

Zeitschrift: Kriminologie / Schweizerische Arbeitsgruppe für Kriminologie SAK = Criminologie / Groupe Suisse de Criminologie GSC = Criminologia / Gruppo Svizzero di Criminologia GSC

Herausgeber: Schweizerische Arbeitsgruppe für Kriminologie

Band: 37 (2020)

Artikel: Instrumente der Risikoeinschätzung als Grundlage für die Wiedereingliederung : Stand der Forschung und Entwicklungspotential

Autor: Endrass, Jérôme / Noll, Thomas / Rossegger, Astrid

DOI: <https://doi.org/10.5169/seals-1051426>

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften. 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. [Siehe Rechtliche Hinweise.](#)

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. [Voir Informations légales.](#)

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. [See Legal notice.](#)

Download PDF: 08.02.2025

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

Instrumente der Risikoeinschätzung als Grundlage für die Wiedereingliederung – Stand der Forschung und Entwicklungspotential

JÉRÔME ENDRASS* / THOMAS NOLL** / ASTRID ROSSEGER***

Inhaltsverzeichnis

Zusammenfassung	69
1. Einleitung	70
2. Stand der Forschung: Überlegenheit mechanischer Risikobeurteilungen	71
3. Stellenwert der forensischen Expertise	74
4. Abklärung des Interventionsbedarfs anstelle einer Risikoquantifizierung.....	75

Zusammenfassung

Risikobeurteilungen haben sich in der forensischen Psychologie und Psychiatrie fest etabliert. Seit knapp 25 Jahren sind mehrere Instrumente verfügbar um das Rückfallrisiko standardisiert zu erfassen. Validierungsstudien im deutschsprachigen Raum haben einen Zusammenhang zwischen Rückfällen und Punktwerten ausgewiesen. Einige dieser Verfahren wurden aufgrund von epidemiologischen Ansätzen entwickelt und erlauben, gemäss der Entwickler der Verfahren, eine Schätzung der Rückfallquote für Subpopulationen von Straftätern. Diese Rückfallquoten konnten im deutschsprachigen Raum allerdings nicht repliziert werden. Von den mittlerweile weit über 500 publizierten Verfahren, wurde zudem nur sehr wenige validiert übersetzt und kommen für einen Einsatz in der Schweiz in Frage. Wenige Verfahren adressieren spezifisch die Faktoren, die für eine forensisch psychotherapeutische Fallkonzeption notwendig sind. Ähnlich sieht es im Bereich der Vollzugslockerungen aus. Die wenigen

* Stabschef, Amt für Justizvollzug, Zürich

** Leiter der Fachstelle Forensik des Amtes für Justizvollzug, Zürich

*** Stellvertretende Leiterin der Forschungs- und Entwicklungsabteilung des Justizvollzugs Zürich

Verfahren, die für die Einschätzung der Möglichkeiten der erfolgreichen Wiedereingliederung hilfreich sind, wurden für die Schweiz nicht abschliessend validiert und können nur unter Vorbehalt eingesetzt werden. Möglichkeiten diese Lücke zu schliessen sowie neue Ansätze in der forensischen Beurteilung werden kurz vorgestellt und in Grundzügen diskutiert.

1. Einleitung

Seit Jahren gilt in den einschlägigen Mindestanforderungen bzw. Empfehlungen für Prognosegutachten¹ sowie der Rechtsprechung die Konvention, dass standardisierte Risikobeurteilungsinstrumente immer nur als Hilfsmittel, aber nicht als alleinige Grundlage einer Risikobeurteilung zu Rate gezogen werden dürfen.² Solche Instrumente seien insofern von Nutzen, als dass sie *«Anhaltspunkte über die Ausprägung eines strukturellen Grundrisikos eines Betroffenen liefern»*. Jedoch seien solche Instrumente *«für sich allein nicht geeignet, eine fundierte individuelle Gefährlichkeitsprognose tragfähig zu begründen. Zur Erstellung einer individuellen Prognose bedarf es über die Anwendung derartiger Instrumente hinaus daher zusätzlich einer differenzierten Einzelfallanalyse durch den Sachverständigen»*.³ So eingängig diese Praxis klingt, so instabil ist ihr Fundament, wenn man eine tieferschürfende Betrachtung des aktuellen Stands der Forschung berücksichtigt. Es zeigt sich nämlich mit grosser Konstanz, dass die mechanische Beurteilung des Rückfallrisikos anderen Ansätzen überlegen ist.

¹ A. BOETTICHER et al., Mindestanforderungen für Prognosegutachten, NStZ 2006, 537 ff., insb. 541; H.-L. KRÖBER et al., Empfehlungen für Prognosegutachten: Erfahrungswissenschaftliche Empfehlungen für kriminalprognostische Gutachten, Forensische Psychiatrie Psychologie Kriminologie 2019 (online first; 18.10.2019).

² BGer vom 4.12.2015, 6B_424/2015, E. 3.3; BGer vom 19.6.2018, 6B_582/2017, E. 2.2.

³ BGer vom 4.12.2015, 6B_424/2015, E. 3.3; ausführlich bestätigt in BGer vom 19.6.2018, 6B_582/2017, E. 2.2 mit Betrachtung der forensischen Literatur.

2. Stand der Forschung: Überlegenheit mechanischer Risikobeurteilungen

Spätestens seit der zu Beginn der 1970er Jahre veröffentlichten «Baxstrom-Studie» erreichte die forensische Risk-Assessment-Forschung die Mainstream-Journals der akademischen Psychiatrie. STEADMAN/COCOZZA⁴ gingen in ihrer Untersuchung der «Baxstrom»-Patienten der Frage nach, wie hoch die Rückfallrate bei forensischen Psychiatriepatienten ist, die von Klinikern als hochrückfallgefährdet eingestuft worden waren, aber als Folge eines Gerichtsurteils des Amerikanischen Bundesgerichts aus formalen Gründen entlassen werden mussten. Das Ergebnis war ernüchternd. So konnten STEADMAN/COCOZZA zeigen, dass klinische Risikobeurteilungen das Rückfallrisiko von psychisch kranken Straftätern substanziell überschätzen. Von den als hochrückfallgefährdet eingestuften forensischen Psychiatriepatienten, fielen nach der Entlassung lediglich 14 von 98 Patienten mit aggressivem Verhalten auf. Die Autoren schlussfolgerten, dass die klinischen Experten bei ihren Experten weniger das Konstrukt der Gefährlichkeit, als den Krankheitsverlauf – analog zu nicht straffälligen Patienten – beurteilt hatten. Was sind also Kriterien für Gefährlichkeit, die losgelöst vom Krankheitsverlauf beurteilt werden können?

Zur Beantwortung dieser Fragen legten die Autoren bereits damals ein aktuarisches Instrument vor – die «legal dangerousness scale». Diese Skala erlaubte eine spezifischere Beurteilung des Rückfallrisikos als die rein klinische Urteilsfindung. Das statistische Modell der «legal dangerousness scale» identifizierte 11 der 14 Rückfalltäter als Personen mit einem hohen Risiko. Darüber hinaus wies das statistische Modell bei 25 weiteren Personen ein hohes Rückfallrisiko aus (die nicht rückfällig wurden). Die Sensitivität (True Positive Rate) des Modells betrug $TPR = 0.79$, die Spezifität des Modells (True Negative Rate) betrug $TNR = 0.70$. COCOZZA/STEADMAN⁵ schlussfolgerten, dass das von ihnen empirisch ermittelte «statistische Modell» trennscharf zwischen rückfälligen und nicht rückfälligen Tätern unterschied.

⁴ H. J. STEADMAN/J. J. COCOZZA, The criminally insane patient: Who gets out?, *Social psychiatry* 8(4), 1973, 230-238.

⁵ J. J. COCOZZA/H. J. STEADMAN, Some refinements in the measurement and prediction of dangerous behavior, *American Journal of Psychiatry* 131(9), 1974, 1012-1014.

Diese Untersuchung war nicht die erste Studie, die Treffergenauigkeit einer strukturierten mit einer unstrukturierten Beurteilungsmethode verglich. Schon Mitte der 1950er Jahre publizierte PAUL MEEHL⁶ eine systematische Literaturanalyse zum Thema. Der Autor ging der Frage nach, welche Form der Urteilsfindung empirisch die besten Ergebnisse liefert: Das freie Expertenurteil oder die hochstandardisierte Urteilsfindung, die kaum Raum für Einzelfallanpassungen lässt. Das Ergebnis war eindeutig. In 19 von 20 identifizierten Studien war die standardisierte Beurteilung überlegen. Die klinische Expertise wurde somit durch den Algorithmus geschlagen. MEEHL berücksichtigte dabei Studien aus dem ganzen Spektrum der Psychologie.

Das Ergebnis von MEEHL wurde knapp fünfzig Jahre später in einer sehr aufwendigen Meta-Analyse repliziert, die die Ergebnisse aus 56 Jahren Forschung zu Methoden der Urteilsbildung zusammenfasste. Auch in dieser Meta-Analyse⁷ wurden Daten aus dem ganzen Spektrum der Psychologie berücksichtigt. Die Autorinnen und Autoren kamen zum Schluss, dass die klinische Urteilsfindung der algorithmusgestützten Beurteilung deutlich unterlegen ist. Die Wahrscheinlichkeit eines besseren Resultates betrug für die mechanische Methode gegenüber der klinischen Methode 13%. Hervorzuheben gilt es, dass der gefundene Effekt der Überlegenheit der standardisierten Herangehensweise insbesondere für die Beurteilung des Gewaltrisikos gilt.

Dies lässt sich auch anhand einer jüngst veröffentlichten Zürcher Studie aus dem Kontext der Häuslichen Gewalt illustrieren. SEEWALD et al.⁸ verglichen die Trennschärfe der Risikobeurteilung erfahrener forensischen Psychiaterinnen und Psychiater mit der Trennschärfe der Risikobeurteilung von Psychologiestudentinnen. Während die Studentinnen ihr Urteil einzig auf die Anwendung eines Risk-Assessment-Instrumentes abstützten, basierte das Urteil der forensischen Psychiater entweder auf einer freihändigen Beurteilung oder aber der Integration und Einbettung eines

⁶ P. MEEHL (Hrsg.), *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence* Minneapolis, University of Minnesota 1954.

⁷ S. ÆGISDÓTTIR et al., *The Meta-Analysis of Clinical Judgment Project: Fifty-Six Years of Accumulated Research on Clinical Versus Statistical Prediction*, *The Counseling Psychologist* 34(3), 2006, 341-382.

⁸ K. SEEWALD et al., *Assessing the risk of intimate partner violence: Expert evaluations versus the Ontario Domestic Assault Risk Assessment*, *Journal of Forensic Psychology: Research and Practice* 17, 2017, 217-231 (doi:10.1080/24732850.2017.1326268).

Risk-Assessment-Instrumenten in das klinische Urteil. In der Regel wandten die Forensiker Instrumente an, die dem Prinzip des strukturierten klinischen Urteils folgten.

Die Trennschärfe der von den Studierenden vorgenommenen mechanischen Beurteilung war deutlich höher als die Trennschärfe der klinischen Beurteilung der Fachpersonen mit jahrelanger Berufserfahrung. Der Algorithmus schlug die Experten. Die klinischen Urteile der forensischen Psychiater und Psychiaterinnen waren zudem so wenig trennscharf, dass sie nicht einmal besser als der Zufall waren.

Wenngleich spektakulär, ist dieses Ergebnis wenig bemerkenswert, da es sich mehr oder weniger nahtlos in eine umfangreiche Literatur einfügt: Wenn es darum geht, gewalttätiges Verhalten zu beurteilen, ist die klinische Methode, sprich die Expertenmeinung, dem einfachen Algorithmus, sprich der Anwendung eines Instruments, unterlegen. Versuchen Experten den Algorithmus durch ihr Fachwissen zu ergänzen und an den konkreten Einzelfall anzupassen, wird gemäss einer Meta-Analyse von HANSON/MORTON-BURGON⁹ die Prognose zudem verschlechtert.

Bemerkenswert ist, dass es sich bei den forensischen Risk-Assessment-Instrumenten um Instrumente handelt, deren Algorithmen sehr simpel sind. Sie basieren auf sehr einfachen Modellen. Auf künstlicher Intelligenz und auf Maschinenlernen basierende Algorithmen, anhand derer hochkomplexe Aufgaben gelöst werden können, werden hierbei nicht einmal berücksichtigt. Einem solchen Algorithmus ist es kürzlich gelungen, sich selbst das Schachspielen beizubringen und dann den bis dato stärksten Schachcomputer zu schlagen.¹⁰ Dies, obwohl der Algorithmus nicht auf eine Strategiesammlung, die von Menschen angelegt worden ist, zurückgreifen konnte, sondern sich das Spiel gewissermassen «von der Pike» auf selbst aneignen musste. Dies gelang ihm innerhalb von gerade einmal vier Stunden. Anders die in der forensischen Psychologie und Psychiatrie verwendeten Algorithmen. Diese sind sehr rudimentär: Sie enthalten Merkmale, die robust mit Rückfälligkeit korrelieren, so robust, dass die Zusammensetzung der

⁹ R. K. HANSON/K. E. MORTON-BOURGON, The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies, *Psychological Assessment* 21(1), 2009, 1-21 (doi: 10.1037/a0014421).

¹⁰ D. SILVER et al. (2018).

Merkmale austauschbar ist. KRONER/MILLS/ REDDON¹¹ haben in einer sehr breit rezipierten Arbeit die Trennschärfe mehrerer Instrumente untersucht, namentlich Psychopathie-Checkliste (PCL-R), Level of Service Inventory (LSI) und Violence Risk Appraisal Guide (VRAG). Ergänzend haben die Autoren die Trennschärfe künstlicher Skalen berücksichtigt. Diese künstlichen Skalen wurden nach dem Zufallsprinzip erzeugt: Alle Kriterien («Items») der untersuchten Instrumente wurden in einen Topf geworfen. Anschliessend wurden zufällig Kriterien aus dem Topf gezogen und daraus eine künstliche Skala geformt. Eine zufällig generierte Skala setzte sich somit beispielsweise aus Kriterien aus der Psychopathie-Checkliste (PCL-R), solchen aus dem Level of Service Inventory (LSI) und solchem aus dem Violence Risk Appraisal Guide (VRAG) zusammen.

Bemerkenswert war zweierlei: Erstens war die Trennschärfe dieser künstlichen Skalen signifikant besser als der Zufall. Das heisst, dass ein wilder Item-Mix das Rückfallrisiko spezifischer beurteilt als forensische Psychiatern der Zürcher Studie. Zweitens waren die künstlichen Skalen nicht schlechter als die etablierten Prognoseinstrumente. Das heisst, dass es empirisch keine Rolle spielt, ob man die PCL-R allein, den VRAG allein oder eine bunte Mischung aus beidem anwendet.

Das wichtige Zwischenfazit für die forensische Psychiatrie lautet somit: Erfahrene Psychiater werden selbst dann von einfachen Algorithmen bei der Risikobeurteilung geschlagen, wenn die Psychiater ihr Urteil auf eine Reihe von Prognose-Instrumenten abstützen können.

3. Stellenwert der forensischen Expertise

Heisst dies, dass die forensische Expertise nutzlos sind? Nein, wie nachfolgend gezeigt werden soll. Forensische Experten sind schlecht darin, die Frage nach der Rückfallwahrscheinlichkeit zu beantworten. Nicht weil forensische Expertinnen und Experten per se schlechte Prognostiker wären, sondern weil der Umgang mit Wahrscheinlichkeiten Menschen überfordert. Die israelischen Psychologen AMOS TVERSKY und DANIEL KAHNEMAN konnten in verschiedenen Arbeiten zeigen, dass Wahrschein-

¹¹ D. G. KRONER/J. F. MILLS/J. R. REDDON, A Coffee Can, factor analysis, and prediction of antisocial behavior: The structure of criminal risk, *International Journal of Law and Psychiatry* 28(4), 2005, 360-374.

lichkeitsaussagen schnell zu einer Überforderung führen.¹² Selbst Statistiker*innen gelingt es nicht gut, Urteile zu fällen, die auf Wahrscheinlichkeitsüberlegungen beruhen. Wenn es nun aber sogar Statistiker*innen schwerfällt, mit Wahrscheinlichkeitsaussagen umzugehen, wieso ist die Frage der Rückfallwahrscheinlichkeit ein Standardbestandteil des forensischen Fragekatalogs? Diese Frage kann so in der Form von keinem Forensiker im Sinne einer echten Expertise beantwortet werden. Das heisst, dass das Expertenurteil weniger trennscharf als das Ergebnis eines einfachen Algorithmus' ist.

4. Abklärung des Interventionsbedarfs anstelle einer Risikoquantifizierung

Abgesehen von der Rückfallwahrscheinlichkeit gibt es aber eine ganze Reihe von Fragen, die ein Experte mit hinreichender Reliabilität und Validität beurteilen kann: Die Frage nach der Diagnose, nach der Deliktdynamik sowie den dazu passenden risikosenkenden Interventionen und die Frage nach der für die Behandlung angemessene Institution.

Moderne forensische Instrumente wie die vor kurzem eingeführte Octagon-Intervention¹³ verzichten deshalb explizit auf eine Quantifizierung des erwarteten Rückfallrisikos. Dies, weil der Octagon-Ansatz das Fachwissen der forensischen Experten abgreifen will, ohne in die Falle der Wahrscheinlichkeitsproblematik zu tappen. Im Kern handelt es sich bei der Octagon-Intervention um ein Instrument, das sich am strukturierten klinischen Urteil orientiert. Der Ansatz berücksichtigt ferner Erkenntnisse aus der Kognitionspsychologie, indem versucht wird, kognitiven Verzerrungen im Urteilsprozess vorzubeugen. Das erste Ergebnis der

¹² Anstelle vieler: D. KAHNEMAN/A. TVERSKY, Prospect theory: An analysis of decision under risk, in: L. C. MacLean/W. T. Ziemba (Hrsg.), World Scientific Handbook in Financial Economics Series, Handbook of the Fundamentals of Financial Decision Making, 2013, 99–127 ; D. KAHNEMAN/A. TVERSKY, Subjective probability: A judgment of representativeness, Cognitive Psychology 3(3), 1972, 430–454; D. KAHNEMAN/A. TVERSKY, Judgment under Uncertainty: Heuristics and Biases, Science 185(4157), 1974, 1124–1131 (doi: 10.1126/science.185.4157.1124).

¹³ www.octagon-intervention.ch; J. ENDRASS/A. ROSSEGGER, Herausforderungen im Bedrohungsmanagement und das Octagon als neuer Ansatz, format magazine 7, 2017, 36–41.

Anwendung der Octagon-Intervention resultiert in der qualitativen Frage, ob weiterer Interventionsbedarf besteht oder nicht. Wird ein Interventionsbedarf bejaht, geht es in einem zweiten Schritt darum festzustellen, welches die relevante Risikodisposition ist. Darauf aufbauend werden Eskalationsszenarien diskutiert sowie geeignete Massnahmen zur Vermeidung der Eskalation erörtert. Zum Schluss mündet das Vorgehen in konkreten Interventionsempfehlungen.

Der Ansatz der Octagon-Intervention wird seit 2015 von verschiedenen Polizeikorps schweizweit eingesetzt und fortlaufend weiterentwickelt. Es handelt sich dabei um ein qualitatives Verfahren zur Risikoeinschätzung. Mit dem Wechsel von der quantitativen Beurteilung (wie hoch ist das Risiko?) hin zur qualitativen Beurteilung (muss ich was tun und wenn ja was?) hat man das Trennschärfeproblem durch eine Umformulierung der Fragestellung entschärft.

Weiterführende Studien müssen nun den Nachweis erbringen, ob diese Umformulierung der Fragestellung, d.h. der Wechsel vom quantitativen zum qualitativen Risikobegriff, den gewünschten Effekt liefern kann.