

# Risikokategorisierung, Risikokommunikation und risikobezogene Entscheidungsfindung in der forensischen Psychiatrie und Psychologie – über Versuche einer besseren Transparenz und Fairness

*Reinhard Eher, Markus G. Feil & Martin Rettenberger*

## **Zusammenfassung**

Die erkenntnistheoretischen und methodischen Grundlagen und die mit diesen einhergehenden Dilemmata und Herausforderungen gehen im Tun forensischer Versorgungseinrichtungen häufig unter, wenigstens verschwinden sie aus dem Diskurs und Bewusstsein. Gleichzeitig sind sie in den arbeitstäglich häufig zu treffenden Entscheidungen, die oft weitreichende Konsequenzen haben, ständig präsent und wirksam.

Nach einer Darstellung und Erörterung der Herausforderungen des forensischen Alltags wird das Fünfkategorien-Modell zur Risikokommunikation (Hanson et al., 2017) vorgestellt. Dieses füllt insofern eine Lücke, als es die Risikokommunikation über fünf methodisch klar definierte Kategorien vereinheitlicht und Anschlussfähigkeit an bestehende Methoden und Instrumente zur Risikoeinschätzung schafft. Ferner bietet es eine Orientierung bezüglich des in den jeweiligen Kategorien notwendigen Interventionsbedarfs und der Erwartung der durch entsprechende Interventionen erzielbare Effekte. Damit sichert es in der Einrichtung zu treffende Entscheidungen ab und ermöglicht eine rationale Ressourcenallokation.

Diskutiert werden auch wesentliche Kritikpunkte und bestehende Unsicherheiten des Fünfkategorien-Modells. Es stellt aus Sicht der Autoren eine deutliche Innovation dar und stimuliert Innovation und Entwicklung in der Forschung und Praxis.

**Schlüsselwörter:** Risikoeinschätzung, Risikokategorisierung, Fünfkategorien-Modell

## **Risk Categories, Risk Communication and Risk Relevant Decisions in Forensic Psychology and Psychiatry – Moving towards more Transparency and Fairness**

### **Abstract**

Risk relevant decisions are part of everyday life in forensic psychology and psychiatry. However, along with these challenges different dilemmas occur. In this article, the development of a common risk language, the Five-Level Risk and Needs System, is presented and discussed. This system enables a reliable and objective risk communication, assuring the transmitter and the receiver of the risk message may generate the same ideas of risk. Additionally, it offers an objective and quantitative estimate about the advantages and disadvantages of the decision-making.

After introducing this model, benefits and concerns are discussed.

**Keywords:** Risk Assessment, Risk Categorizing, 5-Category Risk and Needs Model

### **Einleitung**

Spätestens seit der Covid-19-Pandemie ist das Dilemma der Risikoerfassung, der Risikokommunikation und des Treffens risikoorientierter Entscheidungen auch der breiteren Öffentlichkeit bekannt. Viele Entscheidungen wirkten in Zeiten der Pandemie – jedenfalls für Laien – als unbegründet, nicht nachvollziehbar, teilweise als willkürlich und sogar mutwillig. Solchen Kritikpunkten und Vorwürfen sieht sich die forensische Psychiatrie und Psychologie seit jeher ausgesetzt, wenn es um ihre risikobezogenen Entscheidungen, wie Einweisungen in eine Maßregel, Lockerungen, Entlassungen etc. geht. Die Herausforderungen sind dabei durchaus vergleichbar: In beiden Konstellationen geht bzw. ging es darum, zu einem bestimmten Zeitpunkt mithilfe eines bestimmten, inhärent beschränkten Wissensstandes für die Zukunft bedeutsame Entscheidungen zu treffen. Dabei muss nicht nur in Kauf genommen werden, dass die Informationen, die als Entscheidungsgrundlage zur Verfügung stehen, unvollständig und deshalb aus methodischer und erkenntnistheoretischer Sicht in ihrer Aussagekraft immer beschränkt sind. Auch muss deren Zusammenhang mit dem späteren Ereignis zwar grundsätzlich belegt sein, dieser ist jedoch weit entfernt von einem hohen Bestimmtheitsgrad. Darüber hinaus geht es nicht um Zusammenhänge in einem geschlossenen System, sondern in einem hochkomplexen stochastischen Umfeld, in dem zu guter Letzt die Entscheidung selbst auf die Wahrscheinlichkeit der späteren

Entwicklung Einfluss nehmen kann und soll<sup>1</sup>. Auf das Risiko bezogene Entscheidungen können – zumindest im Nachhinein – dann vermeintlich falsch oder ungerechtfertigterweise richtig erscheinen. In anderen Worten ist eine auf Prognosen basierende Entscheidung z.B. nicht allein schon deshalb richtig, weil ein Proband mit niedrig attestiertem Rückfallrisiko tatsächlich nicht rückfällig wird. Genauso wenig muss sie von vornherein falsch sein, wenn ein solcher Proband rückfällig wird. Ebenso wenig ist eine prognosebezogene Entscheidung zwangsläufig als richtig einzustufen, wenn ein Proband mit attestiertem hohem Risiko rückfällig wird; umgekehrt genauso wenig von vornherein falsch, wenn dieser Proband nicht rückfällig wird.

### **Aber wer oder was entscheidet, ob eine prognostische Einschätzung richtig ist?**

Da Entscheidungen im forensischen Kontext üblicherweise dichotomer Natur sind (Einweisung/Lockerung/Entlassung etc. ja/nein), entsteht nicht selten – nicht zuletzt beim vom Ergebnis Betroffenen – der Eindruck der Willkürlichkeit, vor allem wenn es um die Definition von Grenzen geht, die als Grundlage für eine Entscheidung in die eine oder andere Richtung herangezogen werden. Wenn beispielsweise aus juristischer Sicht die Grenzen der Einweisungsnotwendigkeit in eine freiheitsentziehende Maßregel vermeintlich klar und eindeutig definiert sind (z. B. die Voraussetzung eines „hohen Risikos“ für die Begehung neuerlicher schwerwiegender Straftaten), so stellt sich diese Konstellation aus erfahrungswissenschaftlicher Sicht weniger eindeutig dar. Begrifflichkeiten wie „hohes Risiko“ sind häufig unbestimmt bzw. nicht transparent, möglicherweise sogar unbestimmbar. So konnte gezeigt werden, dass eine nicht-numerische Risikokommunikation (wie z. B. die Zuordnung „hohes Risiko“) selbst bei Fachleuten zu stark unterschiedlichen Einschätzungen, Formulierungen und somit auch Entscheidungen führen kann (Zoe, Carter, Harris, & Sharpe, 2008). Zu ähnlich ernüchternden Ergebnissen kam eine Studie im deutschsprachigen Raum (Rettenberger et al., 2017). „Hohes Risiko“ wurde von manchen Expertinnen und Experten bereits bei einer 20-prozentigen Rückfallwahrscheinlichkeit angenommen und als solches formuliert, andere kamen erst bei einer 80-prozentigen Rückfallwahrscheinlichkeit zu dieser Einschätzung.

Hält man sich vor Augen, wie bedeutsame Entscheidungen an die Prognose geknüpft sein können und um welch hohe Rechtsgüter (-verletzungen) es unter Umständen geht (lebenslanger Entzug der Freiheit vs. schwerwiegendes Gewaltdelikt), würde man

---

1 In den Wirtschaftswissenschaften wird dieser Effekt als Lucas-Kritik bezeichnet. Der Begriff ist nach dem Wirtschaftsnobelpreisträger Robert E. Lucas benannt. Ihm zufolge kann einer Prognose (zumindest im Falle der Prognose ungünstiger oder negativer Verläufe) die Tendenz folgen, die Prognose zu widerlegen, da aufgrund der Prognose gegen das prognostizierte Ergebnis und einen ungünstigen weiteren Verlauf gearbeitet wird. Analog zur Makroökonomik wird auch in der forensischen Praxis versucht, den tatsächlichen Eintritt einer ungünstigen Prognose durch besonders intensive Interventionsmaßnahmen zu verhindern.

sich eine perfekte Entscheidungsfindung wünschen. Es wäre gut, wenn sich jede Entscheidung am Ende des Tages insofern als richtig herausstellt, als bspw. jeder in den Maßregelvollzug Eingewiesener tatsächlich von der Begehung einer schwerwiegenden Straftat abgehalten wurde, und wenn jeder Nicht-Eingewiesene niemals eine schwerwiegende Straftat begehen wird.

Ein derartiger Wunsch kommt aufgrund der stochastischen Natur prognostischer Entscheidungen nicht nur dem Wunsch nach Hellseherei gleich. Er muss schon an der Unzuverlässigkeit bei der Erhebung risikorelevanter Merkmale scheitern. Selbst im besten Fall, beim Befolgen aller derzeit verfügbaren wissenschaftlichen Erkenntnisse, mithin also bei Anwendung standardisierter Risikoeinschätzungsverfahren durch entsprechend ausgebildete Expertinnen und Experten, können die Messergebnisse abweichen (Reliabilität eines Messverfahrens), woraus zwangsläufig Einschränkungen der (prädiktiven) Validität resultieren. Dabei ist weiter zu konstatieren, dass die prädiktive Validität schon aufgrund des stochastischen Charakters der Risikoeinschätzung von Natur aus beschränkt ist und deshalb niemals perfekt sein kann. Die besten Prognoseergebnisse bewegen sich bei etwa (grob übersetzt) 70 bis 80 Prozent Treffsicherheit (Singh, 2013; Singh, Grann, & Fazel, 2011). Als entscheidendes Effektstärkenmaß für die Übereinstimmung von Risikomerkmale und deren Zusammenhang mit der späteren Rückfälligkeit wird die AUC (Area Under the Curve) herangezogen (Rice & Harris, 2005). Sie beschreibt das Verhältnis zwischen der Sensitivität (Anteil der tatsächlich Rückfälligen, die mit einem bestimmten Grenzwert des angewandten Verfahrens identifiziert werden können) und der Spezifität (Anteil der tatsächlich Nicht-Rückfälligen, denen mit demselben Grenzwert Nicht-Rückfälligkeit attestiert wurde) bei jedem möglichen Punktwert eines Risikoeinschätzungsverfahrens.

Bei 100%iger Sensitivität und 100%iger Spezifität läge ein AUC-Wert von 1.0 vor. Die Treffsicherheit wäre perfekt. Während solche Werte zumindest annähernd möglich und bei bestimmten diagnostischen Verfahren auch gefordert sind (z.B. bei Labortests auf bestimmte Antigene oder Antikörper), sind sie in der forensischen Psychiatrie und Psychologie angesichts des stochastischen Umfelds mit Hellseherei gleichzusetzen (wie oben erwähnt: Die Treffsicherheit der besten Risikoeinschätzungsverfahren liegt in der Regel zwischen 0.7 und 0.8). Risikoassoziierte Entscheidungen gehen somit – abhängig vom herangezogenen Grenzwert – immer entweder auf Kosten der Sensitivität oder auf Kosten der Spezifität. In anderen Worten geht eine prognostische Festlegung mithilfe eines standardisierten Verfahrens immer mit dem folgenden Dilemma bzw. Paradoxon einher: Die Entscheidung ist mit unvermeidbarer Sicherheit in einem gewissen Prozentsatz entweder „zu streng“ (diagnostisch gesprochen liberal) oder „zu wenig streng“ (diagnostisch konservativ). Bei jeder prognostischen Zuordnung, z. B. im Zuge der Frage nach einer Einweisung in den Maßregelvollzug, ist per definitionem ein gewisser Prozentsatz insofern „falsch“, als Probanden nicht eingewiesen werden, die später schwere Gewaltdelikte begehen (falsch Negative). Gleichzeitig werden Probanden eingewiesen, die später auch ohne Einweisung keine schweren Gewaltdelikte begangen hätten (falsch Positive).

## Ist deswegen die Prognose falsch? Nein. Ist eine auf dieser Prognose basierende Entscheidung falsch? Möglich.

Expertinnen und Experten sprechen in diesem Zusammenhang von der *Number Needed to Detain (NND)*, in Anlehnung an die *Number Needed to Treat (NNT)* – also jener Anzahl an Inhaftierten (oder Patienten), die eingewiesen (bzw. behandelt) werden müssen, damit ein Inhaftierter (bzw. Patient) von der Maßnahme insofern profitiert, als er von der Rückfälligkeit abgehalten wird. Diese Zahl stellt den Kehrwert des *Positive Predictive Values (PPV)* dar, also des Verhältnisses der später tatsächlich rückfälligen Personen zu der Gesamtzahl der Probanden, denen anhand eines standardisierten Verfahrens Rückfälligkeit vorausgesagt wurde. Beide Maße sind abhängig von der Basisrate: Je niedriger die Basisrate ist, umso größer ist die NND (unter der Voraussetzung gleichbleibender Sensitivität und Spezifität des Verfahrens).

Hier liegt im Übrigen eines der großen Dilemmata von „Ultima ratio“-Maßnahmen. Weil „Ultima ratio“-Maßnahmen eine einschneidende Bedeutung für das Individuum haben (z.B. langfristiger oder sogar unbefristeter Freiheitsentzug) und Grundrechte geschützt werden müssen, sind sie nur bei Gefahr für besonders gravierende Straftaten in der Zukunft anzuwenden. Aber gerade gravierende Straftaten sind seltene Ereignisse (somit mit niedriger Basisrate), bei deren Vorhersage die Treffsicherheit per definitionem also vergleichsweise gering ist. Je niedriger die Wahrscheinlichkeit solcher Ereignisse (also „die Basisrate“) ist, umso schlechter ist die PPV, und umso größer ist damit ihr Kehrwert, die *Number Needed to Detain*. Es ist also ein unausweichliches Dilemma, dass gerade beim Wunsch der Verhinderung von gravierenden (= seltenen) Ereignissen die Rate der falsch Positiven relativ hoch sein muss. Eine Möglichkeit zum Umgang mit diesem Dilemma ist es, die Prüfungsvoraussetzungen für Kriminalprognosen mittels Formalkriterien (z. B. Vorliegen einer schwerwiegenden psychischen Störung oder einer bestimmten Anzahl einschlägiger Vorstrafen) möglichst streng zu formulieren, um so die absolute Anzahl derjenigen zu senken, bei denen überhaupt eine Kriminalprognose erstellt werden muss, wodurch die Basisrate für schwerwiegende Rückfälle erhöht wird (Eher, Lindemann, Birklbauer, & Müller, 2016). Durch eine Erhöhung der Basisrate – d. h. der Wahrscheinlichkeit schwerwiegender Delikte in dieser vorselektierten Gruppe – erhöht sich zwar nicht die allgemeine Treffsicherheit der angewandten Risikoeinschätzungsverfahren – weil die AUC letztlich unabhängig von der Basisrate ist –, aber es wird die PPV größer und die *Number Needed to Detain* kleiner. Somit würde sich nicht nur der Anteil der falsch Positiven unter den Begutachteten verringern, sondern aufgrund der strengeren Prüfungsvoraussetzungen auch die Gesamtzahl der überhaupt zu Prüfenden. Überlegungen, die angesichts einer „Ultima ratio“-Maßnahme sinnvoll und gerechtfertigt erscheinen.

Darüber hinaus ist die NND naturgemäß stark vom gewählten Cut-Off im Risikoeinschätzungsverfahren abhängig: Je höher der Cut-Off-Wert im Risikoeinschätzungsverfahren gesetzt wird (also je „gefährlicher“ ein Proband sein muss, um eingewiesen

zu werden), umso niedriger ist die NND, wobei dies allerdings mit einer Abnahme der Sensitivität einhergeht. In anderen Worten: Wenn man die Falsch-Positiven-Rate bei Einweisungen geringhalten möchte, steigt somit zwangsläufig die Falsch-Negativ-Rate, also die Zahl jener Rückfälligen, die durch eine Einweisung unter Umständen von einer weiteren Straftat abgehalten hätten werden können.

Deshalb ist es aus Sicht der Autoren umso wichtiger, möglichst wirksame Alternativmöglichkeiten zu den „Ultima ratio“-Maßnahmen in Erwägung zu ziehen und evtl. auch auszubauen. Im Zusammenhang mit der Maßregel-/Maßnahmeneinweisung würde z. B. die – realistische – Zahl von  $NND = 8$  bedeuten, dass mit acht Eingewiesenen sichergestellt ist, dass einer von ihnen eine schwerwiegende Straftat nicht mehr begeht, die er ohne Einweisung begangen hätte. Im Umkehrschluss bedeutet dies, dass die sieben anderen auch ohne Einweisung keine weitere schwere Straftat begangen hätten (Buchanan & Leese, 2001).

Bei Anwendung üblicher standardisierter Risikoeinschätzungsverfahren und sinnvoller Cut-Off-Werte kann von einer 50%igen Sensitivität und einer etwa 70 bis 80%igen Spezifität ausgegangen werden (Buchanan & Leese, 2001). Das heißt: Trifft man risikoassoziierte Entscheidungen mithilfe aktueller kriminalprognostischer Verfahren, „übersieht“ man zunächst die Hälfte aller späteren Rückfälligen, um bei 20 Prozent allerer, die nicht rückfällig werden, ebenso die Maßnahme/Maßregel anzuordnen. Damit nicht genug der schlechten Nachrichten: Dieses Ergebnis wäre nur zu erzielen, wenn man sich strikt an standardisierte Kriminalprognoseverfahren hält. Wendet man sie nicht bzw. nicht konsequent an, ist das Ergebnis noch schlechter (Hanson & Morton-Bourgon, 2009).

## **Wer beurteilt, ob eine prognostische Entscheidung richtig ist bzw. war?**

Bedeutsame Entscheidungen werden in vielen Bereichen der forensischen Psychiatrie und Psychologie getroffen. Nicht nur gerichtliche Entscheidungen über Unterbringungen oder Entlassungen sind dabei relevant. Intramural geht es z. B. um die Verlegung auf weniger oder höher gesicherte Stationen, bei vollzugsöffnenden Maßnahmen um die Vertretbarkeit von Lockerungen bzw. um Entscheidungen über Länge, Ausmaß und konkrete Ausgestaltung von Vollzugslockerungen, um die Art und Dauer von Behandlungsmaßnahmen etc. Auch bei Entscheidungen über eine Nachbetreuung oder Nachbehandlung nach Entlassung stellt sich die Frage einer diesbezüglichen Notwendigkeit. Nicht nur angesichts einer Vielzahl von Probandinnen und Probanden sind solche Entscheidungen mit Ressourcenfragen verknüpft. Die Entscheidungsmöglichkeiten sind auch von den vorhandenen Ressourcen abhängig. Spätestens dann ist die Frage der Risikoeinschätzung und der risikobezogenen Entscheidungen (bzw. Entscheidungsfähigkeit) auch eine politische Frage bzw. eine Frage der politischen

Verantwortungsübernahme, nicht nur bzgl. der Finanzierung und Finanzierbarkeit von Interventionsmaßnahmen.

### **Die Richtigkeit einer risikoassoziierten Entscheidung ist nicht allein eine methodische oder wissenschaftstheoretische Frage.**

Dennoch sind die wissenschaftlichen Grundlagen für die Risikoeinschätzung und darauf basierende Entscheidungen enorm wichtig und durch die Fachleute zu kommunizieren. Entscheidungen, die auf Wahrscheinlichkeitseinschätzungen beruhen, gehen mit einer großen Verantwortung einher, zumal solche ja nie den Beweis ihrer Richtigkeit im Sinne des Eintritts eines bestimmten späteren Ergebnisses antreten können. Risikobezogene Entscheidung müssen auf der Basis des Wissensstandes getroffen werden, der zum Zeitpunkt der Entscheidung zur Verfügung steht und bekannt ist.

Grundsätzlich ebenso wie im Einzelfall ist die Voraussetzung für eine prognostische Entscheidung, dass risikorelevante Merkmale am Individuum erfasst werden, die einen empirisch gesicherten Zusammenhang mit dem späteren unerwünschten (und zu verhindernden) Ergebnis aufweisen. Die Merkmale dürfen nicht willkürlich erhoben sein. Studien über diesen Zusammenhang müssen vorliegen, die Größe und Bedeutung des Zusammenhangs muss quantifiziert sein. Probanden mit mehr oder weniger solchen beobachtbaren Merkmalen müssen sich im Hinblick auf die Rückfallwahrscheinlichkeit reliabel und valide unterscheiden lassen, d. h. die Risikomerkmale müssen in der Lage sein, Gruppen von Probanden mit messbar weniger Rückfällen von Gruppen mit messbar mehr Rückfällen zu unterscheiden. Die meisten gängigen Risikoeinschätzungsverfahren, insbesondere die statistisch-aktuarischen, verfügen über diese Fähigkeit (Singh, 2013). Somit können Probanden mit unterschiedlichem Rückfallrisiko klar voneinander abgegrenzt werden. Diese Grenzen können damit für risikoassoziierte Entscheidungen herangezogen werden. Sie können die Vor- und Nachteile der jeweiligen Entscheidung quantitativ veranschaulichen, das Dilemma der Entscheidungsfindung transparent, und damit die Entscheidung nachvollziehbar machen. Die Person, die die Entscheidung getroffen hat, ebenso wie ihr Gegenüber, kennen damit die Vor- und Nachteile der jeweiligen Entscheidung und können explizit Verantwortung dafür übernehmen.

Die Grenzwerte solcher Risikokategorien können zwischen den verschiedenen Risikoeinschätzungsverfahren bislang allerdings zum Teil stark variieren. In vielen Fällen ist auch die Validität der Grenzziehung im Hinblick auf tatsächliche, statistisch signifikante Unterschiede in Bezug auf die Rückfälligkeit nicht nachgewiesen. So verfügt z. B. die Originalversion des *Violence Risk Appraisal Guide-Revised* (VRAG-R; (Davies, Helmus, & Quinsey, 2022) über neun Risikokategorien (Risk Bins), der *Stable-2007* über drei Risikokategorien (Eher, Matthes, Schilling, Haubner-MacLean, & Rettenberger, 2011) und die aktuelle Version des *Static-99* und *Static-99R* über fünf Kategorien (Eher, Rettenberger, Etzler, Eberhaut, & Mokros, 2019).

## Das Fünfkategorien-Modell des US Department of Justice – ein innovativer Ansatz

Die Verfügbarkeit einer reliablen und validen Kategorienbildung mithilfe messbarer Kriterien ist allerdings die Voraussetzung für eine transparente, verantwortungsvolle und für eine Vergleichbarkeit geeignete Entscheidungsfindung. Den Versuch der Vereinheitlichung und transparenten Kategorienbildung machte eine Arbeitsgruppe um R. Karl Hanson (Kanada). Sie definierten fünf Kategorien, deren Mittelkategorie sich einheitlich an der Basisrate orientiert (an der durchschnittlichen Rückfallrate einer repräsentativen Stichprobe) sowie jeweils zwei darüber- und darunterliegende Kategorien (Hanson, Bourgon, et al., 2017).

Der Vorteil des Fünfkategorien-Modells ist u. a., dass die nominale und numerische Risikokommunikation integriert werden, womit der nominalen Risikokommunikation Objektivität und Transparenz ermöglicht wird. Gleichzeitig steht eine vereinheitlichte nominale Risikokommunikation zur Verfügung, die sich einerseits aus dem relativen Risiko ableitet, sich gleichzeitig aber an einer Größe des absoluten Risikos orientiert, nämlich der Basisrate. Somit wird das Problem der Stichprobenvariabilität bei Angabe des absoluten Risikos gelöst: Das Risiko wird „nominal“ im Sinne eines definierten und transparenten relativen Risikos zur Basisrate angegeben. Für den Fall, dass verlässliche Basisraten zur Verfügung stehen, kann für die entsprechende Kategorie auch das absolute Risiko angegeben werden. Die Stichprobenvariabilität wirkt sich durch die Information über eine entsprechend anzupassende Basisrate nicht mehr störend aus.

Die allgemeine Gültigkeit dieses Modells ist dadurch gegeben, dass die mittlere Risikokategorie („durchschnittliches Risiko“) immer durch dasjenige Risiko definiert ist, das die Gesamtgruppe (im besten Fall eine repräsentative Stichprobe z. B. für alle aus der Strafhaft entlassenen Gewalttäter oder für alle Sexualstraftäter oder für alle Personen, die aufgrund pädosexueller Straftaten verurteilt wurden) aufweist (Basisrate). Somit ist die Bezeichnung „durchschnittliches Risiko“ eine klar definierte Größe, ihre Definition eindeutig und nachvollziehbar. Unter- und oberhalb dieser Mittelkategorie wird dem Fünfkategorien-Modell zufolge jeweils eine Kategorie mit der halben („unterdurchschnittliches Risiko“) bzw. der doppelten („überdurchschnittliches Risiko“) Rückfallbasisrate definiert (Hanson, Babchishin, Helmus, Thornton, & Phenix, 2017), wodurch die entsprechenden Grenzwerte in einem bestimmten Risikoeinschätzungsverfahren festgelegt werden. Jeweils darüber und darunter liegen die beiden Randkategorien mit den definierten Bezeichnungen „hohes Risiko“ (bzw. „deutlich überdurchschnittliches Risiko“ für Sexualstraftäter-Populationen) und „niedriges Risiko“. Mittlerweile wurden für verschiedene Risikoeinschätzungsverfahren die Grenzwerte zur Bildung dieser Kategorien bestimmt und stehen zur praktischen Anwendung bzw. zur Risikokommunikation bereit (Eher et al., 2019; Gaunersdorfer & Eher, 2023; Wirove, Olver, & Haag, 2023). Im vorliegenden Heft findet sich ein Vorschlag zur Kategorisierung des VRAG-R (Rice et al., 2013) im Fünfkategorien-Modell.

Mit dem Fünfkategorien-Modell kann somit der berechtigten Kritik an der unkritischen, gar leichtfertigen Angabe des absoluten Risikos (Rückfallraten in Prozent bei verschiedenen Risikokategorien) zur Untermauerung eines bestimmten Risikos begegnet werden. Es ist allzu leicht, z. B. computergestützt und internetbasiert Risikoeinschätzungsverfahren durchzuführen, die dann auf ein absolutes Risiko verweisen und entsprechende Prozentsätze ausgeben, nicht aber auf die begrenzte Übertragbarkeit der Daten angesichts des bekannten Phänomens der Stichprobenvariabilität hinweisen. Rückfallraten – ob mittels statistischer Verfahren errechnet oder beobachtet – sind Darstellungen des absoluten Risikos, das wiederum stichprobenabhängig sein kann. In anderen Worten heißt das, ein Proband kann in der Jurisdiktion X trotz gleichem VRAG-R-Wert eine höhere oder niedrigere Rückfallwahrscheinlichkeit als ein Proband in der Jurisdiktion Y haben. Auf dieser Grundlage müssten Risikoeinschätzungsverfahren in den jeweiligen Ländern ihrer Anwendung – oder jedenfalls Jurisdiktionen und Sprachräumen – nicht nur validiert sein, sondern auch kategorienspezifische Rückfallraten im Sinne des absoluten Risikos zur Verfügung stehen.

Bei Anwendung des Fünfkategorien-Modells sind Kalibrierungsstudien aus dem eigenen Sprach- bzw. Kulturraum nicht mehr in der Form notwendig; jedenfalls nicht für die grundsätzliche Anwendbarkeit eines bestimmten Risikoeinschätzungsverfahrens. Es reicht die Validierung des Verfahrens vor dem Hintergrund des Fünfkategorien-Modells (im Sinne der Definition und Festlegung der Grenzen von fünf Kategorien) in einer methodisch angemessenen Studie mit repräsentativen Stichproben. In der Folge ist das relative Risiko der Kategorien zueinander definiert, wodurch ein Verfahren angesichts der guten Übertragbarkeit und bekannten Stabilität des relativen Risikos auch in anderen Ländern und Jurisdiktionen angewandt werden kann. Sind ferner die Abstände der jeweiligen Kategorien zueinander definiert bzw. können sie aus der Validierungsstudie entnommen werden, sind Prognoseinstrumente und ihre Risikokommunikationsmöglichkeiten auch stabil und auf andere Bereiche und Delikt- bzw. Probandengruppen übertragbar.

Das Fünfkategorien-Modell verfügt aufgrund seiner Objektivität, seiner Transparenz und seiner breiten Anwendungsmöglichkeit über eine Reihe von weiteren Vorteilen. Hanson et al. (2017) machen nicht nur Vorschläge zur Zahl und statistischen Definition der Risikokategorien, sondern bieten weitere Anwendungen des Risk-Need-Responsivity-Modells (RNR-Modell) an.

Bspw. soll die Kategorisierung nicht nur anhand des Rückfallrisikos („risk“) vorgenommen werden, sondern auch bezogen auf die (Zahl der) kriminogenen Bedürfnisse („needs“). Hanson et al. (2017) beschreiben bspw., dass der Kategorie I („niedriges Rückfallrisiko“) zugeordnete, verurteilte Sexualstraftäter keine oder nur relativ wenige kriminogene Bedürfnisse aufweisen. Selbst wenn diese vorliegen, sind diese zeitlich befristet bzw. situativ begrenzt und liegen in der Regel nicht in schwerwiegender Form vor. Solche Probanden verfügen ferner über persönliche Ressourcen und Stärken. Sie unterscheiden sich dann nicht oder nur geringfügig von Menschen, die bis dato keine Straftaten begangen haben, und ihr Risiko, neuerliche Delikte zu begehen,

ist im Wesentlichen vergleichbar mit der Wahrscheinlichkeit bei Personen, die bis dahin noch nicht einschlägig in Erscheinung getreten sind. Es ist davon auszugehen, dass sie kooperativ gegenüber Regeln und Anweisungen sind. Sie brauchen aber unter Umständen Beratung. Demgegenüber würden bspw. Probanden der Kategorie IV („überdurchschnittliches Risiko“) zugeordnet, bei denen viele, schwerwiegende und chronische kriminogene Bedürfnisse festgestellt werden würden (die sich dementsprechend meist im „risk“, also in vielen und ggf. verschiedenen Arten von Vorstrafen zeigen würden). Hanson et al. (2017) empfehlen bei solchen Probanden z.B. einen Behandlungsumfang von 200 bis 300 Stunden oder mehr sowie eine intensive Überwachung der Einhaltung von Weisungen.

Im Fünfkategorien-Modell werden somit auch Interventionsbedarfe vorgeschlagen, wenigstens indirekt Aspekte der Ansprechbarkeit („responsivity“) berücksichtigt und Vorschläge für deren Umsetzung gemacht. Schließlich werden auch realistische Erwartungen bezüglich der Interventionseffekte formuliert. Bspw. sei in der Kategorie III („durchschnittliches Risiko“) mittelfristig ein deutlicher Behandlungseffekt in Form der Reduktion des Rückfallrisikos zu erwarten, wodurch sich die Probanden im ihrem Rückfallrisiko denjenigen Probanden der Kategorie II annähern könnten.

Somit ist das Fünfkategorien-Modell auch eine evidenzbasierte Hilfestellung für die Ressourcenallokation und Therapieplanung (Hanson, Babchishin, et al., 2017; Hanson, Bourgon, et al., 2017). Es unterstützt und sichert im Alltag forensischer Einrichtungen ständig geforderte, weitreichende Entscheidungen ab. Dadurch, dass es bspw. (übrigens erstmals) eine mithilfe wissenschaftlicher Methoden ermittelte Definition von „niedrigem Risiko“ (Kategorie I des Fünfkategorien-Modells) zur Verfügung stellt, wird es vertretbar, entsprechenden Probanden unter Umständen auch keine Therapie, sondern punktuelle Beratung anzubieten. Diese Entscheidung sollte ggf. dadurch abgesichert sein, dass zwei oder drei unterschiedliche Risikoeinschätzungsverfahren zu einem vergleichbaren Ergebnis kommen, denselben Probanden also derselben Risikokategorie zuordnen. Diese Entscheidung wäre auch mit dem Wissen darüber gerechtfertigt, dass einige wenige dieser Kategorie zugeordnete Probanden (mit oder ohne Therapie) dennoch rückfällig werden.

Ebenso könnte man z.B. fordern, dass durch mindestens zwei standardisierte Risikoeinschätzungsverfahren ein nach dem Fünfkategorien-Modell definiertes „hohes Risiko“ ausgewiesen sein muss, um aus forensisch-fachlicher Sicht die Anordnung einer freiheitsentziehenden Maßregel oder anderer schwerwiegender Einschränkungen empfehlen zu dürfen. Denn Probanden aus der Risikokategorie „hoch“ unterscheiden sich deutlich von Personen der Kategorie „überdurchschnittlich“. Sie weisen – im Gegensatz zu den anderen Kategorien – Defizite in vielen relevanten Bereichen (psychologisch, interpersonell und im Lebensstil) auf, die in der Regel chronifiziert, schwerwiegend und überdauernd sind. Auch sind ihre Ressourcen und Stärken oft begrenzt, sofern überhaupt welche vorliegen (Hanson, Bourgon, et al., 2017).

Die Notwendigkeit einer derartigen größeren Objektivität des Begriffs „hohes Risiko“ zeigt sich in den Ergebnissen einer kürzlich durchgeführten Studie. 52 % aller in der

Maßnahme untergebrachten Sexualstraftäter in Österreich wiesen lediglich ein niedriges oder mittleres Risiko (nach dem Fünfkategorien-Modell) auf. Die juristische Voraussetzung für die Einweisung in den Maßnahmenvollzug ist allerdings eine „hohe Wahrscheinlichkeit neuerlicher Straftaten mit schweren Folgen“ (Eher, 2022).

Diesen Vorteilen des Fünfkategorien-Modells stehen naturgemäß auch einige Limitationen gegenüber. Hanson et al. (2017) weisen selbst darauf hin, dass das Modell bislang auf allgemeine Kriminalität und generelle Rückfälligkeit bezogen ist, seine Adaption an und Anwendung auf spezielle Taten bzw. Tätergruppen (wie sexuelle Kindesmissbraucher; häusliche Gewalt) noch geleistet werden muss (vgl. Hogan, 2021). Ferner benennen Hanson et al. (2017) die Gefahr, dass das Fünfkategorien-Modell ausschließlich zur statistischen Kategorisierung und Benennung des Risikos verwendet wird, ohne dass daraus Entscheidungen und Maßnahmen für eine aktive und eine angemessene Resozialisierung ermöglichende Risikoreduktion abgeleitet und konsequent umgesetzt werden (vgl. Simourd & Olver, 2019).

Die Datenlage zur Anwendung des Fünfkategorien-Modells ist derzeit noch überschaubar. Bei weitem nicht alle strukturierten Risikoeinschätzungsverfahren stellen Algorithmen für seine Anwendung zur Verfügung. Während es bei aktuarischen Verfahren vergleichsweise leicht ist, Grenzwerte für die jeweiligen Kategorien zu bilden, ist das bei klinisch-idiographisch ausgerichteten Verfahren (z. B. Instrumente, die dem Modell des *Structured Professional Judgement* folgen) aus methodischen Gründen schwierig und konzeptionell nicht (leicht) vereinbar. Auch scheinen die Empfehlungen zur Dosierung bzw. Menge von Interventionen von Hanson et al. (2017) durch die in ihrer Veröffentlichung selbst herangezogenen Daten wie durch die generelle Datenlage wenig untermauert und sind insofern noch mit Vorsicht zu betrachten (Simourd & Olver, 2019). Dennoch müssen die Empfehlungen von Hanson et al. (2017) als dem Stand der Forschung entsprechend und konzeptionell plausibel bezeichnet werden.

Darüber hinaus messen unterschiedliche Risikoeinschätzungsverfahren, selbst wenn sie für die gleiche Deliktgruppe und die gleichen Outcome-Kriterien (bspw. Rückfall in einem bestimmten Zeitraum) entwickelt wurden, unterschiedliche Schwerpunkte des Risikos (Barbaree, Langton, & Peacock, 2006). Das ist mit ein Grund dafür, dass die Übereinstimmung verschiedener, für das Fünfkategorien-Modell „geeichter“ Instrumente bei der Zuordnung von Probanden zu den fünf Kategorien bei deutlich unter 50 Prozent liegt (vgl. Feil et al., 2024 (in diesem Heft); Hogan & Olver, 2023).

## Wie also weiter?

Trotz dieser Limitationen und Kritik stellt das Fünfkategorien-Modell aus Sicht der Autoren einen Meilenstein in der Risikokommunikation und für das Risikomanagement von Straftätern dar. Seine Grundlagen und Annahmen sind transparent, objektiv und empirisch fundiert. Es stellt selbst eine Innovation dar. Es hat bereits Forschung und Konzeptentwicklung stimuliert und wird weitere Innovationen in der Versorgungspraxis anstoßen. Es erleichtert nicht nur die einrichtungsinterne Kommunikation und Entscheidungsfindung im Alltag forensischer Versorgungseinrichtungen, sondern sichert deren Praxis u.a. auch ab für das erwartbare Ereignis eines schweren Rückfalls – trotz guter Konzepte und Entscheidungen. Ferner kann es gut als Kommunikationsgrundlage der Versorgungseinrichtungen mit den unmittelbaren Auftraggebern (den Gerichten, also der Judikative) sowie den Auftraggebern und Kostenträgern der Exekutive genutzt werden, sowie für die Darstellung und Erläuterung der Aufgaben und Tätigkeiten in der weiteren Öffentlichkeit.

## Literatur

- Barbaree, H. E., Langton, C. M., & Peacock, E. J. (2006). Different actuarial risk measures produce different risk rankings for sexual offenders. *Sex Abuse, 18*(4), 423-440.
- Buchanan, A., & Leese, M. (2001). Detention of people with dangerous severe personality disorders: a systematic review. *Lancet, 358*(9297), 1955-1959. doi:S0140-6736(01)06962-8.
- Davies, S. T., Helmus, L. M., & Quinsey, V. L. (2022). Improving Risk Communication: Developing Risk Ratios for the VRAG-R. *Journal of Interpersonal Violence, 37*(1-2), 835-862. doi:10.1177/0886260520914555
- Eher, R. (2022). Über die Notwendigkeit von Objektivität, Transparenz und Nachvollziehbarkeit forensisch-psychiatrischer Gutachten von Maßnahmenpatientinnen und -patienten. *Stodertaler Forensiktage; Hinterstoder, 22.-23.9.2023*.
- Eher, R., Lindemann, M., Birklbauer, A., & Müller, J. (2016). Der Gefährlichkeitsbegriff als Voraussetzung für die Verhängung vorbeugender freiheitsentziehender Maßnahmen - eine kritische Betrachtung und Vorschläge de lege ferenda. *Recht und Psychiatrie, 34*(2), 96-106.
- Eher, R., Matthes, A., Schilling, F., Haubner-MacLean, T., & Rettenberger, M. (2011). Dynamic Risk Assessment in Sexual Offenders Using STABLE-2000 and the STABLE-2007: An Investigation of Predictive and Incremental Validity. *Sexual Abuse: A Journal of Research and Treatment, Advance online publication*. doi: 10.1177/1079063211403164.
- Eher, R., Rettenberger, M., Etzler, S., Eberhaut, S., & Mokros, A. (2019). Eine gemeinsame Sprache für die Risikokommunikation bei Sexualstraftätern. Trenn- und Normwerte für das neue Fünf-Kategorienmodell des Static-99 [A common language for communicating risk in sexual offenders – the 5-category model Static-99]. *Recht und Psychiatrie, 37*, 91 - 99.

- Gaunersdorfer, K., & Eher, R. (2023). Die deutschsprachige Version der Violence Risk Scale – Sexual Offense Version (VRS-SO): Prädiktive und konvergente Validität und Kalibrierung des Fünf-Kategorien-Modells. *Monatsschrift für Kriminologie und Strafrechtsreform*, 106(4), 301-313. doi:doi:10.1515/mks-2023-0012
- Hanson, R. K., Babchishin, K. M., Helmus, L. M., Thornton, D., & Phenix, A. (2017). Communicating the results of criterion referenced prediction measures: Risk categories for the Static-99R and Static-2002R sexual offender risk assessment tools. *Psychol Assess*, 29(5), 582-597. doi:10.1037/pas0000371
- Hanson, R. K., Bourgon, G., McGrath, R., Kroner, D., D'Amora, D. A., Thomas, S. S., & Tavaréz, L. P. (2017). A Five-Level Risk and Needs System: Maximizing Assessment Results in Corrections through the Development of a Common Language. Retrieved from [http://sarato-so.org/pdf/A\\_Five\\_Level\\_Risk\\_and\\_Needs\\_System\\_Report.pdf](http://sarato-so.org/pdf/A_Five_Level_Risk_and_Needs_System_Report.pdf)
- Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: a meta-analysis of 118 prediction studies. *Psychol. Assess*, 21(1), 1-21. doi:2009-03401-001 [pii];10.1037/a0014421 [doi]
- Rettenberger, M., Eher, R., Desmarais, S. L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., Dean, K., Doyle, M., Folino, J. O., Godoy-Cervera, V., Grann, M., Ho, R. M. Y., Large, M. M., Pham, T. H., Nielsen, L. H., Rebocho, M. F., Reeves, K. A., de Ruiter, C., Seewald, K., & Singh, J. P. (2017). Kriminalprognosen in der Praxis. *Diagnostica*, 63(1), 2-14. doi:10.1026/0012-1924/a000168
- Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. *Law Hum. Behav*, 29(5), 615-620. doi:10.1007/s10979-005-6832-7 [doi]
- Singh, J. P. (2013). Predictive Validity Performance Indicators in Violence Risk Assessment: A Methodological Primer. *Behavioral Sciences & the Law*, 31(1), 8-22.
- Singh, J. P., Grann, M., & Fazel, S. (2011). A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clin. Psychol. Rev*, 31(3), 499-513. doi:S0272-7358(10)00194-7 [pii];10.1016/j.cpr.2010.11.009 [doi]
- Wirove, R. L., Olver, M. E., & Haag, A. (2023). Discrimination and Calibration Properties of the Violence Risk Appraisal Guide–Revised in a Not Criminally Responsible Provincial Population. *Assessment*, 30(5), 1672-1687. doi:10.1177/10731911221116564
- Zoe, H. N., Carter, A. M., Harris, G. T., & Sharpe, A. J. (2008). Does using nonnumerical terms to describe risk aid violence risk communication? Clinician agreement and decision making. *J. Interpers. Violence*, 23(2), 171-188. doi:23/2/171 [pii];10.1177/0886260507309337 [doi]