anderer Berufsgruppen signifikant ($p = .001$).

Für die Beantwortung der Frage, welches konkrete Instrument wie häufig in der Praxis angewandt wurde, wurde

¹ Um die Vergleichbarkeit zwischen den teilnehmenden Ländern sicherzustellen, verpflichtete sich jeder Kooperationspartner vor Beginn der Datenerhebung dazu, die drei genannten Berufsgruppen über die entsprechenden Berufsverbände zu kontaktieren; dieses Vorgehen wurde aus Gründen der Einheitlichkeit und Vergleichbarkeit selbst dann eingehalten, wenn in einzelnen Ländern bestimmte Berufsgruppen traditionell bei der Erstellung kriminalprognostischer Einschätzungen nur eine untergeordnete Rolle spielen.

zunächst eine vergleichsweise umfangreiche Liste mit international etablierten Prognoseinstrumenten vorgegeben. Darüber hinaus konnten bis zu drei weitere Instrumente zusätzlich in Freitextfeldern angegeben werden. Dabei ist anzumerken, dass im Fokus des Projekts die Prognose des *allgemeinen Gewaltrückfallrisikos* stand, weshalb in der vorgegebenen Liste auch nur Verfahren aufgeführt waren, die speziell für dieses Vorhersagekriterium entwickelt worden waren.² Um vor allem die aktuelle Prognosepraxis in den Mittelpunkt der Ergebnisbetrachtung zu rücken, gibt Tabelle 1 einen Überblick darüber, welche Instrumente wie häufig in den letzten 12 Monaten eingesetzt wurden.³ Dabei zeigt sich, dass die mit Abstand am häufigsten eingesetzten Verfahren die PCL-R und der HCR-20 sind. Ebenfalls vergleichsweise häufig wurden das *Forensische Operationalisierte Therapie-Risiko-Evaluations-System* (FOTRES; Rossegger et al., 2011) und der *Violence Risk Appraisal Guide* (VRAG; Quinsey et al., 2006) genannt. Während sich beim FOTRES hinsichtlich der Benutzungshäufigkeit in den vergangenen 12 Monaten keine signifikanten Unterschiede zwischen den drei Berufsgruppen ergaben, verwendeten signifikant mehr Psychologinnen und Psychologen als Psychiaterinnen und Psychiater den HCR-20 ($p < .05$), die PCL-R ($p < .001$) und den VRAG ($p < .001$)⁴, wohingegen die zuletzt genannte Gruppe signifikant häufiger als die übrigen Berufsgruppen den HCR-20 ($p < .05$) und die PCL-R einsetzte ($p < .05$). Beim VRAG zeigte sich zwischen Psychiatrie und anderen Berufsgruppen kein signifikanter Unterschied. Erwartungsgemäß war die Anwendungshäufigkeit bei HCR-20, PCL-R und VRAG in der Psychologiegruppe signifikant höher als bei den übrigen Berufsgruppen (jeweils $p < .001$).

Der durchschnittliche Zeitaufwand für die kriminalprognostische Einschätzung lag insgesamt bei $M = 10.70$ Stunden ($SD = 8.86$; Spannweite = 1–30), wobei sich die drei Berufsgruppen signifikant unterschieden, $F(2, 63) = 7.94$, $p = .001$: Während zwischen Psychologie und Psychiatrie kein Unterschied bestand, wendeten sowohl Angehörige der Psychologie ($p = .001$) als auch der

Psychiatrie ($p < .05$) mehr Zeit auf als Angehörige anderer Berufsgruppen.

Gefragt, für welchen Zweck die Prognoseinstrumente vorrangig eingesetzt werden, gaben 80 % ($n = 78$) der teilnehmenden Kolleginnen und Kollegen an, dass das Ergebnis des Instruments als Grundlage für die Einschätzung des zukünftigen Gewalt- bzw. Rückfallrisikos diene. Mit 49 % ($n = 47$) gab allerdings auch fast die Hälfte an, dass das bzw. die Prognoseinstrumente zur Therapieindikation bzw. zur Entwicklung von Behandlungs- und Behandlungsplänen eingesetzt wurden. Darüber hinaus nutzten 29 % ($n = 28$) Prognoseinstrumente zur Verlaufsdagnostik im Sinne einer Überprüfung bzw. Evaluation von bereits begonnenen Behandlungs- und Betreuungsmaßnahmen. Bei den drei letztgenannten Fragen zeigten sich lediglich bei der *Erstellung der Gewaltprognose* ein signifikanter Unterschied zwischen den Berufsgruppen, $F(2, 94) = 3.70$, $p < .05$, wobei dies insbesondere auf eine signifikant häufigere Zustimmung innerhalb der Psychologiegruppe im Vergleich zu den anderen Berufsgruppen zurückzuführen war ($p < .05$). Zwischen der Psychologie- und Psychiatriegruppe ergab sich hingegen kein signifikanter Unterschied. Bei den übrigen beiden Anwendungszwecken (*Entwicklung und Überprüfung des Behandlungsplans*, siehe Tabelle 2) zeigten sich keine signifikanten Unterschiede zwischen den Gruppen.

Wahrgenommene Nützlichkeit

Tabelle 2 zeigt, wie nützlich die Prognoseinstrumente für die genannten drei Aufgaben eingeschätzt wurden, wobei nur diejenigen Instrumente Berücksichtigung fanden, deren Verwendung in der vorangegangenen Tabelle 1 von zumindest fünf teilnehmenden Personen angegeben wurde.⁵ Im Hinblick auf die kriminalprognostische Einschätzung zukünftiger Gewalttaten wurde die Mehrheit der verwendeten Instrumente auch als nützlich eingeschätzt. Die einzige Ausnahme bildete das *Short-Term Assessment of Risk and Treatability* (START; Nicholls, Brink, Desmarais,

² Das heißt auch, dass die spezifische Vorhersage von sexuell motivierten Rückfällen zunächst ebenso wenig im Mittelpunkt stand wie die Kriminalprognosen bei jugendlichen bzw. jungen Personen, die Sexualdelikte begingen. Aus diesem Grund fanden sich Instrumente wie zum Beispiel der Static-99 (Hanson & Thornton, 2000), die speziell zur Kriminalprognose bei sexuell motivierten Straftaten konzipiert wurden, nicht in der vorgegebenen Instrumentenliste, konnten allerdings in den Freitextfeldern angegeben werden. Aus Gründen der Übersichtlichkeit werden im vorliegenden Beitrag ausschließlich die Ergebnisse zu den vorgegebenen Instrumenten vorgestellt und diskutiert.

³ Zwei Instrumente, die in der ursprünglichen IRiS-Umfrage enthalten waren, wurden nicht in Tabelle 1 aufgelistet, da sie von keiner Teilnehmerin bzw. keinem Teilnehmer aus Deutschland in den letzten 12 Monaten verwendet wurden; dabei handelte es sich um den HKT-30 (Historische, Klinische, Toekomst-30; Werkgroep Pilotstudy Risicotaxatie Forensische Psychiatrie, 2002) und den SORM (Structured Outcome Assessment and Community Risk Monitoring; Grann et al., 2005). Alle übrigen Instrumente wurden zumindest von einer Teilnehmerin bzw. einem Teilnehmer erwähnt.

⁴ Da bei den übrigen Instrumenten die Gesamtzahl der Anwenderinnen und Anwender relativ niedrig war, wurde auf berufsgruppenspezifische Vergleiche verzichtet.

⁵ Aus Gründen der Übersichtlichkeit wurden in Tabelle 2 sowohl die Antwortkategorien „Nutzlos“ und „Ziemlich nutzlos“ als auch die Kategorien „Nützlich“ und „Ziemlich nützlich“ in jeweils eine Antwortkategorie („Eher Nutzlos“ bzw. „Eher Nützlich“) zusammengefasst.

Tabelle 1. Häufigkeit der Anwendung einzelner Prognoseinstrumente in den letzten 12 Monaten (N = 97)

	Fast nie (1 – 20%)	Selten (21 – 40%)	Manchmal (41 – 60%)	Häufig (61 – 80%)	Sehr häufig (81 – 99%)	Immer (100%)	Gesamt
COVR	0	1 (1%)	0	0	0	0	1 (1%)
FOTRES	2 (2%)	3 (3%)	3 (3%)	3 (3%)	6 (6%)	10 (10%)	27 (28%)
GSIR	0	0	0	0	1 (1%)	0	1 (1%)
HCR-20	4 (4%)	3 (3%)	5 (5%)	10 (10%)	9 (9%)	10 (10%)	41 (42%)
LS/CMI	0	0	1 (1%)	1 (1%)	0	0	2 (2%)
LSI-R	2 (2%)	0	1 (1%)	3 (3%)	2 (2%)	2 (2%)	10 (10%)
OGRS	0	0	0	0	1 (1%)	0	1 (1%)
PCL-R	6 (6%)	3 (3%)	9 (9%)	5 (5%)	10 (10%)	10 (10%)	43 (44%)
PCL:SV	1 (1%)	0	4 (4%)	2 (2%)	4 (4%)	5 (5%)	16 (17%)
RAMAS	0	0	1 (1%)	0	0	0	1 (1%)
RM2000	0	1 (1%)	0	0	0	0	1 (1%)
SAPROF	0	1 (1%)	2 (2%)	2 (2%)	2 (2%)	2 (2%)	9 (9%)
START	1 (1%)	3 (3%)	1 (1%)	0	0	0	5 (5%)
V-RISK-10	0	0	0	0	0	1 (1%)	1 (1%)
VRAG	2 (2%)	3 (3%)	4 (4%)	3 (3%)	8 (8%)	6 (6%)	26 (27%)
VRS	1 (1%)	0	0	0	1 (1%)	0	2 (2%)

Anmerkungen: Die Prozentangaben in Klammern beziehen sich auf den Anteil an der Gesamtanzahl an kriminalprognostischen Einschätzungen, die in den letzten 12 Monaten vorgenommen wurden und in denen das jeweilige Instrument zum Einsatz kam. COVR = Classification of Violence Risk, FOTRES = Forensisches Operationalisiertes Therapie-Risiko-Evaluations-System, GSIR = General Statistical Information on Recidivism, HCR-20 = Historical, Clinical, Risk Management-20, LS/CMI = Level of Service/Case Management Inventory, LSI-R = Level of Service Inventory-Revised, OGRS = Offender Group Reconviction Scale, PCL-R = Psychopathy Checklist-Revised, PCL:SV = Psychopathy Checklist: Screening Version, RAMAS = Risk Assessment and Management Schedule, RM2000 = Risk Matrix 2000, SAPROF = Structured Assessment of Protective Factors, START = Short-Term Assessment of Risk and Treatability, V-RISK-10 = Violence Risk Screening-10, VRAG = Violence Risk Appraisal Guide, VRS = Violence Risk Scale.

Webster & Martin, 2006), ein vergleichsweise neues Instrument, das aktuell nur von wenigen Personen verwendet wird. Im Hinblick auf die Therapieindikation und die Entwicklung von Behandlungs- und Betreuungsplänen wurden die verwendeten Instrumente ebenfalls mehrheitlich als grundsätzlich nützlich eingeschätzt. Für die verlaufdiagnostische Überprüfung bzw. Evaluation von Behandlungs- und Betreuungsplänen wurden vor allem FOTRES, HCR-20 und START mehrheitlich als nützlich eingestuft, PCL-R, SAPROF und VRAG hingegen wurden in diesem Zusammenhang relativ häufig neutral bewertet.⁶

Es passte zur subjektiv wahrgenommenen Nützlichkeit der einzelnen Instrumente, dass jeweils nur eine Minderheit auf die Frage, ob die Verwendung der Instrumente auf institutionelle Vorgaben, persönliche Präferenz oder beides zurückzuführen sei, angab, dass dies ausschließlich aufgrund von institutionellen Vorgaben geschehe. Die höchsten Präferenzangaben erzielte START mit 80 % (beides: 20 %, n = 5), das *Level of Service Inventory-Revised*

(LSI-R; Dahle, Harwardt & Schneider-Njepel, 2012) mit 70 % (Vorgabe: 10 %, beides: 20 %, n = 10), VRAG mit 63 % (Vorgabe: 17 %, beides: 21 %, n = 24) und HCR-20 mit 53 % (Vorgabe: 15 %, beides: 33 %, n = 40). Mittlere Werte zeigten sich für PCL-R (Präferenz: 44 %, Vorgabe: 12 %, beides: 44 %, n = 41), PCL:SV (Präferenz: 33 %, Vorgabe: 27 %, beides: 40 %, n = 15) und SAPROF (Präferenz: 44 %, Vorgabe: 11 %, beides: 44 %, n = 9). Der niedrigste Präferenzwert aller Instrumente wurde für FOTRES ermittelt (Präferenz: 8 %, Vorgabe: 44 %, beides: 48 %, n = 25).

Statistisch-aktuarisch vs. klinisch-strukturiert

Im nächsten Abschnitt wurden unterschiedliche Fragen zu der im Eingangsteil dargestellten Kontroverse zwi-

⁶ Eine detailliertere Analyse zeigte, dass bei den häufiger verwendeten Instrumenten FOTRES, HCR-20, PCL-R und VRAG keine signifikanten Unterschiede zwischen den drei Berufsgruppen hinsichtlich der wahrgenommenen Nützlichkeit vorlagen. Bei den übrigen Instrumenten, die in Tabelle 2 aufgeführt sind, wurden keine weiterführende Datenanalysen vorgenommen, da diese meist nur von einer Berufsgruppe eingesetzt wurden: Das LSI-R (7 von 9), die PCL:SV (10 von 14) und der SAPROF (8 von 9) wurden überwiegend von Psychologinnen und Psychologen, der START (3 von 5) überwiegend von Psychiaterinnen und Psychiatern angewandt.

Tabelle 2. Wahrgenommene Nützlichkeit (NK) der Prognoseinstrumente im Hinblick auf die Einschätzung des zukünftigen Gewalttrisikos sowie auf die Entwicklung und die Überprüfung bzw. Evaluation von Behandlungs- und Betreuungsplänen ($N = 97$)

	NK Erstellung Gewaltprognose			NK Entwicklung Behandlungsplan			NK Überprüfung Behandlungsplan		
	Eher Nutzlos	Neutral	Eher Nützlich	Eher Nutzlos	Neutral	Eher Nützlich	Eher Nutzlos	Neutral	Eher Nützlich
FOTRES	5 (22%)	1 (4%)	17 (74%)	2 (14%)	0	12 (86%)	2 (22%)	0	7 (78%)
HCR-20	4 (11%)	4 (11%)	30 (79%)	3 (13%)	5 (21%)	16 (67%)	2 (13%)	4 (25%)	10 (63%)
LSI-R	0	2 (22%)	7 (78%)	1 (20%)	0	4 (80%)	1 (33%)	1 (33%)	1 (33%)
PCL-R	2 (5%)	5 (12%)	34 (83%)	4 (16%)	10 (40%)	11 (44%)	6 (33%)	9 (50%)	3 (17%)
PCL:SV	2 (14%)	1 (7%)	11 (79%)	1 (11%)	2 (22%)	6 (67%)	2 (33%)	1 (17%)	3 (50%)
SAPROF	0	2 (22%)	7 (78%)	0	0	9 (100%)	0	4 (57%)	3 (43%)
START	1 (20%)	2 (40%)	2 (40%)	0	1 (25%)	3 (75%)	0	1 (25%)	3 (75%)
VRAG	1 (4%)	7 (30%)	15 (65%)	5 (31%)	7 (44%)	4 (25%)	4 (40%)	5 (50%)	1 (10%)

Anmerkungen: Die Prozentangaben in Klammern beziehen sich auf den relativen Anteil an Zustimmung im Verhältnis zur Gesamtanzahl an Personen, die das Instrument anwenden. FOTRES = Forensisches Operationalisiertes Therapie-Risiko-Evaluations-System, HCR-20 = Historical, Clinical, Risk Management-20, LSI-R = Level of Service Inventory-Revised, PCL-R = Psychopathy Checklist-Revised, PCL:SV = Psychopathy Checklist: Screening Version, SAPROF = Structured Assessment of Protective Factors, START = Short-Term Assessment of Risk and Treatability, VRAG = Violence Risk Appraisal Guide.

schen aktuarisch-statistischen Instrumenten einerseits und den klinisch-strukturierten SPJ-Instrumenten andererseits gestellt. Die erste Frage in diesem Abschnitt behandelte die bevorzugte Form der Risikokommunikation: Dabei gaben 11% ($n = 11$) der teilnehmenden Kolleginnen und Kollegen an, probabilistische Risikoangaben zu verwenden⁷; 31% ($n = 30$) bevorzugten eine dichotome Beurteilung⁸ und 71% ($n = 69$) präferierten die den SPJ-Instrumenten innewohnende kategoriale Beurteilung (d.h., das Rückfallrisiko wird als „niedrig“, „moderat“ oder „hoch“ eingestuft). Im Hinblick auf die Form der Risikokommunikation zeigte sich lediglich bei der kategorialen Beurteilung ein signifikanter Unterschied zwischen den drei Berufsgruppen, $F(2, 80) = 6.65, p < .05$, der darauf zurückzuführen war, dass Mitglieder der Psychologiegruppe diese Form der Risikokommunikation signifikant häufiger verwendeten als Angehörige anderer Berufsgruppen ($p < .05$). Mit Bezugnahme auf die zuletzt genannte kategoriale Form der Risikokommunikation wurde in der nächsten Frage untersucht, ab welcher Wahrscheinlichkeit aus der subjektiven Sicht der teilnehmenden Person die Einstufung „hohes Rückfallrisiko“ bezüglich erneuter Gewalttätigkeit gerechtfertigt wäre. Wie Abbildung 1 veranschaulicht, variierte die Angabe diesbezüglich relativ stark und reichte von 5% bis zu 80% ($M = 48.68, SD = 17.50$), wobei sich varianzanalytisch kein signifikanter Unterschied zwischen den drei Berufsgruppen zeigte, $F(2, 75) = .51, p = .61$. Etwas geringer fiel die Variation der Antworten aus, wenn nach der maximalen Wahrscheinlichkeit erneuter

Gewalttätigkeit gefragt wurde, bis zu der die Bezeichnung „niedriges Rückfallrisiko“ noch gerechtfertigt ist (siehe Abbildung 2). Hier reichten die Antworten von 0% bis 51% und lagen im Mittel bei $M = 20.93\%$ ($SD = 13.47$); auch hier lagen keine signifikanten Unterschiede zwischen den Berufsgruppen vor, $F(2, 75) = .48, p = .62$.

Auf die Frage, inwieweit es überhaupt zulässig sei, gruppenstatistische Erkenntnisse auf den Einzelfall anzuwenden, lehnten 27% ($n = 21$) eine Übertragung probabilistischer Angaben auf den individuellen Begutachtungsfall generell ab, 27% ($n = 21$) zeigten diesbezüglich eine neutrale Haltung und 46% ($n = 36$) stimmten einer Anwendung gruppenstatistischer Erkenntnisse bei Einzelfallprognosen zu.⁹ Gefragt nach der Wahrscheinlichkeit, das, mittels eines statistisch-aktuarischen Prognoseinstrumentes, ermittelte Rückfallrisiko nach oben zu korrigieren, falls die klinische Erfahrung ein höheres Rückfallrisiko indiziert, gab mit 87% ($n = 68$) die klare Mehrheit an, dass eine auf klinischer Erfahrung basierende Korrektur des Ergebnisses „wahrscheinlich“ oder „sehr wahrscheinlich“ sei, während lediglich 5% ($n = 4$) der Teilnehmerinnen und Teilnehmer angaben, dass sie eine derartige Korrektur für „unwahrscheinlich“ oder „sehr unwahrscheinlich“ halten würden; die verbleibenden 8% ($n = 6$) gaben diesbezüglich eine neutrale bzw. unklare Haltung an. Weder bei den Fragen zu den prozentualen Grenzen eines hohen bzw. niedrigen Rückfallrisikos, noch bei den zuletzt genannten Fragen zur Anwendung gruppenstatistischer Erkenntnisse im Einzelfall und einer

⁷ Beispiel: Die Rückfallwahrscheinlichkeit, innerhalb von 5 Jahren nach Entlassung erneut aufgrund eines Gewaltdelikts verurteilt zu werden, liegt bei 25%.

⁸ Beispiel: Es ist „wahrscheinlich“ oder „nicht wahrscheinlich“, dass die zu begutachtende Person erneut Gewaltdelikte begehen wird.

⁹ Zur besseren Lesbarkeit wurden die Kategorien „Starke Ablehnung“ und „Ablehnung“ sowie „Starke Zustimmung“ und „Zustimmung“ in jeweils eine Antwortkategorie zusammengefasst.

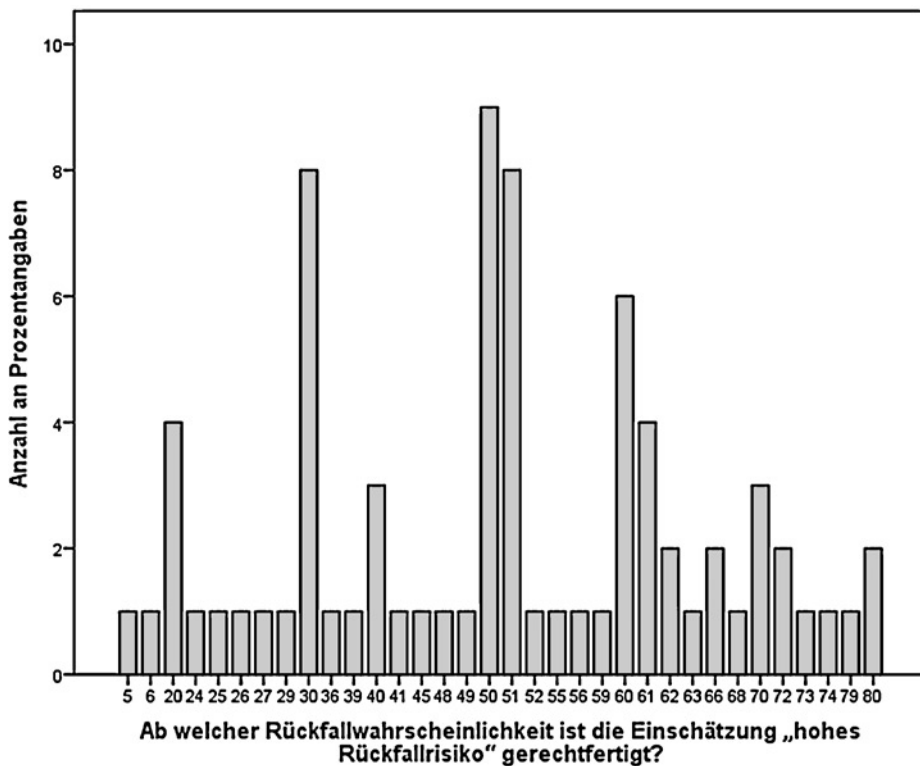


Abbildung 1. Häufigkeitsverteilung der numerischen Angaben der Versuchspersonen auf die Frage, ab welcher prozentualen Rückfallwahrscheinlichkeit die Teilnehmer die kategoriale Einstufung „hohes Rückfallrisiko“ als gerechtfertigt ansehen ($n = 75$).

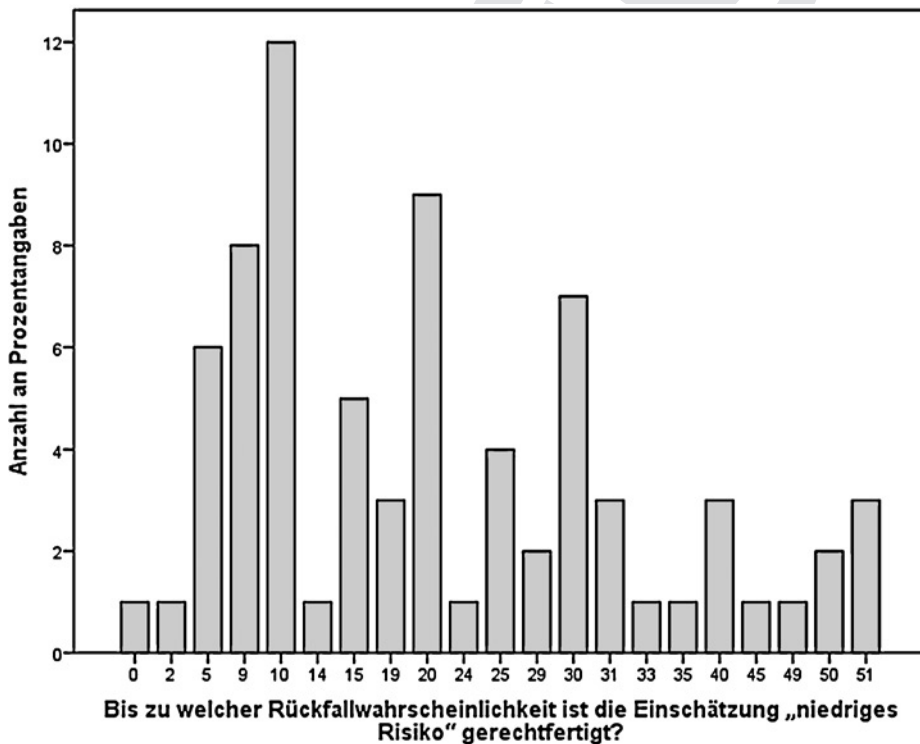


Abbildung 2. Häufigkeitsverteilung der numerischen Angaben der Versuchspersonen auf die Frage, bis zu welcher prozentualen Rückfallwahrscheinlichkeit die Teilnehmer die kategoriale Einstufung „niedriges Rückfallrisiko“ als gerechtfertigt ansehen ($n = 75$).

möglichen Korrektur eines gruppenstatistisch gewonnenen Wahrscheinlichkeitswertes zeigten sich signifikante Unterschiede zwischen den drei Berufsgruppen.

Abschließend wurden statistisch-aktuarische Prognoseinstrumente und klinische SPJ-Instrumente direkt gegenübergestellt und hinsichtlich mehrerer Kriterien ver-

Tabelle 3. Gegenüberstellung statistisch-aktuarischer Prognoseinstrumente und klinischer SPJ-Verfahren ($n = 79$)

Kriterium	Aktuarisch	Klinisch (SPJ)	Beide gleich
Anwendung erfordert mehr Zeit	14 % ($n = 11$)	51 % ($n = 40$)	35 % ($n = 28$)
Höhere Validität	23 % ($n = 18$)	33 % ($n = 26$)	44 % ($n = 35$)
Höhere Reliabilität	27 % ($n = 21$)	28 % ($n = 22$)	46 % ($n = 36$)
Sind von größerem Interesse für Gerichte	17 % ($n = 13$)	29 % ($n = 23$)	54 % ($n = 43$)
Bei forensischen Patienten besser geeignet	8 % ($n = 6$)	47 % ($n = 37$)	46 % ($n = 36$)
Bei Strafgefangenen besser geeignet	20 % ($n = 16$)	35 % ($n = 28$)	44 % ($n = 35$)

Anmerkung: SPJ = Structured Professional Judgment.

glichen (siehe Tabelle 3). Hierbei zeigte sich, dass in der Regel beide Methodengruppen mehrheitlich als gleichwertig eingestuft wurden, wobei zwei Ausnahmen festzustellen waren: Zum einen wurden SPJ-Instrumente im Vergleich zu statistisch-aktuarischen Instrumenten mehrheitlich als zeitaufwändiger eingeschätzt und zum anderen wurden SPJ-Instrumente im Vergleich zu aktuarischen Instrumenten für die kriminalprognostische Einschätzung bei forensischen Patientinnen und Patienten von der Mehrheit als relevanter eingestuft. Die zuletzt genannte Frage wurde von den drei Berufsgruppen als einzige dieses Abschnitts signifikant unterschiedlich bewertet, $F(2, 76) = 3.60$, $p < .05$: Für die Anwendung im Maßregelvollzug favorisierten Psychologinnen und Psychologen im Vergleich zu Mitgliedern anderer Berufsgruppen signifikant stärker aktuarische Prognoseinstrumente gegenüber SPJ-Verfahren ($p < .05$), während die übrigen Gruppenunterschiede keine statistische Signifikanz aufwiesen.

Diskussion

Kriminalprognostischen Einschätzungen über das zukünftige Risiko neuerlicher Straftaten kommt für die psychologische, psychotherapeutische und psychiatrische Versorgung und Betreuung von straffällig gewordenen Personen eine zentrale Bedeutung zu. In den letzten Jahrzehnten wurde eine Vielzahl an unterschiedlichen Instrumenten und Verfahren entwickelt, mit deren Hilfe die kriminalprognostische Praxis unterstützt und auf ein wissenschaftlich fundiertes Niveau gehoben werden sollte. Obwohl eine Vielzahl von Studien zur Entwicklung, Adaptation, Validierung und Normierung vorliegt, ist bis heute allerdings weitgehend unklar, ob bzw. wie die genannten Instrumente und Methoden in der Praxis rezipiert und eingesetzt werden (Viljoen et al., 2010). Vor diesem Hintergrund entstand das IRiS-Projekt mit dem Ziel, weltweit Daten über das kriminalprognostische Vor-

gehen in der Praxis zu erfassen (Singh et al., 2014). In der vorliegenden Arbeit wurden die IRiS-Ergebnisse aus Deutschland ausgewertet, so dass erstmals Informationen darüber gewonnen werden konnten, mit welchen kriminalprognostischen Methoden hierzulande gearbeitet wird und welche Meinungen und Einstellungen gegenüber diesen Methoden vorherrschend sind.

Bei der Betrachtung der Ergebnisse zeigt sich zunächst, dass auch in der deutschen Praxis eine Reihe von international etablierten Instrumenten breite Anwendung findet. Den besonders häufig genannten Instrumenten PCL-R, PCL:SV, HCR-20 und VRAG kam auch in Untersuchungen aus Großbritannien und Nordamerika eine besonders prominente Rolle zu (Archer et al., 2006; Viljoen et al., 2010), wobei zusammenfassend konstatiert werden kann, dass insbesondere die PCL-R, der HCR-20 und der VRAG nicht nur im angloamerikanischen Sprachraum, sondern offenbar zunehmend auch hierzulande zu Standardinstrumenten in der kriminalprognostischen Begutachtung werden. Ebenfalls relativ häufig kommt das FOTRES zur Anwendung, eine spezifisch deutschsprachige Entwicklung, über deren konzeptionelle und psychometrische Eigenschaften jedoch noch einige Unklarheiten herrschen (Keller et al., 2011; zum bisherigen Stand der Validierung des Verfahrens siehe Rossegger et al., 2011). Der LSI-R, der sich international großer Beliebtheit erfreut und weltweit zu den meist verwendeten und am besten validierten Verfahren gehört (Archer et al., 2006; Harwardt & Schneider-Njepel, 2013), wird im deutschsprachigen Raum noch vergleichsweise selten angewandt. Dies mag allerdings zumindest zum Teil auf den Umstand zurückzuführen sein, dass der LSI-R erst kürzlich als offizielle deutschsprachige Version veröffentlicht wurde (Dahle et al., 2012). Setzt man die Ergebnisse der Anwendungspraxis in Bezug zum internationalen Stand der Forschung bezüglich der prognostischen Qualität der einzelnen Instrumente und Verfahren (Yang, Wong & Coid, 2010), ist zu konstatieren, dass die Anwenderinnen und Anwender kriminalprognostischer Verfahren in Deutschland weitgehend auf Instrumente und Verfahren zurück-

greifen, deren prädiktive Validität als empirisch gut abgesichert gelten kann. So gehörten die in der vorliegenden Studie häufig genannten Instrumente PCL-R, HCR-20 und VRAG auch in der Meta-Analyse von Yang et al. (2010) zu den Instrumenten, die die höchsten Validitätsindizes erzielten.

Zur breiten Anwendungspraxis passend werden die Prognoseinstrumente mehrheitlich als nützlich eingestuft, und zwar nicht nur für die *Erstellung der Gewaltprognose* per se, sondern auch hinsichtlich der *Entwicklung* und *Überprüfung von Behandlungsplänen*. Dieses Ergebnis spiegelt die Empirie der forensischen Evaluationsforschung über die Wirksamkeit von Behandlungs- und Betreuungsmaßnahmen bei straffällig gewordenen Personen wider, der zufolge Prognoseinstrumenten sowohl bei der Planung als auch bei Durchführung von Interventionen eine maßgebliche Rolle zukommt (Andrews & Bonta, 2006; Gretenkord, 2013; Hanson et al., 2009). In Übereinstimmung mit der subjektiv wahrgenommenen Nützlichkeit der einzelnen Instrumente gaben die deutschen IRiS-Teilnehmenden mehrheitlich an, dass sie die Instrumente vor allem oder zumindest unter anderem aufgrund persönlicher Präferenz auswählten und verwendeten.

Die Kontroverse zwischen aktuarisch-statistischen und klinisch-strukturierten SPJ-Instrumenten scheint auf den ersten Blick für die überwiegend praktisch tätigen Kolleginnen und Kollegen weniger relevant zu sein als für die jeweiligen Wissenschaftlerinnen und Wissenschaftler, in deren Publikationen intensive und zum Teil auch hitzige Diskussionen über Pro und Kontra der beiden Methodengruppen geführt werden (z. B. Quinsey et al., 2006). Bei den meisten der abgefragten Kriterien wurden aktuarische Instrumente und SPJ-Verfahren als weitgehend gleichwertig eingestuft. Damit werden zum einen vorangegangene Untersuchungen aus Nordamerika bestätigt (Viljoen et al., 2010). Bei näherer Betrachtung zeigen sich jedoch zumindest indirekt gewisse Vorbehalte gegenüber dem statistisch-aktuarischen Vorgehen im Vergleich zu den eher offeneren SPJ-Instrumenten. So lässt sich an der Anwendungshäufigkeit der letzten 12 Monate ablesen, dass mit dem VRAG lediglich ein aktuarisches Instrument unter den als aktuell tatsächlich praxisrelevant einzustufenden Verfahren zu finden ist. Ein Kernmerkmal von aktuarischen Prognoseinstrumenten besteht darin, dass sie der Anwenderin bzw. dem Anwender empirisch ermittelte Rückfallraten pro Punktwert oder Risikokategorie für einen bestimmten Zeitraum zur Verfügung stellen (auch bekannt als *absolute Risikomaße* oder *Normwerte*; Eher et al., 2012). Diese Form der Risikokommunikation wird allerdings lediglich von einer Minderheit (11%) bevorzugt. Ein zentraler Grund für diese skeptische Haltung bei überwiegend praktisch tätigen Sachverständigen wird sicherlich darin begründet sein, dass, wie im Einlei-

tungsteil dargestellt, das deutsche Strafrecht immer eine auf den jeweiligen Einzelfall abzielende individuelle kriminalprognostische Einschätzung fordert. Dies kann allerdings per definitionem durch gruppenstatistisch gewonnene Erkenntnisse wie sie statistisch-aktuarischen Prognoseinstrumenten zugrunde liegen nur bedingt geschehen (Boetticher et al., 2007; Dahle, 2010). Es muss an dieser Stelle allerdings darauf hingewiesen, dass auch die überwiegende Mehrheit vordergründig individuumszentrierter Informationen schlussendlich direkt oder indirekt empirisch aus gruppenstatistischen Beobachtungen abgeleitet wurden – ein Umstand der in der einschlägigen Forschungsliteratur von Anbeginn bemerkt und diskutiert wurde (Meehl, 1954/2013).

Mit einer Zustimmung von über 70% favorisierte die klare Mehrheit eine kategoriale Form der Risikokommunikation, die auch für SPJ-Instrumente die Kommunikationsmethode der Wahl darstellt (von Franqué, 2013). Problematisch an dieser Form der Risikokommunikation ist allerdings, dass sich sowohl psychowissenschaftliche Sachverständige als auch potentielle Auftraggeberinnen und Auftraggeber stark darin unterscheiden können, was sie genau unter einem niedrigen, moderaten oder hohen Rückfallrisiko verstehen (Hilton, Carter, Harris & Sharpe, 2008). Darüber hinaus besteht ein weiteres grundsätzliches Problem darin, dass bisher kaum juristischer Konsens darüber herrscht, wie die vielfältigen Risikobeschreibungen, die in den Gesetzestexten enthalten sind (für einen Überblick im deutschen Rechtssystem siehe Boetticher et al., 2007), in die probabilistisch-empirische Sprache der Sachverständigen übersetzt werden könnten oder sollten (Prentky, Janus, Barbaree, Schwartz & Kafka, 2006). Diese grundsätzlichen Probleme der Risikokommunikation wurden auch durch die vorliegende Studie bestätigt: Während die individuellen Grenzen im Hinblick auf die Einschätzung eines niedrigen Rückfallrisikos zwischen 0% und 51% variierten, reichten die Angaben bei der Kategorisierung eines hohen Risikos sogar von 5% bis 80%. Mit anderen Worten unterscheiden sich sogar psychowissenschaftliche Expertinnen und Experten extrem darin, bis wann sie von einem niedrigen bzw. ab wann sie von einem hohen Risiko sprechen. In Anbetracht der Notwendigkeit eines interdisziplinär geführten Diskurses bei der Einschätzung, Interpretation und Begründung kriminalprognostischer Einzelfallentscheidungen ist zu befürchten, dass die Variabilität im Urteil in der Alltagspraxis eher zu- als abnehmen dürfte. Aus methodischer Sicht ähnlich problematisch ist das Ergebnis zu bewerten, dass mit annähernd 90% die überwiegende Mehrheit angab, aufgrund ihrer klinischen Erfahrung das Ergebnis eines aktuarischen Prognoseinstrumentes zu korrigieren, falls ihnen das notwendig erscheint. Zwar ist die Studienlage innerhalb der forensischen Prognosefor-

schung in dieser Frage noch begrenzt, die bisher verfügbaren Daten lassen aber den Schluss zu, dass eine klinisch-individuelle Korrektur aktuarisch gewonnener Ergebnisse (sogenannter *Clinical Override*) in der Regel zu einer Verschlechterung der Prognoseleistung führt (Hanson & Morton-Bourgon, 2009; Wormith et al., 2012).

Ein Vergleich der Berufsgruppen, der insbesondere vor dem Hintergrund der zuletzt genannten Ergebnisse signifikante Resultate erwarten ließ, ergab bei praktisch tätigen Kolleginnen und Kollegen vergleichsweise geringe Unterschiede. Aufgrund der methodisch-statistischen Ausbildung hätte erwartet werden können, dass Personen mit psychologischer Ausbildung, insbesondere statistisch-aktuarischer Prognoseinstrumente, deutlich positiver bewerten würden als ihre ärztlichen Kolleginnen und Kollegen sowie Angehörige anderer Berufsgruppen. Eine allgemein stärker ausgeprägte Präferenz für strukturierte Verfahren zeigte sich in Ansätzen: So gaben die Psychologinnen und Psychologen häufiger als andere Berufsgruppen an, auf standardisierte Prognoseinstrumente zurückzugreifen und dabei vor allem international etablierte und psychometrisch gut abgesicherte Prognoseinstrumente zu bevorzugen (Singh et al., 2014; Yang et al., 2010). Eine naheliegende Hypothese zur Erklärung dieses Ergebnisses könnte in der akademischen Ausbildung im Bereich der psychologischen Diagnostik liegen, in deren Rahmen die Diskussion von psychometrischen Qualitätskriterien eine zentrale Rolle einnimmt (Kury & Adams, 2010). Bei den weiteren Vergleichen ergaben sich lediglich marginale Unterschiede, aufgrund derer die Ableitung eines systematischen Unterschieds zwischen den Berufsgruppen jedenfalls verfrüht erscheint.

Insgesamt müssen bei der Interpretation der Ergebnisse einige methodische Einschränkungen und Probleme berücksichtigt werden, die die Aussagekraft und Generalisierung schmälern könnten. So sind online-basierte Umfragen besonderen potentiellen Verzerrungen ausgesetzt (Blank, Fielding & Lee, 2008), die im vorliegenden Fall möglicherweise dazu führten, dass eine allgemein positive Haltung gegenüber Prognoseinstrumenten überrepräsentiert war. Personen, die eine eher negative Einstellung zu Prognoseinstrumenten aufweisen, könnten möglicherweise eine Teilnahme generell abgelehnt haben bzw. sich spätestens nach dem Lesen der Eingangserklärung gegen eine Teilnahme entschieden haben. For-

schungen über die Bereitschaft zur Teilnahme an Umfragen zeigten, dass positive Bewertungen des Umfragegegenstands bzw. der als Umfrageinitiatoren vermuteten Personen bei einem substantiellen Teil von potentiellen Teilnehmerinnen und Teilnehmern einen Einfluss auf die Teilnahmebereitschaft besitzt (Cialdini, 2009; Groves, Singer & Corning, 2000). Aus diesem Grund muss angenommen werden, dass die Teilnahmebereitschaft bei Personen, die generell eine positive Haltung gegenüber standardisierten Kriminalprognoseverfahren haben und sie dementsprechend in der Praxis regelmäßig einsetzen, höher ist als bei Personen, die der Anwendung von Prognoseinstrumenten grundsätzlich ablehnend gegenüberstehen. In diesem Zusammenhang ist zu monieren, dass es aus mehreren Gründen nicht möglich war, eine exakte Teilnahme- bzw. Ablehnungsrate zu berechnen: Zum einen liegen keine gesicherten Zahlen darüber vor, wie viele Personen über die E-Mail-Verteiler der Berufsgesellschaften kontaktiert wurden und zum anderen muss als wahrscheinlich angenommen werden, dass ein substantieller Anteil an kontaktierten Personen in mehreren Verteilern Mitglied ist (z. B. sowohl in der Sektion Rechtspsychologie des BDP als auch in der Fachgruppe Rechtspsychologie der DGPs). Des Weiteren war es aus technischen Gründen nicht möglich, eine Abbruchquote unmittelbar nach bzw. während des Lesens des Einführungstextes zu ermitteln. Als Orientierungspunkte für die Teilnahmebereitschaft können an dieser Stelle zunächst frühere Untersuchungen aus Nordamerika genannt werden, die ebenfalls elektronische Umfragen zur kriminalprognostischen Praxis durchführten, und in denen die Teilnahmequote zwischen 30 % und 60 % lag (Archer et al., 2006; Viljoen et al., 2010). Aufgrund der offiziellen Mitgliederzahlen der Berufsverbände, die sich bereit erklärten, den Link zur Umfrage an ihre Mitglieder weiterzuleiten, muss angenommen werden, dass die Teilnahmequote in der vorliegenden Untersuchung am unteren Rand bzw. unter dem angegebenen Referenzbereichs nordamerikanischer Studien liegt.¹⁰ In diesem Zusammenhang ist die Selektivität der Stichprobe als eine weitere methodische Einschränkung der vorliegenden Studie zu nennen. Aufgrund der Vorgaben des internationalen Forschungsverbands, der die IRiS-Umfrage konzipierte (Singh et al., 2014), sollten lediglich die im Methodenteil angegebenen Berufsverbände kontaktiert werden.¹¹ Trotz dieser Ein-

¹⁰ Eine exakte Berechnung wurde schon alleine dadurch ausgeschlossen, dass den Berufsverbänden keine exakten Zahlen darüber vorlagen, wie viele unabhängige Einzelpersonen in den jeweiligen Verteilern aufgenommen waren. Noch schwieriger gestaltete sich die Frage, wie viele Personen überwiegend oder ausschließlich mit kriminalprognostischen Aufgaben betraut war bzw. anders gefragt, wie viele Personen überhaupt keine Berührungspunkte mit Kriminalprognosen haben, weil sie beispielsweise ausschließlich im familienrechtlichen Bereich oder der Glaubhaftigkeit als Sachverständige auftreten.

¹¹ Die über die offiziellen Berufsverbände (DGPs, BDP, DGPPN und Netzwerk Forensik der Bundesfachvereinigung Leitender Krankenpflegepersonen der Psychiatrie) hinaus gehende Verteilung des Umfrage-Links durch die Abteilung Justizvollzug und Kriminologischer Dienst im Nie-

schränkungen liefert die vorliegende Studie wissenschaftlich und praktisch relevante Einblicke in die kriminalprognostische Praxis in Deutschland, da erstmals überhaupt Daten darüber vorliegen, welche Verfahren in welcher Art und Weise von Praktikern angewandt werden. Die Ergebnisse lassen darauf schließen, dass ein großer Anteil an praktisch tätigen Personen im Bereich der Forensischen Psychologie und Psychiatrie bei der Erstellung kriminalprognostischer Einschätzungen auf standardisierte Instrumente und Verfahren zurückgreift und sie als ein nützliches Werkzeug einstuft.

Literatur

- Andrews, D. A. & Bonta, J. (2006). *The psychology of criminal conduct* (4th ed.). Newark, NJ: Anderson Publishing.
- Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V. & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment*, 87, 84–94.
- Basdekis-Jozsa, R., Rettenberger, M. & Briken, P. (2013). Begutachtung bei sexuellen Störungen. In P. Briken & M. Berner (Hrsg.), *Praxisbuch Sexuelle Störungen* (S. 258–271). Stuttgart: Thieme.
- Blank, G., Fielding, N. G. & Lee, R. M. (2008). *The SAGE handbook of online research methods*. London: SAGE.
- 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.
- Burgess, E. W. (1928). Factors determining success or failure on parole. In A. A. Bruce, A. J. Harno, E. W. Burgess & J. Landesco (Eds.), *The workings of the indeterminate sentence law and the parole system in Illinois* (pp. 221–234). Springfield, IL: State Board of Parole.
- Cialdini, R. B. (2009). *Influence: Science and practice* (5th ed.). Boston, MA: Pearson Education.
- Dahle, K.-P. (2005). *Psychologische Kriminalprognose*. Herbolzheim: Centaurus Verlag.
- Dahle, K.-P. (2010). Die Begutachtung der Gefährlichkeits- und Kriminalprognose des Rechtsbrechers. In R. Volbert & K.-P. Dahle (Hrsg.), *Forensisch-psychologische Diagnostik im Strafverfahren* (S. 67–114). Göttingen: Hogrefe.
- Dahle, K.-P. (2013). (Sach-)Verständige Auswahl und Integration von Basisrateninstrumenten und Prognoseinstrumenten der 'dritten Generation'. In M. Rettenberger & F. von Franqué (Hrsg.), *Handbuch kriminalprognostischer Verfahren* (S. 337–346). Göttingen: Hogrefe.
- Dahle, K.-P., Harwardt, F. & Schneider-Njepel, V. (2012). *LSI-R: Inventar zur Einschätzung des Rückfallrisikos und des Betreuungs- und Behandlungsbedarfs von Straftätern. Deutsche Version des Level of Service Inventory-Revised nach Don Andrews und James Bonta*. Göttingen: Hogrefe.
- Eher, R., Schilling, F., Haubner-MacLean, T., Jahn, T. & Rettenberger, M. (2012). Ermittlung des relativen und absoluten Rückfallrisikos mithilfe des Static-99 in einer deutschsprachigen Population entlassener Sexualstraftäter. *Forensische Psychiatrie, Psychologie, Kriminologie*, 6, 32–40.
- Franqué, F. von (2013). Strukturierte, professionelle Risikobeurteilungen. In M. Rettenberger & F. von Franqué (Hrsg.), *Handbuch kriminalprognostischer Verfahren* (S. 357–380). Göttingen: Hogrefe.
- Grann, M., Sturidsson, K., Haggård-Grann, U., Hiscoke, U. L., Alm, P.-O., Dernevik, M. et al. (2005). Methodological development: Structured outcome assessment and community risk monitoring (SORM). *International Journal of Law & Psychiatry*, 28, 442–256.
- Gretenkord, L. (2013). Warum Prognoseinstrumente? In M. Rettenberger & F. von Franqué (Hrsg.), *Handbuch kriminalprognostischer Verfahren* (S. 19–36). Göttingen: Hogrefe.
- Grove, W. M. & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, 2, 293–323.
- Grove, W. M., Zald, D.H., Lebow, B. S., Snitz, B. E. & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, 12, 19–30.
- Groves, R. M., Singer, E. & Corning, A. (2000). Leverage-salience theory of survey participation: Description and an illustration. *Public Opinion Quarterly*, 64, 299–308.
- Hanson, R. K., Bourgon, G., Helmus, L. & Hodgson, S. (2009). The principles of effective correctional treatment also apply to sexual offenders: A meta-analysis. *Criminal Justice and Behavior*, 36, 865–891. [Verw. fehlt]
- Hanson, R. K. & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment*, 21, 1–21.
- 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.
- Hare, R. D. (2003). *Hare Psychopathy Checklist-Revised* (2nd ed.). Toronto: Multi-Health Systems.
- Hart, S. D. (1998) [Verw. fehlt]. The role of psychopathy in assessing risk for violence: Conceptual and methodological issues. *Legal and Criminological Psychology*, 3, 121–137.
- Hart, S. D., Cox, D. N. & Hare, R. D. (1995). *Hare Psychopathy Checklist: Screening Version*. Toronto: Multi-Health Systems.
- Harwardt, F. & Schneider-Njepel, V. (2013). LSI-R – Level of Service Inventory-Revised. In M. Rettenberger & F. von Franqué (Hrsg.), *Handbuch kriminalprognostischer Verfahren* (S. 243–253). Göttingen: Hogrefe.
- Hilton, N. Z., Carter, A. M., Harris, G. T. & Sharpe, A. J. B. (2008). Does using nonnumerical terms to describe risk aid violence risk communication? Clinician agreement and decision making. *Journal of Interpersonal Violence*, 23, 171–188.
- Hurdacas, C. C., Singh, J. P., de Ruiter, C. & Petrila, J. (2014). Violence risk assessment tools: A systematic review of surveys. *International Journal of Forensic Mental Health*, 13, 181–192.
- Jackson, R. L. & Hess, D. T. (2007). Evaluation for civil commitment of sex offenders: A survey of experts. *Sexual Abuse: A Journal of Research and Treatment*, 19, 425–448.
- Keller, F., Kliemann, A., Karanedialkova, D., Schnoor, K., Schütt, U., Keiper, P. et al. (2011). Beurteilerübereinstimmung im Forensischen Operationalisierten Therapie-Risiko-Evaluations-System. *Nervenheilkunde*, 30, 813–817.

dersächsischen Justizministerium ergab sich erst durch die erstmalige Kontaktaufnahme im Rahmen der ersten Aussendung und war zunächst nicht als separate Maßnahme der Stichprobenrekrutierung vorgesehen.

- Kury, H. & Adams, B. (2010). Prognosegutachten im Strafvollzug. *Forum Strafvollzug – Zeitschrift für Strafvollzug und Straffälligenhilfe*, 59, 81–87.
- Meehl, P. E. (1954/2013). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis, MN: University of Minnesota Press.
- Mokros, A. (2013). PCL-R/PCL:SV – Psychopathy Checklist-Revised/Psychopathy Checklist: Screening Version. In M. Rettenberger & F. Franqué, von (Hrsg.), *Handbuch kriminalprognostischer Verfahren* (S. 83–107). Göttingen: Hogrefe.
- Mokros, A., Vohs, K. & Habermeyer, E. (2014). Psychopathy and violent reoffending in German-speaking countries: A meta-analysis. *European Journal of Psychological Assessment*, 30, 117–129.
- Nicholls, T. L., Brink, J., Desmarais, S. L., Webster, C. D. & Martin, M. L. (2006). The Short-Term Assessment of Risk and Treatability (START): A prospective validation study in a forensic psychiatric sample. *Assessment*, 13, 313–327.
- Prentky, R. A., Janus, E., Barbaree, H. E., Schwartz, B. K. & Kafka, M. P. (2006). Sexually violent predators in the courtroom: Science on trial. *Psychology, Public Policy, and Law*, 12, 357–393.
- Quinsey, V. L., Harris, G. T., Rice, M. E. & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk* (2nd ed.). Washington, DC: American Psychological Association.
- Rettenberger, M. & Eher, R. (2012). Möglichkeiten und Grenzen der Verwendung von Prognoseinstrumenten bei Sexualstraftätern. In J. M. Müller, M. Rösler, P. Briken, P. Fromberger & K. Jordan (Hrsg.), *EFPPP Jahrbuch 2012 – Empirische Forschung in der Forensischen Psychiatrie, Psychologie und Psychotherapie* (S. 121–126). Berlin: Medizinisch-Wissenschaftliche Verlagsgesellschaft.
- Rossegger, A., Laubacher, A., Moskvitin, K., Villmar, T., Palermo, G. B. & Endrass, J. (2011). Risk assessment instruments in repeat offending: The usefulness of FOTRES. *International Journal of Offender Therapy and Comparative Criminology*, 55, 716–731.
- Singh, J. P., Desmarais, S. L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., Dean, K. et al. (2014). International perspectives on the practical application of violence risk assessment: A global survey of 44 countries. *International Journal of Forensic Mental Health*, 13, 193–206.
- Singh, J. P., Grann, M. & Fazel, S. [Verw. fehlt] (2011). A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review*, 31, 499–513.
- Viljoen, J. L., McLachlan, K. & Vincent, G. M. (2010). Assessing violence risk and psychopathy in juvenile and adult offenders: A survey of clinical practices. *Assessment*, 17, 377–395.
- Webster, C. D., Douglas, K. S., Eaves, D. & Hart, S. D. (1997). *HCR-20: Assessing the risk for violence (version 2)*. Vancouver: Mental Health, Law, and Policy Institute, Simon Fraser University.
- Werkgroep Pilotstudie Risicotaxatie Forensische Psychiatrie (2002). *Findings of a nationwide pilot study on the HKT-30*. The Hague: Ministerie van Justitie.
- Wormith, J. S., Hogg, S. & Guzzo, L. (2012). The predictive validity of a general risk/needs assessment inventory on sexual offender recidivism and an exploration of the professional override. *Criminal Justice and Behavior*, 39, 1511–1538.
- Yang, M., Wong, S. C. P. & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk-assessment tools. *Psychological Bulletin*, 136, 740–767.

Jun.– Prof. Dr. Martin Rettenberger

Abteilung Sozialpsychologie
Psychologisches Institut
Binger Straße 14–16
55122 Mainz
martin.retttenberger@uni-mainz.de

PhD Jay P. Singh

Singh Institute of Forensic Research
LLC 1170 Plaza America Drive
Suite 810 Reston
VA 20190
jaysingh@sifrllc.com (korrespondierender Autor)