

Der Nutzen Algorithmus gestützter Verfahren in der forensischen Risikoeinschätzung. Eine wissenschaftliche Bestandesaufnahme

Prof. Dr. Jérôme Endrass



Kanton Zürich
Direktion der Justiz und des Innern

NORBERT NEDOPIL, JÉRÔME ENDRASS,
ASTRID ROSSEGER, THOMAS WOLF

PROGNOSE: RISIKOEINSCHÄTZUNG IN FORENSISCHER PSYCHIATRIE UND PSYCHOLOGIE

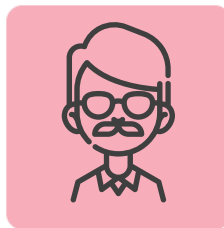
EIN HANDBUCH FÜR DIE PRAXIS

 PABST

METHODEN DER FORENSISCHEN RISIKOEINSCHÄTZUNG

Unstrukturiertes
Klinisches Urteil

Erste Generation



Aktuarische
Verfahren -
Algorithmusbasiert

Zweite Generation



Strukturiertes
klinisches Urteil (SPJ)

Dritte Generation



AKTUARISCHE INSTRUMENTE – ALGORITHMEN IN DER FORENSISCHEN RISIKOEINSCHÄTZUNG

Gewichtung und Kombination von Informationen nach festen und eindeutigen Regeln:

- Bewerten Sie die Aufgaben: Jede Antwort wird mit einer gewichteten Punktzahl versehen
- Berechnen Sie die Gesamtpunktzahl
- Einstufung des Täters in eine Risikokategorie von niedrig bis hoch, wobei jede Risikokategorie eine gruppenbasierte Rückfallwahrscheinlichkeit aufweist



VIOLENCE RISK APPRAISAL GUIDE - VRAG

VIOLENT RECIDIVISM OF MENTALLY DISORDERED OFFENDERS

The Development of a Statistical Prediction Instrument

GRANT T. HARRIS
MARNIE E. RICE
Mental Health Centre, Penetanguishene
VERNON L. QUINSEY
Queen's University, Kingston

Originalarbeit 577

Der Violence Risk Appraisal Guide (VRAG) – ein Instrument zur Kriminalprognose bei Gewalttätern
Übersichtsarbeit und autorisierte deutsche Übersetzung
The Violence Risk Appraisal Guide (VRAG) – a Tool for the Risk Assessment of Violent Offenders
Review and Authorized German Translation

Autoren

A. Rossegger¹, F. Urbanik², C. Danielsson¹, J. Endrass¹

Institute

¹ Psychiatrisch-Psychologischer Dienst, Justizvollzug Kanton Zürich
² Jugendforensische Abteilung der Universität Basel

1	Lived with both biological parents to age 16 (except for death of parent)	3
2	Elementary school maladjustment	2
3	History of alcohol problems	1
4	Marital status	-2
5	Criminal history score for convictions and charges for nonviolent offenses prior to the index offense	-
6	Failure on prior conditional release	3
7	Age at index offense	2
8	Victim injury	1
9	Any female victim	1
10	Meets DSM-III criteria for any personality disorder	-2
11	Meets DSM-III criteria for schizophrenia	1
12	Hare Psychopathy Checklist-Revised score	-
Total score:		10
Total score including prorating:		17

VIOLENCE RISK APPRAISAL GUIDE - VRAG

4

Marital status

(at time of index offense)

- Ever married (or lived common law in the same home for at least 6 months) [-2]
- Never married [1]
- Unclear or incomplete information

7

Age at index offense

(at most recent birthday)

- ≥ 39 [-5]
- 34-38 [-2]
- 28-33 [-1]
- 27 [0]
- ≤ 26 [2]
- Unclear or incomplete information

VIOLENCE RISK APPRAISAL GUIDE - VRAG

VIOLENT RECIDIVISM OF MENTALLY DISORDERED OFFENDERS

The Development of a Statistical Prediction Instrument

GRANT T. HARRIS
MARNIE E. RICE
Mental Health Centre, Penetanguishene
VERNON L. QUINSEY
Queen's University, Kingston

Originalarbeit 577

Der Violence Risk Appraisal Guide (VRAG) – ein Instrument zur Kriminalprognose bei Gewaltstraftätern
Übersichtsarbeit und autorisierte deutsche Übersetzung
The Violence Risk Appraisal Guide (VRAG) – a Tool for the Risk Assessment of Violent Offenders
Review and Authorized German Translation

Autoren A. Rossegger¹, F. Uebanik¹, C. Danielsson¹, J. Endrass¹
Institute ¹ Psychiatrisch-Psychologischer Dienst, Justizvollzug Kanton Zürich
² Jugendforensische Abteilung der Universität Basel

Normative Data for Probability of recidivism

Probability of Violent Recidivism at Two Different Mean Lengths of Opportunity as a Function of Nine Equal-Sized Violence Risk Appraisal Guide (VRAG) Categories

VRAG Total score	Risk category	Probability of recidivism	
		7 years	10 years
≤ -22	1	0%	8%
-21 to -15	2	8%	10%
-14 to -8	3	12%	24%
-7 to -1	4	17%	31%
0 to 6	5	35%	48%
7 to 13	6	44%	58%
14 to 20	7	55%	64%
21 to 27	8	76%	82%
≥ 28	9	100%	100%

AKTUARISCHE INSTRUMENTE – ALGORITHMEN IN DER FORENSISCHEN RISIKOEINSCHÄTZUNG

Weltweit

- Mehr als 100 verschiedene aktuarische Instrumente sind heute im Einsatz
- Meta-Analysen zeigen: Aktuarische Instrumente unterscheiden erfolgreich zwischen Rückfälligen und Nicht-Rückfälligen mit einer mittleren AUC von .72 (Fazel et al., 2013)

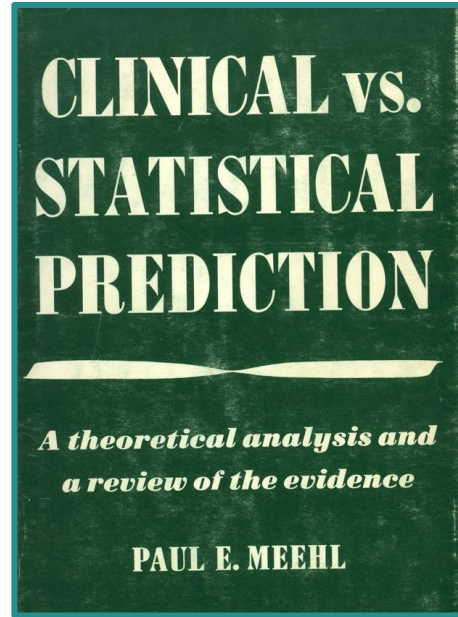
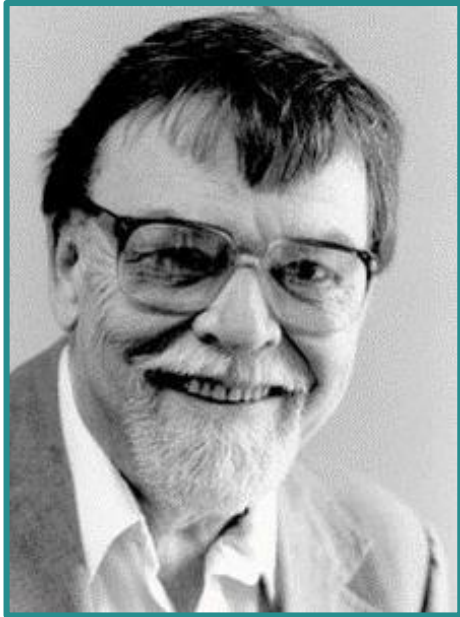
Deutscher Sprachraum

- Mehrere Tools wurden ins Deutsche übersetzt (z. B. VRAG, SORAG, ODARA)
- Mehrere Replikationsstudien zeigten eine vergleichbare Trennschärfe (AUC) wie bei Referenzgruppen (z. B. VRAG, SORAG, ODARA, MnSOST, Static-99, Static-2002)
- Neue aktuarische Instrumente wurden entwickelt (z.B. Radar-iTE)



- recht gute Trennschärfe
- Wird weltweit häufig in der Polizeiarbeit und bei Gutachten eingesetzt

Algorithmen vs. klinische Einschätzung – (Meehl 1954)



- Vergleich von klinischer und statistischer (mechanischer) Methode anhand von 20 Studien.
- In 19 von 20 Arbeiten war die statistische Methode der klinischen überlegen.

Meta-Analyse: Algorithmus vs. Klinisches Urteil

Ergebnis

- Überlegenheit von Algorithmen: in 52% der Studien
- Überlegenheit der klinischen Methode: in 10% der Studien
- Gleichstand: in 38% der Studien
- **Die beste Performance gegenüber der klinischen Methode erzielten Algorithmen bei der Einschätzung des Gewaltrisikos**



- Überlegenheit gegenüber unstrukturierten klinischen Beurteilungen

• Major Contribution

The Meta-Analysis of Clinical Judgment Project: Fifty-Six Years of Accumulated Research on Clinical Versus Statistical Prediction

Stefanía Ægisdóttir
Michael J. White
Paul M. Spengler
Alan S. Maugherman
Linda A. Anderson
Robert S. Cook
Cassandra N. Nichols
Georgios K. Lampropoulos
Blain S. Walker
Genna Cohen
Jeffrey D. Rush
Ball State University

The Accuracy of Recidivism Risk Assessments for Sexual Offenders: A Meta-Analysis of 118 Prediction Studies

R. Karl Hanson
Public Safety Canada

Kelly E. Morton-Bourgon
Department of Justice, Canada

This review compared the accuracy of various approaches to the prediction of recidivism among sexual offenders. On the basis of a meta-analysis of 536 findings drawn from 118 distinct samples (45,398 sexual offenders, 16 countries), empirically derived actuarial measures were more accurate than unstructured professional judgment for all outcomes (sexual, violent, or any recidivism). The accuracy of structured professional judgment was intermediate between the accuracy found for the actuarial measures and for unstructured professional judgment. The effect sizes for the actuarial measures were moderate to large by conventional standards (average d values of 0.67–0.97); however, the utility of the actuarial measures will vary according to the referral question and samples assessed. Further research should identify the psychologically meaningful factors that contribute to risk for reoffending.

Keywords: risk assessment, sexual offenders, meta-analysis

Average Predictive Accuracy of Various Forms of Risk Assessment for Sexual Offenders

Form of risk assessment	Recidivism outcome criteria					
	Sex <i>d</i> (95% CI)	<i>k</i>	Any violence <i>d</i> (95% CI)	<i>k</i>	Any <i>d</i> (95% CI)	<i>k</i>
Measures designed for sexual recidivism						
Empirical actuarial	0.67 (.63, .72)	81	0.51 (.47, .56)	42	0.52 (.48, .56)	43
Mechanical	0.66 (.58, .74)	29	0.40 (.31, .49)	10	0.37 (.30, .43)	19
Structured professional judgment	0.46 (.29, .62) ^a	6	0.31 (.13, .49)	3	0.26 (.11, .41)	4
Measures designed for violent recidivism						
Empirical actuarial	0.39 (.31, .46)	20	0.78 (.70, .86)	15	0.74 (.66, .82)	14
Mechanical	0.33 (.07, .60)	4	0.31 (.07, .56)	3	—	
Measures designed for any recidivism						
Empirical actuarial	0.62 (.49, .75)	9	0.79 (.63, .95)	5	0.97 (.93,1.01)	10
Unstructured professional judgment	0.42 (.32,0.51)	11	0.22 (.15, .29) ^b	7	0.11 (.06, .17)	9

Table 3

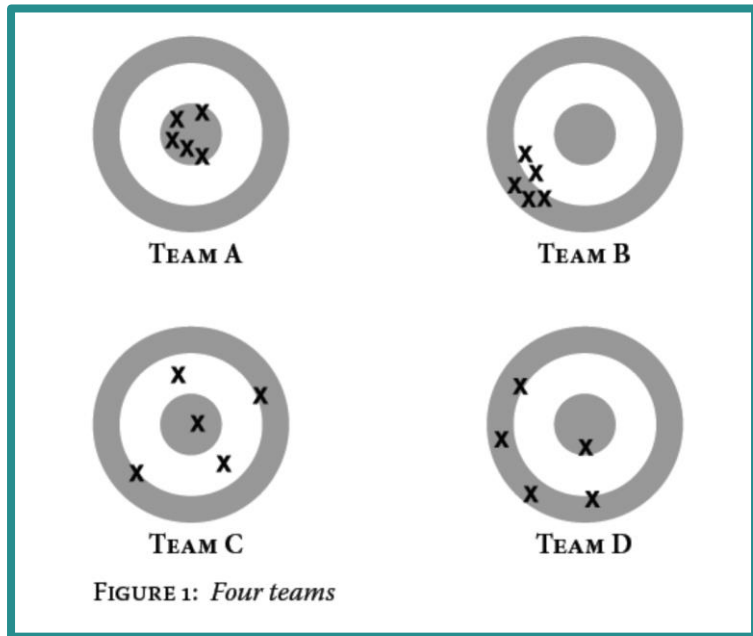
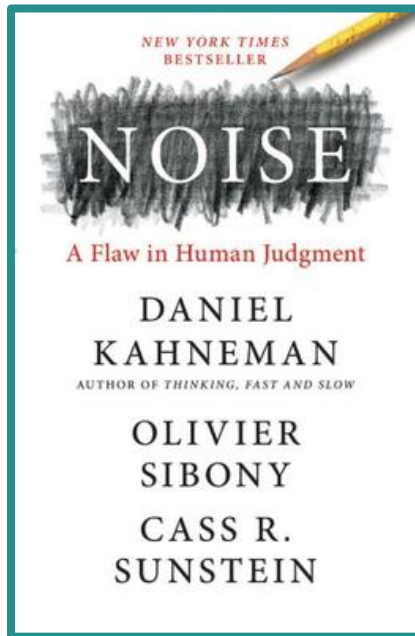
Comparison of Empirical Actuarial With Clinically Adjusted Actuarial Risk Assessments

Study	Measure	Actuarial risk score	Adjusted risk category	Difference	Type of recidivism	Recidivists/ total
Hanson (2007)	Static-99	1.04	.95	-.09	Sexual	34/507
	Static-99	0.91	.78	-.13	Sexual or violent	65/507
Gore (2007)	MnSOST-R	0.50	.31	-.19	Sexual	19/381
Vrana et al. (2008)	LSI-OR	0.90	.54	-.36	Sexual or violent	25/198
	LSI-OR	0.98	.64	-.34	Any	52/198
Unweighted <i>M</i>		0.87	.64	-.22		

Note. MnSOST-R = Minnesota Sex Offender Screening Tool—Revised (Epperson et al., 1995); LSI-OR = Level of Service Inventory—Ontario Revision (Girard & Wormith, 2004).

Daniel Kahnemann

Bias and Noise - (systematische) Verzerrung und Rauschen



Systematische Verzerrung forensischer Instrumente (VRAG)

OPEN ACCESS Freely available online

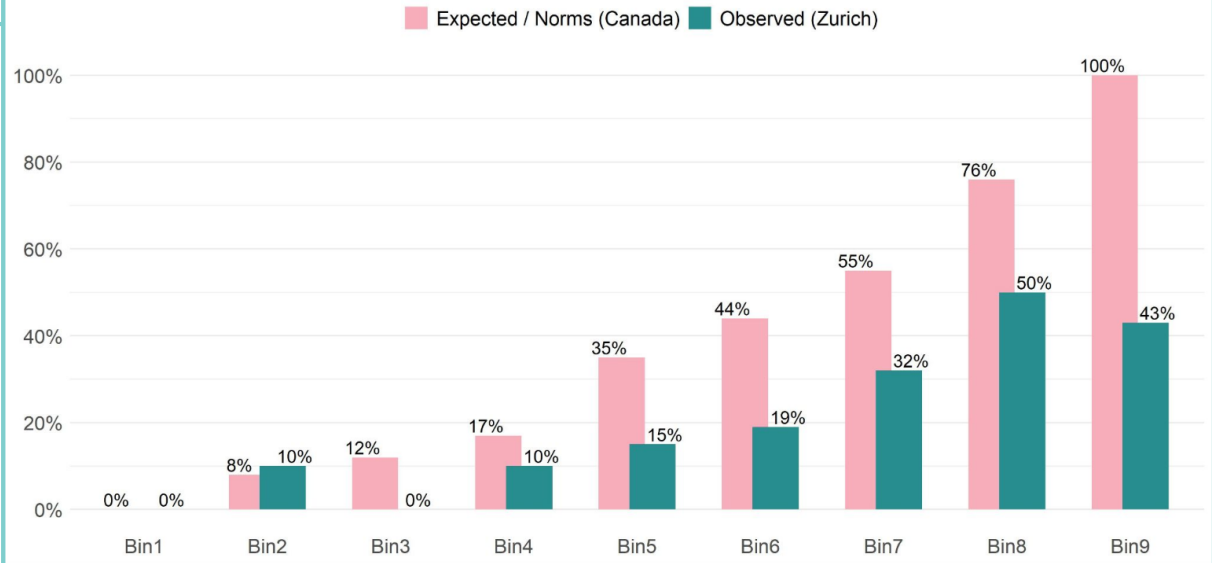


Replicating the Violence Risk Appraisal Guide: A Total Forensic Cohort Study

Astrid Rossegger^{1,2}, Jérôme Endrass^{1,2*}, Juliane Gerth^{1,2}, Jay P. Singh^{1,3}

¹Department of Mental Health Services, Office of Corrections, Canton of Zurich, Zurich, Switzerland, ²Department of Psychology, Germany, ³Institute of Health Sciences, Molde University College, Molde, Norway

7-Year Post-Release Recidivism Rates in the VRAG Development Sample vs. the Swiss Sample



Systematische Verzerrung - PLAUSIBLE ERKLÄRUNG

Swiss Medical Weekly

Formerly: Schweizerische Medizinische Wochenschrift

An open access, online journal • www.smw.ch

Original article | Published 15 March 2021 | doi:10.4414/ismw.2021.20462

Cite this as: Swiss Med Wkly. 2021;151:w20462

Recidivism in Switzerland: the influence of custodial sanctions

Gonçalves Leonel C.^{a,b}, Baggio Stéphanie^c, Weber Michael^{b,d}, Urwyler Thierry^a, Noll Thomas^a, Singh Jay P.^{a,b}, Rossegger Astrid^b, Endrass Jérôme^{b,d}

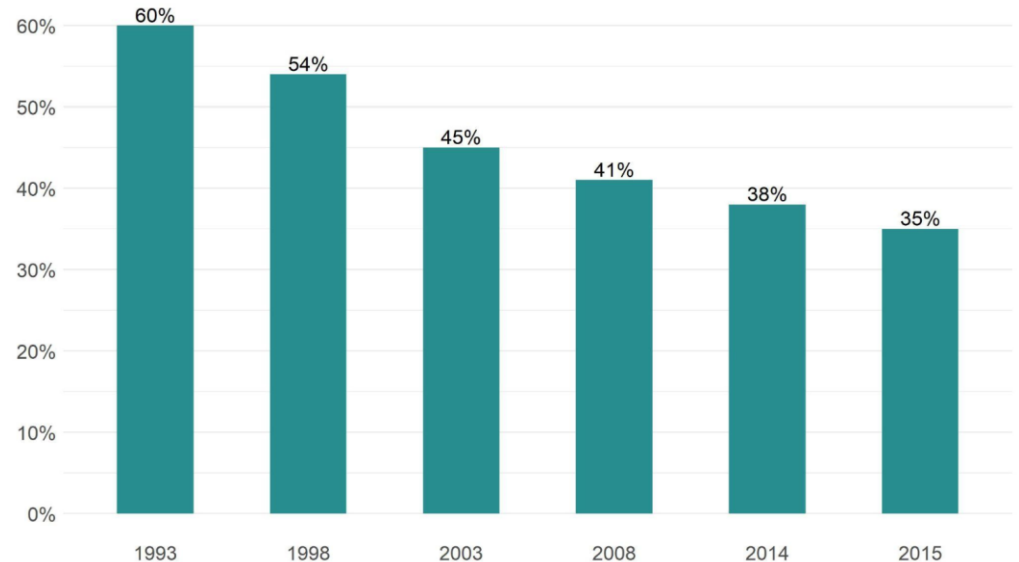
^a Canton of Zurich, Directorate of Justice and the Interior, Office of Corrections, Zurich, Switzerland

^b Department of Psychology, University of Konstanz, Germany

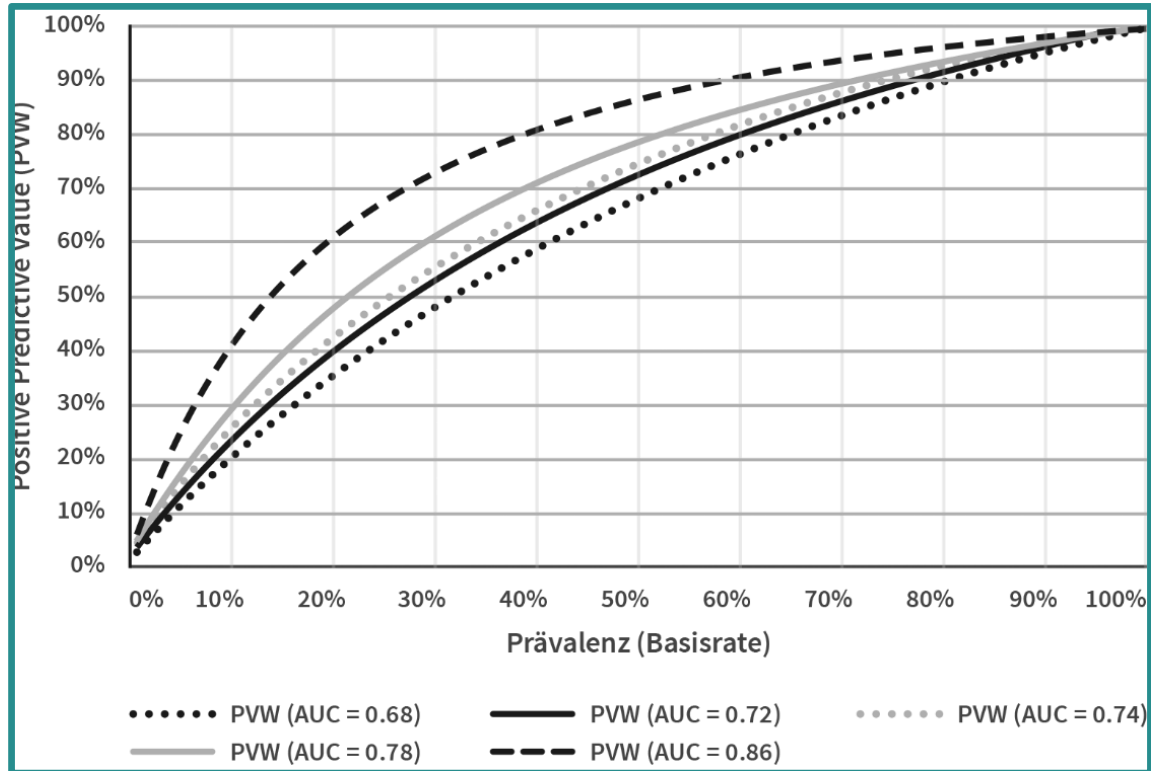
^c Division of Prison Health, Geneva University Hospitals and University of Geneva, Thônex, Switzerland

^d Forensic Psychiatric Hospital, Psychiatric Hospital of the University of Basel, Switzerland

Post-Release Recidivism Rates among Violent Offenders (FSO 2021)



PPV als Funktion von Basisrate und Trennschärfe



Sicherheit von Algorithmen (Spiegelhalter 1983 und Stead 1994)

	Pharmaka	Algorithmen
Phase 1	Sicherheit: Erste Erprobung bei menschlichen Versuchspersonen.	Digitale Testung: Funktionieren bei vorgegebenen Datensätzen.
Phase 2	Proof-of-concept: Wirksamkeit und Zweckmässigkeit bei einzelnen Patienten.	Laboruntersuchung: Vergleich mit menschlicher Entscheidung.
Phase 3	Randomisierte Kontrollgruppenstudie: Vergleich mit dem Goldstandard.	Feldstudie: Konstrollierte Untersuchung der Auswirkungen.
Phase 4	Nachkontrolle: Untersuchung von Langzeit-Effekten	Routine-Einsatz: Prüfen von Schwierigkeiten.

NORBERT NEDOPIL, JÉRÔME ENDRASS,
ASTRID ROSSEGER, THOMAS WOLF

PROGNOSE: RISIKOEINSCHÄTZUNG IN FORENSISCHER PSYCHIATRIE UND PSYCHOLOGIE

EIN HANDBUCH FÜR DIE PRAXIS

Toolbox-Ansatz
(Nedopil, Endrass, Rosseger &
Wolf (2021))

PABST

Wie können neue
Technologien
berücksichtigt werden?

Wie können Algorithmen in
der forensischen Risiko-
einschätzung angemessen
eingesetzt werden?



Bail Hearing Algorithm

(Kleinberg et al. 2017)

Ergebnis

- Ausgangslage – 10 Millionen “bail hearings” pro Jahr.
- Einziges Kriterium: Rückfallrisiko bzw. Fluchtrisiko.
- Empirische Untersuchung nur über a) Falsch-Negativen und b) Richter-Rating möglich (Falsch-Positive nicht untersuchbar).
- Ergebnisse: a) 48.5% der Hochgefährlichen (1er Perzentile) werden entlassen und b) kein Zusammenhang zwischen Richterstrenge und Rückfallquote.
- Interpretation: a) Mit der Hälfte der Inhaftierungsquote könnte eine gleich hohe Rückfallrate sicher gestellt werden bzw. b) mit der gleichen Inhaftierungsquote kann eine 75% höhere Rückfallreduktion erzielt werden.
- Nebenergebnisse: a) Mit einem algorithmengestützten Verfahren werden Minoritäten weniger diskriminiert – bei geringerer Rückfallrate, geringere Inhaftierungsquote. b) Richter schneiden schlechter als ihr paramorphes Modell ab.





Effectiveness of a risk–need–responsivity-based treatment program for violent and sexual offenders: Results of a retrospective, quasi-experimental study

Katharina Seewald¹, Astrid Rossegger^{1,2}, Juliane Gerth^{1,2}, Frank Urbaniok^{1,2}, Gary Phillips³ and Jérôme Endrass^{1,2*} 

¹Department of Psychology, University of Konstanz, Germany

²Department of Mental Health Services, Office of Corrections, Canton of Zurich, Switzerland

³Gary S. Phillips, MAS Center for Biostatistics, Department of Biomedical Informatics, The Ohio State University College of Medicine, Columbus, Ohio, USA

Untersuchungspopulation (Katamnese 7.9 Jahre)



N=171

**6.4 Jahre Katamnese
4.5 Jahre Therapie**

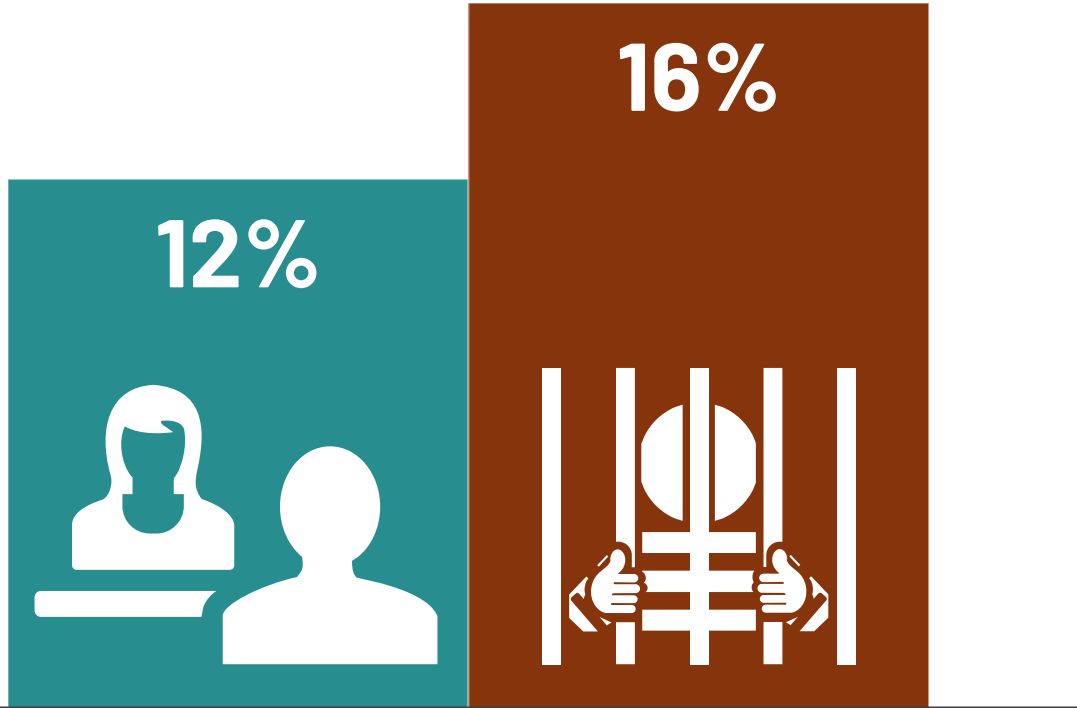


N=241

8.5 Jahre Katamnese



**Kontrolle RF
2013**



The effects of sexual offender treatment on recidivism: an international meta-analysis of sound quality evaluations

Martin Schmucker¹ · Friedrich Lösel^{1,2}

Effectiveness of a risk–need–responsivity-based treatment program for violent and sexual offenders: Results of a retrospective, quasi-experimental study

Katharina Seewald¹, Astrid Rossegger^{1,2}, Juliane Gerth^{1,2}, Frank Urbaniok^{1,2}, Gary Phillips³ and Jérôme Endrass^{1,2*} 

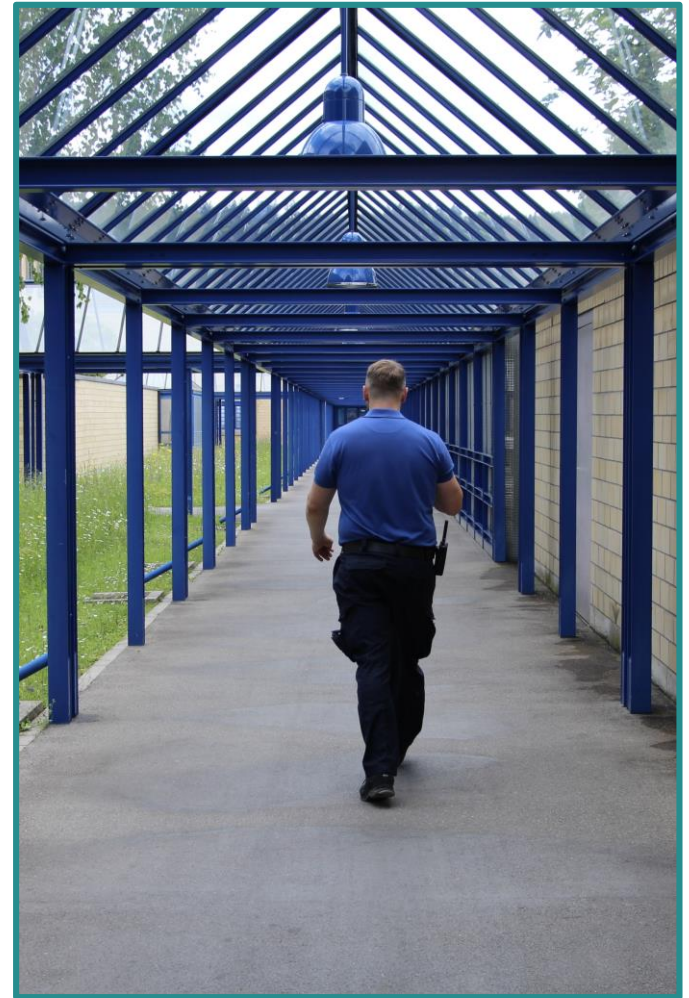
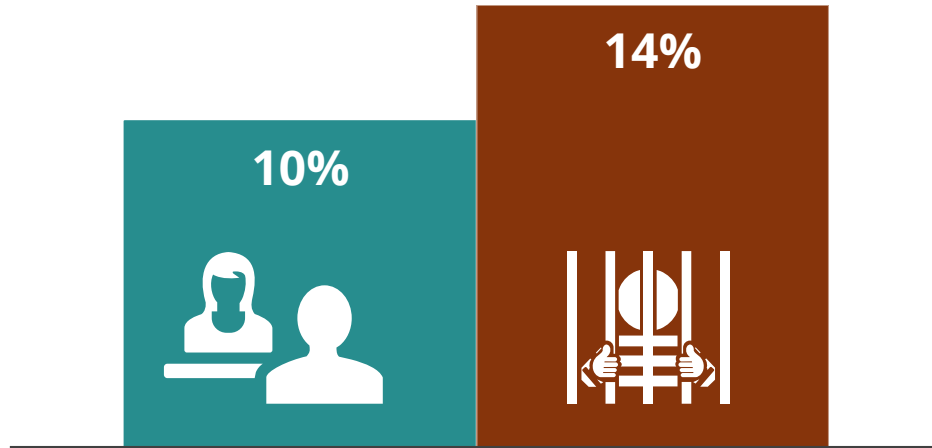
¹Department of Psychology, University of Konstanz, Germany

²Department of Mental Health Services, Office of Corrections, Canton of Zurich, Switzerland

³Gary S. Phillips, MAS Center for Biostatistics, Department of Biomedical Informatics, The Ohio State University College of Medicine, Columbus, Ohio, USA



Es müssen 25 Personen behandelt werden, um einen Rückfall mit einem schweren Gewalt- oder Sexualdelikt zu verhindern



Effectiveness of a risk–need–responsivity-based treatment program for violent and sexual offenders: Results of a retrospective, quasi-experimental study

Katharina Seewald¹, Astrid Rossegger^{1,2}, Juliane Gerth^{1,2}, Frank Urbaniok^{1,2}, Gary Phillips³ and Jérôme Endrass^{1,2*}

¹Department of Psychology, University of Konstanz, Germany

²Department of Mental Health Services, Office of Corrections, Canton of Zurich, Switzerland

³Gary S. Phillips, MAS Center for Biostatistics, Department of Biomedical Informatics, The Ohio State University College of Medicine, Columbus, Ohio, USA

Table 4. Results of Cox proportional hazard regression (treatment vs. control)

	HR	SE	t	p-Value > t	95% CI
Univariate model (non-imputed data set)					
Treatment	0.847	.24	−0.60	.552	0.49–1.46
Multivariable model (multiple imputed data sets)					
Treatment	0.948	.27	−0.19	.853	0.54–1.67
Criminal record with violent or sexual offence	3.265	.99	3.92	<.001	1.81–5.90
Age at start of post-release follow-up	0.942	.02	−3.66	<.001	0.91–0.97
VRAG sum score	1.031	.02	1.91	.056	1.00–1.06

Note. HR = hazard ratio; SE = standard error; CI = confidence interval; VRAG = Violence Risk Appraisal Guide.

Multivariable Cox Regression – (Seewald et al. 2017)

Ergebnis

- Therapie senkt (nicht signifikant) das Rückfallrisiko um 5%.
- Eine einschlägige Vorstrafe verdreifacht das Rückfallrisiko.
- Pro Punktwert auf dem aktuarischen Instrument VRAG steigt das Rückfallrisiko um 3%.
- Die Reifung (Lebensalter) reduziert das Rückfallrisiko um 6% - pro zusätzlichem Lebensjahr.

Alterseffekt – (Seewald et al. 2017)

Ergebnis

- 20-jähriger Straftäter: Therapie ist entscheidend
- 40-jähriger Straftäter: Reifung ist 20-mal wichtiger als Therapie
- 60-jähriger Straftäter: Reifung ist wichtiger als Therapie und Vorstrafe!



ALGORITHMEN IN DER PRAXIS

- (Spiegelhalter 2020)

Fragen von Betroffenen

- Passt mein Fall in den Kompetenzbereich des Algorithmus?
- Wie begründet man die Zuteilung (in den Kompetenzbereich)?
- Was wäre rausgekommen, wenn andere Parameter verwendet worden wären?
- Gab es Informationen, die matchentscheidend waren?
- Wie hoch ist das Ausmass der Ungewissheit?

Voraussetzungen für den Einsatz

- Funktioniert der Algorithmus in unterschiedlichen Kontexten?
- Gibt es ein einfacheres Modell, das gleich gut ist?
- Kann ich die Funktionsweise jeder/jedem problemlos erklären?
- Kann ich den Betroffenen ihr Ergebnis leicht erklären?
- Ist erkennbar, wenn das Ergebnis nicht belastbar ist und wird die Ungewissheit berücksichtigt?
- Wird der Algorithmus richtig und mit der angemessenen Skepsis eingesetzt?
- Ist der Algorithmus in der Praxis nachweisbar hilfreich?