

<b>Zeitschrift:</b>	Schweizerische Zeitschrift für Kriminologie = Revue suisse de criminologie = Rivista svizzera di criminologia = Swiss Journal of Criminology
<b>Herausgeber:</b>	Schweizerische Arbeitsgruppe für Kriminologie
<b>Band:</b>	5 (2006)
<b>Heft:</b>	1
<b>Artikel:</b>	Grundsätzliches zur Prognose und zum Einsatz von Prognoseinstrumenten zur Beurteilung von Rückfallgefahren bei Straftätern
<b>Autor:</b>	Noll, Thomas / Endrass, Jérôme / Urbaniok, Frank
<b>DOI:</b>	<a href="https://doi.org/10.5169/seals-1050873">https://doi.org/10.5169/seals-1050873</a>

### **Nutzungsbedingungen**

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

### **Conditions d'utilisation**

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

### **Terms of use**

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

**Download PDF:** 02.02.2026

**ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>**

Thomas Noll, Jérôme Endrass, Frank Urbaniok

# Grundsätzliches zur Prognose und zum Einsatz von Prognoseinstrumenten zur Beurteilung von Rückfallgefahren bei Straftätern

## Zusammenfassung

Im vorliegenden Beitrag werden die zwei grundsätzlichen Methoden zur Bestimmung des Rückfallrisikos bei Straftätern vorgestellt. Dazu werden zunächst die wichtigsten Termini der Prognosewissenschaften erläutert. Weiter beschäftigt sich der Aufsatz mit den Möglichkeiten und Grenzen der Vorhersehbarkeit menschlichen Verhaltens. Der Präsentation der zwei klassischen Ansätze zur prognostischen Einschätzung von Rückfallgefahren (die klinische und die statistische Methode) mit ihren Vorteilen und Schwächen folgt die Diskussion des Versuchs eines neuen qualitativen Schritts im Rahmen der Weiterentwicklung klinischer Instrumente und Dokumentationssysteme in Form eines neuen Prognoseinstruments.

**Schlüsselwörter:** Prognoseinstrumente – statistische Methode – klinische Methode – Vorhersehbarkeit menschlichen Verhaltens.

## Résumé

L'article suivant présente les deux principales méthodes d'évaluation du risque de récidive des délinquants. Les termes les plus importants des sciences du pronostic sont d'abord expliqués. L'article examine ensuite les possibilités et les limites de la prédiction du comportement humain. La présentation des deux méthodes classiques pour pronostiquer les récidives criminelles – à savoir les méthodes clinique et statistique – en élucide les avantages et les inconvénients. Suit une discussion sur la tentative d'un pas qualitatif dans le cadre de l'évolution des instruments cliniques et des systèmes documentaires sous forme d'un nouvel instrument de pronostic.

**Mots-clés:** instruments de pronostic – méthode statistique – méthode clinique – prédiction du comportement humain.

## Summary

The following article presents the two basic methods of determining the risk of recidivism for offenders. After explaining the most important technical terms of prognostics, the possibilities and limitations of predicting human behaviour are discussed. The illustration of the two classical methods of predicting criminal recidivism – the clinical and the statistical one – including their advantages and weaknesses, is followed by the discussion of a newly developed clinical instrument representing a further qualitative stage in the evolution of prognostic instruments and documentation systems.

**Keywords:** prognostic instruments – statistical method – clinical method – prediction of human behaviour.

## 1. Einleitung

Die Gesellschaft hat im Interesse der öffentlichen Sicherheit das berechtigte Bedürfnis zu wissen, mit welcher Wahrscheinlichkeit Straftäter vermutlich wieder delinquieren werden. Dies gilt besonders bei Gewaltstraftäten. Nicht nur das Gerichtsurteil mit dem Entscheid darüber, ob eine sichernde Massnahme, z.B. eine Verwahrung, nötig sein wird oder nicht, hängt von der Rezidiv-Risikobeurteilung des Täters ab, sondern auch Entscheide über bedingte Entlassung und Modalitäten des Strafvollzugs (Vollzugslockerungen). Zum Teil werden auch «Rückwärtsprognosen», bei denen von der Tat auf wahrscheinliche Tätermerkmale geschlossen wird, zu Fahndungszwecken durchgeführt, was etwa in den USA Aufgabe der so genannten Profiler des FBI ist oder bei uns unter dem Begriff der operativen Fallanalyse (OFA) bekannt ist.

Risikobeurteilungen, insbesondere Gefährlichkeitseinschätzungen bei Gewaltstraftätern nach spektakulären Rückfällen, stehen immer wieder im Fokus der öffentlichen Kritik, und die Diskussionen darüber werden häufig auf einer sehr emotionalen Ebene geführt. Dies hat in der Schweiz die Debatte über die «Verwahrungsinitiative», die eine Verschärfung der sichernden Massnahmen des Schweizerischen Strafgesetzbuchs herbeiführen soll, einmal mehr gezeigt. Das umfassende Medieninteresse ausschliesslich auf ein Bedürfnis der Öffentlichkeit nach Unterhaltung oder auf unreflektierte Vergeltungsmotive der Bevölkerung zurückzuführen, wäre falsch. Die rege Anteilnahme hat auch mit der potentiellen eigenen Betroffenheit der Bürger zu tun, in Zukunft selber Opfer werden zu können – sei dies als direktes Opfer oder als Angehörige/r eines Opfers. Was auch immer unsere aktuelle Rolle sein mag – besorgter Bürger, Opfer, Täter oder involvierte Fachpersonen (z.B. Richter) – wir alle haben das berechtigte Bedürfnis zu wissen, wie Risikoprognosen (Gefährlichkeitseinschätzungen) durchgeführt werden (Andrews and Bonta, 2003). Gefordert ist also vor dem Hintergrund der zum Teil äusserst

schwerwiegenden Folgen der psychiatrischen Rezidiv-Risikoeinschätzung eine grösstmögliche Transparenz.

Ziel der vorliegenden Arbeit ist es, einige wichtige Begriffe der Prognoseforschung zu definieren und einen Überblick über die zwei verschiedenen Arten von Instrumenten zur Risikoeinschätzung von Rückfällen – die statistische und die klinische – darzustellen, wobei die Betrachtung schwerpunktmaßig aus der Perspektive der statistischen Methode erfolgt.

## 2. Definition einiger Begriffe

Trotz der relativen Geläufigkeit der folgenden Begriffe sollen sie kurz definiert werden, da ihnen im Bereich der Prognosewissenschaften eine besondere Bedeutung zukommt.

### *Rezidiv*

Obwohl Rezidive im allgemeinen Sprachgebrauch als «Wiederauftreten einer Krankheit» definiert werden (Lexikonredaktion, 1999), wird im Angelsächsischen «recidivism» als «habitual relapse into crime» umschrieben (Fowler and Fowler, 1970), und auch im deutschsprachigen Raum wird in der Forensik mit «Rezidiv» ein «Rückfall» im Sinne einer erneuten Deliktbegehung verstanden. Dabei kann unterschieden werden, ob es sich um irgendein Delikt (Killias, 2002), um einschlägige Delikte oder sogar um einschlägig-gewalttätige Delikte handelt (Quinsey et al., 2003).

### *Gelegenheit*

Studien über Rezidive müssen auch Kriterien bezüglich der Gelegenheit, die potentielle Rezidivisten zum Rückfälligwerden hatten, beinhalten. Es macht wenig Sinn, die Rückfälligkeit von Straftätern zu untersuchen, solange sie sich noch im geschlossenen Strafvollzug befinden, da sie dort offensichtlich weniger Gelegenheit haben, rückfällig zu werden, als in Freiheit. Weniger eindeutig sind Situationen, in denen eine Person vom geschlossenen Vollzug in die Halbfreiheit oder von einer geschlossenen auf eine offene psychiatrische Abteilung verlegt wird. Beim Studium von Rückfälligkeit muss die Entscheidung darüber, in welchen Fällen Gelegenheit gegeben war und in welchen nicht, bis zu einem gewissen Grad arbiträr getroffen werden. Das Merkmal «Gelegenheit» ist vor allem relevant bei denjenigen Personen, die (offenbar) nicht rückfällig geworden sind. So muss z.B. bei Rückfälligkeitstudien immer auch geprüft werden, ob die betreffende Person nur

deshalb laut Aktenlage nicht rückfällig wurde, weil sie ausgewandert oder gestorben ist. Bei der Gruppe der Rückfälligen ist die Operationalisierung der «Gelegenheit» eher im Hinblick auf die genaue Analyse des Tatverhaltens und der Risikoanalyse von Bedeutung, nicht aber für die Messung der Rückfallraten an sich. Aufgrund der Tatsache der registrierten Rückfälligkeit muss de facto irgend eine Gelegenheit bestanden haben (Quinsey et al., 2003).

### *Gefährlichkeit*

Der Gefährlichkeitsbegriff setzt sich aus zwei Komponenten zusammen: aus einer Tatbestandsaussage und einer Wahrscheinlichkeitsaussage. Eine Gefährlichkeit muss also immer sowohl qualitativ (bzgl. Tatbestand) als auch quantitativ (bzgl. Wahrscheinlichkeit) spezifiziert werden. Der Prognostiker gibt an, welche Straftaten (qualitativ) mit wie grosser (z.B. sehr geringer, geringer, moderater, hoher oder sehr hoher) Wahrscheinlichkeit (quantitativ) durch den Straftäter in Zukunft (kurz-, mittel- oder langfristig) zu erwarten sind (Urbaniok, 2003). Ob diese Kombination aus Wahrscheinlichkeit und Tatbestand unter dem Begriff der Gefährlichkeit zu subsummieren ist, ist eine gesellschaftliche Bewertung, die stellvertretend als rechtsnormative Einordnung durch ein Gericht vorgenommen wird.

### *Risikokalkulation*

Bei einer prognostischen Risikokalkulation werden Persönlichkeitsdispositive erfasst, aus denen heraus strafrechtlich relevante Handlungsmotivationen entstehen und die somit ein bestimmtes Risiko präsentieren. In der Regel werden zusätzlich situative Faktoren beschrieben, durch die Handlungsimpulse in besonderer Weise generiert, verstärkt oder Hemmschwellen gesenkt werden. Daneben sind aber auch protektive Faktoren möglich. Eine Risikokalkulation ist aber vor allem das Erkennen der Art und Ausprägung einer individuellen tatbegünstigenden Persönlichkeitsdisposition und ihrer handlungsrelevanten Konsequenzen (Urbaniok, 2004b).

### *Reliabilität*

Damit wird die Messgenauigkeit (Präzision) eines Instrumentes angegeben (Bortz and Döring, 2002). (Retest-) Reliabel ist eine Messung, wenn bei wiederholter Messung desselben Phänomens dieselben Resultate erzielt werden. Das heisst aber noch nicht, dass die Messungen auch akkurat waren.

Eine Waage, die das Gewicht einer Person bei kurzzeitig hintereinander erfolgenden Messungen mit 75, 65 und 85 Kilo angibt ist nicht reliabel. Andererseits muss eine Waage, die bei derselben Person am selben Tag immer 75 Kilo angibt, also reliabel ist, nicht unbedingt auch akkurat sein, nämlich dann nicht, wenn die betroffene Person in Wirklichkeit z.B. 90 Kilo wiegt.

#### **Validität**

Akkurat ist ein Test, wenn er eine hohe Validität besitzt. Die Erfassung und Überprüfung der Validität eines Tests ist sehr viel aufwändiger als diejenige der Reliabilität. Es gibt verschiedene Arten der Validität. Im Falle der Risikokalkulation bei Straftätern geht es vor allem um die so genannte Kriteriumsvalidität, die sich daran bemisst, ob der Testwert die späteren Verhaltenswahrscheinlichkeiten korrekt vorhergesagt hat (Bortz and Döring, 2002). Weitere Arten von Validität sind Inhaltsvalidität und Konstruktvalidität.

In diesem Zusammenhang ist bei der Konstruktion von Prognoseinstrumenten auf eine besondere Problematik hinzuweisen: Das so genannte Validierungsproblem (Warmuth, 1995). Um ein Prognoseinstrument auf seine Validität zu überprüfen, müssen prognostisch als rückfallgefährdet Eingeschätzte in Freiheit kontrolliert werden. Aufgrund offensichtlicher ethischer Überlegungen können jedoch als Rezidivisten prognostizierte Straftäter, bei denen schwere Straftaten zu befürchten sind, aus einem rein wissenschaftlichen Interesse nicht einer Freiheitsbewährung ausgesetzt werden (welche sie mit der prognostizierten Wahrscheinlichkeit ernsthaft missbrauchen würden). Eine konkrete Überprüfung von Prognosen ist in Situationen möglich, in denen verurteilte und prognostizierte Straftäter im Nachhinein etwa aufgrund formaljuristischer Gegebenheiten entlassen werden müssen, wie dies in den Fällen Baxstrom v. Herold (Steinböck, 1997) und Dixon (Volbert, 1986) geschah.

In der Literatur werden noch weitere ähnliche Fälle genannt. So wurden z.B. lediglich 8% der Patienten, die regulär aus einer Diagnose- und Behandlungseinrichtung entlassen worden waren, rückfällig. Gegen die Empfehlung der Einrichtung entlassene Strafgefangene wiesen hingegen eine Rückfallquote von 34,7% auf (Kozol, Boucher and Garofallo, 1972). In einer anderen Studie zeigte sich eine Differenz in der Rückfälligkeit von 7 zu 46% (Carney, 1974).

Theoretisch sind «ungewollte» Entlassungen aber – neben retrospektiven Untersuchungen – eine Möglichkeit zur Validierung eines Rezidivrisiko-Prognoseinstruments. Insbesondere bei endlichen Freiheitsstrafen müssen die Täter (respektive Gefangenen oder Patienten) spätestens nach Ablauf des gerichtlich ausgesprochenen Freiheitsentzugs wieder entlassen werden, selbst wenn sie (weiterhin) als rückfallgefährdet beurteilt werden. In dieser Problematik (Holenstein, 2004) sieht Warmuth auch eine Chance: Eine Überprüfung des vorhergesagten Rückfallrisikos wird möglich, indem der Verlauf des Freigelassenen bezüglich etwaiger erneuter Straftaten beobachtet wird.

#### **Effektgrösse**

Die Effektgrösse (auch relative Grösse des Effektes, Effekt oder Mass der statistischen Assoziation genannt) gibt Auskunft darüber, wie gross und damit auch wie bedeutungsvoll ein Unterschied oder Zusammenhang ist. Im Gegensatz zur Signifikanz eines Ergebnisses, die eine Aussage über die Wahrscheinlichkeit eines Ereignisses macht, – unter der Annahme, dass die Hypothese eines Zufallsbefundes gültig ist. Die Effektgrösse wird auch als Mass der praktischen Bedeutsamkeit bezeichnet.

*Selection ratio* wird der Anteil der Straftäter genannt, der durch die Festlegung des Wertes C zur Gruppe der zu erwarteten Rezidivisten gezählt wird. Sobald bekannt ist, wie viele der Straftäter rückfällig wurden, kann die Gesamtheit der Straftäter in vier Gruppen eingeteilt werden. Diese vier Gruppen werden häufig mit der so genannten Vier-Felder-Tafel dargestellt. Ihre Anwendung gehört zu den methodischen Grundlagen der Prognosewissenschaft und gewisse Schlussfolgerungen daraus galten bis vor kurzem als unbestritten. Auf die methodische Problematik, die Vier-Felder-Tafel als Grundlage von Validitätsüberlegungen anzusehen, wurde jüngst an anderer Stelle hingewiesen.

**Tabelle: Vier-Felder-Tafel**  
(frei nach Andrews/Bonta, 2003, 228)

	Rezidiv	Kein Rezidiv
Rezidiv prognostiziert	richtig Positive, «Treffer»	falsch Positive, «falscher Alarm»
Kein Rezidiv prognostiziert	falsch Negative, «Verpasste»	richtig Negative, «Treffer»

Die Vier-Felder-Tafel wird klassischerweise dahingehend interpretiert, dass der Prozentsatz der als rückfallgefährdet eingeschätzten Personen in Relation zu allen Rezidivisten als Trefferquote bezeichnet wird (richtig Positive / falsch Negative + richtig Positive). Sie gilt als ein Synonym für die *Sensitivität* des betreffenden Testinstruments. Der Anteil der als rückfallgefährdet eingeschätzten Personen an allen Nicht-Rückfälligen (falsch Positive / richtig Negative + falsch Positive) heisst «Falscher Alarm»-Rate. *Spezifität* wird definiert als 1 minus «Falscher Alarm»-Rate, oder anders ausgedrückt: Je niedriger der Anteil der sogenannten falsch Negativen ist, desto höher ist die Sensitivität der Methode. Je niedriger der Anteil der sogenannten falsch Positiven («falscher Alarm»), desto spezifischer ist die Methode (Nedopil, 2000).

Wenn nun in unserem Beispiel entschieden wird, dass die Anzahl der sogenannten falsch Negativen zu hoch ist, also z.B. zu viele gewalttätige Rezidivisten aufgrund einer günstigen Prognose im Test freigelassen werden, könnte statt des zuvor beschlossenen Cut-off-Scores C ein tieferer Wert K festgelegt (und somit die selection ratio erhöht) werden. Daraus würde gemäss der gegenwärtig vorherrschenden Theorie eine kleinere Anzahl sogenannt falsch Negativer resultieren. Aber auch die übrigen Werte würden sich verändern: Ein weiterer günstiger Effekt dieser Verschiebung des Cut-off-Wertes gegen links wäre die Erhöhung der Trefferquote. Ungünstig wäre dabei aber die ebenfalls daraus folgende Erhöhung der «Falscher Alarm»-Rate und die Reduktion der Anzahl der richtig Negativen. Die Erhöhung des Cut-off-Wertes würde auf der anderen Seite zu einer Reduktion der «falschen Alarme», aber auch zu einer Erhöhung der falsch Negativen führen.

Bei der Anwendung von statistischen Modellen muss zwingend die *Basisrate* des zu prognostizierenden Phänomens berücksichtigt werden. Als Basisrate wird die Häufigkeit eines Ereignisses (z.B. Sexualdelikt) in einer definierten Population verstanden. Die Basisrate wird also immer in Bezug zu einer bestimmten Population gesetzt. Die Auswahl der Bezugspopulation muss in Abhängigkeit von dem interessierenden Ereignis erfolgen. So macht es einen Unterschied, ob ein Erstdelikt oder ein Rückfalldelikt vorhergesagt werden soll. Bei der Prognose der Erstdelikte muss die Allgemeinbevölkerung als Referenzgröße herangezogen werden. Nur sehr wenige Menschen begehen

Sexualdelikte, die Basisrate ist also sehr tief. Dies führt zu einem methodischen Problem. Für jedes nicht-deterministische Modell lässt sich festhalten: Je niedriger die Basisrate, desto ungünstiger wird das Verhältnis zwischen «korrekt Positiven» und «falsch Positiven». Im Beispiel der Prognose von Sexualdelinquenzen würde dies bedeuten: Ginge man von der Allgemeinbevölkerung aus und würde zur Prognose von Sexualdelikten ein statistisches Modell verwendet werden, dann würden sehr viele Personen fälschlicherweise in die als «potenzielle Täter» deklarierte Gruppe fallen. Selbst wenn das Modell auf hohem Niveau gleich sensitiv wie spezifisch ist, hat die sehr tiefen Basisraten zur Folge, dass das Verhältnis zwischen «korrekt Positiven» und «falsch Positiven» sehr ungünstig ist. Dieser Zusammenhang zwischen der Auftretenshäufigkeit eines Ereignisses und der Treffergenauigkeit wird als das *Basisratenproblem* beschrieben. Daraus lässt sich folgende Frage ableiten: Da Sexualdelikte selten sind, bedeutet dies nun, dass diese mit statistischen Methoden nicht prognostiziert werden können? Nein, denn aus der forensisch-psychiatrischen Sicht geht es in den meisten Fällen darum, die Rückfallgefahr einzuschätzen. Das bedeutet, dass bei Rezidivisten nicht die Allgemeinbevölkerung, sondern eine spezifische Straftäterpopulation die Referenzpopulation darstellt. Die Basisrate von Rückfällen (bei Sexual- und Gewaltstraftaten) ist nicht mehr in dem tiefen Bereich, in dem die erwähnten mathematischen Schwierigkeiten auftreten. Insofern kann für die praktisch relevanten Einsatzfelder in vielen Fällen das Basisratenproblem gelöst werden, indem die korrekte Referenzpopulation berücksichtigt wird.

Ein anderes bedeutendes Dilemma der statistischen Prognosemethoden ist der sogenannte «sensitivity-specificity-trade-off». «Falsch Negative» und «falsch Positive» (=«falsche Alarme») sind bei gegebenem Testinstrument in dem Sinn eng miteinander verknüpft, dass das eine nicht reduziert werden kann ohne Erhöhung des anderen. Für primär klinisch (nicht statistisch) generierte Instrumente sollte dieser Effekt allerdings ebenso wie die anderen oben dargelegten methodischen Probleme (z.B. Basisratenproblem) eine geringe oder gar keine Rolle spielen.

### *Schwierigkeit, verschiedene Instrumente miteinander zu vergleichen*

Aus oben Gesagtem wird ersichtlich, welche Schwierigkeiten zu überwinden sind, wenn die Qualität verschiedener Risikoprognose-Instrumente miteinander verglichen werden soll: Wird auch nur ein Wert verändert, z.B. die selection ratio durch Verschieben des Cut-off-Scores, verändern sich sofort alle anderen Werte (außer Effektgrösse).

Damit die Trefferquoten verschiedener Methoden dennoch eingeschätzt und miteinander verglichen werden können, wurde von verschiedenen Autoren (Rice and Harris, 1995; Swets, 1992) folgende Lösung vorgeschlagen: Der «Receiver Operating Characteristics»-Wert (ROC). Bei gewissen Autoren wird der Wert auch «Relative Operating Characteristic» genannt, z.B. bei Rice und Harris (Rice and Harris, 1995), was aber nicht weiter von Bedeutung ist, zumal gemäss S. Lemeshow anlässlich eines Biostatistik-Seminars in der Psychiatrischen Universitätsklinik Zürich vom 29.10. 2002 sowieso unklar sei, weshalb man dafür genau diesen Ausdruck gewählt habe. Ursprünglich wurden die ROC-Techniken für die Kommunikationstechnologie entwickelt, um dann von Mossmann (Mossmann, 1994) auch im Bereich der Vorhersage gewalttätiger Rückfälligkeit angewandt zu werden (ROCs wurden zwar bereits zuvor sporadisch zur Kriminalprognose benutzt, eine standardmässige Erhebung erfolgt allerdings erst seit einigen Jahren).

Wenn es nun von einem Instrumentarium heisst, es habe einen ROC von 0,73, so bedeutet dies, dass eine 73%-ige Wahrscheinlichkeit besteht, dass ein zufälligerweise aus der Gruppe der Rückfälligen Ausgewählter einen höheren Score hatte als ein zufälligerweise aus der Gruppe der Nichtrückfälligen Ausgewählter.

Durch Einführung der relativ einfachen ROC-Berechnung kann nun also eine bestimmte Testmethode statistisch evaluiert und mit anderen verglichen werden,

### *Vorhersehbarkeit menschlichen Verhaltens*

Immer wieder wird darauf hingewiesen, dass menschliches Verhalten nicht oder nur in sehr begrenztem Rahmen vorausgesagt werden könne (Horstkotte, 1986; Jockusch, 1990; Kühl and Schumann, 1989; Laves, 1975; Rasch, 1985). Es wird zum Teil versucht, aktuelle Kriminalprognosen in ihrer Aussagekraft, aber auch bezüglich der ihnen angeblich zugrunde liegenden ethischen Haltung mit den Bemühungen C.

Lombrosos, italienischer Mediziner und Anthropologe des 19. Jahrhunderts, zu vergleichen. Lombroso vertrat in seinem Werk «L'uomo delinquente» oder der von ihm 1880 gegründeten Zeitschrift «Archivio di Psichiatria, Anthropologia Criminale e Scienze Penali» die Auffassung, dass man «Verbrechernaturen» anhand ihrer Physiognomie von «anständigen Menschen» unterscheiden könne, oder auch beispielsweise, indem man das Verhältnis von Körpergrösse zur Spannweite der Arme ausmass, welches bei Verbrechern anders sei als bei der Durchschnittsbevölkerung.

Daneben werden im Zusammenhang mit der Legalprognose immer wieder verschiedene Einwände genannt, die im Folgenden dargelegt werden sollen. So wird darauf hingewiesen, dass es sich bei menschlichem Verhalten um ein subjektives Geschehen handle. Zwar basieren die psychischen Prozesse, die schliesslich zu einem bestimmten Verhalten führen, auf objektiven, in der Regel überprüfbar Fakten. Wie diese objektiven Gegebenheiten aber wahrgenommen und interpretiert werden und in welcher Form sie zu Prämissen des späteren Handelns werden, könnte nicht (allein) von den objektiv beobachtbaren Voraussetzungen abgeleitet, sondern müsse gemeinsam mit dem Betroffenen geklärt werden. Holzkamp schliesst daraus, dass einem ohne Kooperation des Täters also der Zugang zu prognoserelevanten Fragen verbaut bleibe, was in der forensischen Situation die Regel sei (Holzkamp, 1983).

Diese These hält allerdings näherer Prüfung nicht stand. Da das Verhalten einer Person die direkte, kausale Folge seiner inneren Prozesse ist, können aus dem objektiven, beobachtbaren Verhalten eines Menschen in den meisten Fällen Rückschlüsse über seine psychischen Vorgänge gezogen werden. Da Verhaltensweisen immer Ausdruck von Persönlichkeitsmerkmalen beziehungsweise der innerpsychischen kognitiv-affektiven Prozesse sind, stellen Handlungsmuster eine wesentliche Informationsquelle dar. Häufig muss dem Verhalten sogar mehr Gewicht beigemessen werden als dem Gesagten, welches zum Teil durch strategische oder anderweitig subjektiv geprägte Selektionsmechanismen (Verdrängen, Vergessen, subjektive Wahrnehmungen, Aufrechterhaltung eigener Idealbilder etc.) bestimmt ist.

Ein weiterer Grund für die allgemeine Prognose-Skepsis stellt das Objekt der Vorhersage, nämlich gewalttägiges Verhalten, dar. Gewalt-

verbrechen seien sehr selten, d.h. die Basisrate (siehe oben) sei niedrig (um 10% bei ein- bis dreijährigen Follow-ups [Andrews and Bonta, 2003]) – wobei dies sehr von der untersuchten Population abhängt: In diesen Fällen könne es sich ein Prognostiker einfach machen, indem er vorhersagt, dass der Täter nicht rückfällig wird: In 9 von 10 Fällen würde seine Prognose bei einer Basisrate von 10% stimmen. Umgekehrt würde dies bedeuten, dass ein tatsächlicher Rückfall wegen seiner Seltenheit sehr schwer zu prognostizieren sei, vor allem bei kurzem Beobachtungszeitraum, da mit der Länge der Nachuntersuchungszeit auch die Zahl der Rückfälle steigt (Pfäfflin, 1995). Ähnliche Fragestellungen ergeben sich beim Betrieb gefährlicher Unternehmungen oder Gewerbe, bei denen es statistisch selten zu schweren Unfällen kommt, wie z.B. bei Atomkraftwerken oder auch beim Luftverkehr. In der Strafrechtslehre wird hier von «erlaubtem Risiko» (Rehberg, 1962) oder «sozialer Adäquanz» (Roxin, 1983) gesprochen. Die Konstruktion der sozialen Adäquanz ist in der Lehre umstritten (Trechsel und Noll, 1998). Trotz grundverschiedener Ansätze wird in Analogie zur sozialen Adäquanz bestimmter Gefahrenquellen zum Teil auch die Rückfallprognose bei Straftätern als «kalkuliertes Risiko» bezeichnet (Rasch, 1985).

Zu dieser Argumentation ist zunächst festzustellen, dass Rückfälligkeitssbasiraten stark abhängig von der jeweiligen Population sind. So sind die Raten bei den Populationen, die zu prognostizieren sind (Strafgefangene, bestimmte Gewalt- und Sexualstraftätergruppen) erheblich höher als 10%. Gerade bei Hoch-Risiko-Populationen wie Sexualstraftätern wurden in Langzeit-Follow-ups von bis zu zehn Jahren für gewalttätige Rückfälle Basisraten um 50% beobachtet (Harris, Rice and Cormier, 1991, Rice and Harris, 1997, Rice et al., 1990). Außerdem bestünde die Aufgabe des Prognostikers bei einer 10% Basisrate ja gerade nicht im Erkennen der 9 Nicht-Rückfälligen, sondern der einen rückfälligen Person (sofern sich diese in ihren Risiken tatsächlich unterscheiden, also nicht alle ein individuelles Risiko von 10% aufweisen).

Ein weiteres Argument, das gegen die grundsätzliche Vorhersehbarkeit menschlichen Verhaltens ins Feld geführt wird, ist die sog. Zirkularität menschlicher Handlungen. Demnach könne man bei einem Menschen nicht sagen, dass aus A automatisch B folge, denn verschiedene Faktoren beeinflussten sich gegenseitig:

A wirke auf B, B seinerseits auf C, welches wiederum A beeinflusse, A wirke aber auf D, welches ... usw. (Urbaniok, 2003). Dieses Argument verkennt, dass die meisten zu prognostizierenden Phänomene (Physik, Wetter, Biologie etc.) auf in diesem Sinn zirkulären Prozessen beruhen, was hinsichtlich der Prognostizierbarkeit in keiner Weise ein methodisches Hindernis darstellt.

Hingewiesen wird von den Prognoseskeptikern auch auf die Kontingenzen menschlichen Verhaltens. Das Konzept der Kontingenzen geht auf den amerikanischen Soziologen Parsons zurück und besagt, dass der Mensch offen sei gegenüber Handlungsalternativen, da er über einen freien Willen verfüge (Parsons und Shils, 1951). Die Kontingenzen sei ein Kriterium, das menschliches Verhalten vom tierischen unterscheide. Tiere handelten viel instinktiver, seien viel mehr durch angeborene Auslösemechanismen gesteuert als der Mensch. Die Vorstellung von Vorhersagbarkeit menschlichen Verhaltens verkenne diese spezifisch menschliche Qualität der Kontingenzen (Steinböck, 1997). Nun ist gerade das Konzept des «freien Willens» aktuell Gegenstand kontroverser Diskussion. Unabhängig davon liegt es aber auf der Hand, dass sich eine Person nicht in jedem Moment «neu erfindet», sondern Handlungsmöglichkeiten in einem Spektrum individuell determinierter Möglichkeiten und Prägungen bestehen.

### **3. Aktuelle Prognosemethoden**

Die Nomenklatur der verschiedenen Methoden, mit welchen eine Prognose erarbeitet werden kann, ist verwirrend, u.a. weil in der angelsächsischen Literatur unter gewissen Methoden etwas anderes verstanden wird als in der deutschsprachigen Literatur, aber auch, weil die Bezeichnungen der verschiedenen Methoden von Autor zu Autor sehr unterschiedlich sein können. Es können grundsätzlich zwei Ansätze unterschieden werden: die klinische und die statistische Methode (eine sinnvollere Unterteilung könnte in Zukunft in der Abgrenzung von profil- und prozessorientierten Methoden bestehen). Der klinische Ansatz geht vom Prinzip aus, dass der Einzelfall als solcher und nicht in Relation zu einer bestimmten Gruppe betrachtet wird. Dieser Ansatz kann weiter unterteilt werden in die klinisch-intuitive und die klinisch-kriteriengeleitete Methode, sodass sich schliesslich folgende drei Methoden ergeben:

- 1a *Die klinisch-intuitive Methode*, die im englischsprachigen Raum (vor allem Kanada und USA) meist «Clinical Judgement» oder «First Generation of Offender Risk-Assessments» genannt wird.
- 1b *Die klinisch-kriteriengeleitete Methode*, im Englischen ebenfalls «Clinical Judgement» genannt und mittels «Structured Clinical Guides» oder Kriterienkatalogen ermittelt. Im Deutschsprachigen Raum wird vor allem diese Methode angewendet.
- 2 *Die statistische Methode*, deren Instrumente die «Actuarial Risk-Assessments» respektive die «Second and Third Generation Risk-Assessments» sind. Dieses Verfahren hat sich vor allem im angelsächsischen Raum entwickelt.

#### *Die klinisch-intutive Methode*

Bei der klinisch-intuitiven Methode zeichnet sich die Urteilsfindung vor allem durch subjektive und intuitive Kriterien, «gut feelings», die empirisch nicht validiert sind, aus (Andrews and Bonta, 2003). Hier spielen so variable Faktoren mit wie Menschenbild des Gutachters, seine persönliche Einstellung zu Veränderungsmöglichkeiten von Straftätern und die unterschiedliche Gewichtung von Faktoren über den Verlauf der künftigen Unterbringung eine Rolle. Auch bei der Rückfallprognose mittels anderer Prognosemethoden dürften diese Faktoren mitspielen, jedoch aufgrund einer klareren Strukturierung besser kontrollierbar sein. In verschiedenen Studien konnte die Ungenauigkeit der klinisch-intuitiven Methode aufgezeigt werden: 1986 konnten Quinsey und Maguire anhand einer untersuchten Population von 360 Männern (28 Langzeitpatienten, 199 Untersuchungshäftlinge, 47 gegen ihren Willen hospitalisierte Patienten, 86 Zurechnungsunfähige), von denen jeder durch zwei bis fünf erfahrene forensische Psychologen oder Psychiater untersucht wurde, aufzeigen, dass die folgenden Charakteristika von den Examinatoren als gefährlich beurteilt wurden: Niedrige Intelligenz, ein Tötungsdelikt in der Kriminalanamnese, Gewalttätigkeit während Inhaftierung und Hospitalisierung gegen den Willen des Betreffenden. Wie sich herausstellte, wurden dabei die Merkmale «Gewalt während Inhaftierung» und «Tötungsdelikt in der Vergangenheit» überbewertet (im statistischen Prognoseinstrument «Violence Risk Appraisal Guide» gilt ein Tötungsdelikt gar als protektiver

Faktor bezüglich der Begehung künftiger einschlägiger Delikte – wobei dieses Merkmal nur innerhalb des VRAG so bewertet werden darf), Sexualdelikte wurden als Prädiktor unterbewertet, und anstatt als direkt proportional wurde der IQ als invers proportional zur Gefährlichkeit betrachtet (Quinsey and Maguire, 1986).

In einer anderen Studie wurden forensische Kliniken gebeten, Empfehlungen bezüglich Entlassung und spätere Betreuung in der Gemeinschaft einer grossen Gruppe forensischer Patienten abzugeben. Es wurde festgestellt, dass die Kliniken am ehesten bereit waren, diejenigen Patienten sofort und ohne Betreuung zur Entlassung zu empfehlen, die am gefährlichsten waren, während sie die längste Inhaftierung und die intensivste Betreuung für die am wenigsten gefährlichen Patienten forderten (Rice, Harris and Quincey, 1986).

Goldberg kam 1968 in einer Zusammenfassung verschiedener Studien über das klinisch-intuitive Urteil von Psychologen zum Schluss, dass weder der Ort der klinischen Ausbildung, noch professionelle Erfahrung in irgendeinem Zusammenhang zur Treffsicherheit der Prognose stehen. Dazu kam, dass auch die Menge der zur Verfügung stehenden Informationen über Tat und Täter keinen Einfluss auf die Qualität der Prognose hatte, dafür aber auf das Vertrauen der Klinik in das eigene prognostische Urteil (Goldberg, 1968).

In einer Studie von 1993 mit 714 Fällen (Lidz, Mulvey and Gardner, 1993) konnten die Autoren dagegen zeigen, dass Intuition erfahrener Psychiater öfter zu einem korrekten Ergebnis führt als reiner Zufall: Von der Gruppe, die als gefährlich beurteilt wurde, wurden 53% gewalttätig rückfällig, bei der als ungefährlich eingeschätzten Gruppe waren dies nur 36%. Eine Folgestudie (Gardner et al., 1996) kam aber zum Schluss, dass die statistische Methode der reinen Intuition überlegen war.

#### *Die klinisch-kriteriengeleitete Methode*

Bei der klinisch-kriteriengeleiteten Methode werden Kriterienkataloge als Leitfaden zur Unterstützung verwendet (im Gegensatz zur klinisch-intuitiven Methode, wo die Entscheidungskriterien sehr individuell und zum Teil inhaltlich nicht nachvollziehbar sind). Inzwischen gibt es verschiedene Kriterienkataloge, so z.B. den Kriterienkatalog von Nedopil (Nedopil, 1986), den HCR-20 (Webster et al., 1995) oder den Dittmann-Katalog (Dittmann, 2000),

wobei der in der deutschsprachigen Schweiz am häufigsten verwendete zurzeit der Dittmann-Katalog ist. Dieser ist eine Weiterentwicklung der Kriterienkataloge von Rasch (Rasch, 1986) und von Nedopil. Darin werden günstige und ungünstige Prognosefaktoren genannt. Beim Katalog nach Nedopil wurden bei der Analyse von verschiedenen Prognosegutachten acht für die Prognose wichtige Kriterien extrahiert, die je einen nach beiden Seiten offenen Bereich darstellen, in dem der Explorandum einzuordnen ist. Dies verleiht diesem Kriterienkatalog einen dimensionalen Charakter, der bei sonst gewissermassen als Checkliste konzipierten Kriterienkatalogen fehlt. Die acht Bereiche Nedopils sind: Krankheitseinsicht und Therapiemotivation, günstiger sozialer Empfangsraum, Abnahme klinischer Symptomatik, selbstkritische Auseinandersetzung mit der Tat, Offenheit bei der Selbstdarstellung, Bewährung in der bisher erlangten Freiheit, Nachreifung der Persönlichkeit, konkrete und realisierbare Zukunftsperspektiven (Nedopil, 1986).

Kritisiert wird an Kriterienkatalogen u.a., dass sich vor allem der weniger erfahrene Untersucher im konkreten Fall unter Umständen mit einer bestimmten Anzahl positiver und negativer Kriterien konfrontiert sieht, ohne zu wissen, wie diese gegeneinander abzuwegen und zu gewichten sind. Wie andernorts bereits ausgeführt, ist auf das bedeutsame Problem der so genannten Scheinrelevanzen hinzuweisen: Damit ist die günstige oder ungünstige Bewertung von Merkmalen gemäss Kriterienkatalog gemeint, die unter Umständen für den konkreten Einzelfall keine Relevanz aufweisen.

Kriterienkataloge stellen hinsichtlich Strukturierung und Transparenz des Bewertungsvorgangs einen klaren Fortschritt im Vergleich zur intuitiven Methode dar. An Grenzen stossen die derzeitigen Kriterienkataloge angesichts ihrer begrenzten Operationalisierung und Differenzierung sowie fehlender Auswertungsregeln.

#### *Die statistische Methode*

Das im angelsächsischen Sprachraum am häufigsten verwendete Instrument der statistischen Methode ist das «Actuarial Risk-Assessment». In diesen Tests wird verschiedenen Variablen, die einen Einfluss auf das Risiko eines Rezidivs haben, eine spezifische statistische Gewichtung zugeordnet. Actuarial Instruments gehen von der Prämisse aus, dass – falls Genauigkeit der Vorhersage oberstes Ziel sein soll – es am besten

ist, zu beobachten, wie sich die Mitglieder einer Gruppe vergleichbarer Individuen über eine bestimmte Zeit hinweg im Vergleich zueinander verhalten. Dies geschieht unter anderem durch Follow-up-Studien grösserer Gruppen von Probanden. Die dabei gewonnenen Daten bezüglich Gewalt können dann statistisch mit früher erhobenen Daten in Verbindung gebracht werden. So hat etwa Burgess über 3000 auf Bewährung Entlassene untersucht und 21 Faktoren ermitteln können, die die Gruppe der Rückfälligen von den Nichtrückfälligen unterschieden haben. Darauf erhielt jeder der 3000 Probanden pro vorliegenden Faktor einen Punkt. Bei denjenigen mit der höchsten Punktzahl war die Rückfallrate 76%, bei denjenigen mit der kleinsten Punktzahl 1,5%. Dieses Vorgehen wird grundsätzlich noch heute in der Methodologie der «Actuarial Risk-Assessments» bevorzugt (Andrews and Bonta, 2003). Dieses Vorgehen gleicht vom Prinzip her sehr demjenigen, welches die «Actuaries» der Versicherungsgesellschaften bei der Evaluation des Risikos verschiedener Schadensfälle anwenden, um die Versicherungsprämien entsprechend festzulegen. Daher stammt auch der Ausdruck «actuarial». Das entscheidende Merkmal der «Actuarial Risk-Assessments» ist die Tatsache, dass die Werte, die ein einzelnes Individuum im Test erzielt, mit statistisch erhobenen Referenzdaten korrelieren.

Zum Teil wird zwischen «Actuarial Risk-Assessments» zweiter und dritter Generation unterschieden. Nahezu allen Instrumenten zweiter Generation fehlt ein theoretisches Fundament. Sie sind gewöhnlich kurz und bestehen nur aus historischen, also statischen Items. Somit werden die dynamischen Faktoren vernachlässigt. Es wird von den von Andrews und Bonta erwähnten «Big Four» (die besten Prädiktoren für gewalttätige Rezidive könne man alle unter einen der folgenden vier theoretischen Faktoren subsumieren: antisoziale Kollegen, Einstellungen, Persönlichkeit und kriminelle Anamnese) bei Tests zweiter Generation also nur auf die (statische, irreversible) kriminelle Vergangenheit fokussiert (Andrews and Bonta, 2003). Beispiele dieser Instrumente zweiter Generation sind der Burgess Risk-Scale (Burgess, 1928), der Violence Risk-Appraisal Guide (VRAG) (Quinsey et al., 2003) und der Salient Factor Score (Hoffman, 1994). Die Instrumente der dritten Generation beinhalten zusätzlich zu den statischen auch dynamische Faktoren, die potentielle funktionale oder kausale Faktoren kriminellen Verhaltens

seien (Gendreau, Little and Goggin, 1996). Ein typisches Beispiel eines Dritt-Generation-Assessment-Tools ist der Level of Service Inventory-Revised (LSI-R) (Andrews and Bonta, 1995). Die Psychopathy Checklist-Revised (PCL-R) (Hare, 1991) wird z.T. zu den Instrumenten zweiter, z.T. aber auch zu den Instrumenten dritter Generation gezählt, da Uneinigkeit darüber herrscht, inwiefern die darin beschriebenen Persönlichkeitsausprägungen veränderlich, also dynamisch, sind.

Als Kritik an den aktuell praktizierten «Actuarial Risk-Assessments» wird genannt, dass sie in der Regel nur wenige Items enthalten (beim PCL-R sind es z.B. nur 20 Merkmale), so dass der Einzelfall nur grob erfasst wird. Daher könnten mit dieser Methode nur die Hochrisikogruppen erfasst werden, die von erfahrenen Beurteilern ohnehin erkannt würden (Urbanik, 2003).

#### *Kombination verschiedener Methoden?*

In der Literatur wird zum Teil dafür plädiert, verschiedene der oben genannten Methoden miteinander zu kombinieren, da die Beurteilung so am flexibelsten an das Zusammentreffen der verschiedenen Faktoren, die beim Zustandekommen einer Straftat eine Rolle spielen können, angepasst werden könne. So empfiehlt beispielsweise Nedopil, bei der Risikoeinschätzung (s)einen Kriterienkatalog in Verbindung mit der klinisch-intuitiven Methode anzuwenden (Nedopil, 2000). Andere Autoren warnen vor einer solchen Kombination; dass die Verbesserung der Zuverlässigkeit einer der Methoden mittels Intuition möglich sei, sei unwahrscheinlich (Dawes, Faust and Meehl, 1989). Zulässig sei höchstens die Kombination von «Actuarial Risk-Assessment» und Kriterienkatalog (Serin, 1995).

#### **4. Diskussion**

In diesem Aufsatz wurden die zwei grundsätzlichen Methoden zur Evaluation der Rückfallgefahr bei Straftätern dargestellt: die klinische (mit den Untergruppen intuitiv und kriteriengeleitet) und die statistische Methode. Die Methoden haben beide ihre Stärken und Schwächen, die es gegeneinander abzuwägen gilt.

In den verschiedenen Studien, welche die Treffsicherheit klinisch-intuitiver Methoden mit derjenigen der klinisch-kriteriengeleiteten und der statistischen Methoden vergleichen, ergab sich stets eine deutliche Unterlegenheit

der Intuition. Ihre prädiktive Kraft war z.B. bezüglich Gewaltstraftaten allenfalls etwas besser als der Zufall. Zum Teil war sie sogar schlechter, da sich die Prognostiker auf Vorurteile stützten, die sich in statistischen Analysen längst als inkorrekt erwiesen hatten, etwa dass ein Straftäter mit tiefem Intelligenzquotienten eher rückfällig werde als einer mit einem hohen IQ. Aus diesem Grund ist die klinisch-intuitive Methode zumindest im angelsächsischen Sprach- und Kulturraum – zu Recht – obsolet geworden.

Als typische Mängel der Kriterienkataloge – also der derzeitigen Praxis der klinischen Methode – gelten die oben beschriebenen Scheinrelevanzen, die theoretisch auch bei einem aktuarischen (d.h. statistischen) Prognoseinstrument zu einer falschen Prognose führen können. Daneben wird bei den Kriterienkatalogen im Zusammenhang mit ihrer noch geringen Operationalisierung die Schwierigkeit für den unerfahrenen Prognostiker kritisiert, die einzelnen Kriterien korrekt zu gewichten, also im Einzelfall protektive und belastende Faktoren gegeneinander abzuwägen, da diesbezügliche Richtlinien meist fehlen.

In zahlreichen Studien konnte gezeigt werden, dass statistische Instrumente gesamthaft in ihrer Prognose akkurater waren als klinische Instrumente (Dawes, Faust and Meehl, 1989; Gardner et al., 1996; Quinsey et al., 2003), da ihnen ein Grossteil der oben genannten Mängel nicht anhaftet. Aber auch die aktuarischen Instrumente haben Schwächen. Zu nennen wäre, dass sie nur wenige Kriterien beinhalten und aufgrund der stets impliziten Informationsreduktion nicht in der Lage sind, einen Einzelfall differenziert abzubilden. Damit bieten sie praktisch keine Möglichkeit, seltene Merkmale oder Idiosynkrasien einer zu beurteilenden Person zu erfassen. Sie sind vor allem aber kaum geeignet, Straftäter ausserhalb hoher oder typischer Risikoprofile adäquat zu beschreiben, da hier Feindifferenzierungen notwendig wären. Eine weitere Schwäche besteht in der Dominanz historischer Variablen, da statistische Methoden kaum in der Lage sind, Fluktuationen des Risikos im Längsschnitt darzustellen.

Den theoretischen Idealzustand würde eine differenzierte, flexibel auf den Einzelfall anwendbare, in ihrem Entscheidungsgang explizit und transparent darstellbare und in ihrem Ergebnis quantifizierbare Vorgehensweise darstellen. Eine Methodik auf diesem «Idealniveau»

ist weder klinisch, noch statistisch, sondern möglichst genau und differenziert. Fortschritte in Richtung weiterer Optimierung in diesem Sinne können sowohl durch den statistischen, als auch durch den klinischen Ansatz erreicht werden. Die statistischen Ansätze haben den Vorteil klarer Operationalisierung im Sinne eindeutig definierter Kriterien und Auswertungsregeln. Sie kämpfen dafür mit dem Problem der eingeschränkten Reichweite. Die klinischen Verfahren haben durch die Fokussierung auf den individuellen Einzelfall – dessen Risiko ja abzubilden ist – das Potenzial grösstmöglicher Reichweite und Einzelfall bezogener Differenzierung. Sie haben gegenwärtig Schwächen im Bereich der Operationalisierung (Definition der Merkmale, Gewichtungen, Auswertungsregeln) und überlassen vor dem Hintergrund eher grober Kriterien die Feindifferenzierung dem Geschick und der Erfahrung des jeweiligen Beurteilers. Gesamthaft erscheint es sinnvoll, Weiterentwicklungen sowohl über den statistischen als auch über den klinischen Ansatz zu forcieren.

Im Rahmen der Weiterentwicklung klinischer Instrumente und Dokumentationssysteme ist das Forensische Operationalisierte Therapie-Risiko-Evaluations-System (FOTRES, Urbanik, 2004a) der Versuch, einen neuen qualitativen Schritt zu unternehmen. Die damit verbundenen Ziele (Differenzierung, Transparenz, Nachvollziehbarkeit, expliziter Entscheidungsgang, möglichst genaue Annäherung an ein individuelles Risikoprofil usw.) wurden bereits erläutert. In diesem mehr als 700 Einzelitems umfassenden Instrument ist jedes Merkmal genau definiert, es sind einzelfallbezogene Gewichtungen möglich und es existieren klare Auswertungsregeln. Das FOTRES wurde in einem mehrjährigen empirischen Interaktionsprozess anhand der Beurteilung konkreter Fälle entwickelt. Es wurden fortlaufend nach hohem klinischen Standard beurteilte Fälle in der Weise eingearbeitet, dass die Algorithmen des Bewertungssystems dann geändert bzw. differenziert wurden, wenn ein bestimmter Fall in dem System nicht genügend differenziert abgebildet werden konnte. Es liegt ein Handbuch vor, und das System wird seit 4 Jahren von verschiedenen Anwendern benutzt, wobei in weit über 150 Fällen qualitätsgesicherte Beurteilungen vorgenommen wurden. Neben dieser klinisch-empirischen Validierung sind zur statistischen Validierung auch langjährige Studien an grossen Stichproben geplant.

Bei den statistischen Instrumenten, die sich vor allem als Screeningverfahren eignen, steht zukünftig die Entwicklung und Validierung komplexer regressionsanalytischer Modelle im Zentrum des Interesses. Hervorzuheben ist hierbei unter anderem der VRAG. Diesem in verschiedenen internationalen Studien validierten Instrument liegen moderne regressionsanalytische Methoden zugrunde und es bildet verschiedene Risikoklassen ab. Es ist mit einer ROC von 0,73 in der Lage, gewalttätige Rückfälle zu prognostizieren. Dieses Verfahren kommt im deutschsprachigen Raum bislang noch kaum zum Einsatz, weil hierzulande derzeit noch Validierungsstudien fehlen. Ergebnisse aus einer laufenden Untersuchung zur Validität des VRAG an einer deutschsprachigen Population sind demnächst zu erwarten.

### Literatur

- Andrews D.A., Bonta J., *The psychology of criminal conduct*, Anderson, Cincinnati, 2003, 226–234.
- Andrews D.A., Bonta J., *The Level of Service Inventory-Revised*, Multi-Health Systems, Toronto, 1995.
- Bortz J., Döring N., *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*, Springer, Berlin, 2002, 195–199.
- Burgess E.W., Factors determining success or failure on parole, *Journal of Criminal Law and Criminology*, 19/1928, 241.
- Carney F.L., The indeterminate sentence at Patuxert, in: M. Riedel (Hrsg.), *Crime and Delinquency*, Praeger, New York, 1974, 135–143.
- Dawes R.N., Faust D., Meehl P.E., Clinical versus actuarial judgement, *Science*, 243/1989, 1668–74.
- Dittmann V., Was kann Kriminalprognose heute leisten?, in: Bauhofer St., Bolle T.-H., Dittmann V. (Hrsg.), *Gemeingefährliche Straftäter?*, Rüegger, Chur, 2000, 67–95.
- Faust D., Ziskin J., The expert witness in psychology and psychiatry, *Science*, 241/1988, 31–5.
- Fowler F., Fowler H., *The pocket Oxford dictionary of current English*, Oxford University Press, London, 1970, 686.
- Gardner W., Lidz C.W., Mulvey E.P., Shaw E.C., Clinical versus actuarial predictions of violence in patients with mental illnesses, *Journal of Consulting and Clinical Psychology*, 64/1996, 602–9.
- Gendreau P., Little T., Goggin C., A meta-analysis of the predictors of adult offender recidivism: What works!, *Criminology*, 34/1996, 575–607.
- Goldberg L.R., Simple models or simple processes?, *American Psychologist*, 23/1968, 483–96.
- Hare R.D., *Manual for the Hare Psychopathy Checklist – Revised*, Multi-Health Systems, Toronto, 1991.
- Harris G.T., Rice M.E., Cormier C.A., Psychopathy and violent recidivism, *Law and Human Behavior*, 15/1991, 625–37.
- Hoffman P., Twenty years of operational use of a risk prediction instrument: The United States Parole Commission's Salient Factor Score, *Journal of Criminal Justice*, 22/1994, 477–94.
- Holenstein P., Todesstrafe light, *Weltwoche* 3/2004.

- Holzkamp K., *Grundlegung der Psychologie*, Campus, Frankfurt a.M., 1983.
- Horstkotte H., Strafrechtliche Fragen zur Entlassungs-praxis nach § 67d Abs. 2 StGB, *Monatsschrift für Kriminologie und Strafrechtsreform*, 69/1986, 332–40.
- Jockusch U., Therapeutischer Massregelvollzug – Leis-tungsfähigkeit und Grenzen, *Recht und Psychiatrie*, 1/1990, 14–8.
- Killias M., *Grundriss der Kriminologie. Eine europäische Perspektive*, Stämpfli, Bern, 2002, 488.
- Kozol H. L., Boucher R. J., Garofalo R. F., The diagnosis and treatment of dangerousness, *Crime and Delinquency*, 18/1972, 371–92.
- Kühl J., Schumann K. F., Prognosen im Strafrecht – Prob-leme der Methodologie und Legitimation, *Recht und Psychiatrie*, 7/1989, 126–48.
- Laves R. G., The prediction of «dangerousness» as a cri-terion for involuntary civil commitment, *Journal of Psychiatry and Law*, 3/1975, 291–326.
- Lexikonredaktion Meyers, *Meyers Grosses Taschenlexi-kon*, B.I.-Taschenbuchverlag, Mannheim, 1999, 278.
- Lidz C. W., Mulvey E. P., Gardner W., The Accuracy of Pre-dictions of Violence to Others, *Journal of the American Medical Association*, 269/1993, 1007–111.
- Mossmann D., Assessing predictions of violence. Being accurate about accuracy, *Journal of Consulting and Clinical Psychology*, 62/1994, 783–92.
- Nedopil N., Kriterien der Kriminalprognose bei psychiatri-schen Gutachten. Eine Bestandsaufnahme aufgrund praktischer Erfahrungen, *Forensia*, 7/1986, 167–83.
- Nedopil N., *Forensische Psychiatrie*, Thieme, Stuttgart, 2000, 243 ff.
- Parsons T., Shils E., *Some fundamental categories of the theory of action: A general statement. Toward a general theory of action*, Harvard University Press, Cambridge, 1951, 3 ff.
- Pfäfflin F., Rückfallprognose bei Sexualdelinquenz, *Recht und Psychiatrie*, 13/1995, 106–14.
- Quinsey V. L., Ambtman R., Variables affecting psychia-trists' and teachers' assessment of the dangerousness of mentally ill offenders, *Journal of Consulting and Clinical Psychology*, 47/1979, 353–62.
- Quinsey V. L., Cyr M., Perceived dangerousness and trea-tability of offenders: The effects of internal versus ex-ternal attributions of crime causality, *Journal of Inter-personal Violence*, 1/1986, 458–71.
- Quinsey V. L., Harris G., Rice M., Cormier C. A., *Violent off-enders: Appraising and managing risk*, APA, Washingt-on DC, 2003, 44 ff.
- Quinsey V. L., Maguire A., Maximum security psychiatry patients: Actuarial and clinical prediction of dange-rousness, *Journal of Interpersonal Violence*, 1/1986, 143–71.
- Rasch W., Die Prognose im Massregelvollzug als kalku-lierbares Risiko, in: Schwind H. D. (Hrsg.), *Festschrift für Günter Blau*, Walter Gruyter Verlag, Berlin, New York, 1985, 309 ff.
- Rasch W., *Forensische Psychiatrie*, Kohlhammer, Stuttgart, 1986, 291 ff.
- Rehberg J., *Zur Lehre vom erlaubten Risiko*, Diss., Zürich, 1962.
- Rice M. E., Harris G. T., Violent recidivism: Assessing pre-dictive validity, *Journal of Consulting and Clinical Psy-chology*, 63/1995, 737–48.
- Rice M. E., Harris G. T., Cross validation and extension of the Violence Risk Appraisal Guide for child molesters and rapists, *Law and Human Behavior*, 21/1997, 231–41.
- Rice M. E., Harris G. T. et al., Recidivism among male insa-nity acquittees, *Journal of Psychiatry and Law*, 18/1990, 379–403.
- Rice M. E., Harris G. T., Quinsey V. L., Treatment for foren-sic patients, in: *Mental health and law: Research policy and services*, Carolina Academic Press, New York, 1986, 141 ff.
- Roxin C., Bemerkungen zur sozialen Adäquanz im Straf-recht, in: Günter Kohlmann (Hrsg.), *Festschrift Ulrich Klug*, Köln, 1983, 303–313.
- Serin R. C., Treatment responsiveness in criminal psycho-paths, *Forum on Correction Research*, 7/1995, 23–6.
- Steinböck H., Das Problem schwerer Gewalttaten und deren Prognostizierbarkeit, *Recht und Psychiatrie*, 15/1997, 67–77.
- Swets J., The science of choosing the right decision thres-hold in highstake diagnostics, *American Psychologist*, 47/1992, 522–32.
- Trechsel S., Noll P., *Schweizerisches Strafrecht, Allgemei-ner Teil I*, Schulthess, Zürich, 1998, 137.
- Urbanik F., *Was sind das für Menschen – was können wir tun. Nachdenken über Straftäter*, Zytglogge Verlag, Bern, 2003, 248.
- Urbanik F., *Forensisches Operationalisiertes Therapie-Risiko-Evaluations-System*, Zytglogge Verlag, Bern, 2004 (zit. Urbanik, 2004a).
- Urbanik F., Validität von Risikokalkulationen bei Straftä-ttern – Kritik an einer methodischen Grundannahme und zukünftige Perspektiven, *Fortschritte für Neurologie und Psychiatrie*, 72/2004, 260–9 (zit. Urbanik, 2004b).
- Volbert R., Zwischenfälle im Massregelvollzug: Wie kal-kulierbar ist das Risiko?, *Monatsschrift für Kriminal-biologie und Strafrechtsreform*, 69/1986, 341–7.
- Warmuth M., Die Prognose von Urlaub und Ausgang – ein Vergleich zwischen Massregelpatienten und Strafge-fangenen, *Recht und Psychiatrie*, 13/1995, 15–9.
- Webster C. D. et al., *HCR-20: Assessing risk for violence*, Simon Fraser University, Vancouver 1995.

**Thomas NOLL, Dr. med., lic. iur.**

*Psychiatrisch-Psychologischer Dienst  
Amt für Justizvollzug  
Feldstrasse 42  
8090 Zürich  
thomas.noll@ji.zh.ch*

**Jérôme ENDRASS, Dr. phil.**

*Psychiatrisch-Psychologischer Dienst  
Amt für Justizvollzug  
Feldstrasse 42  
8090 Zürich*

**Frank URBANIK, Dr. med.**

*Psychiatrisch-Psychologischer Dienst  
Amt für Justizvollzug  
Feldstrasse 42  
8090 Zürich*