

**Risikoorientierte Interventionen bei Gewalt- und
Sexualstraftätern:
Voraussetzungen, Konzepte und Evaluation**

Dissertation

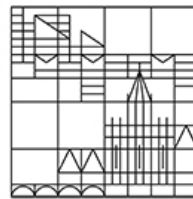
zur Erlangung des akademischen Grades der Doktorin der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von

Katharina Seewald

an der

Universität
Konstanz



Mathematisch-Naturwissenschaftliche Sektion
Fachbereich Psychologie

Tag der mündlichen Prüfung: 02. Oktober 2015

1. Referent: apl. Prof. Dr. Jérôme Endrass

2. Referent: Prof. Dr. Thomas Elbert

DANKSAGUNG

Ein besonderes Dankeschön geht an Jérôme Endrass für die Übernahme der Betreuung dieser Arbeit und seine multi-kontextuelle Unterstützung in den vergangenen Jahren. Danke auch an Thomas Elbert und Frank Urbaniok für ihre Bereitschaft, Teil der Prüfungskommission zu werden.

Herzlich danken möchte ich meinen (ehemaligen) wissenschaftlichen Kolleginnen und Kollegen für den intensiven Austausch während der gemeinsamen Zeit in der Arbeitsgruppe Forensische Psychologie an der Universität Konstanz und im Psychiatrisch-Psychologischen Dienst im Amt für Justizvollzug in Zürich:

Astrid Rossegger, Elisabeth Bauch, Diana Fries, Juliane Gerth, Gilda Giebel, Cornel Gmür, Franziska Loock, Frauke Rösch, Simone Schlotthauer und Elena Yundina;

sowie Anna-Lena Braun vom Lehrstuhl für Kriminologie, Jugendstrafrecht und Strafvollzug der Universität Gießen für das offene Ohr bei kriminologischen Fragestellungen und Seena Fazel vom Department of Psychiatry der University of Oxford, UK für die Bestärkung im Vorhaben dieser wissenschaftlichen Qualifikation.

Ein großes Dankeschön geht schließlich an meine Familie und an meinen Freundeskreis für ihre Unterstützung jeglicher Art.

INHALTSVERZEICHNIS

VORBEMERKUNG.....	VIII
ZUSAMMENFASSUNG.....	IX
1 EINFÜHRUNG.....	1
1.1 Psychische Störungen im forensischen Kontext.....	1
1.1.1 Epidemiologie psychischer Störungen in der forensischen Population.....	1
1.1.2 Psychische Störungen und Gefährlichkeit.....	6
1.2 Valide Risikobeurteilungen als Voraussetzung für risikoorientierte Interventionen.....	10
1.2.1 Methoden der Risikobeurteilung	10
1.2.2 Validierung von Risikobeurteilungen.....	14
1.3 Konzeptuelle Annahmen eines risikoorientierten Behandlungsrahmens	17
1.3.1 Risikoorientierung in forensischen Interventionen.....	17
1.3.2 Interventionen bei jugendlichen Straftätern.....	20
1.4 Evaluation der Wirksamkeit forensischer Interventionen	22
1.4.1 Randomized Controlled Trials in der Forensik – der Goldstandard?.....	22
1.4.2 Quasi-experimentelle Alternativen	26
1.5 Ausblick.....	31
2 VERÖFFENTLICHUNGEN UND MANUSKRIPTE.....	34
2.1 Fazel S, Seewald K (2012) Severe mental illness in 33,588 prisoners worldwide: Systematic review and meta-regression analysis. <i>The British Journal of Psychiatry</i> 200(5), 364-373.....	34
2.1.1 Abstract	34
2.1.2 Introduction.....	35
2.1.3 Methods	36
2.1.4 Results.....	40
2.1.5 Discussion.....	50
2.1.6 Funding & Acknowledgements	54

2.2 Seewald K, Rossegger A, Endrass J (2015). Risikoeinschätzungen bei Sexualstraftätern im institutionellen Bereich. In JM Fegert & M Wolff (Eds.) <i>Kompendium „Sexueller Missbrauch in Institutionen“: Entstehungsbedingungen, Prävention und Intervention</i> (pp 561-573). Weinheim: Verlagsgruppe Beltz.	57
2.2.1 Risk-Assessment bei Gewalt- und Sexualstraftätern	57
2.2.2 Sexuelle Gewalt im Strafvollzug.....	63
2.2.3 Sexuelle Übergriffe gegen Kinder und Jugendliche im institutionellen Setting.....	66
2.2.4 Fazit.....	68
2.3 Seewald K, Rossegger A, Urbaniok F, Endrass J (under review). Assessing the risk of intimate partner violence: Expert evaluations versus the Ontario Domestic Assault Risk Assessment. <i>International Journal of Offender Therapy and Comparative Criminology</i>	69
2.3.1 Abstract	69
2.3.2 Background.....	69
2.3.3 Methods	72
2.3.4 Results.....	76
2.3.5 Discussion.....	80
2.3.6 Acknowledgements	83
2.4 Seewald K, Yundina A, Endrass J (in press). Juvenile Corrections. In Kerley KR, Copes H, De Li S, Lane J, Sharp SF (Eds.) <i>The Encyclopedia of Corrections</i>. Wiley Blackwell.	84
2.4.1 Abstract	84
2.4.2 Interrupting criminal careers with correctional interventions	84
2.4.3 Correctional interventions with juvenile offenders.....	85
2.4.4 Intervention programs with no or potentially harmful effects.....	88
2.5 Seewald K, Rossegger A, Endrass J (manuscript ready for submission). Examining the effectiveness of an offense-focused treatment program in violent and sexual offenders using propensity score estimation procedures and cox regression model	91
2.5.1 Abstract	91
2.5.2 Introduction.....	92
2.5.3 Methods	96
2.5.4 Results.....	104
2.5.5 Discussion.....	112
EIGENLEISTUNG.....	116
LITERATURVERZEICHNIS	117

TABELLENVERZEICHNIS

Table 1: Pooled prevalences for psychosis and major depression in prisoners.....	44
Table 2: Metaregression analyses of sources of heterogeneity in the prevalence of psychosis and major depression in prisoners.....	44
<i>Supplementary Table 1:</i> Update: Details of included studies since 2001	55
<i>Supplementary Table 2:</i> Reported comorbidity in prisoners.....	56
Table 3: Item content of the Ontario Domestic Risk Assessment (ODARA)	74
Table 4: Differences in prematched groups on key characteristics	100
Table 5: Stepwise binary logistic regression modeling of group assignment to estimate propensity score.....	102
Table 6: Descriptive study characteristics by group.....	105
Table 7: Stepwise binary logistic regression modeling of the outcome criterion for regression adjustment.....	107
Table 8: Treatment effects assessed with different procedures based on estimation of propensity score.....	108
Table 9: Differences in characteristics predicting PS	108
Table 10: Post-estimation Survival Analysis: Cox Regression Results	111

ABBILDUNGSVERZEICHNIS

Figure 1: Flow diagram showing the different steps in searching of relevant publications (2001-2010).....	37
Figure 2: Meta-analysis of the prevalence of psychotic illnesses in prisoners by country group (low-middle income vs. high income).	45
Figure 3: Prevalence of psychotic illness in prisoners over time	46
Figure 4: Meta-analysis of the prevalence of major depression in prisoners by country group (low-middle income vs. high income).	48
Figure 5: Prevalence of major depression in prisoners over time.....	49
Figure 6: Distribution of cases assigned to UCJ nine-point Likert scale of recidivism risk and observed recidivism per scale point.....	78
Figure 7: Distribution of cases assigned to ODARA risk bins.....	78
Figure 8: ODARA violent recidivism norm rates and observed violent recidivism rates per risk bin in the present study.....	79
Figure 9: Observational period and post-release follow-up.....	98
Figure 10: Methodological steps following propensity score estimation.....	104
Figure 11: Overview of effects of offense-focused treatment.....	108
Figure 12: Post-estimation Survival Analysis: Kaplan-Meier survival function	110
Figure 13: Post-estimation Survival Analysis: Kernel-smoothed hazard function.....	110
Figure 14: Post-estimation Survival Analysis: Schoenfeld residuals.....	110

VORBEMERKUNG

Die Publikationen und Manuskripte der vorliegenden Dissertationsschrift sind in unterschiedlichen Kontexten der Tätigkeit der Autorin in der forensisch-psychiatrisch/psychologischen Forschung entstanden. Die Eingangsarbeit als meta-analytische Betrachtung psychiatrischer Morbidität im Strafvollzug wurde während eines Forschungsaufenthaltes im Department of Psychiatry an der Oxford University, UK unter Supervision von Prof. Dr. Seena Fazel erarbeitet.

Die übrigen Publikationen und Manuskripte entstanden im Rahmen der wissenschaftlichen Tätigkeit der Autorin im Bereich Evaluation und Entwicklung des Psychiatrisch-Psychologischen Dienstes im Amt für Justizvollzug, Kanton Zürich, Schweiz unter Supervision von Prof. Dr. Jérôme Endrass, stv. Hauptabteilungsleiter des Dienstes und Leiter der Arbeitsgruppe Forensische Psychologie an der Universität Konstanz, sowie in Zusammenarbeit mit den jeweiligen Co-AutorInnen.

ZUSAMMENFASSUNG

Für die erfolgreiche Rehabilitation von Gewalt- und Sexualstraftätern gelten risikoorientierte Interventionen als „best practice“. Voraussetzung dafür, dass sich ein Behandlungsprogramm am Rückfallrisiko eines Täters orientieren kann, ist eine valide Risikobeurteilung. Wenn auch, wie in der ersten Publikation dieser Dissertationsschrift verdeutlicht wird, psychische Störungen in der forensischen Population deutlich häufiger repräsentiert sind als in der Allgemeinbevölkerung, spielen diese für eine solche Risikobeurteilung eine untergeordnete Rolle. Sie bilden allerdings eine wichtige Grundlage für eine gelungene Diversion psychisch kranker Straftäter in ein adäquates forensisches Setting, das die nötigen Rahmenbedingungen für eine erfolgreiche Behandlung schaffen kann. Psychiatrische Diagnosen sollten jedoch kein zentrales Element für eine Gefährlichkeitsbeurteilung darstellen. Vielmehr muss der Gefährlichkeitsbegriff unabhängig vom Krankheitsbegriff und stattdessen durch eine profunde Beurteilung des Rückfallrisikos operationalisiert werden. Verschiedene Methoden des Risk-Assessments werden im Rahmen der zweiten Publikation dieser Arbeit zusammengefasst und für die Anwendung bei Sexualstraftätern im institutionellen Bereich erläutert. Welche Methoden der Risikobeurteilungen in welchen Kontexten zu favorisieren sind, ist seit mehreren Jahrzehnten substantieller Bestandteil der forensischen Forschung. Grundsätzlich unterscheiden sich die verschiedenen Methoden im Grad ihrer Strukturierung und Standardisierung betreffend der Identifikation von Risikofaktoren, deren Messung sowie Kombination und hinsichtlich des finalen Beurteilungsergebnisses. Im Rahmen von Risikoeinschätzungen bei häuslicher Gewalt stellt die dritte Arbeit der vorliegenden Dissertationsschrift zwei dieser methodischen Ansätze gegenüber. Das klinische Urteil, dessen Grad an Strukturierung am geringsten ist und ein aktuarisches Risk-Assessment Instrument (*Ontario Domestic Assault Risk Assessment* [ODARA]), das den höchsten Grad an Strukturierung und Standardisierung, bis hin zur Angabe spezifischer Rückfallwahrscheinlichkeiten für jedes mögliche Ergebnis aufweist. Die Resultate stehen im Einklang mit bisherigen Ergebnissen zum Vergleich der Trennschärfe der Vorhersage von Rückfälligkeit zwischen den unterschiedlichen Methoden der Risikobeurteilung. Die Überlegenheit der mechanischen Urteilsbildung, sowie die professionsunabhängige

Anwendbarkeit des ODARA werden verdeutlicht. Es muss allerdings diskutiert werden, in welchem Maße sich unter Umständen ein mediiender Einfluss wirksamer Interventionen auf die Genauigkeit der klinischen Vorhersage auswirkt.

Forensische Interventionen gelten dann als wirksam, wenn sich ein reduzierender Effekt auf die Rückfälligkeitsraten nachweisen lässt. Welche Formen von Interventionen bei delinquenten Jugendlichen die beste Wirksamkeit aufweisen, fasst die vierte Publikation der vorliegenden Dissertationsschrift zusammen. In dieser Übersichtsarbeit wird deutlich, dass es vor allem multisystemische und multiprofessionelle Programme sind, die eine Reduzierung der Rückfälligkeit jugendlicher Straftäter zur Folge haben und so einer kriminellen Karriere entgegen wirken können. Dies im Gegensatz zu strafenden und/oder abschreckenden Maßnahmen, die entweder keinerlei oder einen negativen Effekt im Rahmen einer Erhöhung von Verurteilungsraten aufweisen.

Bei erwachsenen Gewalt- und Sexualstraftätern erweisen sich risikoorientierte, dem „Risk-Needs-Responsivity (RNR)“-Modell folgende Interventionen als bislang wirksamste Form der Straftäterbehandlung. Diese orientieren sich in ihrer Intensität am Risikolevel des Straftäters (*risk*) und identifizieren sogenannte „kriminogene Bedürfnisse“ (*needs*), die deliktrelevanten Charakter haben und mit der Behandlung adressiert werden sollen. Zuletzt richtet sich die Intervention nach der individuellen Erreichbarkeit und Ansprechbarkeit im Rahmen der Fähigkeiten des Straftäters (*responsivity*). Die deliktpräventiven Therapien, die im Psychiatrisch-Psychologischen Dienst im Kanton Zürich durchgeführt werden, orientieren sich an diesem Modell. Das letzte Manuskript der vorliegenden Dissertationsschrift berichtet Ergebnisse der Evaluation der Wirksamkeit dieser Therapien. Nachdem sich randomisierte Kontrollgruppenstudien zur Untersuchung der Wirksamkeit von Interventionen als im forensischen Setting besonders schwer umsetzbar darstellen, bedient sich die Studie eines quasi-experimentellen Designs und findet statistische Auswege, Effektverzerrungen aufgrund der Verletzung der randomisierten Gruppenzuteilung zu kontrollieren. Eine Kohorte von Gewalt- und Sexualstraftätern mit gerichtlich angeordneter Therapie wurde mit einer unbehandelten Kontrollgruppe aus einer Stichtagserhebung im Kanton Zürich im Hinblick auf Unterschiede in der Legalbewährung verglichen. Dabei zeigte sich ein robuster Behandlungseffekt in der Interventionsgruppe, der für die Effektivität des untersuchten Behandlungsansatzes spricht.

1 EINFÜHRUNG

1.1 Psychische Störungen im forensischen Kontext

Die forensische Relevanz psychischer Störungen lässt sich unter zwei Gesichtspunkten beleuchten: Zum einen im Hinblick auf die Epidemiologie psychischer Störungen in der forensischen Population, sowohl im Gefängnis- als auch im stationär-psychiatrischen Kontext und zum anderen in Bezug auf die Rolle psychischer Störungen bei Risikobeurteilungen und der Implementierung von an diesem Risiko orientierten Managementstrategien.

1.1.1 Epidemiologie psychischer Störungen in der forensischen Population

Epidemiologische Aspekte psychischer Störungen im forensischen Kontext umfassen im Wesentlichen zwei Perspektiven, die Rückschlüsse auf ein besonderes Behandlungsbedürfnis dieser Population ermöglichen: Den generellen Vergleich mit Prävalenzraten, wie sie in der Allgemeinbevölkerung auftreten einerseits, und die Betrachtung spezifischer epidemiologischer Verteilungen in Abhängigkeit des Delikttyps andererseits.

Im Vergleich zur Allgemeinbevölkerung finden sich Diagnosen von Persönlichkeitsstörungen in der forensischen Population deutlich häufiger wieder. Schätzungen der Weltgesundheitsorganisation gehen von einer Prävalenz von 6.1% von Persönlichkeitsstörungen (nach DSM-IV-Kriterien) in der Allgemeinbevölkerung aus (Huang et al., 2009). Die Gesamtprävalenz von Persönlichkeitsstörungen in der Gefängnispopulation liegt nach meta-analytischen Befunden indes mit 65.0% bei männlichen und 42.0% bei weiblichen Straftätern/Straftäterinnen weit höher (Fazel & Danesh, 2002). Die Prävalenz der spezifischen anti- oder dissozialen Persönlichkeitsstörung in der Gefängnispopulation wird auf 47.0% bei männlichen und 21.0% bei weiblichen Straftätern/Straftäterinnen geschätzt (Fazel & Danesh, 2002). In der umgekehrten Betrachtung konnte gezeigt werden, dass die Chancen einer (initialen, nicht-rückfälligen) gewalttätigen Handlung bei Personen mit der Diagnose einer Persönlichkeitsstörung in der Allgemeinbevölkerung um ein dreifaches höher sind als in der nicht mit einer Persönlichkeitsstörung diagnostizierten allgemeinen

Vergleichsgruppe. Die spezifische Diagnose einer antisozialen Persönlichkeitsstörung erhöht diese Chancen wiederum auf das 12.8-fache (Yu, Geddes, & Fazel, 2012).

Neben anti- oder dissozialen Persönlichkeitsstörungen, sind auch psychotische Störungen, sowie Störungen aus dem schizophrenen Formenkreis mit einer deutlich höheren Prävalenz in der forensischen Population vertreten. Für die allgemeine (männliche) Gefängnispopulation beschreiben meta-analytische Ergebnisse eine Prävalenz von 3.7% psychotischer Störungen (Fazel & Seewald, 2012). Morgan, Fisher, Duan, Mandracchia, und Murray (2010) berichten von 22.2% mit Schizophrenie diagnostizierten männlichen Tätern in einer psychiatrischen Einheit im Strafvollzug und Habermeyer, Wolff, Gillner, Strohm, und Kutscher (2010) veröffentlichen nahezu identische Zahlen (22.3%) für Erkrankungen aus dem schizophrenen Formenkreis im Rahmen der Unterbringung in einer deutschen Maßregelvollzugsklinik. In Kohortenstudien mit schizophrenen Patienten und gesunden Kontrollen zeigte sich, dass die Chancen auf Verurteilungen aufgrund von Gewalttaten in der schizophren erkrankten Gruppe 2.3-fach höher sind als in der nicht erkrankten Gruppe. Bei der Betrachtung der Komorbidität mit Substanzmissbrauch/-abhängigkeit ergab sich, dass diese einen deutlichen Effekt modifizierenden Einfluss hat: Odds Ratios (ORs) von 4.4 bei schizophren erkrankten Personen mit komorbider Substanzmissbrauchsdiagnose werden im Vergleich zu ORs von 1.2 für schizophren erkrankte Personen ohne diese Komorbidität berichtet. Dieses Ergebnis zeigt, dass die Erhöhung der Chancen auf eine Verurteilung für ein Gewaltdelikt allein aufgrund einer schizophrenen Erkrankung an sich gering ist (Fazel, Långström, Hjern, Grann, & Lichtenstein, 2009).

Im Rahmen affektiver Erkrankungen sind die Prävalenzen in der Gefängnispopulation nur minimal erhöht. Ausgehend von WHO-Erhebungen werden Lebenszeitprävalenzen in der Allgemeinbevölkerung für Depressionen zwischen 11.1%-14.6% (Kessler & Bromet, 2013) und für Störungen aus dem bipolaren Spektrum von 2.4% berichtet (Merikangas, Jin, He, & et al., 2011). In der forensischen Population sind depressive Erkrankungen mit 10.2% bei männlichen Gefängnisinsassen und mit 14.1% bei weiblichen Gefängnisinsassinnen vertreten (Fazel & Seewald, 2012). Eine höhere Prävalenz von 21.6% wird hingegen aus einer psychiatrischen Einheit im Strafvollzug berichtet (Morgan et al., 2010). Bipolare Störungen finden sich mit einer Häufigkeit von 2.6% in der allgemeinen Gefängnispopulation (Baillargeon et al., 2009) und mit einer

Häufigkeit von 17.0% im forensisch-psychiatrischen Kontext wieder (Morgan et al., 2010).

Eine kürzlich erschienene Übersichtsarbeit zum Verhältnis depressiver Erkrankungen und delinquentem Verhalten kommt zu dem Schluss, dass Personen, die an depressiven Symptomen leiden, mit einer dreifach erhöhten Wahrscheinlichkeit für gewalttätige (wie auch für nicht-gewalttätige) Straftaten verurteilt werden (Fazel et al., 2015). Wird für familiäre Störvariablen durch Vergleich mit (Halb-)Geschwistern, wie auch vorgeschichtlichen Substanzmissbrauch und frühere Kriminalität kontrolliert, bleibt eine signifikante Erhöhung der Chancen in der depressiv erkrankten Gruppe (OR=1.2 für Halbgeschwister; OR=1.5 für Geschwister) bestehen. Diskutiert und weiterhin empirisch untersucht werden muss in diesem Zusammenhang der medierende Effekt von (unter Umständen subdiagnostisch relevanten) Persönlichkeitsstörungen und dem Einfluss antidepressiver Medikation (Fazel et al., 2015). Personen mit der Diagnose einer bipolaren Störung weisen im Vergleich zur nicht erkrankten Allgemeinbevölkerung um den Faktor 2.3 erhöhtes Risiko auf, Gewalttaten zu begehen (Fazel, Lichtenstein, Grann, Goodwin, & Langstrom, 2010). In der Gruppierung nach komorbidem Substanzmissbrauch zeigt sich allerdings auch in dieser klinischen Subgruppe der deutliche Einfluss der Komorbidität mit dieser Störung: Bipolar Erkrankte ohne Komorbidität mit Substanzmissbrauch weisen ein um den Faktor 1.3 erhöhtes Risiko für Gewalttaten auf, während in der bipolaren Gruppe mit Komorbidität das dieses Risiko auf den Faktor 6.4 erhöht ist (Fazel et al., 2010).

Besondere epidemiologische Verteilungen in Abhängigkeit des Delikttyps

Kenntnisse über das Vorliegen bestimmter psychischer Störungen bei wiederum bestimmten Formen von Delinquenz sind im Rahmen des phänomenologischen Verständnisses, aber auch im Rahmen der Identifikation von Risikofaktoren oder protektiven Faktoren sinnvoll. Nach einer schwedischen Kohortenstudie von Fazel und Grann (2004) leidet etwa jeder vierte Täter eines Tötungsdelikts unter Substanzmissbrauch/-abhängigkeit (24.2%) sowie jeder fünfte unter einer psychotischen oder schizophrenen Erkrankung (19.1%). Persönlichkeitsstörungen sind mit 14.0% auch in dieser Subgruppe prävalenter als in der Allgemeinpopulation, affektive Störungen sind indes deutlich seltener: Bipolare Störungen lagen bei 3.1% und Angststörungen 1.7% der untersuchten Täter vor.

Im Kontext von häuslicher Gewalt berichten Shorey, Febres, Brasfield, und Stuart (2012) ebenfalls von einer besonderen Verteilung psychischer Belastungen unter den Tätern. Die höchst prävalente Diagnose in diesem Zusammenhang stellte Alkoholmissbrauch/-abhängigkeit mit 39.1% dar. Etwa jeder vierte Täter häuslicher Gewalt (26.2%) leidet zudem unter einer Posttraumatischen Belastungsstörung, sowie jeweils jeder fünfte (19.9%; 19.5%) an depressiven Erkrankungen oder generalisierter Angststörung.

Auch in der Gruppe der Sexualstraftäter (unabhängig von der Art des Sexualdelikts) finden sich laut den Ergebnissen einer „case-control“-Studie im Vergleich zur Allgemeinpopulation erhöhte Wahrscheinlichkeiten für psychische Erkrankungen. Diese Täter weisen mit einer um den Faktor 4.8 erhöhten Wahrscheinlichkeit eine schizophrene Erkrankung auf, sowie mit einer um den Faktor 3.4 erhöhten Wahrscheinlichkeit eine bipolare Störung. Persönlichkeitsstörungen sind fast um ein 30-faches (OR: 29.8) wahrscheinlicher als in der nicht-delinquenten Allgemeinpopulation, Alkohol- und Drogenmissbrauch jeweils um ein vierfaches (Fazel, Sjostedt, Langstrom, & Grann, 2007).

Auf dem Hintergrund dieser erhöhten Prävalenzraten wird deutlich, warum auf eine diagnostische Einschätzung eines Gewalt- und/oder Sexualstraftäters nicht verzichtet werden kann, auch wenn etwaige Diagnosen im Rahmen von Risikobeurteilungen nur eine untergeordnete Rolle spielen (Kapitel 1.1.2). Die psychiatrische Diagnostik ist in der forensischen Praxis zwingend notwendig für die Klärung der Frage nach der Unterbringung des Täters. Es gilt dabei zu untersuchen, ob eine psychische Störung – ganz unabhängig von ihrem Erklärungswert für das Tatverhalten oder ihres inkrementellen Beitrags zur Risikoprognose – so schwerwiegend ist, dass der Täter engere psychiatrische und psychotherapeutische Betreuung oder flankierende Maßnahmen zur Unterstützung benötigt, die im Rahmen der Straftat nicht umgesetzt werden können. Eine gelungene Diversion dieser Täter kann nicht nur das Risiko von (gewalttätigen) Zwischenfällen im Gefängnis reduzieren (Schenk & Fremouw, 2012), sondern auch die erhöhten Suizidraten psychisch kranker Gefängnisinsassinnen und -insassen eindämmen (Baillargeon et al., 2009). Der forensischen Praxis (im Gefängnis- wie auch im stationär-psychiatrischen Setting) kommt somit die Aufgabe zu, relevante Störungsbilder zu identifizieren, sowie passende Rahmenbedingungen in einer Form zu gestalten, die eine erfolgreiche Behandlung des psychiatrischen Störungsbildes bei gleichzeitigem Management des Rückfallrisikos gewährleistet.

Fazel S, Seewald K (2012) Severe mental illness in 33,588 prisoners worldwide: systematic review and meta-regression analysis. *British Journal of Psychiatry* 200(5), 364-373.

Die Meta-Analyse, die den Einstieg in die vorliegende Dissertationschrift bildet, diente der Aktualisierung einer wichtigen und außerordentlich gut rezipierten Arbeit von Fazel & Danesh zur psychiatrischen Epidemiologie bei Gefängnisinsassinnen und -insassen aus dem Jahre 2002. Eine Dekade nach deren Erscheinen wurden im Hinblick auf Trends in der psychiatrischen Belastung von Strafgefangenen neue Veröffentlichungen und empirische Befunde identifiziert und in Zusammenhang zu jenen der ersten Veröffentlichung gestellt. Zusätzlich wurde nach dem Klassifizierungsmodell der Weltbank untersucht, inwiefern sich Prävalenzen psychotischer und depressiver Erkrankungen in Gefängnispopulationen zwischen sogenannten „high income“ Ländern und „low-middle income“ Ländern unterscheiden. Dieser Fokus ergab sich aus der Tatsache heraus, dass „low-middle income“ Länder einen deutlich höheren Anteil an weltweit inhaftierten Personen abbilden (Walmsley, 2009).

Die Identifikation relevanter Publikationen erfolgte nach PRISMA-Kriterien (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009). Einschlusskriterium war die Diagnosestellung der relevanten Störungen nach ICD-10- oder DSM-IV-Kriterien durch ein psychiatrisches Interview oder ein validiertes diagnostisches Instrument. Berücksichtigt wurden sowohl Stichproben von Männern und Frauen aus Untersuchungs- als auch aus Strafhaft. Allgemeine Prävalenzraten beliefen sich auf 3.7% für psychotische Störungen und 11.4% für depressive Erkrankungen, ohne signifikante Unterschiede zwischen den Geschlechtern. Wie erwartet konnten signifikant höhere Prävalenzen in „low-middle income“ Ländern festgestellt werden: Psychotische Störungen 3.5% vs. 5.5%; Depressive Erkrankungen 10.0% vs. 22.5%. Ein Anstieg an psychotischen und depressiven Erkrankungen in der Gefängnispopulation über die Zeit hinweg ließ sich – entgegen der initialen Annahme – weltweit zwar nicht, in der US-amerikanischen Gefängnispopulation allerdings doch finden.

Die Ergebnisse verdeutlichen zum einen die besondere psychiatrische Belastung inhaftierter Personen im Strafvollzug sowie die Notwendigkeit einer adäquaten Umleitung psychisch kranker Straftäter und Straftäterinnen in andere

Einrichtungen des Justizvollzuges. Zum anderen unterstreichen sie den besonderen Bedarf an Implementierung gefängnispsychiatrischer Versorgungssysteme in „low-middle income“ Ländern, die mit steigenden Inhaftierungszahlen konfrontiert sind.

1.1.2 Psychische Störungen und Gefährlichkeit

Bevor sich allerdings die Gefängnispsychiatrie konkret ihren Herausforderungen stellen muss, sind im strafprozessualen Verlauf Fragestellungen zu klären, die von psychiatrischer Diagnosestellung nicht unabhängig sind. Die strafrechtliche Aufgabe der Suche nach der geeigneten Bestrafung und/oder der geeigneten Unterbringungsart und -dauer, lässt sich im Spektrum schwerer Fälle von Gewalt- und Sexualdelinquenz nicht ohne Einschätzung psychiatrisch/psychologischer Sachverständiger bewältigen. Dies begründet sich darin, dass strafrechtliche Lösungen dieser Problemstellungen gesetzlich an den psychischen Gesundheitszustand der zu begutachtenden Person geknüpft sind. So soll etwa in Fragen zur Exkulpation in Abhängigkeit des „symptomatischen Zusammenhangs“ zwischen dem psychischen Gesundheitszustand und der Anlasstat entschieden werden, welche juristische Würdigung ebensolche Anlasstat erfährt (Van Gemmeren, 2012). Festgehalten wird dabei, dass dieser „symptomatische Zusammenhang“ nicht nur zwischen dem „Zustand“ und der Anlasstat, sondern auch zwischen „Zustand“ und künftiger Gefährlichkeit gegeben sein muss, um (nach deutschem Strafrecht) eine Unterbringung in einer stationären forensisch-psychiatrischen Einrichtung zu rechtfertigen (Van Gemmeren, 2012).

Verknüpfung des Gefährlichkeits- und des Krankheitsbegriffes

Die Einschätzung der Gefährlichkeit wird also maßgebliche Voraussetzung für die Festsetzung der strafrechtlichen Folgen der Tat. Für die Gefährlichkeitsbeurteilung gelten dabei juristisch drei Maßstäbe: (1) die Erheblichkeit der zu erwartenden Taten, (2) die „Symptomtizität“ als „Kausalität der psychischen Störung für Anlasstat wie für befürchtete Taten“, sowie (3) der Grad der Wahrscheinlichkeit der Wiederholung (Pollähne, 2013). Während sich die juristische Profession im Rahmen des ersten Maßstabes bei der Einschätzung der Erheblichkeit der zu erwartenden Taten die Orientierung an aktuelle gesellschaftspolitische Strömungen, Risikobereitschaften und kriminalpolitische Tendenzen verordnet, obliegt die diagnostische Einschätzung der

psychischen Störung (zu 2) dem/der psychiatrisch/psychologischen Sachverständigen. Zeitgleich wird jedoch der juristische Krankheitsbegriff als maßgeblich für letztere Einschätzung beschrieben (Pollähne, 2013). Aufgabe der Praxis der psychiatrisch/psychologischen Sachverständigen ist es folglich, den Krankheitsbegriff der eigenen Disziplin mit der zu beantwortenden juristischen Fragestellung zu vereinbaren. Dabei unterscheidet sich der medizinische Krankheitsbegriff vom strafrechtlichen grundsätzlich. Der strafrechtliche Krankheitsbegriff konstruiert sich vorrangig durch die Überschreitung einer normativ gesetzten Schwelle und sieht das sich daraus determinierende Ausmaß der Funktionseinschränkung im Rahmen der Einsichts- und Steuerungsfähigkeit als qualifizierendes Merkmal für Schuldfähigkeit (Nedopil & Müller, 2012) und somit als den bereits erwähnten „symptomatischen Zusammenhang“ zur Gefährlichkeit (Pollähne, 2013). Der medizinische Krankheitsbegriff hingegen nutzt Ursachen, Verlauf und Therapierbarkeit von Störungsbildern als Definitionskriterien und operationalisierte Klassifikationssysteme (DSM: Saß, Wittchen, Zaudig, & Houben, 2003; ICD: World Health Organization, 2010) zur präzisen und einheitlichen Beschreibung von Symptomen, sowie von deren Konstellationen als Syndrome und Krankheitsbilder (Nedopil & Müller, 2012). Aufgrund dieses definitorischen Unterschiedes der Krankheitsbegriffe beider Disziplinen steht die Praxis forensischer Sachverständiger vor der Herausforderung, den zu begutachtenden Täter in beide definitorischen Räume einzuordnen, oder aber, sich von diesen Krankheitsbegriffen, seien sie strafrechtlich oder medizinisch, zu entfernen und alternative Krankheitsbegriffe zu prägen. Im Versuch der Einordnung in beide Definitionen spricht sich Saß (1985) für ein psychopathologisches Referenzsystem aus, das ausgehend von der forensischen Relevanz psychopathologischer Symptome den strafrechtlichen Krankheitsbegriff zu füllen versucht. Einen alternativen Krankheitsbegriff prägen hingegen Rasch und Konrad (1986). Ihr strukturell-sozialer Krankheitsbegriff beschreibt die forensische Relevanz von Störungen so, dass diese im Sinne ihrer Struktur Krankheitswert haben und gleichzeitig die soziale Kompetenz beeinträchtigen müssen (Nedopil & Müller, 2012). Eine dritte Variante des Umgangs mit den Krankheitsbegriffen im forensischen Kontext ist wiederum die, den Zusammenhang zur einzuschätzenden Gefährlichkeit von der „gesund vs. krank“-Kategorisierung und von psychiatrischen Diagnosestellungen zu entkoppeln und zu einem profunden Risk-Assessment (einer Risikobeurteilung) überzugehen (Urbaniook, 2003). Dies impliziert

auch die Abkehr vom rein medizinischen Charakter forensischer Expertise. Die binäre Klassifizierung in „kranker Täter vs. „gesunder“ Täter erschließt dem psychiatrischen Sachverständigen nur Aussagemöglichkeiten über den (psychiatrisch) „kranken“, nicht über den „gesunden“ Täter. Der/die forensisch-psychiatrische Gutachter/in müsste sich lediglich darauf beschränken, Diagnosen und Therapieprognosen für eben solche „kranken“ Straftäter zu formulieren (Schauer, 2006). Praktisch gesehen kann sich der forensische Gutachter/die forensische Gutachterin nicht ausschließlich auf die durch nosologische Systeme klassifizierbaren Fälle beschränken und ist in der Folge verpflichtet – auch dem juristischen Auftraggeber gegenüber, andere Formen der Erklärung des delinquenten Verhaltens zu wählen, so etwa deliktrelevante Problembereiche, die gleichzeitig Anhaltspunkte für Resozialisierungsstrategien geben (Urbaniok, 2003). Dies impliziert die Forderung nach einem methodisch getrennten Weg für „kranke“ und „gesunde“ Straftäter, da sich die Art der zu treffenden Aussagen im gutachterlichen Prozess zwischen Diagnosen und Therapieprognosen beim „kranken“ im Gegensatz zu Resozialisierungsempfehlungen beim „gesunden“ Täter unterscheidet (Schauer, 2006), während letztere ebenso therapeutische Interventionen oder niederschwellige Managementangebote enthalten können und sollen.

Dass die Diskussionen um Krankheitsbegriffe und die strafrechtliche Würdigung psychischer Störungen immer wieder neu entzündet werden und gleichzeitig hoch aktuell sind, zeigt auch die Einführung des Begriffs der „unsound mind“ (gem. § 5 Abs. 1 S. 2e EMRK) durch die Europäischen Menschenrechtskonvention. Diese Formulierung hält den Begriff der psychischen Störungen flexibel und an gesellschaftliche Veränderungen anpassbar, was sich im Hinblick auf regelmäßige Revisionen internationaler Klassifikationssysteme als äußerst sinnvoll erweist (Höffler & Stadtland, 2012).

Die Rolle psychischer Störungsbilder bei forensischen Risikobeurteilungen

Die forensische Praxis operationalisiert den Begriff der Gefährlichkeit also durch die Beurteilung des Rückfallrisikos nach einem vorausgegangenem Anlassdelikt oder auch durch das Ausführungsrisiko nach ausgesprochener Drohung. Inwiefern psychiatrische Diagnosen für eine solche Risikobeurteilung relevant sein können, kann anhand von Vergleichsstudien mit verschiedenen Kontrollgruppen aufgezeigt werden.

Die Chance eines Rückfalls bei Personen mit einer antisozialen Persönlichkeitsstörung ist etwa verglichen mit psychiatrisch nicht anderweitig erkrankten Straftätern fast doppelt so hoch (OR=1.9), und verglichen mit einer anderweitig psychiatrisch erkrankten Straftätergruppe zweieinhalb mal so hoch (OR=2.5; Yu et al., 2012). Dass schwere psychische Störungen (Achse-1-Störungen, DSM-IV: Saß et al., 2003) allerdings an sich keinen robusten Zusammenhang mit Rückfälligkeit aufweisen, wurde schon mehrfach empirisch nachgewiesen (Bonta, Law, & Hanson, 1998; Elbogen & Johnson, 2009; Grann, Danesh, & Fazel, 2008). Schwere psychische Erkrankungen sind nie alleine Auslöser für (erneutes) delinquentes Verhalten (Elbogen & Johnson, 2009; Fazel, Gulati, Linsell, Geddes, & Grann, 2009; Fazel, Långström, et al., 2009; Kingston, Olver, Harris, Wong, & Bradford, 2015). Während die alleinige Diagnose einer schizophrenen Erkrankung gar einen negativen Zusammenhang zum Rückfallrisiko aufweist (Bonta et al., 1998; Quinsey, Harris, Rice, & Cormier, 2006), gilt die Komorbidität mit Substanzmissbrauch insgesamt als robuster, erhöhender Moderator des Gewalttrisikos bei schweren psychischen Erkrankungen (Achse-1-Störungen) wie auch bei Persönlichkeitsstörungen (Baillargeon et al., 2010; Elbogen & Johnson, 2009; Fazel, Långström, et al., 2009; Van Dorn, Volavka, & Johnson, 2012; Volavka & Swanson, 2010; Walter, Wiesbeck, Dittmann, & Graf, 2011).

Die inkrementelle Trennschärfe wiederum, die durch die Hinzunahme von Diagnosen wie Substanzmissbrauch oder Persönlichkeitsstörungen zusätzlich zu statischen prognostisch relevanten Variablen wie Alter, Geschlecht und kriminelle Vorgeschichte entsteht, ist überraschend gering (Grann et al., 2008). Ein direkter Zusammenhang zwischen schweren psychischen Erkrankungen, wie sie in aktuellen Klassifikationssystemen beschrieben werden und dem (Gewalt-)Risiko bleibt zum einen fraglich und zum anderen kaum generalisierbar (Volavka & Swanson, 2010). Profunden Risikobeurteilungen müssen folglich alternative Konzepte zugrunde liegen. Wenn nun die Abkehr von psychopathologischen Klassifikationssystemen im Rahmen dieser Risikobeurteilungen gefordert wird (Urbaniok, 2003), bleibt jedoch die Frage offen, wie sich dies in der konkreten Praxis der Risikobeurteilung zeigen kann und vor allem wie subdiagnostisch relevante psychopathologische Merkmale einfließen können.

Im Rahmen der mechanischen Urteilsbildung (siehe 1.2.1) wird aktuell in der Weiterentwicklung aktuarischer Instrumente auf Items mit psychiatrischen Diagnosen verzichtet. Eine neue Version des Violence Risk Appraisal Guides (VRAG: Quinsey et al.,

2006) der VRAG-R (Rice, Harris, & Lang, 2013) enthält im Gegensatz zu seiner Vorgängerversion weder die (in der Originalversion als protektiver Faktor einfließende) Diagnose einer Schizophrenie, noch die einer Persönlichkeitsstörung. Stattdessen wurde die Abfrage des Alkoholmissbrauchs um illegale Substanzen erweitert, um diesen Einfluss differenzierter und subdiagnostisch abzubilden (Rice et al., 2013).

Auch in der sogenannten „case formulation“ (siehe 1.2.1), wird zunehmend auf die Verbindung zwischen pathologischen Persönlichkeitsaspekten und Delinquenz, statt auf diagnostisch starre Persönlichkeitsstörungen gesetzt (Logan & Johnstone, 2010). Diese Bildung individueller Delikthypothesen findet auch nach FOTRES (Forensisches Operationalisiertes Therapie-Risiko-Evaluations-System: Urbaniok, 2007) anhand von unter Umständen subdiagnostisch ausgeprägten, prognostisch jedoch relevanten Merkmalen statt (Endrass & Rossegger, 2012). Damit ist ein nosologiefreier, phänomenologischer Zugang zur Täterpersönlichkeit und zum Tatverhalten gewährleistet, der eine spezifische, hypothesengerichtete Intervention mit dem Ziel der Senkung des Rückfallrisikos zur Konsequenz haben kann.

1.2 Valide Risikobeurteilungen als Voraussetzung für risikoorientierte Interventionen

1.2.1 Methoden der Risikobeurteilung

Forensische Risikobeurteilungen können nicht nur den initialen Therapiebedarf, sondern auch Vollzugsverlauf, Lockerungsmaßnahmen und bedingte Entlassungen bestimmen. Man unterscheidet im Wesentlichen drei verschiedene methodische Herangehensweisen, die sich im Ausmaß ihrer Strukturierung unterscheiden: (1) Das intuitive/klinische Urteil, (2) die Anwendung mechanischer Risk-Assessment Instrumente und (3) das strukturierte klinische Urteil.

Beim intuitiven, unstrukturiert-klinischen Urteil („unstructured clinical judgment“ [UCJ]) basiert die Risikobeurteilung auf dem Experten- und Erfahrungswissen der beurteilenden Person. Die Auswahl der zu berücksichtigenden Kriterien, sowie deren Gewichtung folgen keinen standardisierten Vorgaben (Instrumenten oder Checklisten). Es handelt sich um eine ideographische, am Einzelfall orientierte Einschätzung (Murray & Thomson, 2010; Rossegger, Endrass, & Gerth, 2012a). Mechanische Risk-Assessment Instrumente hingegen, sind dadurch gekennzeichnet, dass sie Risikofaktoren enthalten,

die nach gruppenstatistischen Analysen einen prädiktiven Wert für das vorherzusagende Kriterium aufweisen. Die Abbildung der Risikofaktoren im Instrument ist durch geschlossene Antwortkategorien höchst standardisiert. Ebenso erfolgt die Integration der Informationen über die Risikofaktoren nach einem festgelegten Auswertungsalgorithmus, und wird schließlich in eine Wahrscheinlichkeitsaussage überführt (Douglas, Cox, & Webster, 1999; Fazel, Singh, Doll, & Grann, 2012; Rossegger, Endrass, & Gerth, 2012b).

Die Nachteile sowohl der klinischen, als auch der mechanischen Urteilsbildung liegen jedoch gleichsam auf der Hand. Während das klinische Urteil subjektiven Urteilsverzerrungen unterliegt – und zahlreich empirisch belegt die geringste Trennschärfe aufweist (Ægisdóttir, White, Spengler, & Maugherman, 2006; Grove, Zald, Lebow, Snitz, & Nelson, 2000; Meehl, 1954) – sind unter den mechanischen Instrumenten vor allem die aktuarischen Instrumente aufgrund ihrer Unsensibilität gegenüber Spezifika des Einzelfalls (Cooke & Michie, 2010), wie auch gegenüber dynamischer Risikoveränderungen in der Kritik (Sreenivasan, Kirkish, Garrick, Weinberger, & Phenix, 2000). Eine wesentliche Kritik besteht ebenfalls darin, dass die Überführung in Wahrscheinlichkeitsaussagen schlicht für den Einzelfall keinerlei Bedeutung haben kann, da die zugrunde liegenden Normrückfallraten gruppenstatistisch entstanden sind und der Messfehler einer einzelnen Person zu gross sei, um diese sinnvoll einer Rückfallwahrscheinlichkeit zuzuordnen (Hart, Michie, & Cooke, 2007). Die dritte Methode als strukturiert klinisches Urteil („structured professional judgment“[SPJ]) bedient in vielerlei Hinsicht die Kritikpunkte an den vorher genannten Ansätzen. Bei ihrer Anwendung werden zu berücksichtigende Kriterien vorgegeben und standardisiert definiert. Die Gewichtung dieser Kriterien allerdings, die Integration der Informationen, sowie das Ableiten der (nicht mit einer Wahrscheinlichkeitsaussage assoziierten) Risikokategorie aus diesen wird mittels eines klinischen Urteils vorgenommen (Douglas et al., 1999; Fazel et al., 2012).

Geht man davon aus, dass einer Risikobeurteilung vier Stufen zugrunde liegen, nämlich (1) die Identifikation von Risikofaktoren, (2) die Messung der Risikofaktoren, (3) die Kombination der Risikofaktoren und (4) die finale Beurteilung, stellen Skeem und Monahan (2011) die verschiedenen Methoden sehr gelungen gegenüber. Während das unstrukturiert-klinische Urteil über keinerlei strukturierende Komponente, auf keiner der vier Stufen verfügt, werden beim strukturierten klinischen Urteil, am Beispiel des

HCR-20 (Webster, Douglas, Eaves, & Hart, 1997) zumindest die ersten beiden Stufen strukturiert durchgeführt. Mechanische Instrumente, wie der LSI-R (Andrews, Bonta, & Wormith, 2004) strukturieren allerdings die ersten drei Prozessstufen, also die Kombination der Risikofaktoren eingeschlossen und lassen die finale Beurteilung klinisch offen. Andere mechanische Instrumente hingegen sind über alle Stufen hinweg durchstrukturiert (am Beispiel einer aktuarischen Variante, dem VRAG: Quinsey et al., 2006).

Aus der praktischen Perspektive haben die meisten mechanischen Instrumente den Nachteil, überwiegend statische, unveränderbare Merkmale zu erfassen. Dadurch wird eine Verlaufsbeurteilung unmöglich und das erfasste Risikolevel bleibt unsensibel für Reduktionen beispielsweise durch einen Therapieerfolg (Rossegger, Endrass, et al., 2012b). Einige wenige mechanische Instrumente schließen diese Lücke, indem sie auch dynamische Risikofaktoren miteinbeziehen: FOTRES 2.0 (Urbaniok, 2007); LSI-R (Andrews et al., 2004); Stable-2007 (Hanson, Harris, Scott, & Helmus, 2007). Was jedoch auch bei der Berücksichtigung dynamischer Risikofaktoren als Kritikpunkt bleibt, ist, dass diese ebenfalls aus gruppenstatistischen Analysen ausgewählt wurden, die für den spezifischen Einzelfall unter Umständen keinerlei Relevanz haben müssen (Cooke & Michie, 2010; Hart, Sturmey, Logan, & McMurrin, 2011). Mit sogenannten „case formulations“, also individuellen Fall- bzw. Delinquenzhypothesen käme man hingegen der spezifischen Relevanz der abgebildeten Risikokonstellation wieder näher (Hart et al., 2011). Dabei werden in einem ersten Schritt relevante Problembereiche identifiziert, deren Erklärungswert beurteilt und Therapieziele formuliert (Hart et al., 2011; Moore & Drennan, 2013). FOTRES 2.0 (Urbaniok, 2007) folgt bereits diesem Ansatz: Der Deliktmechanismus wird individuell anhand prognostisch relevanter Problembereiche beschrieben, die gleichzeitig als in der Therapie zu adressierende Themen betrachtet werden und somit Ansatzpunkte für ein spezifisches Risikomanagement darstellen (Endrass & Rossegger, 2012). Die Prüfung wissenschaftlicher Gütekriterien stellt jedoch insgesamt für „case formulation“-Ansätze eine Herausforderung dar (Hart et al., 2011) und ist insofern auch für FOTRES 2.0 – trotz vielversprechender erster Ergebnisse (Rossegger et al., 2011) – noch nicht abgeschlossen (Endrass & Rossegger, 2012).

Seewald K, Rossegger A, Endrass J (2015). Risikoeinschätzungen bei Sexualstraftätern im institutionellen Bereich. In JM Fegert & M Wolff (Eds.) Kompendium „Sexueller Missbrauch in Institutionen“: Entstehungsbedingungen, Prävention und Intervention (pp 561-573). Weinheim: Verlagsgruppe Beltz.

Anhand von Risikobeurteilungen bei Sexualstraftätern werden in diesem Buchkapitel die verschiedenen Methoden der Risikobeurteilungen verdeutlicht und im spezifischen Rahmen sexueller Übergriffe in Institutionen gegenübergestellt. Als mechanische Risk-Assessment Instrumente werden der *Static-99* (Hanson & Thornton, 1999), der *Sex Offender Risk Appraisal Guide* (SORAG; Quinsey et al., 2006) und der sowohl bei Gewalt- als auch bei Sexualstraftätern anwendbare *Violence Risk Appraisal Guide* (VRAG; Quinsey et al., 2006) eingeführt und für die statische Beurteilung des Rückfallrisikos für Sexual- und/oder Gewaltdelikte empfohlen. Für eine dynamische Beurteilung eignet sich im Rahmen des Assessments von Sexualstraftätern der *STABLE-2007* (Hanson et al., 2007). Diese Instrumente sind allerdings nur im Kontext solcher Fälle anzuwenden, in denen die Handlungsschwelle zur Tat bereits überschritten wurde. Die Identifikation von Risikofaktoren im Vorfeld, also im Sinne einer Primärprävention wird dadurch nicht ermöglicht. Dieser Identifikation kommt daher eine besonders zentrale Rolle zu, wenn es darum geht, sexuelle Übergriffe in Institutionen zu verhindern. Im Rahmen von sexueller Gewalt im Strafvollzug werden Risikofaktoren mit der sogenannten *Inmate Level of Risk Screening Measure* (Gonsalves, Walsh, & Scalora, 2012) zusammengestellt. Diese bildet sowohl das Risiko einer Person, Opfer eines sexuellen Übergriffs zu werden ab, als auch das, einen solchen zu begehen. Für sexuelle Übergriffe gegenüber Kindern in Institutionen steht ein solches Instrument (noch) nicht zur Verfügung. Als mögliche Risikofaktoren, Täter zu werden, oder multiple Taten zu begehen, konnten allerdings frühe Verhaltensauffälligkeiten, eigene Viktimisierungserfahrungen, sowie vermehrte Äusserungen von Drohungen identifiziert werden.

1.2.2 Validierung von Risikobeurteilungen

Eine Hauptaufgabe der forensisch-psychologisch/psychiatrischen Wissenschaften ist es, verschiedene Risikobeurteilungssysteme (wie in Kapitel 1.2.1 erläutert) auf ihre Qualität hin zu überprüfen und Anwendungsempfehlungen zu geben. Die Überprüfung der Qualität schließt zum einen die Validierung auf Kriteriumsebene, also die Überprüfung der trennscharfen Identifikation von potentiellen Rückfalltätern und Nicht-Rückfalltätern ein, sie beschäftigt sich aber auch mit der passenden Kalibrierung von Risk-Assessment Instrumenten. Kalibrierung beschreibt in diesem Zusammenhang die Replizierbarkeit der, den Risikokategorien zugrunde liegenden Normwerten der Rückfallwahrscheinlichkeit in der Untersuchungspopulation. Validierungs- und auch Kalibrierungsstudien zu den verschiedensten Risk-Assessment Instrumenten sind zentraler Bestandteil der forensischen Literatur. Unter methodischen Gesichtspunkten betrachtet, stehen diese allerdings vor gewissen Herausforderungen. Die in der Validierung von Risk-Assessment Instrumenten als Goldstandard für Aussagen zur trennscharfen Unterscheidung zwischen Rückfalltätern und Nicht-Rückfalltätern verwendete, basisratenunabhängige *Receiver Operating Characteristic* (Rosenfeld & Penrod, 2011) ist nicht unstrittig. Urbaniok, Rinne, Held, Rossegger, und Endrass (2008) führen aus, dass die damit zugrunde gelegte Dichotomisierung des Rückfallrisikos in rückfällige/nicht-rückfällige Täter einen geringen praktischen Nutzen aufweist. Für die forensische Praxis ist weniger relevant, ob der Rückfalltäter einen signifikant höheren Wert in einem Instrument erreicht als ein Nicht-Rückfalltäter, sondern mehr, wie hoch das Risiko sein könnte. Es bleibt folglich schwierig, anhand der Höhe von AUC-Werten die Qualität verschiedener Methoden von Risikobeurteilungen einzuschätzen (Mossman, 2013; Skeem & Monahan, 2011), da dieser schlicht nicht mehr aussagen kann, als ob die trennscharfe Unterscheidung im bivariaten Kriterium besser war als der Zufall, oder eben nicht ($AUC \pm 0.5$). Vielmehr muss die Wahrscheinlichkeit eines Rückfalls, bzw. das Risikolevel für den Anwender/die Anwenderin bekannt sein, um zu entscheiden, ob und wie risikoorientierte Interventionen zu empfehlen sind (Mossman, 2013; Urbaniok et al., 2008). Da nicht-numerische Risikolevel zu den unterschiedlichsten Interpretationen führen können (Hilton, Carter, Harris, & Sharpe, 2008), geben die meisten mechanischen Risk-Assessment Instrumente in ihren Entwicklungsstudien Normwerte für die Rückfälligkeit in den Risikokategorien und nach unterschiedlichen Zeiträumen des Follow-Ups an (Rossegger, Endrass, et al., 2012b). Dies ermöglicht die Ausformulierung

des Risikos anhand des Vergleichs des individuellen Ergebnisses des Täters mit der Normstichprobe. Fraglich bleibt allerdings, inwiefern Risk-Assessment Instrumente soweit generalisierbar sind, dass der spezifisch einzuschätzende Täter damit charakterisiert werden kann. Diese Generalisierbarkeit lässt sich mithilfe von Replikationsstudien untersuchen (Rossegger, Gerth, Seewald, et al., 2013). In einer Übersichtsarbeit legen Rossegger, Gerth, Seewald, et al. (2013) Kriterien fest, die diese Replikationsstudien erfüllen müssen, um valide Aussagen über die Generalisierbarkeit treffen zu können. Diese betreffen unter anderem Stichprobenmerkmale, Follow-Up-Zeiträume, Rückfalldefinitionen und Anwendungsrichtlinien der jeweiligen Instrumente. Keine der 84 in die Untersuchung eingeflossenen Studien über drei verschiedene Instrumente konnte alle Qualitätskriterien erfüllen.

Da die Mehrheit der Replikationsstudien jedoch signifikante Zusammenhänge mit Rückfälligkeit zum Ergebnis hat, gelten sie im Allgemeinen als empirischer Nachweis für die Qualität des im Instrument abgebildeten Risikomodells (Rossegger, Gerth, Seewald, et al., 2013). Dieser Schlussfolgerung muss folglich vorsichtig begegnet werden. Eine über den simplen Zusammenhang zwischen Ergebnissen eines Risk-Assessment Instruments und Rückfälligkeit hinaus gehende Fragestellung ist für die Praxis mindestens gleichbedeutend: Lassen sich die Rückfälligkeitsraten der Normstichprobe auf die Täterpopulation übertragen, zu welcher der einzuschätzende Täter gehört? Diese Frage der Kalibrierung der Instrumente, also inwiefern beobachtete Rückfallraten denen der Normstichprobe entsprechen, wird in den Untersuchungen zu Risk-Assessment Instrumenten immer noch zu wenig berücksichtigt (Hart et al., 2007; Rossegger, Gerth, Singh, & Endrass, 2013).

Zuletzt stellt sich in der Untersuchung der Validität von Risk-Assessment Instrumenten noch die Frage, ob der Rückfall an sich wirklich das geeignete Merkmal zur Überprüfung der Kriteriumsvalidität ist. Vor allem bei der Berücksichtigung dynamischer Variablen, die durch Interventionen veränderbar sind, kann die Genauigkeit des Risk-Assessments, das den Rückfall als Kriterium ausweist, geschmälert werden (Harris & Rice, 2015). Der Intervention kommt dabei ein mediierender Effekt zu (Belfrage & Strand, 2012; Storey, Kropp, Hart, Belfrage, & Strand, 2014), der in der Kriteriumsvalidierung anhand der Rückfälligkeit unbeachtet bleibt. Vereinfacht betrachtet bedeutet das, dass ein als hoch eingeschätztes Risiko dann als nicht zutreffend eingestuft wird, wenn der Täter nach Ablauf eines bestimmten Follow-Ups nicht rückfällig wird. Dass der beurteilende

Praktiker/die beurteilende Praktikerin dieser Einschätzung folgend allerdings Risikomanagementstrategien empfiehlt oder umsetzt, die am Ende zur Verhinderung eines Rückfalls führen könnten wird dabei außer Acht gelassen. Konsequenterweise sollte also vor allem bei der Evaluierung von Risk-Assessment Instrumenten, die dynamische Komponenten beinhalten, das Kriterium, an dem deren Präzision ermittelt wird überdacht werden.

Seewald K, Rossegger A, Urbaniok F, Endrass J (under review). Assessing the risk of intimate partner violence: Expert evaluations versus the Ontario Domestic Assault Risk Assessment. *International Journal of Offender Therapy and Comparative Criminology*

Anhand von Rückfallrisikobeurteilungen bei Fällen von häuslicher Gewalt wurde die gutachterliche, dem klinisch-intuitiven Modell folgende Risikobeurteilung der aktuarischen Risikobeurteilung durch das *Ontario Domestic Assault Risk Assessment* (ODARA) gegenüber gestellt. Damit vergleicht diese Studie zwei Herangehensweisen, die einerseits keine und andererseits alle vier der unter 1.2.1 erläuterten Komponenten einer Risikobeurteilung strukturieren (Skeem & Monahan, 2011).

Aus den forensisch-psychiatrischen Gutachten wurde das Risikolevel auf einer neun-stufigen Likert-Skala abgetragen, sowie auf der gutachterlichen Informationsgrundlage das ODARA von (im Instrument geschulten) Bachelorstudierenden durchgeführt. Wenngleich nicht alle methodischen Herausforderungen, die an Untersuchungen zu Risikobeurteilungen gestellt werden, in dieser Studie erfolgreich gelöst werden konnten, wurden doch einerseits zusätzlich zur ROC-Analyse Odds Ratios anhand logistischer Regressionsanalysen berechnet, und zum anderen die erwarteten den beobachteten Rückfälligkeitsraten der ODARA-Risikokategorien gegenüber gestellt. Ziel der Studie war weniger die Validierung oder Kalibrierung des ODARAs als Instrument, mehr jedoch die Gegenüberstellung beider Methoden im Rahmen der Beurteilung des Risikos bei häuslicher Gewalt. Das ODARA wies dabei die statistisch signifikant bessere AUC auf (0.78 vs. 0.44), wie auch eine etwa zweieinhalbfach erhöhte Chance auf einen Rückfall mit steigender Risikokategorie (OR=2.4). Das nicht signifikante OR des unstrukturierten Urteils von 0.7 könnte vorsichtig als Tendenz in eine Richtung interpretiert werden, die

mit einer größeren Stichprobe noch untermauert werden müsste. In dieser Interpretation bleibt offen, ob von einem – mit diesem Design allerdings nicht untersuchbaren – mediiierenden Effekt der Intervention ausgegangen werden muss, da keiner der, von den Gutachtern als „hoch“ oder „sehr hoch“ Rückfall gefährdet eingestuften Personen während der Beobachtungszeit rückfällig wurde. Unabhängig davon bleibt das ODARA ein professionsübergreifend einsetzbares Instrument für die Rückfälligkeit im Rahmen von häuslicher Gewalt auffällig gewordener Täter.

1.3 Konzeptuelle Annahmen eines risikoorientierten Behandlungsrahmens

Die überwiegende Mehrheit der Gewalt- und Sexualstraftäter wird nach Verbüßung ihrer Haftstrafe oder nach Beendigung einer Maßnahme/Maßregel wieder in die Gesellschaft entlassen. Die Notwendigkeit von therapeutischen Interventionen zur Rehabilitation und Risikosenkung liegt daher zum einen aus opferpräventiven Gesichtspunkten auf der Hand, und verankert sich zum anderen auch in ideologischen Ansichten betreffend des gesamtgesellschaftlichen Umgangs mit Delinquenz. Zusätzlich finden sich ebenfalls im positiven Kosten-Nutzen-Verhältnis von therapeutischen Interventionen Befürworter wieder: Die durch wirksame Interventionen verhinderten Rückfälle sparen Kosten ein, welche die Kosten der Behandlung am Ende aufwiegen (Aos, Phipps, Barnoski, & Lieb, 2001; Endrass, Rossegger, & Kuhn, 2012).

1.3.1 Risikoorientierung in forensischen Interventionen

In der Frage, was eine wirksame Intervention im forensischen Kontext auszeichnet, bzw. welchem Modell wirksame Interventionen folgen sollen, zeigt sich das von Andrews et al. erstmals 1990 veröffentlichte *Risk-Needs-Responsivity (RNR)*-Modell als bislang empirisch weitreichend in seiner Wirksamkeit belegt (Andrews & Bonta, 2010a, 2010b; Hanson, Bourgon, Helmus, & Hodgson, 2009; Hart & Logan, 2011; Koehler, Lösel, Akoensi, & Humphreys, 2013; Smith, Gendreau, & Swartz, 2009).

Das *Risk-Needs-Responsivity (RNR)*-Modell

Das RNR-Modell ist ein Modell zur Risikoeinschätzung und Rehabilitation von Straftätern, das aus drei Kernprinzipien besteht: *risk principle*, *need principle* und

responsivity principle. Die Wirksamkeit von Interventionsprogrammen erhöht sich dabei in Abhängigkeit davon, wie viele der Prinzipien den Programmen zugrunde liegen (Andrews & Bonta, 2010a; Hanson et al., 2009).

Das *risk principle* setzt eine valide Risikobeurteilung voraus. Die Intervention soll sich am eingeschätzten Risikolevel orientieren. Tätern, die ein höheres Risiko aufweisen, soll in diesem Rahmen eine intensivere Intervention zukommen, als Tätern, die ein geringeres Risiko aufweisen (Andrews & Bonta, 2010a). Das *need principle* beschreibt die für eine Reduzierung des Rückfallrisikos in der Intervention zu adressierenden Ziele, die als sogenannte kriminogene Bedürfnisse bezeichnet werden. Diese können auch als dynamische Risikofaktoren verstanden werden, deren Veränderung eine Auswirkung auf das Rückfallrisiko des Täters hat. Den von Andrews und Bonta (2010a) postulierten kriminogenen Bedürfnissen, liegen die „central eight“ zugrunde. Als „central eight“ werden acht Faktoren verstanden, die mit delinquentem Verhalten assoziiert sind. Diese gliedern sich in die „big four“ und moderate four“. Die „big four“ umfassen (1) früheres kriminelles Verhalten, (2) antisoziale Persönlichkeitsmuster, (3) antisoziale Einstellungen und Werte und (4) Verbindungen zu einer kriminellen Subkultur. Unter „moderate four“ werden (5) familiäre delinquenzfördernde Bedingungen, (6) geringe Leistungsfähigkeit und Engagement im schulischen oder Arbeitskontext, (7) gering ausgeprägte delinquenzferne Freizeitaktivitäten und (8) Substanzmissbrauch zusammengefasst. Unter dem *responsivity principle* wird schließlich die Anpassung der Intervention an die individuellen kognitiven Lern- und Leistungsfähigkeiten, sowie an Beeinflussbarkeit des Täters und dessen Ansprechbarkeit auf die Behandlung verstanden. In Abhängigkeit der Ausrichtung der Intervention (kognitiv-behavioral, tiefenpsychologisch, psychoanalytisch) können diese Elemente jeweils neu definiert werden (Andrews & Bonta, 2010a; Bonta & Andrews, 2007).

Praktische Umsetzungen des RNR-Modells

Zu Beginn der praktischen Umsetzung des RNR-Modells steht eine valide Risikobeurteilung zur Einschätzung der nötigen Intensität der Interventionen. Die meisten Risk-Assessment Instrumente (siehe 1.2.1) stehen im Allgemeinen jedoch in der Kritik, zwar eine Risikoeinschätzung vorzunehmen, jedoch keine Anhaltspunkte für Therapieinhalte zu geben, deren Adressierung eine Reduktion dieses Risikos zur Folge hat (Hart & Logan, 2011; Skeem & Monahan, 2011). Weiter bleibt die prognostische

Relevanz alternativer Ansätze wie der „case-formulation“-Strategien unklar (siehe 1.2.1). Somit ist eine initiale Beurteilung des Rückfallrisikos anhand der verbreiteten, standardisierten Risk-Assessment Instrumente eine Grundvoraussetzung für die Implementierung von auf dem RNR-Modell basierenden Programmen. In einem zweiten Schritt können dann durch die Formulierung individueller Delinquenzhypothesen unter der Verwendung kriminogener Bedürfnisse und (subdiagnostischer) psychopathologischer Merkmale Behandlungsfoki bestimmt werden. Diesem Vorgehen folgen auch die im Psychiatrisch-Psychologischen Dienst im Amt für Justizvollzug des Kantons Zürich durchgeführten deliktpräventiven Therapien (Rossegger, Endrass, Urbaniok, & Borchard, 2012), deren Wirksamkeitsevaluation Teil der vorliegenden Dissertationsschrift ist (siehe Kapitel 1.4.2 und 2.5).

Kritik am RNR-Modell

Kritische Aspekte, die im Rahmen des RNR-Modells formuliert werden, konzentrieren sich auf das Fehlen von Schutzfaktoren im Modell (Hart & Logan, 2011; Ward & Brown, 2004). In ihrer Entwicklung eines Gegenmodells, des „Good Lives Models (GLM)“ argumentieren Ward und Brown (2004) für die Fokussierung positiver Aspekte im Rahmen der Straftäterbehandlung. Eine an Stärken und Ressourcen der Täter orientierte Intervention soll das Ziel haben, sozial akzeptierte, nicht-delinquente Strategien zur Erreichung und Sicherung von Grundbedürfnissen zu etablieren und somit künftige Kriminalität zu verhindern. Auch wenn auf theoretischer Ebene die Diskussionen darum bereits entbrannt sind (Andrews, Bonta, & Wormith, 2011; Andrews & Dowden, 2009; Birgden, 2009; Ward, Yates, & Willis, 2012), ist die Wirksamkeit von Interventionen, denen das GLM zugrunde liegt, in der Praxis noch kaum evaluiert. Neben Fallstudien (Gannon, King, Miles, Lockerbie, & Willis, 2011), beziehen sich aktuellste Publikationen auf das Maß an korrekter Operationalisierung des Modells in Behandlungseinrichtungen (Willis, Ward, & Levenson, 2014). Erste Vergleichsergebnisse der Anwendung beider Modelle zeigen keine signifikanten Unterschiede in der Veränderung von im Rahmen der Therapie adressierten Problembereiche („treatment change“; Harkins, Flak, Beech, & Woodhams, 2012). In Bezug auf die Veränderung konkreter psychologischer Konstrukte wie deliktnaher Kognitionen oder sozio-affektiver Merkmale konnten ebenfalls keine deutlichen Unterschiede je nach Zuteilung zu entweder einer dem RNR-Modell, oder dem GLM

folgenden Interventionsgruppe gezeigt werden (Barnett, Manderville-Norden, & Rakestrow, 2014). Die Autorinnen formulieren jedoch als mögliche Interpretation des minimalen Vorteils des GLMs, dass dieses eine geeignete Alternative zur Aufrechterhaltung der Funktionalität für jene Täter sein könnte, die bereits im Vorfeld der Therapie ein höheres funktionales Niveau aufwiesen. Weitere Evaluierungen, vor allem im Hinblick auf die rückfallpräventive Wirkung müssen dringend Gegenstand künftiger Untersuchungen sein. Da sich allerdings in der konkreten praktischen Umsetzung auch Überlappungen beider Modelle zeigen, lassen sich mögliche Synergieeffekte am Ende nicht ausschließen (Ward et al., 2012).

1.3.2 Interventionen bei jugendlichen Straftätern

Interventionen bei jugendlichen oder jungen erwachsenen Straftätern nehmen insofern eine besondere Rolle innerhalb des forensisch-psychiatrisch/psychologischen Feldes ein, als dass sie mit der Verhinderung etwaiger krimineller Karrieren eine gewichtige Aufgabe in der Kriminalprävention erfüllen (Gmür, Gerth, & Graber, 2012). Wirksamkeitsstudien zu Interventionen bei jugendlichen Straftätern zeigen einen besonders robusten positiven Effekt (Koehler et al., 2013; Lipsey & Cullen, 2007). Dies allerdings nur, wenn sie rehabilitative Elemente aufweisen und nicht rein auf Abschreckung ausgelegt sind (Aos et al., 2001; Lipsey & Cullen, 2007). Programme, die die Prinzipien des RNR-Modells (Kapitel 1.3.1) einhalten, weisen dabei auch bei Jugendlichen den stärksten positiven Effekt auf (OR=1.9; Koehler et al., 2013). „Community-based“ Interventionen, also solche, die nicht in einem Inhaftierungskontext stattfinden, erweisen sich ebenso als signifikant effektiv (OR=1.5; Koehler et al., 2013). Die jugendlichen Täter, deren delinquentes Anlassverhalten und Ausgangsrisiko keine Inhaftierung notwendig machen, werden präferiert in einem als wirksam und kosten-nutzen-effizient klassifizierten Programm untergebracht, welches einem systemischen und multi-professionellen Ansatz folgt. So wird die gesellschaftliche Einbindung der Jugendlichen sichergestellt und an den, in Familienstrukturen inhärenten Risikofaktoren angesetzt. Programme dieser Art werden unter dem Oberbegriff der „Blueprints“-Programme subsumiert (Mihalic & Irwin, 2003) und vom Institute of Behavioral Sciences der University of Colorado, US zertifiziert.

Seewald K, Yundina A, Endrass J (in press). Juvenile Corrections. In Kerley KR, Copes H, De Li S, Lane J, Sharp SF (Eds.) *The Encyclopedia of Corrections*. Wiley Blackwell.

In der Übersichtsarbeit wird deutlich, warum die Betrachtung der Wirksamkeit von Interventionen bei jugendlichen Straftätern von besonderer Bedeutung ist. Die Verhinderung krimineller Karrieren steht dabei im Vordergrund. Konzeptuelle Unterschiede zwischen bestrafenden und rehabilitativen Interventionen bei jugendlichen Straftätern werden erläutert und deren Wirksamkeit betrachtet. Die „Blueprints“-Programme (1) „Multidimensional Treatment Foster Care (MTFC)“ (Fisher & Chamberlain, 2000), (2) „Functional family therapy (FFT)“ (Sexton & Alexander, 2003) und (3) der „Multisystemic Therapy (MST)“-Ansatz (Henggeler, 2011) werden dabei in den Vordergrund gestellt.

Während die Jugendlichen im Rahmen des MTFC in geschulten Pflegefamilien untergebracht, sowie von „case managern“ begleitet und wieder in das System eingebettet werden, folgt das FFT-Programm dem Ansatz, dysfunktionale interpersonelle Verhaltensmuster in der Herkunftsfamilie des Täters zu fokussieren (Sexton & Alexander, 2003). Das MST-Programm wiederum folgt der Annahme, dass Dysfunktionalität im gesamten sozialen Netzwerk der Jugendlichen Ursache deren kriminellen Verhaltens ist, und wird folglich in allen möglichen Bereichen implementiert, in Familie, Schule und im sozialen Umfeld. Fokussiert wird dabei vor allem die (Wiederherstellung der) Funktionalität interpersoneller Beziehungen und Kommunikationsstrukturen sowie die Bearbeitung des Delikts (Henggeler, 2011). Wirksamkeitsstudien zeigen Wiederverurteilungsraten von 5% bei Teilnehmenden im MFTC-Programm im Vergleich zu 24% bei den Nicht Teilnehmenden (Eddy, Whaley, & Chamberlain, 2004), sowie das beste Kosten-Nutzen-Verhältnis im Rahmen verschiedener Interventionen mit jugendlichen Tätern (Drake, Aos, & Miller, 2009). FFT-Programme weisen einen rückfallsenkenden Effekt von 18% auf (Aos et al., 2001). Die Reduktion der Rückfälligkeit in MST-Gruppen wird zwischen 16% und 46% eingestuft (Henggeler, 2011; Lipsey & Cullen, 2007) und in methodisch qualitativ hochwertigen Studien zusätzlich als besonders nachhaltig beschrieben (Sawyer & Borduin, 2011).

In der Übersichtsarbeit wird ebenso ausführlich betrachtet, welche Formen der Interventionen bei jugendlichen Täterinnen und Tätern keinen oder gar einen schädigenden Effekt haben können. Kritisch gesehen werden müssen an dieser Stelle sogenannte „Scared Straight“-Programme, die im Allgemeinen auf dem Prinzip der Abschreckung durch Gefängnisbesuche der Jugendlichen beruhen. Meta-analytische Befunde gehen statt eines delinquenzreduzierenden Effekts von einem Anstieg kriminellen Verhaltens um rund 6% aus (Drake et al., 2009). Auch „Boot Camps“ zeigen keinen signifikanten Effekt der Senkung von Rückfälligkeit auf. Im Gegenteil, es können Studien identifiziert werden, die eine Erhöhung der Rückfälligkeit nach „Boot Camp“-Aufenthalt um rund 10% zeigen (Aos et al., 2001; Meade & Steiner, 2010). In diesen Camps sollen Jugendliche durch militärischen Drill zu Disziplin, Struktur und Unterordnung gezwungen und zu moralischem Verhalten und Teamwork gebracht werden (Gmür et al., 2012). Nichtbefolgung der Regeln wird strikt bestraft und Autoritätenghorsam verstärkt (Lilienfeld, 2007).

1.4 Evaluation der Wirksamkeit forensischer Interventionen

Vor allem nach schwerwiegenden Rückfällen von Tätern, die in einem forensisch-therapeutischen Setting angeschlossen sind oder waren, intensivieren sich gesellschaftspolitische Diskurse über Notwendigkeit, Wirksamkeit, Kosten und das geeignete Setting forensischer Interventionen. Neben der Erforschung der Güte forensischer Risikobeurteilungen, sind diese Fragestellungen im Rahmen der Straftäterbehandlung ein zweites zentrales Thema in der forensisch-psychologisch/psychiatrischen Wissenschaft. Die Wirksamkeitsevaluation der Straftäterbehandlung birgt allerdings eine Reihe methodischer und ethischer Herausforderungen (Hollin, 2008; Ioannidis et al., 2001; Marshall & Marshall, 2007, 2008; Persons & Silberschatz, 1998; Seto et al., 2008; Sherman et al., 1998; Slade & Priebe, 2001).

1.4.1 Randomized Controlled Trials in der Forensik – der Goldstandard?

Aus der medizinischen Wirksamkeitsforschung kommend, wird die Anwendung von randomisierten Kontrollgruppenstudien („Randomized Controlled Trials [RCTs]“),

weitestgehend als Goldstandard für Wirksamkeitsstudien im Allgemeinen betrachtet. Während sich die Wirksamkeit simpler und eindeutiger medizinischer Behandlungsansätze (wie beispielsweise medikamentöser Wirkstoffe) valide durch ein RCT-Design abbilden lässt, muss die zweifelsfreie Anwendbarkeit dieses Designs auf Evaluationen weit komplexerer psychotherapeutischer Verfahren überdacht und diskutiert werden (Persons & Silberschatz, 1998). Nicht nur, weil die Messung von Veränderungen in psychologischen Variablen auf ebenso explizite Art wie die biologischer Variablen nach einer medikamentösen Behandlung deutlich herausfordernd ist (Slade & Priebe, 2001), sondern auch, weil nicht alle psychotherapeutischen Kontexte die Grundannahmen des RCT-Designs umsetzen lassen (Westen, Novotny, & Thompson-Brenner, 2004).

Durch ihr streng experimentelles Studiendesign lassen RCT-Studien zwar zu, die Veränderung in der Kriteriumsvariable lediglich auf die zu testende Intervention zurückführen zu können und zeigen folglich eine hohe interne Validität auf. Die Sterilität und Künstlichkeit des Designs allerdings gestaltet eine Übertragung auf reale therapeutische Situationen im Allgemeinen, in forensischen, und nicht-forensischen Kontexten schwierig. Betrachtet man die beiden Grundprinzipien der RCT-Designs, zum einen die randomisierte Zuordnung von Versuchspersonen zu Experimental- und Kontrollgruppe und zum andern die sogenannte „treatment fidelity“ oder auch „treatment integrity“, das Maß der Umsetzung objektivierter Interventionspläne (Marshall & Marshall, 2007), gehen mit der Implementierung von RCT-Designs im forensischen Kontext verschiedene Probleme einher.

Randomisierte Gruppenzuordnung

Die zufällige Zuordnung zu beiden Gruppen (Kontroll- vs. Behandlungsgruppe, oder zweier Behandlungsgruppen) stellt sicher, dass beide Gruppen sich hinsichtlich zentraler Merkmale, die unter Umständen einen Einfluss auf das Ergebnis der Interventionsevaluation haben, nicht unterscheiden. Allfällige Störvariablen werden durch die Randomisierung bedient und die interne Validität wird hoch gehalten. Somit können Unterschiede in der „post-study“-Messung rein auf den Einfluss der Intervention an sich zurückgeführt und eine Konfundierung durch andere Variablen kontrolliert werden. Wird nun als Misserfolgskriterium der forensischen Intervention die Rückfälligkeit gewählt, ist zu beachten, dass eine Vielzahl statischer, historischer und

dynamischer Variablen einen Einfluss auf das Kriterium der Rückfälligkeit haben (Beech, Friendship, Erikson, & Hanson, 2002; Hanson et al., 2007; Quinsey et al., 2006). Obwohl das „Matching“ hinsichtlich dieser Variablen zwischen Experimental- und Kontrollgruppe den explanatorischen Gehalt der Studie erhöhen würde, da das allgemeine Rückfallrisikolevel zwischen beiden Gruppen konstant gehalten wird, findet man dieses Vorgehen selten in Wirksamkeitsstudien bei forensischen Populationen vor (Marshall & Marshall, 2007). Dies vor allem deshalb nicht, da sich die, ohnehin schon geringen, forensischen Fallzahlen dadurch noch zusätzlich verringern würden, was sich wiederum negativ auf die statistische Gesamtpower der Studie auswirken würde (Hollin, 2008). Zusätzlich zeigen Ergebnisse der bislang größten RCT-Untersuchung von Therapiewirksamkeit bei Sexualstraftätern („Sexual Offender Treatment Evaluation Project“ (SOTEP); Marques, Wiederanders, Day, Nelson, & Van Ommeren, 2005), dass Randomisierung schließlich nicht garantiert, dass Experimental- und Kontrollgruppe hinsichtlich aller Variablen mit korrelativen Zusammenhängen zum Rückfallrisiko ausreichend gut „gematched“ und daher äquivalent in ihrem Ausgangsrisiko sind.

Eine Herausforderung praktischer Natur in der Umsetzung randomisierter Kontrollstudien im forensischen Kontext, ist die der Teilnahmemotivation und Freiwilligkeit. Straftäter in forensischen Settings gehen, trotz Aufklärung, meist von einem positiven Einfluss der Studienteilnahme auf ihren Haft-/Vollzugsverlauf aus (z.B. frühere Vollzugslockerungen, sonstige Bevorteilungen), sowie vom gegenteiligen Effekt bei Verweigerung der Teilnahme an der Studie, oder bei früherem Ausstieg (Gondolf, 2001; Marshall & Marshall, 2007). Bislang ergibt sich noch kein klares Bild des Einflusses der Motivation und Veränderungsbereitschaft von Straftätern auf den Therapieeffekt (Grady, Edwards, Pettus-Davis, & Abramson, 2012; McMurrin & Ward, 2010; Polaschek & Ross, 2010). Bis zur besseren empirischen Untermauerung muss daher davon ausgegangen werden, dass auch motivationale Faktoren eine Störvariable im experimentellen Design sein könnten.

Die bereits genannten theoretischen oder praktischen Schwierigkeiten stehen im Kontext von RCTs, welche die Wirksamkeit zweier Therapieformen vergleichen. Die Untersuchung, ob Therapie überhaupt rückfallpräventiv wirkt, also im Vergleich zur jeglichen Haftunterbringung einen Zugewinn darstellt, ist im Kontext gesetzlich angeordneter Therapien in einem experimentellen, randomisierten Studiendesign schwierig bis gar nicht zu untersuchen. Neben all den bereits genannten methodischen

Schwierigkeiten, ergeben sich in diesem Zusammenhang auch vielfach ethische und moralische Fragestellungen, die es in der Wirksamkeitsforschung im forensischen Kontext zu beantworten gilt (Farrington & Welsh, 2005; Hollin, 2008; Lösel & Schmucker, 2005; Marshall & Marshall, 2007). Auf der Hand liegend etwa die Tatsache, dass die Zurückhaltung von Therapie für Gewalt- und Sexualstraftäter, um den eigentlichen, alleinigen Effekt von Behandlung auf Rückfälligkeit zu überprüfen, mittels einer untherapierten Kontrollgruppe, den Schutz potentieller Opfer negiert (Lösel & Schmucker, 2005), da die Notwendigkeit einer präventiven therapeutischen Maßnahme durch die gesetzliche Anordnung einer Behandlung bereits festgestellt wurde (Marshall & Marshall, 2007). Auch auf der Täterseite ergeben sich moralische Dilemmata: Welche Täter dürfen Therapie bekommen, und welche nicht? Welche bekommen eine bestimmte Form der Therapie und welche werden randomisiert zu welchem spezifischen, zu untersuchenden Programm zugeteilt? Stellen sich Probanden im forensischen Setting zur Teilnahme an einer Therapiewirksamkeitsstudie zur Verfügung und würden, hypothetisch betrachtet, der unbehandelten Kontrollgruppe zugewiesen werden, bleibt für das sensible Gefängnisklima die Herausforderung des Managements persönlicher Enttäuschungsreaktionen, Ärger oder Frustration.

Objektivierte Interventionspläne

Die zweite wichtige Voraussetzung von RCT-Studien ist die Sicherstellung der korrekten, manualgetreuen Umsetzung von Interventionsplänen bei allen Personen in der Interventionsgruppe, bzw. den Interventionsgruppen, wenn zwei Therapieformen verglichen werden („treatment integrity/fidelity“). In Kapitel 1.3 wurde bereits ausgeführt, dass bislang die besten Resultate in der Straftätertherapie erzeugt werden können, wenn sich die Behandlung am individuellen Rückfallrisiko, ebenso wie an spezifischen kriminogenen Bedürfnissen (einer individuellen Delinquenzhypothese folgend), sowie an der eigenen Responsivität des Täters orientiert. Kernelement einer wirksamen Straftätertherapie ist somit die flexible Ausgestaltung, die nicht mit manualisierten Interventionsplänen erfasst werden kann (Andrews & Bonta, 2010a). Das RCT-Design wiederum, mit seiner Vorgabe der radikalen „treatment integrity“ erlaubt keine responsiven, individuellen Anpassungen im Rahmen der Behandlungsumsetzung. Marshall und Marshall (2007) proklamieren daraufhin, mit Bezug auf die Ergebnisse der SOTEP-Studie (Marques et al., 2005), dass die Umsetzung

der Kernprinzipien des RCT-Designs eine effektive Straftäterbehandlung per se unmöglich macht, und somit unweigerlich falsche Ergebnisse erzeugt.

Diese substantiellen Schwierigkeiten bei der Durchführung von RCT-Studien legen die begründete Annahme nahe, dass RCT-Designs, in ihrer bislang umgesetzten Form, im Rahmen forensischer Therapiewirksamkeitsstudien schlicht nicht als das einzige Studiendesign gelten können, das Aufschluss über die Effektivität der Behandlung gibt (Hollin, 2008; Marshall & Marshall, 2007).

1.4.2 Quasi-experimentelle Alternativen

Sofern es um die rückfallpräventive Wirksamkeit von Interventionen, und nicht um prozessorientierte Veränderung innerhalb therapeutischer Behandlungen geht, müssen wissenschaftliche Studien also auf quasi-experimentelle Alternativen zurückgreifen. Diese bieten die Lösung für Herausforderungen der Randomisierung durch adäquate Kontrollgruppenbildung und für ethische Dilemmata. Beech, Mandeville-Norden, und Goodwill (2012) vergleichen beispielsweise „treatment responders“ mit „treatment non-responders“ zur Evaluierung der Unterschiede im Bezug auf Rückfälligkeitsraten. Der Vergleich der Interventionsgruppe mit einer vor Beginn des Behandlungsprogramms zuvor entlassenen Tätergruppe wurde ebenfalls bereits als Auswegmöglichkeit aufgezeigt (Hanson, Broom, & Stephenson, 2004). Als statistische Lösung der Problematik der (unbehandelten) Kontrollgruppe bieten sich zusätzlich verschiedene Matching-Methoden an, so etwa das „Propensity Score Matching“, die Bildung einer durch logistische Regression ermittelten Größe zur Prädiktion der Gruppenzugehörigkeit (Duwe & Goldman, 2009; Pascual-Leone, Bierman, Arnold, & Stasiak, 2011), oder die Kontrolle für potentielle Störvariablen wie des Rückfallrisikos in beiden Vergleichsgruppen (Smid, Kamphuis, Wever, & Van Beek, 2014; Travers, Wakeling, Mann, & Hollin, 2013). In kleineren Stichproben bietet sich ebenfalls an, mittels „case-to-case“-Matching, nach verschiedenen Kriterien passende Einzelfälle zu vergleichen (Polaschek, Wilson, Townsend, & Daly, 2005).

Aussagekraft und Qualität quasi-experimenteller Wirksamkeitsstudien

Ist man sich also der zu gewissen Teilen unlösbaren Aufgabe eines randomisierten Designs in der forensischen Therapieevaluation bewusst, stellt sich die Frage, ob eine Abweichung in der gemessenen Effektgröße aufgrund des quasi-experimentellen

Designs zu erwarten ist. In einer meta-analytischen Betrachtung dieser Fragestellung, kommen Ioannidis et al. (2001) zu dem Resultat, dass die Ergebnisse von RCT-Studien mit denen quasi-experimenteller Studien hoch korreliert sind. Obwohl nicht-randomisierte Untersuchungen einen signifikant höheren Behandlungseffekt aufwiesen, wurde der absolute Unterschied als klinisch irrelevant beschrieben. Die Ergebnisse zeigten ebenfalls, dass die Behandlungseffekte von äußeren Variablen, wie Stichprobengröße, Publikationsbias und der zeitlichen Einordnung der Veröffentlichung beeinflusst waren. Wenn solche Verzerrungen die Effektgröße derart deutlich beeinflussen, ist es aus wissenschaftstheoretischer Perspektive sinnvoll, das Ausmaß dieser Beeinflussung abzubilden, um die Ergebnisse unter diesen Prämissen korrekt zu interpretieren – oder aber auch, von der anderen Seite betrachtet, auszuweisen, mit welcher Wahrscheinlichkeit es zu keinen Verzerrungen in der Effektmessung gekommen ist. Als statistisches Maß für die Eigenschaft eines Verfahrens, den „wahren“ Wert, das „wahre“ Ergebnis, frei von Verzerrungen geliefert zu haben eignet sich die Kennzahl des „positive predictive value (PPV)“. Ioannidis (2005) nimmt dabei verschiedene Zusammenhänge zwischen Verzerrungen im Stichprobendesign und dem PPV an. Diese betreffen unter anderem die Stichprobengröße, die erwartete Effektgröße, die Hypothesenkonformität der Analysen, aber auch finanzielle Interessen der Wissenschaftlerinnen und Wissenschaftler, sowie konkurrierendes Verhalten, die Ergebnisse anderer Arbeitsgruppen widerlegen zu wollen. Nach diesen Annahmen zeigte sich in der statistischen Simulation eine lediglich teilweise Bekräftigung des Goldstandard-Postulates für RCT-Designs im Allgemeinen: RCT-Studien mit eindeutiger Ergebniserwartung, sowie einer angemessenen statistischen Power (erreicht durch substantiell grosse Stichproben) und durch vollständige Standardisierung gering gehaltenen Verzerrungen schneiden mit einem PPV von 0.85 ab und scheinen das Idealdesign darzustellen. Andere RCT-Studien jedoch, die diese Eigenschaften nicht erfüllen, können nach Ioannidis (2005) lediglich PPVs zwischen 0.17 und 0.23 erreichen. Unter Betrachtung der Kritik, Studien mit RCT-Design könnten dem sich selbst auferlegten Grad an Standardisierung im forensischen Kontext nie gerecht werden (Hollin, 2008; Marshall & Marshall, 2007), wird die Erreichung der 85%igen „Wahrheits“-Wahrscheinlichkeit des Idealdesigns schwer möglich. Wird also von dem realistischeren PPV von 0.23 für solche RCT-Designs, die Ergebnisse eben auch nicht

ohne Verzerrungen abbilden ausgegangen, minimiert sich der Unterschied zu quasi-experimentellen Studien mit ausreichender Stichprobengröße ($PPV=0.20$) substantiell. Als Möglichkeit zur Einschätzung der Qualität wissenschaftlicher Wirksamkeitsstudien bietet sich die weit verbreitete *Maryland Scientific Methods Scale* (Sherman et al., 1998) an. Die Skala unterscheidet fünf aufsteigende Qualitätslevel (1) Korrelation zwischen einem Interventionsprogramm und delinquentem Verhalten oder Risikofaktoren für delinquentes Verhalten; (2) Unterschiede in nachfolgender Delinquenz zwischen einer Behandlungs- und einer nicht äquivalenten Kontrollgruppe; (3) Vergleich zwischen zwei oder mehr hinsichtlich relevanter Variablen vergleichbarer Gruppen, mit und ohne Behandlung; (4) Vergleich zwischen Behandlungs- und Kontrollgruppe(n) unter statistischer Kontrolle von Gruppenunterschieden oder individuellem Matching; (5) randomisierte Zuordnung und Analyse vergleichbarer Behandlungs- und Kontrollgruppe(n). Stufe 3 beschreibt dabei eine quasi-experimentelle Studie geringer Qualität, während Stufe 4 eine quasi-experimentelle Studie hoher methodischer Qualität abbildet (Wilson, Bouffard, & Mackenzie, 2005). Lösel und Schmucker (2005) ordneten in ihrer einflussreichen Meta-Analyse die eingeschlossenen Studien zur Wirksamkeit von Sexualstraftäterbehandlungen den vorgeschlagenen Stufen der *Maryland Scientific Methods Scale* zu und kamen zu dem Ergebnis, dass die Mehrheit der Studien (60.0%) die zweite Stufe erreicht. Als quasi-experimentelle Studien mit geringer Qualität (Stufe 3) wurden 23.8%, als solche mit hoher Qualität (Stufe 4) 8.8% der Studien eingeordnet; 7.5% der Studien erreichten als RCT-Studie die fünfte Stufe. Hanson et al. (2009) konstruierten sich in ihrer Meta-Analyse zur Effektivität von Behandlungsprogrammen für Sexualstraftäter ebenfalls ein, auf den Guidelines des *Collaborative Outcome Data Committee* (CODC; Beech et al., 2007) beruhendes vierstufiges System zur Qualitätsbewertung der eingeschlossenen Studien. Im Vergleich zur *Maryland Scientific Methods Scale* wurde diese explizit zur Qualitätseinschätzung der Wirksamkeit von Behandlungsprogrammen für Sexualstraftäter entwickelt (Hanson et al., 2009). Die CODC-Guidelines beinhalten Bewertungskategorien wie Kontrolle der unabhängigen Variable, Ergebniserwartung, Stichprobengröße, Stichprobenschwund, Vergleichbarkeit der Behandlungs- und Kontrollgruppe, Definition der Kriteriumsvariable und Angemessenheit des Vergleichs. Nachdem 81% der Studien nach diesen Kriterien mit ungenügender Qualität beurteilt und wieder ausgeschlossen werden mussten, fanden sich am Ende nur fünf quasi-experimentelle Studien mit einer Qualitätsbewertung als

„gut“ in der Analyse. In diesen Studien kann davon ausgegangen werden, dass nur wenig ausgeprägte Verzerrungen das Ergebnis beeinflussen und Gruppenunterschiede weitestgehend ausgeschlossen wurden. Als von „starker“ Qualität“ konnte keine der den Einschlusskriterien entsprechenden Studie klassifiziert werden.

Auch wenn theoretisch mit einer systematischen Überschätzung der Effektgröße in nicht-randomisierten Studien zu rechnen ist (Ioannidis et al., 2001), zeigt sich in meta-analytischen Betrachtungen forensischer Interventionen, die sowohl quasi-experimentelle, als auch RCT-Studien im forensischen Kontext miteinschlossen, nur ein geringer, bis nicht vorhandener Einfluss des Designs auf die Effektgröße der Wirksamkeit (Babcock, Green, & Robie, 2004; Hanson et al., 2009; Lipsey & Cullen, 2007; Lösel & Schmucker, 2005).

Die Wahl eines quasi-experimentellen Designs im Rahmen forensischer Wirksamkeitsforschung scheint folglich durchaus angemessen. Nach Lösel und Schmucker (2005) und Hanson et al. (2009), lagen bis dato allerdings wenige qualitativ hochwertige quasi-experimentelle Studien vor. Seither sind jedoch einige Bemühungen verschiedener Designalternativen zu finden (Beech et al., 2012; Duwe & Goldman, 2009; Lussier & Gress, 2014; Olver, Nicholaichuk, Gu, & Wong, 2013; Pascual-Leone et al., 2011; Smid et al., 2014; Travers et al., 2013). Die letzte empirische Arbeit der vorliegenden Dissertationsschrift leistet einen Beitrag zur quasi-experimentellen Wirksamkeitsevaluation forensischer Interventionen anhand der Untersuchung der deliktpräventiven Behandlung des Psychiatrisch-Psychologischen Dienstes im Kanton Zürich.

**Seewald K, Rossegger A, Endrass J (manuscript ready for submission).
Examining the effectiveness of an offense-focused treatment program in
violent and sexual offenders using propensity score estimation procedures
and cox regression model**

Seit 1997 werden im Psychiatrisch-Psychologischen Dienst im Amt für Justizvollzug des Kantons Zürich deliktpräventive Therapien durchgeführt (siehe 1.3.1). In der vorliegenden Studie wird die Kohorte, deren Therapie zwischen 1997 und 2009 begann auf den Therapieerfolg hin untersucht (n=255). Kriterium des Therapieerfolges war dabei die Rückfallfreiheit während des Beobachtungszeitraums bis zum Herbst 2013. Als Vergleichsgruppe diente eine unbehandelte Kontrollgruppe, die im Rahmen der Zürcher Forensikstudie mit

Stichtag im August 2000 bestimmt wurde. Dabei handelte es sich um (schuldfähige) Gewalt- und/oder Sexualstraftäter, die ein Mindeststrafmaß von zehn Monaten erhalten hatten, und an keiner freiwilligen oder angeordneten therapeutische Maßnahme teilnahmen (n=285). Da es sich um eine quasi-experimentelle Studie handelte wurden verschiedene Strategien verfolgt, um den beiden verletzten Grundprinzipien von RCT-Designs (1.4.1) Rechnung zu tragen:

Treatment Integrity: Deliktpräventive Therapien im Psychiatrisch-Psychologischen Dienst sind auf individuellen Delinquenzhypothesen aufgebaut und folgen keinem standardisierten Manual. Die „treatment integrity“ wird allerdings operationalisiert durch striktes Qualitätsmanagement, Fallsupervisionen und Prozess- und Entwicklungsüberprüfungen durch Vorgesetzte und jährliche Therapie-, sowie Abschlussberichte. Alle Therapeutinnen und Therapeuten im Dienst verfügen zusätzlich über eine forensische Zertifizierung.

Randomisierung: Da keine randomisierte Gruppenzuteilung erfolgen konnte, wurde auf Basis eines logistischen Regressionsmodells der Gruppenzugehörigkeit ein *propensity score* (PS) gebildet. Dieser wurde sowohl für die Berechnung der Behandlungseffekte („treatment effects“: teffects; StataCorp, 2013), als auch für die folgende Survival-Analyse zugrunde gelegt.

Die Studie lässt sich damit unter der *Maryland Scientific Methods Scale* auf Stufe 4 als qualitativ hochwertige quasi-experimentelle Studie einordnen. Zusätzlich erreichten in der Meta-Analyse von Lösel und Schmucker (2005) nur 17.5% der eingeschlossenen Studien eine Stichprobengröße in dieser Höhe, zwischen 200 und 500 Personen. In der Einordnung nach Hanson et al. (2009) kann die Studie als von „guter“ Qualität beschrieben werden, was dort am Ende nur auf etwa jede fünfte (21.7%) der eingeschlossenen Studien zutraf.

Behandelte Straftäter wiesen signifikant niedrige Rückfallraten auf, als unbehandelte Straftäter (7.1% vs. 13.1%). Nach PS-Matching und inverser PS-Gewichtung ergaben sich signifikante Behandlungseffekte von 7% bzw. 8% Reduktion der Rückfälligkeit. Cox-Regressionsmodelle nach PS-Matching, sowie inverser PS-Gewichtung konnten bei der alleinigen Betrachtung der Behandlungsvariable keine signifikanten Hazard Ratios (HRs) zeigen. Im

multivariablen Modell zeigten sich jedoch Alter, Vorstrafen und VRAG-Summenwert als signifikante Erklärungsvariablen für die Wahrscheinlichkeit der Rückfälligkeit über die Zeit.

Die deliktpräventiven Interventionen im Psychiatrisch-Psychologischen Dienst im Kanton Zürich zeigen sich also robust im Sinne einer Senkung der Rückfälligkeit, selbst bei strengster statistischer Kontrolle möglicher Verzerrungen. Weitergehende Untersuchungen vor allem für die Zeitspanne nach ca. elf Jahren, wenn sich die Wirksamkeit im Rahmen der vorliegenden Datenlage nicht mehr zeigen lässt (kreuzende Hazard-Funktionen) sind mit einer größeren Stichprobe notwendig.

1.5 Ausblick

Für die Umsetzung risikoorientierter Interventionen sind Risikobeurteilungen von zentraler Wichtigkeit. Die Fragen danach, wie eine valide Einschätzung des Risikos aussieht, welche Rolle psychiatrische Diagnosen dabei spielen, ob statische oder dynamische Merkmale miteinbezogen werden oder wie die Integration der Informationen und das finale Urteil aussehen sollen, sind bei Weitem noch nicht abschließend beantwortet. Im Hinblick auf die Bedeutung psychiatrischer Störungsbilder scheint zumindest ein Trend erkennbar, der sich in Richtung einer Abkehr von psychiatrischen Diagnosen und Nosologien hin zur Verwendung subdiagnostischer, deliktrelevanter Problembereiche zeigt. Was die Berücksichtigung statischer oder dynamischer Risikofaktoren angeht, wird einerseits gefordert, letztere in Risikobeurteilungen zu berücksichtigen, um relevante, in der Intervention zu adressierende Problembereiche zu identifizieren und eine Verlaufsabbildung zu ermöglichen (Hart et al., 2011), andererseits aber wird verdeutlicht, dass dynamische Risikofaktoren weit differenzierter betrachtet werden müssen, als bisher geschehen (Harris & Rice, 2015). Es wird dabei nicht nur in Frage gestellt, inwiefern sich verbreitete dynamische Risikofaktoren überhaupt mit Rückfälligkeit in Verbindung bringen lassen, sondern auch, ob Veränderungen im Rahmen dynamischer Items wirklich abbilden, was eine Verlaufsabbildung abbilden soll. Es kann schließlich nicht ausgeschlossen werden, dass eine Veränderung dynamischer Risikofaktoren nicht doch eine Funktion des statischen Ausgangsrisikos ist, und sich etwa Täter mit höheren

statischen Risikowerten auch mit geringerer Wahrscheinlichkeit im Bezug auf dynamische Risikofaktoren positiv entwickeln.

Unabhängig von dieser (fundamentalen) Diskussion erschweren dynamische Risikofaktoren die Validierung von Risk-Assessment Instrumenten am Rückfälligkeitskriterium. Einen aktuellen Lösungsvorschlag machen Lasher, McGrath, Wilson, und Cumming (2015): Davon ausgehend, dass eine Übereinkunft über Behandlungsziele im Allgemeinen zu besseren Ergebnissen einer Therapie führt, untersuchen sie die Güte eines Instruments zur Behandlungsplanung, das ebenso prädiktiven Wert für das Rückfallrisiko aufweist anhand der Übereinstimmung zwischen der Einschätzung des Täters über sich selbst und der Einschätzung der Therapeutin/des Therapeuten. Solche Formen einer eher prozessorientierten Untersuchung von Methoden der Risikobeurteilung und der Identifikation kriminogener Bedürfnisse müssen künftig weiterentwickelt und methodisch untermauert werden.

Als wirksame Interventionen im forensischen Setting haben sich jene etabliert, die dem RNR-Modell nach Andrews und Bonta (2010a) folgen. Auch die Effektivität der deliktpräventiven Therapien im Kanton Zürich, die sich ebenfalls im RNR-Modell fundieren, konnte aufgezeigt werden. Das Modell bleibt jedoch nicht ohne Kritik. Diese bezieht sich im Wesentlichen auf das Fehlen protektiver Faktoren und Ressourcenorientierung, und so wird sich zukünftige forensische Therapieforschung auch mit möglichen Alternativen auseinandersetzen müssen. Die Evaluation von Behandlungsprogrammen, die in erster Linie auf dem GLM basieren, ist dabei die eine Aufgabe (Willis et al., 2014), die Identifikation möglicher Synergieeffekte beider Ansätze die andere. Nach bisherigen Ergebnissen vorstellbar wären beispielsweise unterschiedliche Schwerpunkte im Verlauf einer therapeutischen Intervention, da sich für das GLM bei Personen mit (bereits) höherer Funktionalität bessere Ergebnisse zeigen (Barnett et al., 2014).

Was die Wirksamkeitsevaluation forensischer Interventionen in den letzten Jahrzehnten betrifft, sind die Diskussionen um RCT-Designs als Goldstandard immer wieder entbrannt. Trotz aller Schwächen quasi-experimenteller Designs, sind diese gefordert, die bestmögliche Alternative zu einem randomisierten Idealdesign darzustellen, da jenes letztlich in seiner Perfektion im forensischen Kontext nie ausreichend umgesetzt werden wird (Hollin, 2008). Aktuell wird die wissenschaftliche Akzeptanz quasi-experimenteller Untersuchungen größer, was sich in verstärkter Repräsentanz nicht-

randomisierter Studien in wissenschaftlichen Zeitschriften innerhalb der letzten Jahre zeigt. Gleichwohl unterscheiden sich diese Studien hinsichtlich ihrer methodischen Qualität deutlich. Es wird auch in Zukunft nötig sein, Interventionen mit einer Varianz an quasi-experimentellen Designs zu untersuchen. So können die Stärken der einen Evaluation die Schwächen der anderen bedienen. Ein erweitertes Bild auf die Wirksamkeit der zu untersuchenden Intervention unter gleichzeitiger Erhaltung einer ausgeprägteren externalen Validität als sie RCT-Designs leisten können, wird dadurch möglich. Als Ansatzpunkt zur Evaluation der Wirkung forensischer Interventionen bieten sich jedoch nicht nur Untersuchungsdesigns an, welche die Rückfallfreiheit der Täter nach der Behandlung als Erfolgskriterium verwenden. Auch hier kommt eine Prozessorientierung ins Spiel. Durch Prä- und Postmessungen lassen sich Veränderungen in deliktrelevanten Problembereichen abbilden, die der Wirkung der Intervention zugeschrieben werden (Klepfisz, O'Brien, & Daffern, 2014; Stinson, Becker, & McVay, 2015). Eine in ihren eigenen Grenzen weitestgehend umfassende Rückfallprävention ist also auch zukünftig auf eine valide Delinquenzhypothese oder „case formulation“ angewiesen, die diese deliktrelevanten Problembereiche im Vorfeld identifiziert. Ob diese Veränderungen schließlich tatsächlich auf die Intervention zurückzuführen sind, stellt hohe Anforderungen an Untersuchungsdesigns zur Kontrolle möglicher Einflussfaktoren wie spezifischer Persönlichkeitsmerkmale des zu Behandelnden und seiner Therapiemotivation, Gruppenkonstellationen, der therapeutischen Beziehung und Aspekte des Behandlungsklimas sowie der Arbeitszufriedenheit des Betreuungspersonals oder der Struktur des sozialen Umfelds bei ambulanten Maßnahmen.

Für politische Entscheidungsträgerinnen und Entscheidungsträger werden es jedoch weiterhin Reduktionen in Rückfallraten sein, die als Erfolgskriterium von Behandlungsprogrammen für Straftäter gelten und deren Implementierung möglich oder unmöglich machen. Daher ist neben aller Notwendigkeit einer prozessorientierten Evaluation forensischer Interventionen methodische und statistische Kreativität gefordert, mit quasi-experimentellen Designs die Effektivität von Straftäterbehandlung am Rückfälligkeitskriterium nachzuweisen und den Korpus an wissenschaftlicher Evidenz dazu substantiell auszubauen.

2 VERÖFFENTLICHUNGEN UND MANUSKRIPTE

2.1 Fazel S, Seewald K (2012) Severe mental illness in 33,588 prisoners worldwide: Systematic review and meta-regression analysis. *The British Journal of Psychiatry* 200(5), 364-373.

2.1.1 Abstract

Background: High levels of psychiatric morbidity in prisoners have been documented in many countries, but it is not known whether rates of mental illness have been increasing over time or whether the prevalence differs between low-middle income nations compared with high income countries.

Aims: To systematically review prevalence studies for psychotic illness and major depression in prisoners, provide summary estimates, and investigate sources of heterogeneity between studies using metaregression.

Method: Studies from January 1, 1966 to December 31, 2010 were identified using 10 bibliographic indexes and reference lists. Keywords included mental*, psych*, prevalence, disorder, prison*, inmate, and jail. Inclusion criteria were unselected prison samples, and that clinical examination or semi-structured instruments were used to make DSM or ICD diagnoses of the relevant disorders. Two researchers independently extracted numbers of prisoners with the relevant disorders and characteristics of the studies.

Results: We identified 109 samples including 33 588 prisoners in 24 countries. Data was meta-analyzed using random effects models, and we found a pooled prevalence of psychosis of 3.6% (95% CI: 3.1–4.2) in male prisoners, and 3.9% (95% CI: 2.7–5.0) in female prisoners. There were high levels of heterogeneity, some of which was explained by studies in low-middle income countries reporting higher prevalences of psychosis (5.5%, 95% CI: 4.2–6.8; $p=0.035$ on metaregression). The pooled prevalence of major depression was 12.4% (95% CI: 8.8–11.7) in male prisoners and 14.1% (95% CI: 10.2–18.1) in female prisoners. The prevalence of these disorders did not appear to be increasing over time, apart from depression in the US ($p=0.008$).

Conclusions: High levels of psychiatric morbidity are consistently reported in prisoners of many countries over four decades. Further research is necessary to confirm whether

higher rates of mental illness are found in low and middle income nations, and examine trends over time within nations with large prison populations.

2.1.2 Introduction

There are over 10 million prisoners worldwide (Walmsley, 2009) a population that has been growing by about one million per decade. In 2008, the US has the largest number of people imprisoned at 2.3 million and highest rate per head of population (at 756 per 100,000 people compared to a median of 145 per 100,000 worldwide), and China, Russia, Brazil, and India had more than a quarter of a million prisoners each (Walmsley, 2009). It has been widely reported that prisoners have elevated rates of psychiatric disorders than the general population, including for psychosis, depression, personality disorder, and substance abuse, which are risk factors for elevated suicide rates (Baillargeon, Penn, et al., 2009; Fazel, Cartwright, Norman-Nott, & Hawton, 2008), premature mortality on release from prison (Kariminia et al., 2007), and increased reoffending rates (Baillargeon, Binswanger, Penn, Williams, & Murray, 2009; Fazel & Yu, 2009). It is estimated that suicide rates within prison are increased 4-5 times (Fazel, Grann, Kling, & Hawton, 2010) and deaths within the first week of release 29 fold higher than the general population (Binswanger et al., 2007). Further, a recent review found that reoffending rates are increased by 40% in offenders with psychotic disorders compared with non-mentally ill offenders (Fazel & Yu, 2009).

A previous systematic review estimated that the prevalence of psychosis was typically 4% in prisoners of both genders, and major depression 10% in men and 12% in women (Fazel & Danesh, 2002). However, this review is now a decade old, and, as many psychiatric institutions have continued to reduce their bed numbers (Priebe et al., 2008), a number of commentators have suggested that rates of severe mental illness have been increasing over time in prisoners (Dressing, Kief, & Salize, 2009) although empirical evidence in support is inconsistent (Bradley-Engen, Cuddeback, Gayman, Morrissey, & Mancuso, 2010) and experts have suggested that measures introduced by the World Health Organization and other international humanitarian agencies have improved prison care (Levy, 2011). In addition, there has been no review, to our knowledge, of the mental health of prisoners in low-middle income countries, although the vast majority of prisoners now live in such countries (Walmsley, 2009). As there is a substantial body of new evidence (Moher & Tsertsvadze, 2006), we have conducted a

new systematic review and meta-analysis of the prevalence of psychosis and major depression, and used subgroup analyses and metaregression to explore possible sources of heterogeneity between studies. We hypothesize that there has been an increase in the rates of psychosis and major depression over time, and that low-middle income countries have higher prevalences of these conditions due to their less resourced community and prison health care services (Saxena, Thornicroft, Knapp, & Whiteford, 2007).

2.1.3 Methods

We identified publications estimating the prevalence of psychotic disorders (including psychosis, schizophrenia, schizophreniform disorders, manic episodes) and major depression among prisoners which were published between 1 January 1966 and 31 December 2010. For the period 1 January 1966 to 31 December 2000, methods are described in a previous systematic review conducted by one of the authors (SF; Fazel & Danesh, 2002). For the update and expanded review, from 1 January 2001 to 31 December 2010, we used the following databases: PsycINFO, Global Health, MEDLINE, Web of Science, PubMed, National Criminal Justice Reference Service, EMBASE, OpenSIGLE, SCOPUS, Google Scholar, scanned references, and corresponded with experts in the field (figure 1). Key words used for the database search were the following: mental*, psych*, prevalence, disorder, prison*, inmate, jail and also combinations of those. Non-English articles were translated. PRISMA criteria were followed (Moher et al., 2009).

Inclusion criteria were: (1) study population was sampled from a general prison population; (2) diagnoses of the relevant disorders were made by clinical examination or by interviews using validated diagnostic instruments; (3) diagnoses met standardized diagnostic criteria for psychiatric disorders based on the International Classification of Mental and Behavioural Disorders (ICD) or the Diagnostic and Statistical Manual of Mental Disorders (DSM) and (4) prevalence rates for the relevant disorders in the previous 6 months.

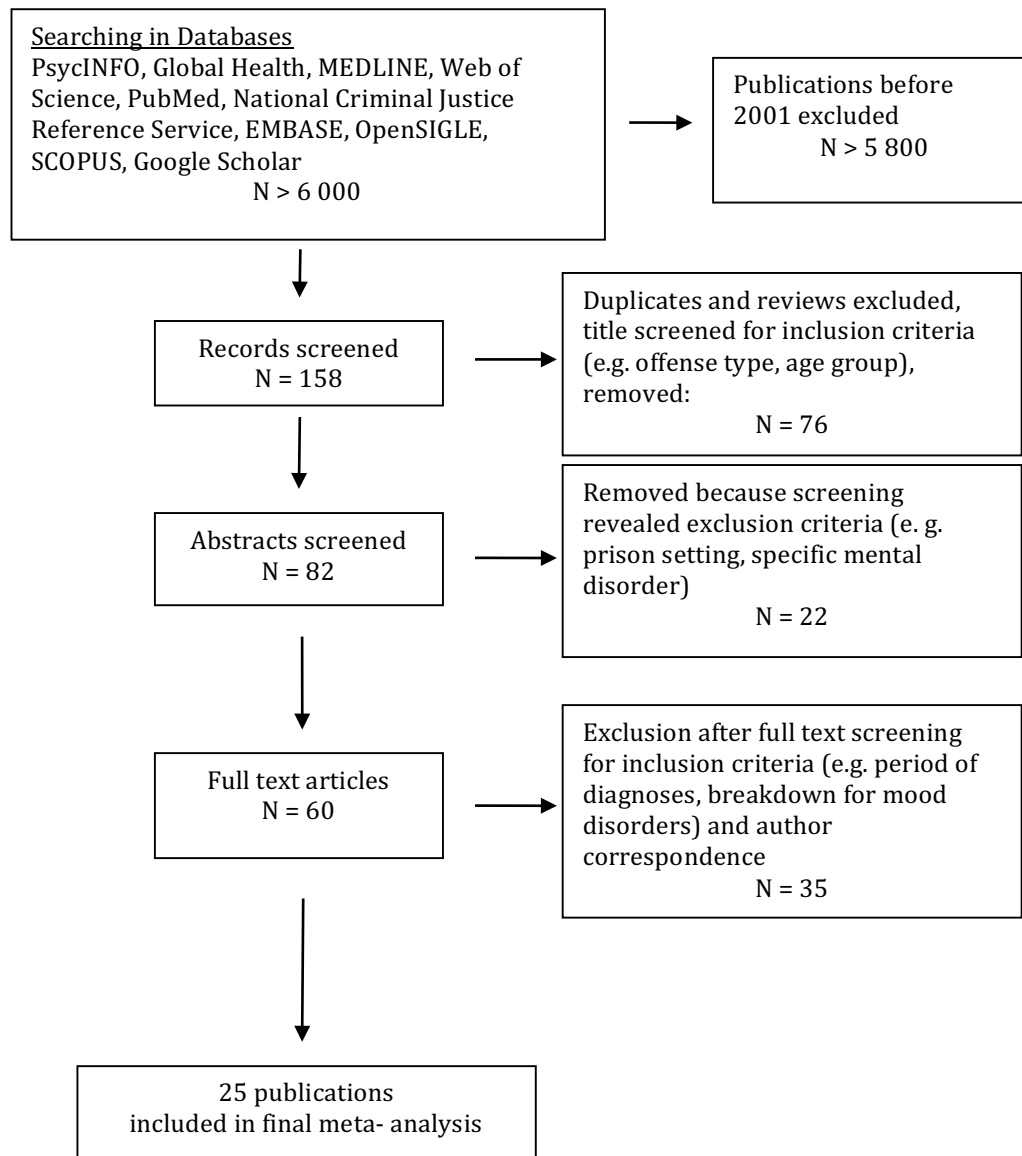


Figure 1: Flow diagram showing the different steps in searching of relevant publications (2001-2010).

In order to include unselected, representative, and generalizable prison samples we only selected studies that conducted a diagnostic interview with a general prison population. We excluded those studies that used a screening tool before conducting the diagnostic interview (Baillargeon, Binswanger, et al., 2009; O'Keefe & Schnell, 2007; Steadman, Osher, Robbins, Case, & Samuels, 2009), as this may lead to underestimate if the screen had poor sensitivity or overestimates if the screen had poor specificity. We also excluded sampled selected populations, e.g. by offense type (Kanyanya, Othieno, &

Ndetei, 2007; Kugu, Akyuz, & Dogan, 2008), as there is evidence that selecting some offender groups may also lead to overestimates which is particularly the case for murder and attempted murder (Fazel & Grann, 2004); age group (Fazel, Hope, O'Donnell, Piper, & Jacoby, 2001); including solely juvenile prisoners (Ajiboye, Yussuf, & Issa, 2009) or prisoners who were in health care settings (Morgan et al., 2010). For example, one study that used a screening tool to identify mentally ill prisoners and was excluded reported a prevalence of 32% for schizophrenia (Lamb, Weinberger, Marsh, & Gross, 2007). Studies that did not separate sentenced from remand prisoners in their report (Segagni Lusignani et al., 2006) and duplicates were excluded. In this update, we did not include personality disorder due to the high heterogeneity reported in the previous work (Fazel & Danesh, 2002). For substance abuse and post-traumatic stress disorder, there are more recent reviews (Fazel, Bains, & Doll, 2006; Goff, Rose, & Purves, 2007) and a substantial body of new work has not emerged since their publication (Abdalla-Filho, De Souza, Tramontina, & Taborda, 2010).

Data extraction

We extracted information on the year of interview, geographical location, gender of prisoners, remand/detainee (including jail inmates) or sentenced prisoner, average age, method of sampling, sample size, participation rate, type of interviewer, diagnostic instrument, diagnostic criteria (ICD vs. DSM), numbers diagnosed with psychotic disorders (ICD 10: F20.xx-F29.xx, F30.xx; DSM IV: 293.xx, 295.xx), and major depression (ICD 10, F 32.xx, F33.xx; DSM IV: 296.2x, 296.3x). Where there was schizophrenia and other psychotic disorders reported separately, we combined them to a single prevalence. For one publication (Falissard, Loze, Gasquet, Duburc, & de Beaurepaire, 2006), we collated „Dépression endogène-Mélancolie”, „Etat dépressif chronique” and „Symptômes psychotiques contemporains des épisodes thymiques” as indicating major depression. The data from each of the identified publications was subdivided into up to four samples (men vs. women, remand/detainee vs. sentenced prisoners). In contrast to our previous review, we included data on low-middle income countries (The World Bank, 2010) and whether the clinical diagnostic interview was conducted within 2 weeks of arrival into the prison (which may influence prevalence rates and also provides an estimate of mental health needs on reception to prison). For the update, we examined rates of comorbidity with substance abuse. The data extraction was done by

two researchers independently (KS and KW). For further clarification about specific studies, we corresponded with authors.

Data analysis

We analyzed sources of heterogeneity by subgroup and metaregression analysis using dichotomous and continuous variables. The year that the interview was conducted and the average age of the prisoners were analyzed as continuous variables. Sample size and response rate were analyzed as both dichotomous and continuous variables. As the median of reported response rates was 81%, we defined „low” as $\leq 80\%$ versus „high” as $>80\%$). The following were analyzed only as dichotomous variables: gender, prisoner type (detainees/remand vs. sentenced prisoners), reception status (interviewed in the first two weeks of reception vs. the rest), type of interviewer (psychiatrist vs. non-psychiatrist), diagnostic instrument (clinical examination vs. semi-structured interview using a diagnostic tool) and classification criteria (ICD vs. DSM). Geographical location was analyzed as low-middle income vs. high income country (The World Bank, 2010). Here we included the US studies within the high income country group. Also, we conducted an additional separate analysis of US studies (vs. rest of world and vs. rest of high income countries) for three reasons: first, there are over 2 million prisoners there (around a fifth of the world prisoner population), second, they constituted 30% of the included studies in the review, and third, mentally disordered prisoners in the US are less likely to be diverted due to judicial and legal reasons, and hence this may contribute to higher prevalence rates (Baillargeon, Binswanger, et al., 2009).

We used a recent method for further examination of heterogeneity, which involves removing up to four outliers and testing whether this reduces I^2 values to below 50%, and then investigating in more detail the study characteristics of these outliers (Patsopoulos, Evangelou, & Ioannidis, 2008).

We calculated pooled prevalence estimates and their 95% confidence intervals (CIs) and transformed the zero cells to 0.5 in order to calculate prevalences as per standard methods (Sweeting, Sutton, & Lambert, 2004). Meta-analyses for prevalences were conducted by gender and prisoner status. We measured the heterogeneity between studies with Cochran’s Q (reported with a χ^2 -value and p-value) and the I^2 statistic (with 95% CIs; Ioannidis, Patsopoulos, & Evangelou, 2007) and used random effects models for summary statistics as heterogeneity was high ($I^2 > 75\%$; Higgins & Thompson, 2002;

Higgins, Thompson, Deeks, & Altman, 2003). The I^2 is an estimate of the proportion of the total variation across studies that is beyond chance. In situations with high between-study heterogeneity, the use of random effects models is recommended as it produces study weights that primarily reflect the between-study variation and thus provide close to equal weighting. Univariate and multivariate metaregression analyses were used to explore possible sources of heterogeneity among studies (Ioannidis, Patsopoulos, & Rothstein, 2008). Factors in univariate metaregression with p-values of <0.1 were included in the final model. We also conducted a test of funnel plot asymmetry (Egger's test) for publication bias using the publication (rather than the sample) as the unit of measurement. A funnel plot is a plot of the estimated prevalence against the sample size of the included studies. Egger's test can reveal a symmetric or asymmetric funnel plot. The latter indicates the existence of a significant publication bias or a systematic heterogeneity between studies (Egger, Smith, Schneider, & Minder, 1997). All analyses were done in STATA statistical software, version 11.1 (StataCorp, 2009).

2.1.4 Results

Study characteristics

The final data set consisted of 81 publications, 56 based on the previous review from the period 1966 - 2001 (Agbahowe, Ohaeri, Ogunlesi, & Osahon, 1998; Andersen, Sestoft, Lillebaek, Gabrielsen, & et al., 1996; Bartholomew, Brain, Douglas, & Reynolds, 1967; Birmingham, Mason, & Grubin, 1996; Bland, Newman, Dyck, & Orn, 1990; Bluglass, 1966; Brinded et al., 1999; Brooke, Taylor, Gunn, & Maden, 1996; B. Bulten, 1998; Cooke, 1994; Daniel, Robins, Reid, & Wilfley, 1988; Davidson, Humphreys, Johnstone, & Owens, 1995; Denton, 1995; DiCataldo, Greer, & Profit, 1995; Eyestone, 1994; Faulk, 1976; Fido & al-Jabally, 1993; Ghubash & El-Rufaie, 1997; Gibson et al., 1999; Gunn, Maden, & Swinton, 1991; Gunn, Robertson, Dell, & Way, 1978; Guy, Platt, Zwerling, & Bullock, 1985; Haapasalo, 2000; Harper & Barry, 1979; Herrman, McGorry, Mills, & Singh, 1991; Hurley & Dunne, 1991; Hurwitz & Christiansen, 1983; Hyde & Seiter, 1987; James, Gregory, Jones, & Rundell, 1980; Jordan, Schlenger, Fairbank, & Caddell, 1996; Joukamaa, 1993, 1995; Krefft & Brittain, 1983; Levander, Svalenius, & Jensen, 1997; A. Maden, Taylor, Brooke, & Gunn, 1996; T. Maden, Swinton, & Gunn, 1994; Mohan, Scully, Collins, & Smith, 1997; Motiuk & Porporino, 1992; Neighbors et al., 1987; Powell, Holt, & Fondacaro, 1997; Poythress et al., 1998; Rasmussen, Storsaeter, & Levander, 1998;

Robertson, 1990; Robins & Reiger, 1991; Roesch, 1995; Schoemaker & VanZessen, 1997; Schuckit, Hermann, & Schuckit, 1977; Simpson A, Brinded P, Laidlaw T, Fairley N, & Malcolm F, 1999; C. Smith, O'Neill, Tobin, Walshe, & Dooley, 1996; Swank & Winer, 1976; L. A. Teplin, 1994; Linda A. Teplin, Abram, & McClelland, 1996; van Panhuis, 1997; Washington & Diamond, 1985; Watt, Tomison, & D, 1993; Wilkins & Coid, 1997) and 25 new ones (see supplementary table 1: Alevizopoulos & Igomenou, 2011; Arnab, Prativa, & Ray, 2009; Assadi et al., 2006; Bermúdez, Romero Mendoza, Rodríguez Ruiz, Durand-Smith, & Saldívar Hernández, 2007; E. Bulten, Nijman, & van der Staak, 2009; Butler & Allnutt, 2003; Curtin et al., 2009; Dudeck, Kopp, & Kuwert, 2009; Duffy, Linehan, & Kennedy, 2006; Falissard et al., 2006; Fotiadou, Livaditis, Manou, Kaniotou, & Xenitidis, 2006; Gunter et al., 2008; Linehan et al., 2005; Parsons, Walker, & Grubin, 2001; Piselli, Elisei, Murgia, Quartesan, & Abram, 2009; Ponde, Cruz Freire, & Santos Mendonca, 2011; Sharma, Nijhawan, Sharma, & Sushil, 2010; Stompe, Brandstätter, Ebner, & Fischer-Danzinger, 2010; Tort, 2009; Trestman, Ford, Zhang, & Wiesbrock, 2007; von Schoenfeld et al., 2006; Watzke, Ullrich, & Marneros, 2006; Wright et al., 2006; Zahari et al., 2010; Zoccali, Muscatello, Bruno, Cambria, & Cavallaro, 2008). These publications provided data on 109 samples that included a total of 33 588 prisoners. Of these, 28 361 (84.4%) were male. The overall weighted mean age was 30.5 years. The studies were conducted in 24 different countries, 8 of which are classified as low-middle income countries: Brazil (Ponde et al., 2011), Dubai (Ghubash & El-Rufaie, 1997), India (Arnab et al., 2009; Sharma et al., 2010), Iran (Assadi et al., 2006), Kuwait (Fido & al-Jabally, 1993), Malaysia (Zahari et al., 2010), Mexico (Bermúdez et al., 2007), and Nigeria (Agbahowe et al., 1998).

There were 72 studies from high income countries. There were 25 studies from the US (Birmingham et al., 1996; Daniel et al., 1988; DiCataldo et al., 1995; Eyestone, 1994; Gibson et al., 1999; Gunn et al., 1991; Gunn et al., 1978; Gunter et al., 2008; Guy et al., 1985; Harper & Barry, 1979; Hyde & Seiter, 1987; James et al., 1980; Jordan et al., 1996; Krefft & Brittain, 1983; Neighbors et al., 1987; Powell et al., 1997; Poythress et al., 1998; Robins & Reiger, 1991; Schuckit et al., 1977; Simpson A et al., 1999; Swank & Winer, 1976; L. A. Teplin, 1994; Linda A. Teplin et al., 1996; Trestman et al., 2007; Washington & Diamond, 1985), three from Canada (Bland et al., 1990; Motiuk & Porporino, 1992; Roesch, 1995), five from Australia (Bartholomew et al., 1967; Butler & Allnutt, 2003; Denton, 1995; Herrman et al., 1991; Hurley & Dunne, 1991), and one from New Zealand

(Brinded et al., 1999). The remaining studies were conducted in Europe including seven in England and Wales (Brooke et al., 1996; Faulk, 1976; A. Maden et al., 1996; T. Maden et al., 1994; Parsons et al., 2001; Robertson, 1990; Watt et al., 1993; Wilkins & Coid, 1997), six in Ireland, (Curtin et al., 2009; Duffy et al., 2006; Linehan et al., 2005; Mohan et al., 1997; C. Smith et al., 1996; Wright et al., 2006) and a number in the Netherlands (B. Bulten, 1998; E. Bulten et al., 2009; Schoemaker & VanZessen, 1997; van Panhuis, 1997), Finland (Haapasalo, 2000; Joukamaa, 1993, 1995), Germany (Dudeck et al., 2009; von Schoenfeld et al., 2006; Watzke et al., 2006), Scotland (Bluglass, 1966; Cooke, 1994; Davidson et al., 1995) Denmark (Andersen et al., 1996; Hurwitz & Christiansen, 1983), Italy (Piselli et al., 2009; Zoccali et al., 2008), Greece (Alevizopoulos & Igomenou, 2011; Fotiadou et al., 2006), Austria (Stompe et al., 2010), France (Falissard et al., 2006), Norway (Rasmussen et al., 1998), Spain (Tort, 2009), and Sweden (Levander et al., 1997).

Nine studies reported results from interviews carried out within 2 weeks of arrival into the prison (E. Bulten et al., 2009; Butler & Allnutt, 2003; Curtin et al., 2009; Duffy et al., 2006; Gunter et al., 2008; Parsons et al., 2001; Stompe et al., 2010; Trestman et al., 2007; Wright et al., 2006), two of them without giving information about the prisoner type (remand/detainee or sentenced; Butler & Allnutt, 2003; Wright et al., 2006).

Psychotic Illnesses

We identified 99 samples from 74 studies that reported rates of psychotic illnesses and included a total of 30 635 prisoners (Agbahowe et al., 1998; Alevizopoulos & Igomenou, 2011; Andersen et al., 1996; Arnab et al., 2009; Assadi et al., 2006; Bartholomew et al., 1967; Birmingham et al., 1996; Bland et al., 1990; Bluglass, 1966; Brinded et al., 1999; Brooke et al., 1996; B. Bulten, 1998; E. Bulten et al., 2009; Cooke, 1994; Curtin et al., 2009; Daniel et al., 1988; Davidson et al., 1995; Denton, 1995; DiCataldo et al., 1995; Dudeck et al., 2009; Duffy et al., 2006; Falissard et al., 2006; Faulk, 1976; Fido & al-Jabally, 1993; Fotiadou et al., 2006; Gibson et al., 1999; Gunn et al., 1991; Gunn et al., 1978; Gunter et al., 2008; Guy et al., 1985; Harper & Barry, 1979; Herrman et al., 1991; Hurley & Dunne, 1991; Hurwitz & Christiansen, 1983; Hyde & Seiter, 1987; James et al., 1980; Joukamaa, 1993, 1995; Krefft & Brittain, 1983; Levander et al., 1997; Linehan et al., 2005; A. Maden et al., 1996; T. Maden et al., 1994; Mohan et al., 1997; Motiuk & Porporino, 1992; Neighbors et al., 1987; Parsons et al., 2001; Piselli et al., 2009; Ponde et

al., 2011; Powell et al., 1997; Poythress et al., 1998; Rasmussen et al., 1998; Robertson, 1990; Robins & Reiger, 1991; Roesch, 1995; Schoemaker & VanZessen, 1997; Schuckit et al., 1977; Sharma et al., 2010; Simpson A et al., 1999; C. Smith et al., 1996; Stompe et al., 2010; Swank & Winer, 1976; L. A. Teplin, 1994; Linda A. Teplin et al., 1996; Tort, 2009; Trestman et al., 2007; van Panhuis, 1997; von Schoenfeld et al., 2006; Watt et al., 1993; Watzke et al., 2006; Wilkins & Coid, 1997; Wright et al., 2006; Zahari et al., 2010; Zoccali et al., 2008). Overall, we calculated a random effects pooled prevalence of 3.6% (95% CI: 3.1–4.2) in male prisoners (1 120 of 26 814 individuals), and 3.9% (95% CI: 2.7–5.0) in female prisoners (182 of 3 821 individuals). There was significant heterogeneity among these studies in the men ($\chi^2=416$, $p<0.0001$; $I^2=83\%$, 95% CI: 79-86) and women prisoners ($\chi^2 = 86$, $p<0.0001$; $I^2 = 68\%$, 95% CI: 54-79).

There was a significant difference in the prevalences in low-middle income countries (5.5%, 95 % CI: 4.2–6.8) compared with high income countries (3.5%, 95 % CI: 3.0–3.9) (figure 2), confirmed by metaregression (table 2; $\beta=0.0204$, $se(\beta)=0.0095$, $p=0.035$). We did not find any difference in prevalences between male and female prisoners, between detainees/remand and sentenced prisoners (table 1), and no statistically significant change in prevalence over time ($\beta=-0.0001$, $se(\beta)=0.0002$, $p=0.84$; figure 3). When we looked specifically at US studies (17 samples), there also appeared to be no change over time ($\beta=-0.0006$, $se(\beta)=0.0005$, $p=0.24$). There was evidence of an asymmetric funnel plot (Egger's test; $t=239.32$, $se(t)=0.0044$, $p<0.001$).

Table 1: Pooled prevalences for psychosis and major depression in prisoners

Variable	Psychosis, % (95% CI)	Major Depression, % (95% CI)
Overall	3.7 (3.2–4.1)	11.4 (9.9–12.8)
Gender of inmates		
Male	3.6 (3.1–4.2)	10.2 (8.8–11.7)
Female	3.9 (2.7–5.0)	14.1 (10.2–18.1)
Prisoner status		
Sentenced prisoners	3.7 (3.0–4.2)	10.5 (8.8–12.1)
Remand prisoners (detainees)	3.5 (4.2–6.8)	12.3 (9.5–15.1)
Country		
Low/middle income	5.5 (4.2–6.8)	22.5 (10.6–34.4)
High income	3.5 (3.0–3.9)	10.0 (8.7–11.2)

Table 2: Metaregression analyses of sources of heterogeneity in the prevalence of psychosis and major depression in prisoners

Study characteristic	Psychosis			Major Depression			
	β	se (β)	p-value	β	se (β)	p-value	
Sex of inmates male vs. female	0.0016	0.063	0.800	0.0323	0.0222	0.151	
Mean age of inmates Continuous	-0.0009	0.0009	0.334	0.0011	0.0033	0.735	
Year of study Continuous	<0.0001	0.0003	0.889	0.0015	0.0013	0.340	
Country	Low/middle vs. high income	0.0204	0.0095	0.035	0.1157	0.0318	0.001
	US vs. rest of the world	0.0007	0.0060	0.902	-0.0043	0.0241	0.859
	within the US, over time	-0.0006	0.0038	0.504	0.0038	0.0013	0.008
Prisoner status Sentenced prisoners vs. detainees	0.0025	0.0038	0.504	0.0136	0.0159	0.396	
On reception First two weeks of reception vs. rest	-0.0016	0.0058	0.778	0.0131	0.0241	0.588	
Participation rate	Continuous	-0.0184	0.0333	0.585	-0.0778	0.0998	0.441
	Low ($\leq 80\%$) vs. high ($> 80\%$)	-2.637	3.0964	0.400	-0.2786	1.1458	0.809
Sample size	Continuous	<0.0001	0.0000	0.337	<0.0001	0.0000	0.577
	≤ 500 vs. >500	0.0088	0.0062	0.156	-0.0147	0.0275	0.594
Interviewer Psychiatrist vs. other	-0.0015	0.0053	0.784	0.0199	0.0231	0.391	
Diagnostic Criteria ICD vs. DSM	-0.0021	0.0055	0.706	-0.0590	0.025	0.021	

Significant associations ($p < 0.05$) are in bold. For comparisons the reference category is given first.

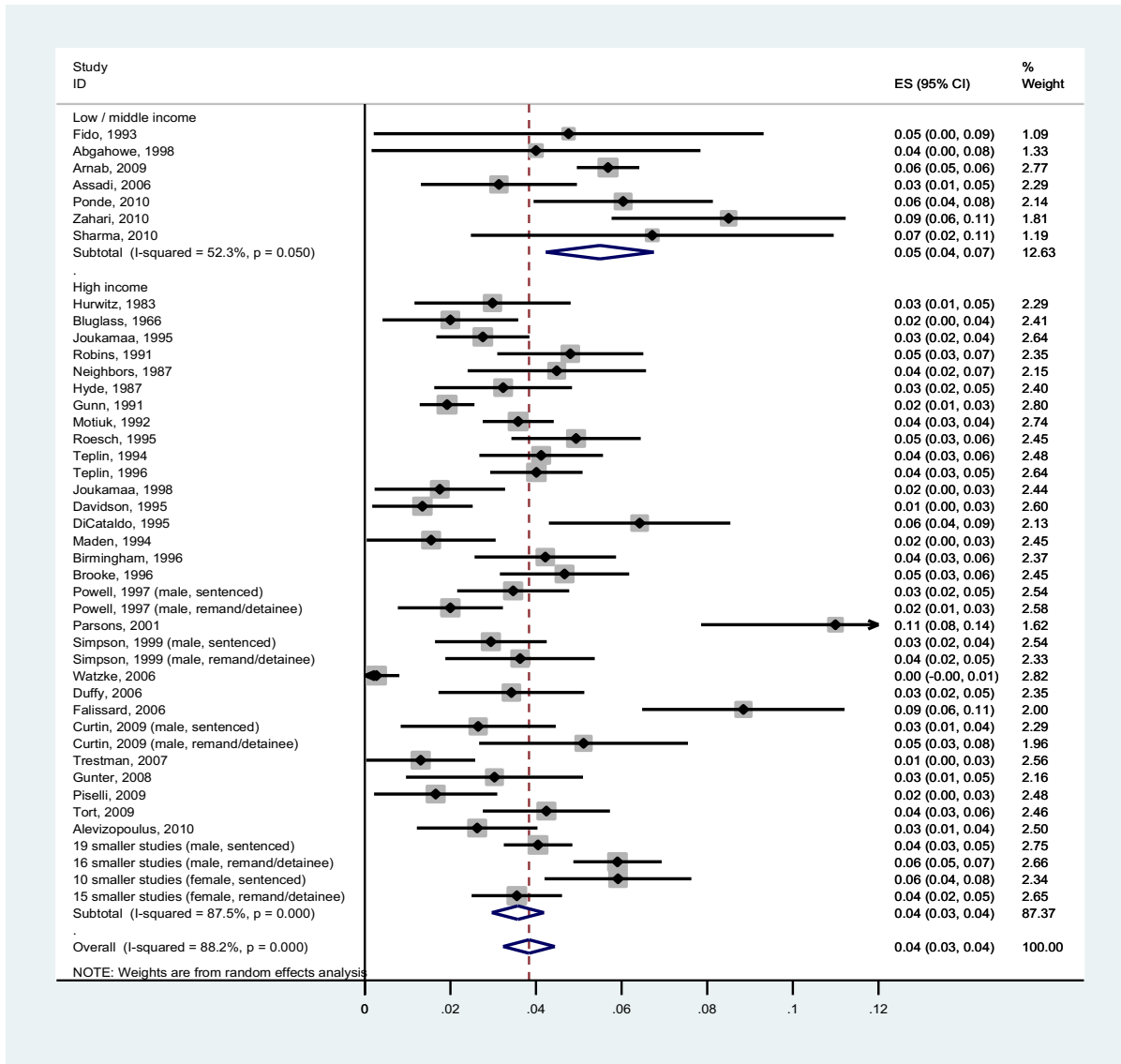


Figure 2: Meta-analysis of the prevalence of psychotic illnesses in prisoners by country group (low-middle income vs. high income).

Weights are from random-effects analysis. Smaller studies: n<250. ES, prevalence.

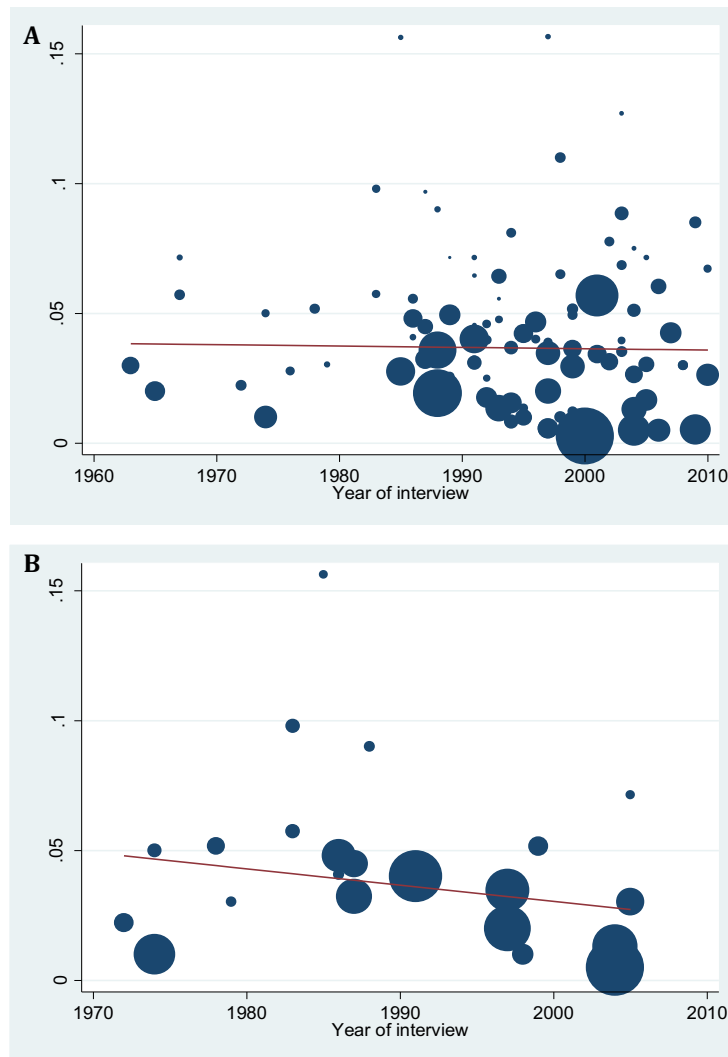


Figure 3: Prevalence of psychotic illness in prisoners over time in (a) individual studies from all countries (including the USA) and (b) studies conducted in the USA only.

The size of the circles is proportional to the sample size of each study.

Major Depression

We identified 54 publications that reported rates of major depression in 20 049 prisoners (Agbahowe et al., 1998; Alevizopoulos & Igomenou, 2011; Andersen et al., 1996; Assadi et al., 2006; Bermúdez et al., 2007; Bland et al., 1990; Brinded et al., 1999; Brooke et al., 1996; B. Bulten, 1998; E. Bulten et al., 2009; Butler & Allnutt, 2003; Curtin et al., 2009; Daniel et al., 1988; Denton, 1995; DiCataldo et al., 1995; Dudeck et al., 2009; Duffy et al., 2006; Eyestone, 1994; Falissard et al., 2006; Faulk, 1976; Fido & al-Jabally, 1993; Fotiadou et al., 2006; Ghubash & El-Rufaie, 1997; Gibson et al., 1999; Gunter et al., 2008; Haapasalo, 2000; Herrman et al., 1991; Hurley & Dunne, 1991; Hyde & Seiter, 1987; Jordan et al., 1996; Linehan et al., 2005; Mohan et al., 1997; Neighbors et al., 1987; Parsons et al., 2001; Piselli et al., 2009; Ponde et al., 2011; Powell et al., 1997; Poythress

et al., 1998; Rasmussen et al., 1998; Robins & Reiger, 1991; Roesch, 1995; Schoemaker & VanZessen, 1997; Schuckit et al., 1977; Simpson A et al., 1999; Stompe et al., 2010; Linda A. Teplin et al., 1996; Tort, 2009; von Schoenfeld et al., 2006; Washington & Diamond, 1985; Watt et al., 1993; Watzke et al., 2006; Wilkins & Coid, 1997; Wright et al., 2006; Zahari et al., 2010). Overall, 12.4% (95% CI: 8.8–11.7) of male prisoners (1 686 of 16 021 individuals), and 14.1% (95% CI: 10.2–18.1) of female prisoners (605 of 4 028), were diagnosed with major depression (figure 4). There was significant heterogeneity among these studies in males ($\chi^2=541$, $p<0.0001$; $I^2=91\%$, 95% CI: 89-93) and also in females ($\chi^2=307$, $p<0.0001$; $I^2=93\%$, 95% CI: 90-94). Even after the exclusion of four outliers in both genders, the I^2 remained above 50%.

There was no significant difference in the prevalence of depression between men and women. However, there appeared to higher prevalences in those studies using DSM criteria and in low-middle income countries confirmed on univariate metaregression (table 2). While there was no evidence for rates of major depression changing over time in the non-US samples, the prevalence of depression appeared to be increasing over time in the US samples, of which there were 17 from 1970 to 2010 ($\beta=0.0038$, $se(\beta)=0.0013$, $p=0.008$; figure 5).

In a multivariate metaregression analysis combining both income group and classification criteria, the finding of a higher prevalence rate in low-middle countries remained significant only for women prisoners and was based on a single Mexican study (Bermúdez et al., 2007). In the US studies, multivariate metaregression was not possible as all the samples that reported information on classification criteria used DSM. However in the non-US and high income samples, classification criteria still remained significant when income group was included in the model and DSM studies reported higher prevalences of depression ($\beta=-0.0645$, $se(\beta)=0.0282$, $p=0.026$). There was evidence of an asymmetric funnel plot ($t=27.78$, $se(t)=0.0452$, $p<0.001$).

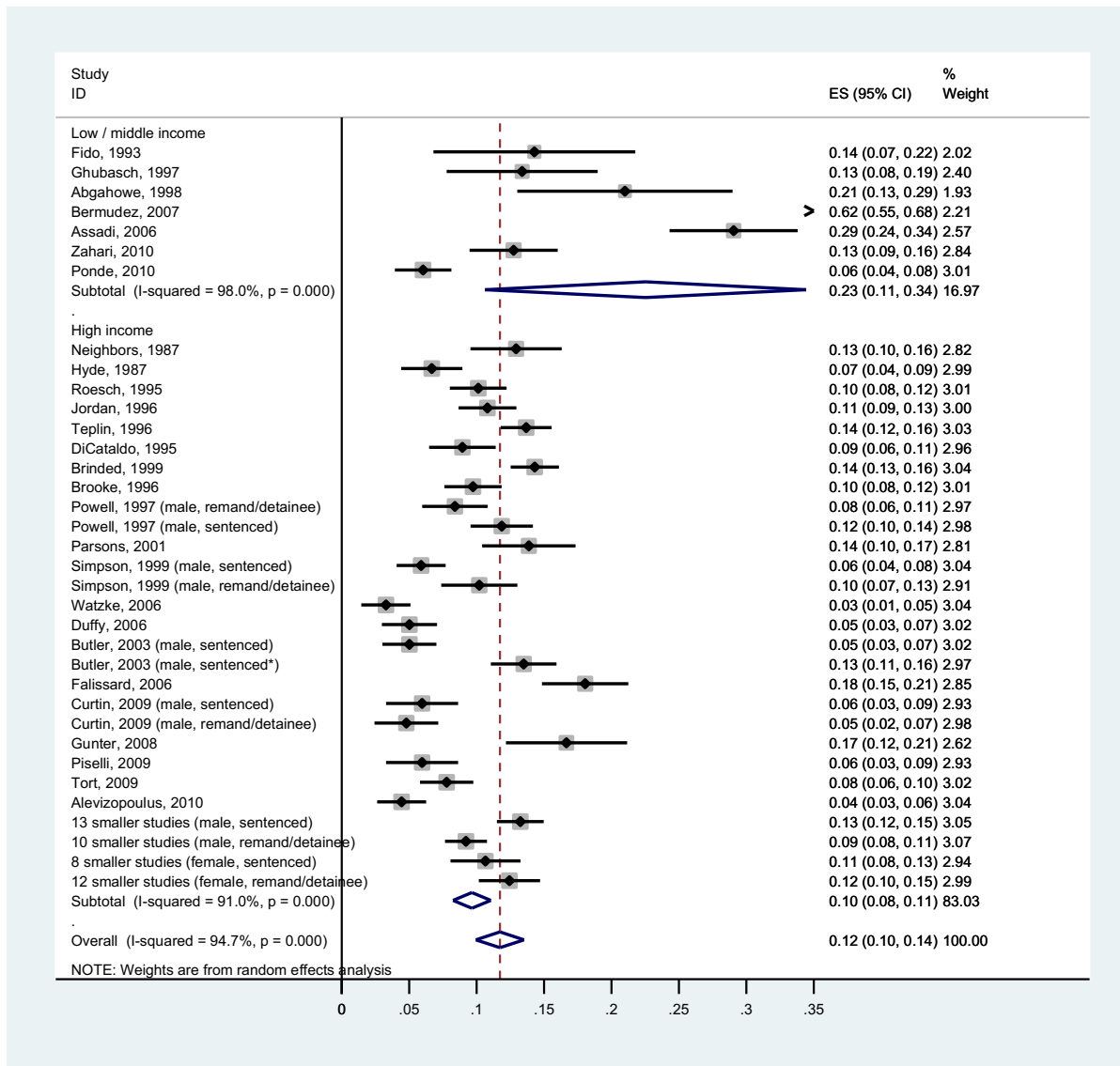


Figure 4: Meta-analysis of the prevalence of major depression in prisoners by country group (low-middle income vs. high income).

Weights are from random-effects analysis. Smaller studies: n<250. ES, prevalence. a. On early reception.

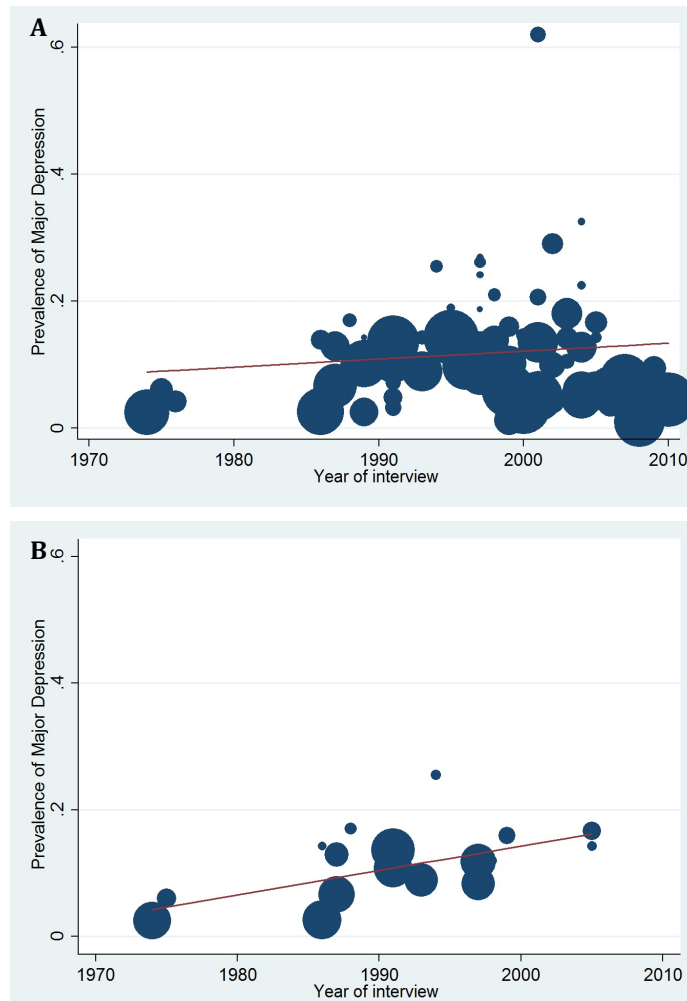


Figure 5: Prevalence of major depression in prisoners over time in (a) individual studies from all countries (including the USA) and (b) studies conducted in the USA only.

The size of the circles is proportional to the sample size of each study.

Comorbidity

There were five publications since 2001 that reported rates of comorbidity in prisoners (Assadi et al., 2006; Bermúdez et al., 2007; Butler & Allnutt, 2003; Duffy et al., 2006; Piselli et al., 2009). These rates ranged from 20.4% to 43.5% in those with any mental disorder who had comorbid substance abuse, from 13.6% to 95.0% in prisoners with psychotic illnesses with comorbid substance abuse, and 9.2% to 82.5% in individuals with mood disorders and major depression with concurrent substance abuse (supplementary table 2).

2.1.5 Discussion

We report a systematic review of the prevalence of psychosis and depression in prisoners based on 109 separate samples (from 81 publications) based on 33 588 prisoners. In addition, we have, for the first time to our knowledge, reviewed research in low and middle income countries (based on 5 792 prisoners), and employed metaregression analyses to explore sources of heterogeneity between studies. In particular, we have examined whether rates of mental illness in prisoners have been increasing over time.

Our main findings were that rates of psychosis in prisoners were significantly higher in low and middle income countries than high income ones (5.5% in low/middle vs. 3.5% in high income nations).

Contrary to expert opinion, there were no significant differences in rates of psychosis and depression between men and women prisoners (World Health Organization, 2009) and between detainees (or remand) and sentenced prisoners. In the 17 US samples included, there appeared to be an increasing prevalence of depression over the 31 years covered by these particular studies (1974-2005). In addition, we found no differences in depression rates between men and women, detainees (or remand) and sentenced prisoners, or other study characteristics that may have explained heterogeneity. The overall prevalences of 3.7% of male and female prisoners with a psychotic illness, and 11.4% with major depression have not materially changed since a 2002 review based on 56 publications of mental illness (Fazel & Danesh, 2002).

In contrast to one of our initial hypotheses, we did not find an increase in rates of psychosis and depression over time. The reasons for this are unclear but improvements in psychiatric care in prison, increased diversion of mentally disordered offenders from prison to hospital, and better living conditions may have contributed (Skeem, Manchak, & Peterson, 2011). International organizations over the last two decades in improving prison health has also been suggested to have played a role (Fraser, Møller, & van den Bergh, 2011).

Three main implications arise from these findings. First, the substantial burden of treatable psychiatric morbidity is confirmed by these findings. One in seven prisoners has depression or psychosis, and treatment may confer additional benefits such as reducing risks of suicide (Fazel et al., 2008) and self-harm (Lohner & Konrad, 2007)

within custody, suicide (Bird, 2008; Pratt, Appleby, Piper, Webb, & Shaw, 2010) and drug-related deaths on release (Kariminia et al., 2007) and reoffending (Baillargeon, Binswanger, et al., 2009; Fazel & Yu, 2009). As reoffending rates are high (at 50% in the US and UK within 2 years of release; 2002; Langan & Lewin) treatment of prisoners may have a potentially large impact on public safety. In this context, the lack of good quality treatment evidence remains concerning (Fazel & Baillargeon, 2011). The role of diversion away from prison at early stages of the criminal process and other collaborations between mental health and the justice system is underscored by our findings (Hassan et al., 2011; Morrissey, Fagan, & Cocozza, 2009; Steadman et al., 2009) particularly as repeat incarcerations are associated with mental illness (Baillargeon, Binswanger, et al., 2009).

Second, the higher prevalence of psychosis in prisoners in low and middle income countries is notable as rates of imprisonment are increasing in more of these countries than high income ones (Walmsley, 2009), and possibly faster, and service provision is likely to be worse. Health services in such countries can potentially use the estimates reported in this review in developing prison medical services, particularly in countries where resources are unlikely to allow for local prevalence studies to be conducted. In poorer countries, the role of explicit mental health budgets in ongoing health programmes could be considered, particularly for marginalized populations such as prisoners (Prince et al., 2007). Our report does not provide information on the causes for higher prisoner rates of psychosis in low and middle income nations but possibilities include fewer opportunities and services for diverting offenders to health services, a stronger relationship between mental illness and criminality, and different socio-cultural factors that mean that more mentally ill persons end up in prison. Poorer legal representation for the mentally ill may be one such factor. The increased comorbidity with opioids in prisoners found in some countries that form part of the illegal drug trade may be another (Assadi et al., 2006).

A final implication from this review is that, although internationally the prevalence of depression does not appear to be increasing in prisoners, in the US, which has the largest prison population worldwide, the rate of depression appears to have been increasing over time. This was not found for psychosis in prisoners internationally or in the US, which may be partly due to the fact that the incidence of psychotic disorders has not increased in the general population either (McGrath et al., 2004). In relation to

increased depression in US prisoners, further work could investigate the possible contributions of the closure of large psychiatric hospitals, the provision of community care, the funding of mental health, and the reported increase in major depression rates in the general population (Kessler et al., 2003). Whatever the causes, the US houses more than three times more mentally ill persons in prison than in all psychiatric hospitals (Torrey, Kennard, Eslinger, Lamb, & Pavle, 2010), and under-treatment for mental illness in US prisons exacerbates these problems (Wilper et al., 2009). Simple measures, including having policies and guidelines for the transfer of the severely mental ill to psychiatric hospitals, training of prison staff, and discharge planning, may improve these rates (Torrey et al., 2010).

The high levels of heterogeneity between the studies are to be expected as the studies were conducted by different groups in a large variety of prisons using differing methods (Higgins, 2008), and this may simply reflect real differences in prevalences over time and by region. This may also be an explanation for the asymmetric funnel plots we reported in addition to possible publication bias (Egger et al., 1997). In addition, publication bias may also explain the small number of studies in low and middle income countries, and publication bias is thought to contribute to this in all mental health research from these countries (D. Singh, 2006).

Our approach to this was to identify causes of heterogeneity, and two possible explanations were found. In depression, we found that studies using DSM criteria had higher rates than those using ICD criteria. Although such differences have occasionally been found in community studies, and a lower congruence between the two diagnostic systems for depression diagnoses compared to some other psychiatric disorders has also been reported (Cheniaux, Landeira-Fernandez, & Versiani, 2009), particular reasons for this difference in prisoners are unclear. Possibilities include that in the diagnostic systems, fatigability is included in the core criteria for depression in ICD, but an associated feature in DSM. In addition, it may be that the distinction between melancholic and non-melancholic forms of depression (Parker, 2000) is more important in prisoners as the overlap between sadness and clinical depression is more difficult to determine.

The strengths of this review include the large number of samples and prisoners included, and therefore the ability to examine prevalences by clinically relevant subgroups with some degree of precision. However, we identified only eight studies in

low and middle income countries, and our findings should be interpreted with caution. Furthermore, we have examined heterogeneity using subgroup analyses and metaregression, which allowed us to investigate dichotomous and continuous variables such as age, sample size, and the date in which the study was conducted. One of the limitations of the review is there may be other explanations for the heterogeneity that we did not test, such as comorbidity with other mental disorders but systematic data was lacking on this. Furthermore, the statistical power of testing trends within nations was limited, and even our findings on US trends were based on 17 studies.

A number of research implications arise from this review. First, studying the epidemiology of mental illness and criminality in low and middle income countries and how it compares to high income countries may provide some reasons for the difference in psychosis prevalence. A recent review found no such studies in low and middle income countries.(Fazel, Gulati, et al., 2009) More research into the treatment of mentally ill prisoners and the most effective models of service delivery is pressing, and further comparison of novel approaches needs further examination (Raimer & Stobo, 2004).

Future prison surveys should include information on comorbidity and past psychiatric history, suicide attempts within custody, treatment received in prison and compliance, and length of custody. In addition, the relationship between mental illness in prisoners and recidivism rates needs further examination.

In summary, prison provides a unique public health opportunity to treat mental illnesses that otherwise may not be treated in the community. Almost all prisoners return to their communities of origin, and effective treatment of mentally ill prisoners will have potentially substantial public health benefits and possibly reduce reoffending rates.

2.1.6 Funding & Acknowledgements

Funding

KS was funded by the German Friedrich-Ebert Stiftung.

Acknowledgements

S Agbahowe, H Andersen, L Birmingham, R Bland, G Cote, M Davidson, B Denton, R Ghubash, J Haapsalo, H Herrman, W Hurley, K Jordan, M Joukamaa, T Maden, D Mohan, B Morentin, W Narrow, K Northrup, T Powell, K Rasmussen, C Schoemaker, N Singleton, C Smith, and G Walters kindly provided additional data from their studies for the initial review. We are grateful to T Butler, B Falissard, H Kennedy, M Pereira Pondé, M Piselli and R Quartesan, CE von Schoenfeld, R Trestman and V Tort for providing further information about their studies for the update. In addition, K Abram, D Black, C James, O Nielssen, MI da Rosa, and K Wada helpfully responded to queries that led to their studies being excluded from the review. We are grateful to J Baillargeon for comments on a previous draft. Kat Witt assisted with being the second data extractor.

Supplementary Table 1: Update: Details of included studies since 2001

Study	Year of interview	Country	Income group	Sampling	Instrument	Diagnostic criteria	Psychiatric Interviewer	Mean age (years)	Participation Rate (%)
Alevizopoulos, 2010	2010	Greece	High	Systematic	ISPI	DSM	N	26.9	Not stated
Arnab, 2009	2001	India	Low-middle	Not stated	SCID	DSM-IV	Y	34.8	Not stated
Assadi, 2006	2002	Iran	Low-middle	Stratified random	SCID	DSM-IV	Y	32.7	88
Bernudez, 2009	2001	Mexico	Low-middle	Convenience	MINI	DSM-IV	N	30.6	Not stated
Bulten, 2009	2009	Netherlands	High	Simple random	MINI	DSM-III-R	N	30.4	81
Butler, 2005	2001	Australia	High	Consecutive	CIDI	DSM-IV	N	29.6	85
Curtin, 2009	2004	Ireland	High	Consecutive	SADS-L	ICD-10	N	29.8	92
Dudeck, 2009	2006	Germany	High	Not stated	SCID	DSM-IV	Y	31.2	Not stated
Duffy, 2006	2001	Ireland	High	Systematic	SADS-L	ICD-10	Y	37.1	64
Fallsarad, 2006	2003	France	High	Stratified random	MINI	DSM-IV	Y	37	63
Fotiadou, 2004	2001	Greece	High	Systematic	MINI	DSM-IV	N	36.5	85
Gunter, 2008	2005	US	High	Simple random	MINI	DSM-IV	N	31.1	Not stated
Linehan, 2009	2002	Ireland	High	Stratified random	SADS-L	ICD-10	N	29.6	70
Parsons, 2001	1998	UK	High	Population	SADS-L	ICD-10	N	28.1	89
Piselli, 2009	2005	Italy	High	Consecutive	SCID	DSM-IV	Y	35.5	72
Ponde, 2010	2006	Brazil	Low-middle	Not stated	MINI	DSM-IV	N	33	Not stated
Sharma, 2010	2010	India	Low-middle	Simple random	Clinical interview	Not stated	N	Not stated	Not stated
Stompe, 2010	2008	Austria	High	Consecutive	SCAN	ICD-10	N	Not stated	Not stated
Tort, 2009	2007	Country	High	Simple random	SCID	DSM-IV	N	Not stated	90
Trestman, 2007	2004	US	High	Systematic	SCID	DSM-IV	N	32	Not stated
von Schoenfeld, 2006	2003	Germany	High	Not stated	SCID	DSM-IV	N	34	82
Watzke, 2006	2000	Germany	High	Not stated	SCAN	ICD-10	Y	31.6	Not stated
Wright, 2006	2002	Ireland	High	Population	SADS-L	ICD-10	Y	27.4	76
Zahari, 2010	2009	Malaysia	Low-middle	Stratified random	SCID	DSM-IV	Y	35	Not stated
Zoccali, 2008	2003	Italy	High	Population	SCID	DSM-III-R	Y	35.8	Not stated

CIDI = Composite International Diagnostic Interview; *ISPI* = Iowa Structured Psychiatric Interview; *MINI* = Mini-International Neuropsychiatric Interview; *SADS-L* = Schedule for Affective Disorders and Schizophrenia; *SCAN* = Schedules for Clinical Assessment in Neuropsychiatry; *SCID* = Structured Clinical Interview for DSM-IV; *DSM-IV* = Diagnostic and Statistical Manual of Mental Disorders (by the American Psychiatric Association); *ICD-10* = International Statistical Classification of Diseases and Related Health Problems (by the World Health Organization).

Supplementary Table 2: Reported comorbidity in prisoners

Group of diagnosis		Comorbidity (%)	
		Substance use disorder	Total Sample (N)
Major psychiatric disorders			
Butler, 2003	Any mental disorder	43.5	586
Piselli, 2009	Psychiatric disorders	20.4	302
Psychotic illnesses			
Assadi, 2006	Psychotic disorder	95.0	351
Butler, 2003	Psychosis (12-month)	13.7	586
Duffy, 2006	Lifetime psychosis	90.0	435
Mood disorders			
Assadi, 2006	Mood disorder	82.5	351
Bermudez, 2007	Major Depressive Episode	78.0	213
Butler, 2003	Major Depression (12-month)	19.5	586

2.2 Seewald K, Rossegger A, Endrass J (2015). Risikoeinschätzungen bei Sexualstraftätern im institutionellen Bereich. In JM Fegert & M Wolff (Eds.) *Kompendium „Sexueller Missbrauch in Institutionen“: Entstehungsbedingungen, Prävention und Intervention* (pp 561-573). Weinheim: Verlagsgruppe Beltz.

2.2.1 Risk-Assessment bei Gewalt- und Sexualstraftätern

Risikobeurteilungen bei Gewalt- und Sexualstraftätern fokussieren primär auf die Beurteilung des Rückfallrisikos von in Freiheit entlassener Straftäter. In der Praxis stellt sich aber auch immer wieder die Frage, mit welcher Wahrscheinlichkeit grenzverletzendes Verhalten während einer institutionellen Unterbringung zu erwarten ist. Welche Faktoren bei einer Risikobeurteilung berücksichtigt, und über welches Vorgehen die Aussage über das Rückfallrisiko einer Person generiert werden sollte, ist seit Jahren Gegenstand wissenschaftlicher Diskussionen (Ægisdóttir et al., 2006; Grove et al., 2000; Meehl, 1954). Es lassen sich grundsätzlich drei Herangehensweisen unterscheiden: (1) Das intuitive/klinische Urteil, (2) das strukturierte klinische Urteil und (3) die Anwendung mechanischer Risk-Assessment Instrumente. Von einem intuitiven/klinischen Urteil wird dann gesprochen, wenn die für die Risikobeurteilung berücksichtigten Kriterien sowie die Gewichtung dieser von Einzelfall zu Einzelfall festgelegt werden. Die Beurteilung basiert auf dem Experten- und Erfahrungswissen des Beurteilers und erfolgt ohne Zuhilfenahme von standardisierten Instrumenten oder Checklisten (Rossegger, Endrass, et al., 2012a). Demgegenüber greift das strukturierte klinische Urteil auf gewisse Vorgaben zurück. Es liegen dabei Checklisten vor, die die zu berücksichtigenden Kriterien definieren. Die Gewichtung dieser Kriterien sowie das Ableiten des Rückfallrisikos aus diesen wird allerdings mittels eines klinischen Urteils vorgenommen (Douglas et al., 1999; Fazel et al., 2012). Die Anwendung mechanischer Risk-Assessment Instrumente ist dadurch gekennzeichnet, dass die anwendende Person Fragen mit geschlossenen Antwortkategorien beantwortet, und diese Antworten nach einem vordefinierten Auswertungsalgorithmus in eine Wahrscheinlichkeitsaussage überführt werden (Rossegger, Endrass, et al., 2012b). Während ein klinisches Urteil an spezifische Bedingungen des Einzelfalls angepasst werden kann, folgt die Anwendung mechanischer Instrumente einem starren Algorithmus und erlaubt keine Anpassung der Risikobeurteilung an spezifische Charakteristika des Einzelfalls (Rossegger, Endrass, et

al., 2012a). Die empirische Befundlage zu der Frage, welches zu den trennschärfsten Beurteilungen führt, ist eindeutig: Bei der Risikobeurteilung von Gewalt- und Sexualstraftätern bewährt sich der Einsatz von standardisierten Instrumenten. Es ist ein robuster Befund, dass mit standardisierten Instrumenten vorgenommene Risikobeurteilungen eine höhere Trennschärfe aufweisen, als Risikobeurteilungen, die anhand eines klinischen Urteils vorgenommen worden sind (Ægisdóttir et al., 2006; Grove et al., 2000; Quinsey et al., 2006).

Bis heute wurde für die Beurteilung des Rückfallrisikos von Gewalt- und Sexualstraftätern eine Reihe entsprechender mechanischer Risk-Assessment Instrumente entwickelt. Zu den in der Praxis am weitesten verbreiteten Instrumenten gehören nach Archer, Buffington-Vollum, Stredny, und Handel (2006) der *Static-99* (Hanson & Thornton, 1999), der *Sex Offender Risk Appraisal Guide* (SORAG; Quinsey et al., 2006; Rossegger, Gerth, Urbaniok, Laubacher, & Endrass, 2010) und der *Violence Risk Appraisal Guide* (VRAG; Quinsey et al., 2006; Rossegger, Urbaniok, Danielsson, & Endrass, 2009). Alle drei Instrumente wurden in Kanada entwickelt. Während der *Static-99* und der *SORAG* ausschließlich bei Sexualstraftätern angewendet werden dürfen, wurde der *VRAG* für die Anwendung bei Gewalt- und Sexualstraftätern entwickelt (Rossegger, Endrass, et al., 2012b). Die Kriteriumsvalidität dieser Instrumente wurde in verschiedenen Untersuchungen repliziert, darunter auch Studien aus Deutschland, Österreich und der Schweiz (Eher, Rettenberger, Schilling, & Pfaefflin, 2008; Rettenberger, Matthes, Boer, & Eher, 2009; Rossegger et al., 2010; Rossegger et al., 2009; Stadtland & Nedopil, 2007). Die insgesamt über viele Studien hinweg erzielten Trennschärfequotienten reichen von $AUC=0.60$ bis $AUC=0.82$ (Rossegger, Endrass, et al., 2012b). Studien, die die Trennschärfe der Instrumente an unterschiedlichen Gruppen von Sexualstraftätern untersuchten, weisen darauf hin, dass sowohl der *Static-99* als auch der *SORAG* in der Gruppe von Tätern, die ein Kind sexuell missbraucht haben eine höhere Trennschärfe erzielen, als in der Gruppe von Vergewaltigern (Ducro & Pham, 2006; G. T. Harris et al., 2003; Rossegger et al., 2010). Nachfolgend werden die drei in der Praxis am weitesten verbreiteten Risk-Assessment Instrumente eingehender vorgestellt:

Static-99

Der Static-99 kann bei erwachsenen männlichen Sexualstraftätern angewendet werden und weist die Wahrscheinlichkeit erneuter Verurteilungen mit einem Sexualdelikt aus (Hanson & Thornton, 1999). Die Risikobeurteilung bezieht sich auf Täter von Kontakt-Sexualdelikten (Vergewaltigung, sexuelle Nötigung, sexuelle Handlungen an nicht-verwandten Kindern, Inzest) oder Hands-off-Sexualdelikten wie Exhibitionismus oder Voyeurismus. Der Static-99 enthält zehn Fragen (Items), die einen überwiegend statischen, d. h. unveränderlichen Charakter, haben. Zu den abgefragten Inhalten gehören deliktbezogene, opferbezogene und demographische Merkmale: (1) das Alter des Täters, (2) die Frage nach dem Zusammenleben mit einer Partnerin für mindestens zwei Jahre, (3) mindestens eine Verurteilung für nicht-sexuelle Gewaltstraftaten im Indexdelikt, (4) mindestens eine Verurteilung für nicht-sexuelle Gewaltstraftaten im Vorfeld des Indexdelikts, (5) frühere Anklagen oder Verurteilungen für Sexualdelikte, (6) Anzahl der Verurteilungen vor dem Indexdelikt, (7) mindestens eine Verurteilung für ein sogenanntes „Hands-off“-Sexualdelikt (z.B. Exhibitionismus), (8) mit dem Opfer verwandt (9) mit dem Opfer bekannt, sowie (10) das Geschlecht des Opfers. Neun der Items folgen einem dichotomen Aufbau und werden mit ja/nein bzw. 0/1 zu kodiert. Item 5 (frühere Anklagen oder Verurteilungen für Sexualdelikte) bietet vier anstelle von zwei Antwortkategorien an und wird mit null bis drei Punkten bewertet. Der Static-99 sieht vor, dass nur Item 2 wegen fehlender Information ausgelassen werden darf. Fehlen die Informationen, um eines der anderen Items zu bewerten, ist die Auswertung nicht mehr zulässig. Für die Auswertung wird über die in den Item-Antworten erzielten Punktwerte der Summenwert gebildet. Dieser kann in eine von vier Risikokategorien überführt werden, für die Normwerte für Rückfallraten hinterlegt sind (Eher & Rettenberger, 2013; G. T. Harris et al., 2003).

Sex Offender Risk Appraisal Guide

Der *Sex Offender Risk Appraisal Guide* (SORAG; Quinsey et al., 2006) ist ebenfalls ein Instrument, das spezifisch für die Anwendung bei Sexualstraftätern entwickelt worden ist. Im Gegensatz zum Static-99, der auch bei „hands-off“ Sexualdelikten angewendet werden kann, sieht der SORAG ausschließlich die Anwendung bei Tätern vor, die wegen eines Kontakt-Sexualdelikts (Vergewaltigung, sexuelle Nötigung, sexuelle Handlungen an nicht-verwandten Kindern, Inzest) angeklagt oder verurteilt worden sind. Der SORAG

unterscheidet sich vom Static-99 auch im Hinblick auf die Definition der Kriteriumsvariable: Er bildet im Gegensatz zum Static-99 die Wahrscheinlichkeit erneuter Anklagen oder Verurteilungen wegen eines Gewalt- und/oder Sexualdelikts innerhalb eines Zeitraums von 7-10 Jahren ab (Quinsey et al., 2006; Rossegger et al., 2010). Das Ergebnis des SORAGs weist somit nicht die Wahrscheinlichkeit einschlägiger Rückfälle mit einem Sexualdelikt aus, sondern die Wahrscheinlichkeit, dass der Täter in einem Spektrum von Gewalt- und Sexualdelikten in Erscheinung tritt. Der SORAG besteht aus 14 Items und liegt in einer validierten und autorisierten deutschen Übersetzung vor (Rossegger et al., 2010). Er umfasst folgende Angaben: (1) bis zum 16. Lebensjahr mit beiden biologischen Eltern gelebt, (2) mangelhafte Anpassung in der Grundschule, (3) Alkoholprobleme in der Vorgeschichte, (4) Zivilstand, (5), nicht-gewalttätige Vorgeschichte, (6) gewalttätige Vorgeschichte, (7) Anzahl Verurteilungen für frühere Sexualstraftaten, (8) Vorgeschichte von Sexualstraftaten an Mädchen unter 14 Jahren, (9) Versagen bei früherer Entlassung, (10) Alter beim Indexdelikt, (11) Vorliegen einer Persönlichkeitsstörung, (12) Vorliegen einer Schizophrenie, (13) Phallometrische Testergebnisse (Item ersetzbar), (14) Punktwert auf der Psychopathie-Checkliste (PCL-R; Hare, 2003). Im Gegensatz zum Static-99 sind alle Items unterschiedlich gewichtet. Das größte Gewicht kommt Item 13 (Ergebnis auf der Psychopathie-Checkliste) zu. Bis zu vier der 13 Fragen dürfen wegen fehlender Informationen ausgelassen werden, was die Anwendungsmöglichkeiten in der Praxis auch bei geringer Informationsgrundlage erhöht. In diesen Fällen wird für die ausgelassenen Fragen ein Ersatzwert gebildet und der Summenrohwert entsprechend korrigiert. Der erzielte Summenwert kann in eine von neun Risikokategorien überführt werden, für die wiederum Normwerte mit Rückfallraten hinterlegt sind (Rossegger et al., 2010).

Violence Risk Appraisal Guide

Der *Violence Risk Appraisal Guide* (VRAG; Quinsey et al., 2006) ist ein Instrument, das bei Gewalt- und Sexualstraftätern das Risiko erneuter Anklagen und Verurteilungen wegen Sexual- und/oder Gewaltstraftaten einschätzt. Er besteht aus 12 Items, die deswegen den Items des SORAG sehr ähneln, weil dieser auf Grundlage des VRAGs entwickelt wurde. Bei der Entwicklung des SORAG wurden einzelne Opfervariablen, die im VRAG abgefragt wurden ersetzt durch spezifische Items für Sexualstraftäter

(phallometrische Testergebnisse, sexuelle Devianzen). Konkret wird beim VRAG folgendes abgefragt: (1) bis zum 16. Lebensjahr mit beiden biologischen Eltern gelebt, (2) mangelhafte Anpassung in der Grundschule, (3) Alkoholprobleme in der Vorgeschichte, (4) Zivilstand, (5), nicht-gewalttätige Vorgeschichte, (6) Versagen bei früherer Entlassung, (7) Alter beim Indexdelikt, (8) Verletzungsgrad des Opfers, (9) Irgendein weibliches Opfer, (10) Vorliegen einer Schizophrenie, (11) Vorliegen einer Persönlichkeitsstörung, (12) Punktwert auf der Psychopathie-Checkliste (PCL-R; Hare, 2003). Der VRAG liegt ebenso wie der SORAG in einer validierten und autorisierten deutschen Übersetzung vor (Rossegger et al., 2009). Alle Items werden ebenfalls gewichtet gewertet, wobei auch hier das größte Gewicht dem Ergebnis auf der Psychopathie-Checkliste zukommt. Bis zu vier der 12 Fragen dürfen wegen fehlender Informationen ausgelassen werden, bei drei Items besteht zudem die Möglichkeit, diese durch hoch korrelierte andere Skalen zu ersetzen. Beim VRAG liegen ebenfalls neun Risikokategorien mit hinterlegten Normwerten vor, in die der Summenwert überführt werden kann (Rossegger et al., 2009).

Allen drei Instrumenten ist gemein, dass sie primär statische Informationen abfragen. Für den Einsatz der Instrumente in der Praxis bedeutet das, dass sich der Einsatz der Instrumente insbesondere dann empfiehlt, wenn eine Risikobeurteilung vor Beginn einer Intervention (z.B. Therapie, institutionelle Unterbringung) durchgeführt wird. Soll hingegen das aktuelle Rückfallrisiko eines Straftäters im Verlauf einer Intervention ausgewiesen und die Wirksamkeit der Intervention im Hinblick auf Veränderungen im Rückfallrisiko ausgewiesen werden, eignet sich das Instrument nicht (Rossegger, Endrass, et al., 2012b). Für eine Beurteilung des Rückfallrisikos, bei der auch dynamische Faktoren berücksichtigt werden empfiehlt sich der Einsatz des STABLE-2007 (Hanson et al., 2007). Der STABLE-2007 baut auf der Anwendung des Static-99 auf. Der STABLE-2007 besteht aus 13 dynamischen Items/Risikobereichen (Eher, Matthes, & Rettenberger, 2012). Während der Static-99 allein auf der Grundlage von Akten angewendet werden kann, sieht der STABLE-2007 als Informationsgrundlage ein halbstrukturiertes Interview vor, in dem Informationen zu den Risikobereichen (z.B. Frauenfeindlichkeit, fehlende Empathie, deviant-sexuelle Vorlieben, etc.) abgefragt und deren Ergebnisse in eine von drei Risikokategorien überführt werden (Hanson et al., 2007; Rossegger, Endrass, et al., 2012b).

Die konsequente Anwendung von mechanischen Risk-Assessment Instrumenten kann einen wichtigen Beitrag zur Rückfallprävention leisten, da mithilfe dieser Instrumente im Durchschnitt trennschärfer zwischen Rückfalltätern und Nicht-Rückfalltätern unterschieden werden kann, als wenn andere Methoden der Risikobeurteilung verwendet werden. Die Instrumente haben zudem den Vorteil der ökonomischen und nicht auf spezifische Berufsgruppen beschränkten Anwendung. Im Falle des Static-99 wurde dies bereits bei der Entwicklung des Instruments berücksichtigt, die Zielgruppe der Anwendenden, für die das Instrument entwickelt wurde beschränkte sich nicht auf forensisch-psychiatrisches oder psychologisches Fachpersonal, sondern auf Personen aller Professionen, die Erfahrung mit kriminalprognostischen Einschätzungen haben (A. Harris, Phenix, Hanson, & Thornton, 2003).

Ein Nachteil der genannten Instrumente ist, dass bei ihrer Entwicklung die Kriteriumsvalidität im Vordergrund stand und die Inhaltsvalidität weniger berücksichtigt wurde. Bei der Entwicklung der Instrumente wurden aus einem Pool von potenziellen Prädiktorvariablen für Rückfälligkeit jene Variablen ausgewählt und in Items überführt, die empirisch am stärksten mit Rückfälligkeit korrelierten. Aus diesem rein korrelativen Zusammenhang kann jedoch nicht auf einen kausalen Zusammenhang zwischen den in den Items abgefragten Informationen und der Kriteriumsvariable (Rückfälligkeit) geschlossen werden. Die Zusammenstellung der Items weist zudem eine gewisse Beliebigkeit auf: Die in dem Instrument enthaltenen Items bilden kein theoretisches Konstrukt ab (Rossegger, Endrass, et al., 2012b). Das hat wichtige Implikationen für die Praxis: Erstens wurden die mechanischen Risk-Assessment Instrumente für die Anwendung in einem spezifischen Kontext entwickelt und die Übertragbarkeit in andere Kontexte muss angezweifelt werden. Die Ergebnisse empirischer Untersuchungen sprechen dafür, dass Instrumente, die bei entlassenen Straftätern trennscharf zwischen rückfälligen und nicht rückfälligen Tätern unterscheiden können, nur unzureichend geeignet sind, die Wahrscheinlichkeit für grenzverletzendes Verhalten während einer institutionellen Unterbringung zu beurteilen (Endrass, Rossegger, Frischknecht, Noll, & Urbaniok, 2008).

Zweitens lässt sich aus den Ergebnissen in den Risk-Assessment Instrumenten kein individuelles Risikoprofil ableiten. Ein solches individuelles Risikoprofil, das für den Einzelfall zentrale deliktfördernde Reaktions-, Denk- und Verhaltensmuster des Täters beschreibt, bildet allerdings die Grundlage aller deliktpräventiven therapeutischen

Interventionen. Die deliktpräventive Therapie rückt schließlich diese definierten Reaktions-, Denk- und Verhaltensmuster ins Zentrum der Intervention. Beispielhaft lassen sich die Risikoprofile von zwei Tätern gegenüberstellen, die beide ein Kind sexuell missbraucht haben: Während Täter A sich in präpubertäre Kinder verliebt, wird der sexuelle Missbrauch bei Täter B dadurch angestoßen, dass er im Kontakt mit erwachsenen Frauen unsicher und unbeholfen ist und er sich im Kontakt mit Mädchen knapp unterhalb der Schutzaltersgrenze kompetenter erlebt. Ein solches Risikoprofil lässt sich nicht aus den genannten mechanischen Risk-Assessment Instrumenten ableiten. Ansatzpunkte für eine Intervention müssen somit losgelöst von den Instrumenten definiert werden.

Drittens ist es nicht möglich, aus den in den Instrumenten abgefragten Informationen Rückschlüsse auf Risikofaktoren für erstmalige sexuelle Übergriffe abzuleiten. Der Einsatz dieser Instrumente ist deutlich auf die Anwendung bei Personen begrenzt, die bereits die Handlungsschwelle einer Sexual- oder Gewaltstraftat überschritten haben und als Täter identifiziert wurden. Für die Primär- als auch für die Sekundärprävention kommt aber der Identifikation von spezifischen Risikofaktoren bzw. Risikokonstellationen eine besondere Bedeutung zu.

2.2.2 Sexuelle Gewalt im Strafvollzug

Während Gewalt im Strafvollzug seit Jahren intensiv wissenschaftlich untersucht wird, lagen bis vor wenigen Jahren nur sehr wenige Studien über die Prävalenz von und Risikofaktoren für spezifisch sexuelle Gewalt im Strafvollzug vor (Jones & Pratt, 2008). Gesetzesinitiativen und politische Vorstöße - wie beispielsweise 2003 der „Prison Rape Elimination Act“ (PREA) in den USA - versuchten diese Lücke zu schließen. Seit der Einführung des „Prison Rape Elimination Act“ arbeiten Behörden und Fachpersonen aus Wissenschaft und Strafvollzug gemeinsam daran, dem Thema sexueller Übergriffe im US-amerikanischen Strafvollzug mehr Aufmerksamkeit zu verleihen und das Vorkommen entsprechender Übergriffe zu reduzieren (Gonsalves et al., 2012). Die Initiierung des „Prison Rape Elimination Act“ erfolgte damals entgegen der Überzeugung von Aufsehern, dass es sich bei sexuellen Übergriffen im Strafvollzug um ein Randphänomen handelt, welches nur selten vorkommt (Hensley & Tewksbury, 2005). Diese Auffassung wird durch die Ergebnisse epidemiologischer Studien kontrastiert: 18% der befragten männlichen Insassen eines Hochsicherheitsgefängnisses in den USA

gaben an, dass ihnen bereits einmal sexuelle Übergriffe angedroht worden waren, 9% berichteten bereits Opfer eines sexuellen Übergriffes geworden zu sein (Hensley, Koscheski, & Tewksbury, 2005). Das „National Inmate Survey“ befragte über 90'000 Insassen US-amerikanischer Gefängnisse. Von den befragten Insassen gaben 4% an, in den letzten 12 Monaten Opfer eines sexuellen Übergriffes gewesen zu sein (Beck, Berzofsky, Caspar, & Krebs, 2013). Forschungsergebnisse weisen zudem darauf hin, dass die Prävalenz bei inhaftierten Frauen höher als jene bei inhaftierten Männern ist: In der Untersuchung von Blackburn, Mullings, und Marquart (2008) gaben 17% der 436 befragten inhaftierten Frauen eines amerikanischen Frauengefängnisses an, Opfer eines sexuellen Übergriffes und 3% Opfer einer Vergewaltigung während der Unterbringung geworden zu sein. Diese Zahlen zeigen auf, dass sexuelle Übergriffe hinter Gefängnismauern kein seltenes Ereignis darstellen und – wie vom „Prison Rape Elimination Act“ angedacht – die Entwicklung und Implementierung von Präventionsstrategien sinnvoll sind. Eine erfolgreiche Präventionsstrategie zur Vermeidung sexueller Übergriffe innerhalb eines institutionellen Settings setzt die Identifikation entsprechender Risikofaktoren voraus. Sind diese Risikofaktoren identifiziert, sollte idealerweise in einem nächsten Schritt ein Instrument entwickelt werden, anhand dessen die Wahrscheinlichkeit ermittelt werden kann, mit der ein Insasse einen sexuellen Übergriff begeht oder – als zweites Ergebnis der Risikobeurteilung – wie wahrscheinlich es ist, dass ein Insasse selber Opfer eines Übergriffes wird.

Bisherige Forschungsarbeiten zum Thema weisen darauf hin, dass ein wichtiger Risikofaktor für sexuell übergriffiges Verhalten das Altersgefüge in der Institution ist (Felson, Cundiff, & Painter-Davis, 2012). Auf der Grundlage von Daten des „National Incident Based Reporting System“ (NIBRS), konnten Felson et al. (2012) aufzeigen, dass die Täter sexueller Übergriffe in Justizvollzugseinrichtungen im Durchschnitt sieben Jahre älter sind als ihre Opfer (35 Jahre vs. 28 Jahre). Ein weiterer Risikofaktor ist der soziokulturelle Hintergrund der Insassen. Während in US-amerikanischen Untersuchungen die Mehrheit der Täter schwarzer Hautfarbe waren, war die Mehrheit der Opfer weißer Hautfarbe (Gonsalves et al., 2012). Ferner sind die Täter den Opfern hinsichtlich der Körperstatur deutlich überlegen (Gonsalves et al., 2012). Mit knapp 90% wurde die überwiegende Mehrheit der untersuchten sexuellen Übergriffe in Gefängnispopulationen von Männern verübt. Entsprechend liegen weniger belastbare

Informationen über Risikofaktoren für institutionelle Gewalt bei weiblichen Insassen vor. Untersuchungen bei weiblichen Insassen haben keine Hinweise darauf ergeben, dass die Korrelate von sexueller Gewalt (soziokultureller Hintergrund, physisches Erscheinungsbild, Alter) geschlechtsspezifisch wären (Gonsalves et al., 2012). Auch für Insassinnen in US-amerikanischen Strafanstalten gilt, dass die Täterinnen in der Regel schwarzer Hautfarbe, den Opfern körperlich überlegen und älter als ihre Opfer waren (Gonsalves et al., 2012). Ob die sexuelle Orientierung der Opfer das Risiko erhöhen könnte, Opfer zu werden, kann nicht abschließend beurteilt werden. Je nach Untersuchung, gaben 41-78% der Opfer von Übergriffen an, homosexuell zu sein.

The Inmate Level of Risk Screening Measure

Gonsalves et al. (2012) stellten eine auf der Grundlage empirischer Literatur eine Checkliste zusammen, anhand derer die Wahrscheinlichkeit intramural-carceraler sexueller Gewalt systematisch zum Zeitpunkt des Eintritts in die Institution beurteilt werden kann. Die Checkliste „*The Inmate Level of Risk Screening Measure*“ setzt sich aus zwei Teilen zusammen. In ersten Teil wird beurteilt, wie hoch das Risiko eines spezifischen Insassen ist, sexuell übergriffiges Verhalten zu zeigen; im zweiten Teil der Checkliste hingegen das Risiko, das der Insasse Opfer eines entsprechenden Verhaltens wird.

Das Risiko einer Person, sexuell übergriffiges Verhalten zu zeigen, wird anhand der folgenden acht Merkmale beurteilt: (1) Mehrere frühere Inhaftierungen, (2) frühere intramurale Gewalt, (3) aggressive oder einschüchternde Haltung bei der Aufnahme in die Institution, (4) wirkt vertraut mit /sozialisiert in der Gefängnisumgebung, (5) wirkt emotional kalt, (6) zeigt ein Muster von zielgerichteter Gewalt oder impulsivem Verhalten, (7) Vorstrafen wegen Sexualstraftaten, (8) Sexualstraftat gegenüber einem Kind oder einer verletzlichen Person begangen („Hate-crime“).

Zur Beurteilung des Viktimisierungsrisikos werden die folgenden Informationen abgefragt: (1) Minderintelligenz, (2) soziale Zurückgezogenheit, Passivität, (3) Unerfahrenheit im Gefängnisumfeld, (4) Opfererfahrungen, (5) schwache physische Statur, (6) Schüchternheit, geringes Selbstvertrauen, Ausstrahlung von Angst und Schwäche, (7) Äußern von Sorge wegen sexuell aufdringlichem bzw. sexuell übergriffigem Verhalten anderer, (8) sexuelle Orientierung, die (von potentiellen

Tätern) als Schwäche eingestuft werden könnte (z.B. Homosexualität), (9) vulnerabel wegen der Art des Anlassdeliktes.

Die Person, die die Checkliste anwendet ist aufgefordert, aus den strukturiert gesammelten Informationen ein globales Gesamturteil zu bilden und das Risiko auf einer dreistufigen Skala (gering, mittelgradig, hoch) darzustellen. Wie bei der strukturierten klinischen Urteilsbildung üblich, steht dem Beurteiler frei, einzelne Fragen stärker zu gewichten und Grenzwerte festzulegen.

Die Validität der Checkliste wurde in einem US-amerikanischen Gefängnis (315 weibliche und 1'842 männliche Insassen) validiert. Die Aufseher beurteilten bei 94% der männlichen Population das Risiko zur Begehung sexueller Übergriffe als gering, bei 6% als mittelgradig und bei 0.3% als hoch. Bei den weiblichen Insassen wurde eine ähnliche Verteilung gefunden (88% gering, 11% mittelgradig, 1% hoch). Die Wahrscheinlichkeit, Opfer eines sexuellen Übergriffes in einer US-amerikanischen Strafanstalt zu werden, wurde für die meisten Insassen als gering eingestuft (94% für männliche und 87% für die weiblichen Insassen). Multiple Regressionsanalysen ergaben eine hohe Trefferquote des Instruments, wobei die Autoren keine Angaben zur Sensitivität und Spezifität lieferten (Gonsalves et al., 2012). Zusammengefasst kann gefolgert werden, dass das Instrument über eine gewisse Augenschein- und auch Kriteriumsvalidität verfügt, jedoch weitere umfangreiche Replikationsstudien in hiesigen Strafvollzugssystemen notwendig sind, bevor eine Empfehlung für den Einsatz des Instrumentes im deutschsprachigen Raum abgegeben werden kann.

2.2.3 Sexuelle Übergriffe gegen Kinder und Jugendliche im institutionellen Setting

Kinder, Jugendliche und Heranwachsende machen nicht nur über ein Drittel (36,2%) der Opfer von Straftaten gegen die sexuelle Selbstbestimmung unter Gewaltanwendung oder Ausnutzen eines Abhängigkeitsverhältnisses aus, sie sind auch überproportional häufiger gefährdet, Opfer zu werden, als Personen im erwachsenen Alter: Die Opfergefährdungszahlen (die innerhalb eines Jahres polizeilich bekannt gewordene Anzahl an Opfern) von Kindern, Jugendlichen und Heranwachsenden im Bereich von Straftaten gegen die sexuelle Integrität lagen laut der Polizeilichen Kriminalstatistik in Deutschland 2012 bei über 200 Fällen pro 100'000 Einwohnern (PKS, 2013). Mehr als die Hälfte aller sexuellen Übergriffe wird von Tätern begangen, die aus dem familiären

oder einem dem Kind bekannten Umfeld stammen (PKS, 2013). Darunter fallen auch Übergriffe, die Kinder und Jugendliche in sie betreuenden Institutionen wie Heimen, Internaten, denen eigentlich eine schützende Aufgabe zukommt, oder in Sportvereinen erfahren. Charakteristika von Tätern, die aus ihrer beruflichen Funktion heraus sexuelle Übergriffe in einer Institution begangen haben, wurden nur in wenigen Studien systematisch untersucht. Diese Studien fokussieren mehrheitlich auf sexuellen Missbrauch durch Priester und Ordensleute der katholischen Kirche (D'Alton, Guilfoyle, & Randall, 2013; Langevin, Curnoe, & Bain, 2000; Perillo, Mercado, & Terry, 2008; Randall, Carr, Dooley, & Rooney, 2011). Perillo et al. (2008) untersuchten in einer umfangreichen empirischen Studie Charakteristika von über 4'000 katholischen Priestern, die durch sexuell übergriffiges Verhalten aufgefallen waren. Den Autoren gelang es, Risikofaktoren für spezifische Konstellationen der sexuellen Übergriffe identifizieren. Sie zeigten auf, dass sich Priester, die mehrere Opfer missbrauchten von jenen, die ein einziges Opfer missbraucht haben dahingehend unterscheiden, dass diese Täter mit multiplen Opfern häufiger jünger sind, bereits früher Verhaltensauffälligkeiten gezeigt hatten, einen missbräuchlichen Konsum von Substanzen aufweisen, selbst sexuelle Viktimisierungserfahrungen gemacht haben, sowie gegenüber ihren vorwiegend männlichen Opfern Drohungen ausgesprochen haben. Täter mit ausschließlich männlichen Opfern waren wahrscheinlicher in ihrer Kindheit selbst sexuell missbraucht worden, fingen schon früh mit sexuell missbräuchlichem Verhalten an, hatten eher ältere Opfer und konsumierten eher psychotrope Substanzen. Bei Tätern, die eine enge Beziehung zum Opfer hatten, waren diese Opfer meist weiblich, im Durchschnitt jünger und multipler Missbrauch kam seltener vor. Diese Täter setzten ebenfalls häufiger spirituelle Manipulationen gegenüber ihren Opfern ein.

Studien über die Psychopathologie von Priestern und deren Delinquenzmustern stützen zudem die Annahme, dass Personen, die im Rahmen ihrer Tätigkeit im kirchlichen Kontext sexuelle Übergriffe verüben, sich nicht von anderen Gruppen sexuellen Missbrauchstätern unterscheiden: Langevin et al. (2000) verglichen klerikalen und nicht-klerikalen sexuelle Missbrauchstäter und konnten aufzeigen, dass sich zwischen den beiden Gruppen hinsichtlich psychiatrischer Belastungen oder Persönlichkeitsdispositionen keine Unterschiede feststellen ließen. Klerikale sexuelle Missbrauchstäter weisen ebenso häufig sexuelle Devianzen, Substanzmissbrauch, eine

psychiatrische Vorgeschichte, sowie gleich selten eine antisoziale Persönlichkeitsstörungen auf, wie nicht-klerikale sexuelle Missbrauchstäter.

2.2.4 Fazit

Sexuelle Übergriffe im institutionellen Rahmen, ausgehend von Betreuungspersonal, aber auch innerhalb der Gruppe der in der Institution untergebrachten Personen, sind ein Phänomen, das aufgrund seiner Häufigkeit einer adäquaten Präventionsstrategie bedarf. Risikobeurteilungen sollten sowohl das Risiko, Täter, als auch das Risiko, Opfer zu werden beinhalten. Für Risikobeurteilungen im Sinne der Primärprävention im intramuralen Kontext wurde ein vielversprechendes strukturelles Beurteilungssystem entwickelt („*The Inmate Level of Risk Screening Measure*“), das im deutschsprachigen Raum noch erprobt werden muss. Im Rahmen der Einschätzung des Erstrisikos sexueller Übergriffe gegenüber Kindern und Heranwachsenden im institutionellen Setting liegen bisher noch keine solchen Checklisten oder Instrumente vor. Diesbezüglich können nur deskriptive empirische Studien Hinweise auf Tätercharakteristika geben. Für die Risikobeurteilung im Sinne der Sekundärprävention der Begehung erneuter Taten gegen die sexuelle Selbstbestimmung kann der Einsatz erprobter mechanischer Risk-Assessment Instrumente (Static-99, SORAG) empfohlen werden.

2.3 Seewald K, Rossegger A, Urbaniok F, Endrass J (under review).
Assessing the risk of intimate partner violence: Expert evaluations
versus the Ontario Domestic Assault Risk Assessment. *International
Journal of Offender Therapy and Comparative Criminology*

2.3.1 Abstract

While primary prevention of intimate partner violence targets preliminary stages to dating violence and focuses on educational aspects, secondary prevention aims at identifying perpetrators who are at high risk of re-assaulting their (former) female intimate partners. A reliable and valid assessment of those perpetrators who put their (former) female partners at risk of being a victim of severe violence is needed. To maximize outreach, the method of assessment should be independent of specific professional qualifications. The purpose of the present study was to examine the accuracies of two violence risk assessment methods in identifying high-risk perpetrators in IPV, comparing forensic experts with psychology students. For 30 cases of IPV, violence risk was assessed by certified forensic psychiatrists in an extensive expert evaluation using unstructured clinical judgment (UCJ) and by undergraduate and graduate research assistants scoring the actuarial Ontario Domestic Assault Risk Assessment (ODARA) instrument. After a mean follow-up period of 8.0 years the base rate of violent recidivism was 20.0%. Students were more accurate in assessing long-term violent (including sexual) recidivism, with an area under the curve (AUC) of 0.78. Raters without extensive clinical training could differentiate those spouses who carried on assaulting their intimate partner from those who desisted from violent behavior.

2.3.2 Background

According to the World Health Organization (WHO), one out of four women (25.4%; 95% CI: 20.9-30.0%;) in Europe and a slightly higher percentage (29.8%; CI: 25.8-33.9) in the Americas has been physically assaulted by her intimate partner at least once (World Health Organization, 2013). A recent EU-wide survey presents similar rates (22%) of women being assaulted in an intimate relationship and points out that violence against women does not occur as a single incident: More than 30% of all women who had been raped by their intimate partner report that this happened repeatedly for at

least six times ("Violence against women: An EU-wide survey," 2014). Specifically for the United States, where rates of 35.6% of women being violently victimized are reported, more than one out of three of those women experienced multiple occurrences of violent assaults (Black et al., 2011). These prevalence rates stimulate an important discussion regarding the protection of women against their violent partners.

Protection can be achieved using different prevention strategies. The „primary prevention” strategy addresses the first time occurrence of domestic violence. Typical primary prevention programs for domestic violence are educational programs; target teenage populations; and include concepts of providing knowledge about dating and partner violence, conflict solving strategies, communication skills, and sex roles (Staggs & Schewe, 2011; Whitaker et al., 2006). „Secondary prevention” strategies focus on those partners who are the most likely to carry on with their violent behavior. In this context, recidivism rates are reported from 15.3-60.0% in Austria, Canada, and the United States (Hilton, Harris, & Rice, 2001; Hilton et al., 2004; Kingsnorth, 2006; Klein & Tobin, 2008; Rettenberger & Eher, 2013). Hence, the risk that a spouse with a history of violent behavior continues to assault his intimate partner is substantial enough to warrant an emphasis on secondary prevention (Hilton et al., 2001; Hilton et al., 2004; Kingsnorth, 2006; Klein & Tobin, 2008; Rettenberger & Eher, 2013). That type of secondary prevention requires the valid and reliable identification of perpetrators who repeatedly put their families at risk. The reliable and valid risk assessment of those individuals who pose a substantial risk for re-offending and/or exacerbating the pattern of violence is a key element of effective secondary prevention.

Over the past two decades, there has been intense debate regarding the most accurate assessment methods for secondary prevention (Ægisdottír et al., 2006). This discussion has emphasized three different assessment methods: the „unstructured clinical judgment” (UCJ), the „structured professional judgment” (SPJ), and the use of actuarial risk assessment instruments (ARAIs; Ægisdottír et al., 2006; Grove et al., 2000; Hanson & Morton-Bourgon, 2009). The three methods can be considered as three different generations of risk assessment while representing UCJ the first generation, actuarial methods the second generation and SPJ as the combination of those as the third generation (Doyle & Dolan, 2002). The UCJ method of evaluating violence risk is exemplified by the use of intuition guided by personal experience and an individualized combination of perceived risk factors (Kropp, 2008; Murray & Thomson, 2010). SPJ is an

alternative method that uses checklists or instruments containing empirically derived risk factors and integrating additional clinical information in the decision-making process. ARAIs integrate empirically derived risk factors in a mechanical way using a pre-specified algorithm and produce a probabilistic estimate of violence risk (Douglas et al., 1999; Fazel et al., 2012). There is substantial evidence that actuarial risk assessment instruments are more accurate than unstructured clinical judgment (Ægisdottír et al., 2006; Grove et al., 2000; Hilton, Carter, et al., 2008) and do also outperform SPJs in risk assessment accuracy in sexual offenders (Hanson & Morton-Bourgon, 2009).

While several SPJ instruments (e.g., Spousal Assault Risk Assessment (SARA; Kropp & Hart, 2000), Danger Assessment (DA; Campbell, 1995)) have been developed to assess the risk of intimate partner violence (IPV), only two equivalent ARAIs – the Ontario Domestic Assault Risk Assessment (ODARA; Hilton et al., 2004) and the related Domestic Violence Risk Appraisal Guide (DVRAG; Hilton, Harris, Rice, Houghton, & Eke, 2008) – have been published in the past decade (Kilvinger, Rossegger, Urbaniok, & Endrass, 2012; Kropp, 2008; Messing & Thaller, 2013). A recent study showed that the ODARA slightly outperformed SPJ instruments. However, neither ODARA and neither one of the SPJ instruments reached an area under the curve (AUC) above 0.70 (Messing & Thaller, 2013). AUCs are the quantification of Receiver Operating Characteristic Analysis, a statistical way of plotting sensitivity of a classification measure against its specificity (Rosenfeld & Penrod, 2011). In the present case, an AUC measures the probability that a randomly selected individual who recidivated received a higher risk classification than a randomly selected individual who did not recidivate (J. P. Singh, 2013).

Despite empirical evidence demonstrating the superior accuracy of ARAIs in assessing IPV risk, the use of these instruments in clinical practice is uncommon. Over 50% of practitioners make judgments about violence risk in an unstructured manner (J. P. Singh et al., 2014). Whereas the superiority of ARAIs compared with UCJ concerning different offender populations and outcomes has repeatedly been shown (Ægisdottír et al., 2006; Bengtson & Langstrom, 2007; Grove et al., 2000; Quinsey et al., 2006), differences in assessment accuracy between UCJ and ARAIs have rarely been investigated in the specific case of IPV. The present study aimed to address this gap by examining the accuracy of expert judgment regarding the long-term risk for IPV and also examining results of the ODARA, which was administered by undergraduate and graduate students.

Following previous results on the accuracy of actuarial risk assessment instruments (Douglas & Ogloff, 2003; Hanson & Morton-Bourgon, 2009), we hypothesized that the ODARA would outperform the UCJ regarding the assessment of long-term risk for IPV in a sample of high-risk IPV offenders in Switzerland. Assuming the superiority of the ODARA compared to UCJ this study further aimed at the encouragement of practitioners in all professional groups working with women victimized by IPV not to hesitate in doing their own violence risk assessment using an actuarial instrument like the ODARA.

2.3.3 Methods

Sample

The majority of cases of threats and intimate partner violence in the Canton of Zurich (Switzerland) are handled by law enforcement without a forensic expert evaluation. A forensic expert is usually only mandated to assess cases considered as high risk and/or if there is evidence that the alleged perpetrator is or could be mentally ill. There is no particular law existing for domestic or intimate partner violence in Switzerland. Existing laws are applied to incidents in the domestic context. Intimate partner violence in this study follows the index criteria of the ODARA: (1) being a male perpetrator and (2) having physically and/or sexually assaulted a (former) female intimate partner or having threatened to kill a (former) female intimate partner while holding a weapon.

Between April 2001 and April 2007, six forensic psychiatrists employed by the Department of Mental Health Services, Office of Corrections, Canton of Zurich, Switzerland assessed the violence risk of 31 men who were charged with threatening or assaulting their (former) female intimate partner. The purpose of the expert evaluations was to assess the risk of severe forms of IPV within the next 18 months, as well as to suggest strategies to manage the risk of such future violent behavior. For this study, perpetrators who died or were deported during the follow-up period were excluded, resulting in a final sample of 30 perpetrators.

The present study aimed to investigate differences in accuracy between UCJ performed by forensic experts and the ODARA scored by psychology students. All included cases met the index criteria for the ODARA.

Measures and materials

Offender and offense characteristics

All demographic information available on the offender (such as sex, age at time of offense, marital status, employment information, previous criminal records) and also about the index offense (e.g. victim characteristics, use of weapons) were collected by research assistants from case files.

UCJ

All forensic experts (n=6) who assessed the risk of perpetrators' recidivism in the present study were experienced psychiatrists who had years of experience working with forensic patients and in assessing individuals' violence risk in court-mandated risk evaluations. Besides the risk assessment experts also recommended risk management strategies for each case. Each psychiatrist assessed from one to 14 offenders. The written expert opinions were between 30-60 pages long and each case took the experts 30-35 hours to complete. The evaluations were based on file information as well as interviews with the perpetrators and victims. The experts had access to complete case files that included medical and criminal records, police reports, witness statements, charges, previous court decisions and previous expert evaluations or institutional records from previous hospitalizations and imprisonments if existing. During the interviews, the experts explored offender's biography, their offending behavior along with affective components and also post-offense behavior and attitudes towards their offense.

ODARA

The ODARA was initially developed as a frontline instrument for police officers (Hilton et al., 2004) and is designed to assess the risk of IPV in male perpetrators who (sexually) assaulted their current or former female intimate partner (or threatened to do so). In the instrument's development sample, the ODARA showed good discrimination between individuals that recidivated and those that did not, with a statistically significant discriminative receiver operating characteristic (ROC) as an area under curve (AUC) of 0.77 (Hilton et al., 2004). In cross-validation studies, AUCs from 0.66 to 0.80 have been reported (Hilton & Harris, 2009; Messing & Thaller, 2013). Results from Austria revealed an AUC of 0.71, suggesting good assessment accuracy in German-speaking

countries (Rettenberger & Eher, 2013). As far as it is known to the authors there is no meta-analytic result for AUCs of the ODARA.

The ODARA contains 13 dichotomous items (see table 3) that can be scored with 0 for „no” or 1 for „yes,” resulting in a total sum score ranging from 0 to 13. This sum score is then transferred to a risk bin. There are seven risk bins, each referring to a certain percentage of violence recidivism (7%, 17%, 22%, 34%, 39%, 53%, 74%) that was found in the instrument’s developmental sample during an average follow-up of 4.9 years (Hilton, Harris, & Rice, 2010). In this manner, each scoring result of the ODARA provides information about the percentage of the developmental sample that had the identical sum score and recidivated. The higher the ODARA sum score the higher is the percentage of offenders in the developmental sample that recidivated in this certain follow-up period. The outcome includes charges and convictions of violence against a (former) female intimate partner. Items of the ODARA are empirically based and use information that is routinely accessible for frontline police officers to make quick decisions about the assaulters’ risk levels. The authors of the ODARA suggest a further and more detailed risk evaluation by an expert if perpetrator’s risk meets a certain cut-off score (Hilton et al., 2004).

Table 3: Item content of the Ontario Domestic Risk Assessment (ODARA)

Item	Description
1	Prior domestic incident
2	Prior nondomestic incident
3	Prior custodial sentence of 30 days or more
4	Failure on prior conditional release
5	Threat to harm or kill at the index assault
6	Confinement of the partner at the index assault
7	Victim concern
8	More than one child
9	Victim’s biological child from a previous partner
10	Violence against others
11	Substance abuse
12	Assault on victim when pregnant
13	Barriers to victim support

Procedure

This study aimed at investigating the differences in two underlying processes of information integration in decision making that lead to a final judgment: the clinical vs. mechanical approach. Whereas the ODARA as an actuarial instrument only contains static variables and combines the information in a mechanical way the extensive expert

evaluations integrated static and dynamic information about a case to a final judgment without uncovering their decision process. To compare the clinical with the mechanical approach of integrating information the final judgment was extracted out of the expert evaluations and compared with the resulting sum score of the ODARA.

UCJ

The forensic psychiatrists did not employ an operationalized terminology to formulate their estimations of recidivism risk. To quantify these estimates, three undergraduate and graduate research assistants extracted the experts' judgment using a nine-point Likert scale (*very low, very low to low, low, low to moderate, moderate, moderate to high, high, high to very high, very high*). This nine-point Likert scale did not correspond to the ODARA risk bins but demonstrates the clinical scales generally used to describe risk levels in the Department of Mental Health Services. The students who screened expert evaluations were trained by senior research staff in understanding structure and composition of evaluations. They were not involved in the process of assessing the perpetrators in person or in writing the court evaluations. KS and AR reviewed the extraction result of each case and reached 100% agreement.

ODARA

The same graduate and undergraduate students who extracted the experts' judgments scored the ODARA based on the information given in the expert evaluations. The inter-rater agreement on ODARA sum score was satisfactory, with $\kappa=0.62$ (Gwet, 2001). All raters were blind to the recidivism outcome.

Follow-up and Recidivism

Follow-up started with date of expert evaluation and ended with date of recidivism or date of criminal record retrieved in September 2012. Information on recidivism during follow-up was collected by screening criminal records for any new charge or conviction. Violent recidivism was defined as any new charge or conviction for a violent offense that occurred after the expert opinion was completed. This definition included murder, assault, robbery, coercion, kidnapping, rape, sexual abuse, child molesting, threats, extortion, exhibitionism, illegal pornography, and arson. Besides, general recidivism was defined as any new charge or conviction for any offense including (sexual) violent and all other re-offenses.

Statistical analyses

Measurements of differences in discrimination validity between two methods of risk assessments are usually done by conducting Receiver Operating Characteristics (ROC) analyses and calculating areas under the curve (AUCs). AUCs in predictive accuracy research have the advantage of reporting a result that is independent from base rates (Rosenfeld & Penrod, 2011). Even though in clinical practice the value of AUCs in terms of validating risk assessment instruments is often questionable (Szmukler, Everitt, & Leese, 2012; Urbaniok, Rinne, et al., 2008) AUCs are calculated in the present study to stay in line with previous studies evaluating the predictive accuracy of the ODARA. For the ODARA the risk bin was used to put in ROC-analyses as this corresponded better with clinical levels of risk used to describe experts' assessments than the continuous ODARA sum score. Furthermore, binary logistic regression was used to determine how the ODARA ratings and the results of expert evaluations are associated with the outcome of violent recidivism. Therefore, two different regression models were calculated for each of the two risk measures. For all analyses violent recidivism was the dependent variable (coded with „0” for no recidivism and with „1” for recidivism). Independent variable in case of expert evaluations was the value of nine-point Likert scale coding and in case of the ODARA the value of sum score.

Details of ethics approval

Since descriptive and follow-up data from own clients of Department of Mental Health Services, Office of Corrections in the Canton of Zurich were used and pseudonymized for further data analyses, no additional ethical approval was required. Authors are part of the Research and Development Division within the Department and officially mandated to evaluate the Department's performance.

2.3.4 Results

Sample characteristics

According to police reports, incidents meeting the ODARA index criteria were threatening behavior (n=23, 76.7%), assault (n=8, 26.7%), and rape (n=2, 6.7%) (multiple elements of an offense were possible in a single incident). Half of the perpetrators (50%, n=15) used a weapon (gun or knife) during the index incident.

Alcohol and illicit drugs were involved in 20.0% (n=6) and 13.3% (n =4) of all cases, respectively. One-fourth (26.7%, n=8) of the perpetrators owned a gun.

The mean age of the perpetrators was 40.5 years (SD=11.4; range 21.2-64.2) at the time of the index incident. Almost half of the perpetrators (43.3%, n=13) were not Swiss nationals. This percentage is higher than in the general population of the Canton of Zurich, where a rate of 26.7% is reported (BFS, 2013). One third (33.3%, n=10) of perpetrators were unemployed at the time of the index incident, which is also a higher rate than in the general male adult Swiss population with only 3.9% (BFS, 2013). The majority of the sample (70%, n=21) was married, again, at a higher rate than the general Swiss population with 43.3% (BFS, 2013). A fourth (26.7%, n=8) of the perpetrators had a criminal record, and 12.9% (n=4) had committed a violent (including sexual) offense prior to the index-incident. Of the perpetrators, 36.7% (n=11) had been in psychiatric and/or therapeutic treatment at least once prior to the index incident. One perpetrator (3.3%) was diagnosed with a psychotic disorder, 26.7% (n=8) with a personality disorder, and 23.3% (n=7) with substance abuse or addiction. In 50.0% of the cases (n=15), the index incident occurred while the perpetrator lived separate from the victim.

Recidivism

The base rate of violent recidivism was 20.0% (n=6), and the base rate of general recidivism (i.e., any offense) was 36.7% (n=11). Newly charged or convicted violent offenses included serious threats, assault, robbery, coercion, and child sexual abuse. The mean follow-up was 8.0 years (SD=2.8, range: 1.4-11.6) and 7.1 years (SD=3.3, range: 0.6-11.6) for violent and general recidivism, respectively.

Risk assessment

UCJ

Two third (n=20) of perpetrators were judged to be from „moderate to high risk” up to „very high risk” to re-offend (n=14, 46.7%). Figure 6 displays the distribution of bin assignment based on Likert scale scoring and the observed recidivism on each scale point. None of those individuals assigned to the „high risk” or „very high risk” category recidivated during follow-up.

ODARA

The mean ODARA sum score was 4.2 (SD=1.8, range: 1-7). The risk bin that contained most of the individuals (n=8, 26.7%) was bin six (see figure 7), referring to a recidivism rate of 53%. A fourth of individuals assigned to the highest ODARA risk bin, referring to a percentage of recidivism of 74%, went on to recidivate. Figure 8 shows the expected rates of recidivism according to the ODARA developmental sample and the recidivism rates observed in the present study's sample by risk bin.

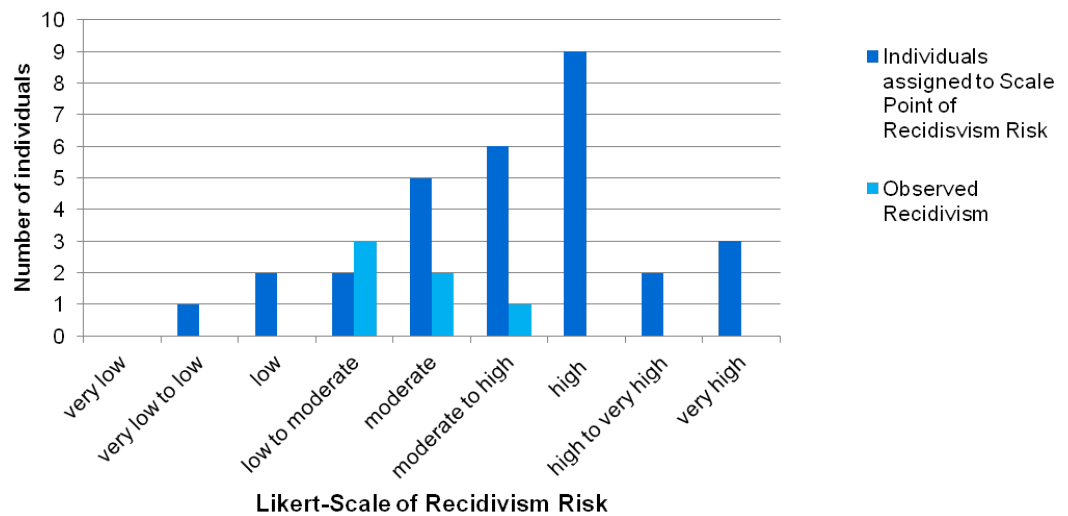


Figure 6: Distribution of cases assigned to UCJ nine-point Likert scale of recidivism risk and observed recidivism per scale point.

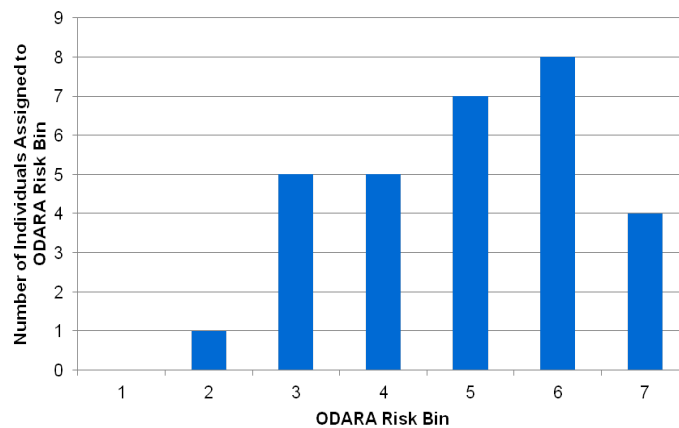


Figure 7: Distribution of cases assigned to ODARA risk bins.

Findings on discrimination validity

UCJ

Exact logistic regression analysis yielded an OR of 0.7 (95% CI: 0.2-2.1), which was not significant ($p>0.05$) using the nine risk categories to assess violent recidivism. An OR of less than one indicates an incorrect discrimination of binary outcome (in our case recidivism vs. non-recidivism) by the tested diagnostic tool. This result cautiously (because of being not significant) implies that a one unit increase in the Likert scale score reduces the odds of recidivism by 30%, meaning that an expert’s estimation of higher risk reduces the chance of recidivism in the assessed offender population. Non-parametric ROC analysis revealed that UCJ risk had an AUC of 0.44 (95% CI: 0.21-0.67).

ODARA

Exact logistic regression analysis yielded an OR of 2.4 (95% CI: 1.0-8.6, $p=0.05$) for the ODARA risk bins assessing violent recidivism. The odds of being classified within a higher risk bin are 2.4 higher for recidivists than for non-recidivists. The ODARA is correctly discriminating the binary outcome, specifically for every one unit increase in the ODARA sum score, the odds of recidivism increases by 140%. Non-parametric ROC analysis revealed that ODARA risk bins had an AUC of 0.78 (95% CI: 0.62-0.94). Differences in AUCs were significant with $\chi^2=5.22$ and $p<0.05$.

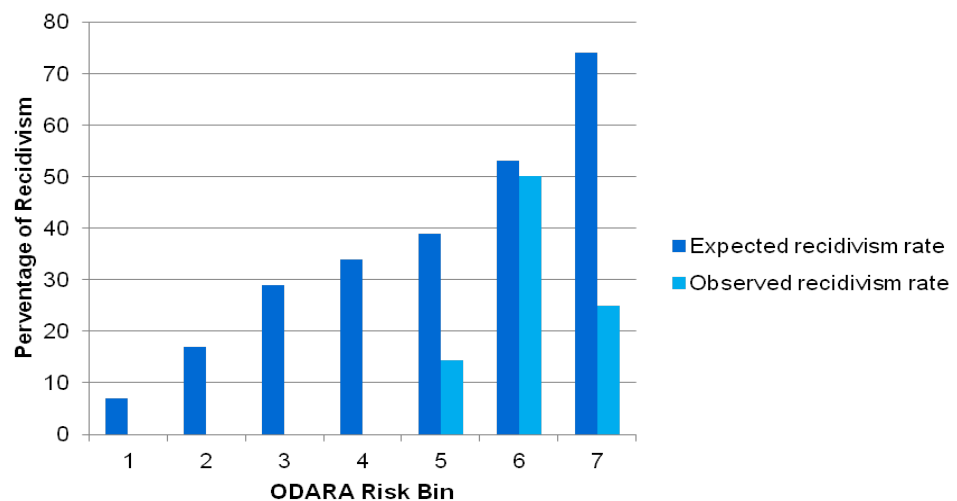


Figure 8: ODARA violent recidivism norm rates and observed violent recidivism rates per risk bin in the present study.

Violent recidivists

All offenders that recidivated were married, two out of six (33.3%) had a previous criminal record of violent offense, half ($n=3$, 50.0%) were diagnosed with a psychiatric disorder, and one (16.7%) had previously been institutionalized in a psychiatric hospital. Their mean follow-up was 5.5 years ($SD=2.5$; range: 2.5-9.0). Three out of six (50%) individuals that recidivated had been assessed by the same expert.

2.3.5 Discussion

One out of four women in Europe and one out of three women in the U.S. is physically assaulted by their intimate partner and also more than 30% in both regions experience these assaults repeatedly (Black et al., 2011; World Health Organization, 2013). Focusing on secondary prevention as is the attempt to identify those offenders at risk to re-offend their intimate partners, the present study aimed to investigate which method of risk assessment in cases of IPV most accurately identifies high-risk perpetrators. Two strategies were compared: risk judgments made by clinical forensic experts and the ODARA scored by undergraduate and graduate students. The study sample included perpetrators who either threatened to and/or actually physically harmed their female intimate partner, also including being sexually violent.

The findings of the present study firstly revealed that students' ODARA scoring outperformed forensic experts in the accuracy of their assessments of long-term violent behavior (differences in AUCs were significant with $\chi^2=5.22$ and $p<0.05$). Thus, in the context of IPV prevention, there is currently no empirical evidence for solely relying on the experts' clinical decisions when assessing the long-term risk for recidivism. Our findings suggest that the ODARA can help professionals without clinical training identify those spouses who are at risk of exacerbating their violent behaviors (AUC of 0.78). Although there are some empirical findings that can serve as a preliminary explanation for the performance of the clinical experts, it remains unclear what exactly led to the lower accuracy rate in the present study. Several explanations for the lower accuracy of UCJs compared to ARAIs have been offered by researchers in different areas of psychological research (Ægisdottír et al., 2006; Arkes, 1981; Borum, Otto, & Golding, 1993; Kingsnorth, 2006; Klein & Tobin, 2008). An increased amount of available information during the decision-making process has been shown to decrease clinicians' judgment accuracy (Ægisdottír et al., 2006), and clinical decision making is also

sensitive to context effects, such as the order of presented cases (Hilton, Carter, et al., 2008) and presentation of early information on the case by others (Otto, 2013). Nevertheless there also some critical points regarding the general use of actuarial instruments to predict individual's risk for further violent acts to be mentioned: First, actuarial instruments are developed on a group level. The prediction of an individual risk to commit further crimes by those instruments can only be done by referring to the score of a group of individuals with similar characteristics. Clinical practitioners criticize that the focus on the individual goes missing and secondly, in this context, with a missing focus on the individual, potentially relevant factors that could also be mediating the risk of an individual offender to re-offend are never considered (Hart et al., 2007). However regarding predictive accuracy empirical evidence is still in favor for actuarial instruments compared to clinical judgments (Ægisdóttir et al., 2006; Grove et al., 2000; Hanson & Morton-Bourgon, 2009; Hilton, Carter, et al., 2008).

The second important finding is based on the fact that the ODARA scoring originated from written expert evaluations. Therefore, the scoring was conducted using information that forensic experts had previously extracted from original case files and evaluated as important for the individual case and for their decision-making process. Students who scored the ODARA did not have any further information than this. This part of the study design highlights an interesting effect: the experts' risk evaluations were not less accurate due to inadequate collection of information. Consequently, it was most likely the strategy used to integrate the extracted information that lacked validity. Thus, the better performance of the ODARA may be attributed to its rule-based and mechanical integration of information, again confirming the superiority of mechanical instruments to clinical judgment.

Practical implication

Regarding the practical relevance of the present study, there is one major implication for all kinds of practitioners working in women's health prevention: The satisfactory discriminative validity of the ODARA coded by undergraduate and graduate students who did not have specialized forensic training supports the use of this instrument to assess IPV risk by all professionals in the field. Furthermore, assessing the risk for IPV is not limited to specific professions, such as psychiatry or psychology.

Based on the findings of the present study, we recommend use of the ODARA as a tool for secondary prevention of IPV. Since the instrument does not have specific prerequisites regarding the educational background of the IPV assessors the ODARA can be used by professionals from various disciplines (e.g., policemen, social workers, victim advocates, women's shelter workers) as a brief but accurate frontline instrument. Nevertheless it should be pointed out that with the recidivism criterion chosen in the present study not all forms of distressing situations for women in difficult relationships are sufficiently displayed. The ODARA can reflect a specific risk of IPV that is accurately defined but certainly not covering all aspects of the problem of domestic violence and the substantial role of dynamic relational aspects.

Limitations and future research

The design of the present study had some limitations. First and foremost, the clinicians were asked to assess the risk of recidivism for the next 18 months and to recommend a risk management strategy. Since there was no case of re-offense within the time period, that the clinicians had to estimate the risk for, the assessments of the clinicians were useful in the first place and not at all are useless in the context of short-term risk assessment in cases of intimate partner violence. Additionally the small sample size of the present study is an obvious limitation as it reduces statistical power and hence impedes the detection of statistically significant effects. However, it provides important preliminary evidence that the ODARA, scored by raters without clinical backgrounds, provides a more accurate assessment for IPV than unstructured experts' judgments made by forensic experts. Generalizability is though not possible until the comparison has been done with a greater sample and a more representative sampling strategy.

Second, the outcome criterion of the assessment was limited to violent (including sexual) recidivism. Whether or not the location of the re-offense was in a domestic setting could not be determined by the method used to code violent recidivism. Considering these points, future research should address this limitation and employ the correct outcome the instrument is supposed to assess at the proper time of risk according to its manual.

Third, whereas the ODARA only assessed the risk of further violence, experts additionally provided recommendations of risk management strategies in their evaluations. Those strategies varied in intensity and were most likely implemented by

the law enforcement agencies and the office of penal corrections, respectively. As discussed in context of other risk assessment methods of IPV, generally, a mediating effect of such an intervention must at least be assumed (Storey et al., 2014). Hence a low AUC for UCJ can also mean that risk management strategies were successfully implemented and high risk offenders did in the end not recidivate due to a specific intervention. This also corresponds to the (still not statistically significant) OR less than one indicating that the odds for an offender in a higher clinical risk category to re-offend are lower. UCJ of a higher risk may have resulted in the implementation of a management strategy that has actually reduced the risk and lead to a recidivism free follow-up-time. This fact limits the interpretation of the comparison of both methods. Our results support the use of the ODARA as an instrument to assess the long-term risk of IPV. Furthermore, there is evidence that the ODARA can be used by various professionals working in women's health prevention. Further research should be done to validate and calibrate this instrument with larger Swiss samples to improve statistical power and also to examine potential mediating effects of interventions in risk assessment.

2.3.6 Acknowledgements

We would like to thank Diana Fries, MSc. for supervising data collection. We thank Claire Reeves, PhD who provided medical writing services on behalf of BioScience Writers, LLC (Houston TX, USA).

2.4 Seewald K, Yundina A, Endrass J (in press). Juvenile Corrections. In Kerley KR, Copes H, De Li S, Lane J, Sharp SF (Eds.) *The Encyclopedia of Corrections*. Wiley Blackwell.

2.4.1 Abstract

Juvenile offenders are rarely incarcerated in prison. Instead, they are placed in correctional facilities providing various forms of interventions to reduce their risk of reoffending and future behavioral problems. In general, those interventions can be divided into rehabilitative and deterrent approaches. Numerous systematic reviews and meta-analytic results suggest a much higher effectiveness in reducing future criminal behavior of interventions following the rehabilitative approach compared to programs following the deterrent approach (such as boot camps or „Scared Straight” programs). Rehabilitative programs following the „Blueprints”-Guidelines such as the Multidimensional Treatment Foster Care (MTFC) Program, Functional Family Therapy (FFT), and Multisystemic Therapy (MST) show substantial effect in reduction of crime rates in delinquent juveniles. Furthermore cost-benefit-analyses also show their efficacy with positive cost-benefit-ratios.

2.4.2 Interrupting criminal careers with correctional interventions

The term „juvenile corrections” summarizes various types of interventions, services, programs and correctional facilities all focusing on the incarceration and rehabilitation of delinquent juveniles. This entry will give an overview about most effective and ineffective correctional programs and interventions with juvenile offenders.

Empirical evidence shows that early delinquency is a risk factor for later adult sexual and violent offending (Laubacher et al., 2014). Although the overwhelming majority of offenders will eventually make a conscious decision to desist from offending (Laub & Sampson, 2001), correctional interventions can accelerate this process when used in juvenile offenders. Generally these interventions can be divided into rehabilitative and deterrent approaches. In the context of general correctional programs, imprisonment is the usual form of deterrence. However, rates of incarcerated youth have been going down in the last decades. In 2012 juveniles under the age of 18 represented less than

1% of the general prison population in the US and in Europe, whereas young adults aged between 18 and 21 compose about 5% of the prison population in Europe.

Although other forms of deterrence such as „Scared Straight” or „Boot Camp” programs were popular for juvenile offenders among policy makers in the late 1970s, multiple other forms of interventions were developed and delivered to delinquent youth. Interventions that aim to improve sociodemographic characteristics (e.g., educational programs), mental health (e.g., psychiatric or psychological interventions), and criminogenic needs (e.g., sexual offender treatment) are considered rehabilitative forms of correctional interventions in juvenile as well as in adult offenders.

Extensive meta-analyses have revealed that correctional interventions incorporating at least some form of rehabilitation are associated with lower recidivism rates in offenders of all age groups (Lipsey & Cullen, 2007). More specifically, correctional interventions based on a multidimensional treatment approach can prevent criminal careers if used early for delinquent youths and adolescents and reduce recidivism rates among adult violent and sexual offenders. Thus rehabilitative interventions play an important role in accelerating the desistance process for both juvenile and adult offenders, and their use should be considered, keeping specific target populations, expected outcome, and known effect sizes in mind.

2.4.3 Correctional interventions with juvenile offenders

Not only correctional intervention programs are variously present in the criminal justice system, also scientific studies reporting the efficacy for such interventions are more numerous for juvenile offenders than for adult offenders. Also, effect sizes of interventions with juvenile offenders are higher than those of interventions with adult offenders. On a more juridical level, adolescent diversion projects and teen courts show a significant decreasing effect on recidivism in juvenile offenders. Besides that, on a basic level, educational programs and counseling also have a decreasing effect on criminal behavior in juveniles. Looking at more specific programs and those especially designed for juveniles in the correctional system, currently, the most effective programs with the best cost-benefit ratios are the Multidimensional Treatment Foster Care (MTFC) Program, Functional Family Therapy (FFT), and Multisystemic Therapy (MST). These interventions have been included in the so-called „Blueprints” programs, which are evaluated and certified by the Institute of Behavioral Sciences at the University of

Colorado in the US (Mihalic & Irwin, 2003). „Blueprints" programs are characterized by the following three main characteristics: (1) clearly identified program outcomes (ensuring intervention specificity), (2) valid and reliable evaluation findings (ensuring good evaluation quality), (3) positive intervention effects for a minimum of 12 months after the intervention has ended, and dissemination of results

Concerning the actual effect of the programs the Blueprints Advisory Board whereas considers as most important that the effect has been obtained with a strong research design, has a sustained character and has been replicated in multiple settings.

Multidimensional Treatment Foster Care (MTFC) program

The Multidimensional Treatment Foster Care (MTFC) program is a community-based multi-modal treatment program for teenagers with histories of chronic and severe criminal behavior. This program aims to decrease those patterns and increase developmentally appropriate normative and prosocial behavior (Fisher & Chamberlain, 2000). Adolescents taking part in the program are placed in carefully chosen and psychologically trained foster families. A structured network of social workers, therapists, teachers, biological parents, and foster parents supports the treatment process. The teenagers are closely supervised and provided with clear limits and fair and predictable consequences for deviant behavior (Fisher & Chamberlain, 2000). The MTFC is a proactive approach in which case managers pay close attention to the daily behavioral patterns of the teenagers and monitor peer associations. The participants are also encouraged to develop academic skills and positive working habits. Detailed daily activity plans and feedback about their progress facilitates learning and structuring processes. Furthermore, parents and family members receive support and training in parenting styles to decrease family conflicts and dysfunctional relationships. Parents and other family members should also be able to support the youth during the critical aftercare phase, to generalize the gains achieved in therapy to other settings (Fisher & Chamberlain, 2000).

Results of an RCT to evaluate treatment effects showed that the MTFC program considerably decreased reconvictions for violent offenses during the 2-year follow-up. Reconvictions in the experimental group were with 5% significantly lower than 24% reconviction in the control group (Eddy et al., 2004). Results of a meta-analysis by Drake et al. (2009) showed that the MTFC program decreased crime outcomes which were

defined as total arrests, total convictions, felony arrests by 17.9%, with cost-benefits of \$88,953 per participant. This benefit was the highest reported for correctional interventions with juvenile offenders.

Functional Family Therapy (FFT)

Functional Family Therapy (FFT) describes a systematic and responsive family-focused intervention for delinquent but not necessarily violent youth. FFT sees the entire family as the “client” (Sexton & Alexander, 2003; p. 327) and assumes that dysfunctional interpersonal and behavioral patterns in the family are displayed in the antisocial or criminal behavior of the youth. However, family patterns can have protective effects as well as risk-enhancing effects. Focusing on „relational functions within the family” (Sexton & Alexander, 2003; p. 330), FFT has three phases: (1) engagement, where the youth and the family are empowered to participate in therapy and receive psychoeducational training; (2) behavior change to reduce antisocial behavior and adapt new behavioral patterns and problem-solving skills to deal with challenging situations; and (3) generalization of learned strategies to overcome critical situations without a therapist (Sexton & Alexander, 2003). FFT has been evaluated in numerous quasi-experimental and RCT studies, with a reported crime outcome decrease of 18.1% and benefits of \$50,000 to \$57,341 per participant (Aos et al., 2001; Drake et al., 2009).

Multisystemic Therapy (MST)

Multisystemic Therapy (MST) is a family- and community-based treatment to decrease antisocial behavior and promote prosocial behavior in juveniles with serious antisocial behavior and also in young sex offenders (Henggeler, 2011). MST views the family as „key to effective behavior change” (Henggeler, 2011; p. 355) and therefore aims to enhance family functioning. The program is implemented in all social surroundings of the antisocial youth: home, school, and community. All institutions and people close to the adolescent work together, creating a network of interconnected systems. Parents receive support regarding childcare and education to break down dysfunctional family settings and establish sustainable and functional family relationships (Henggeler, 2011). Interventions focus on specific offense/antisocial behaviors and their circumstances and triggers. Additionally, cognitive-behavioral techniques and contingency management are integrated into the broader MST treatment model. To lower barriers and simplify scheduling, sessions can take place at home, school, or in the community (Henggeler,

2011). A quality assurance and improvement system enables constant improvement and quality monitoring of the processes and standardized treatment in all countries where MST programs are implemented. ,

In the past three decades more than 30 outcome studies have evaluated MST, and two-thirds of them were RCTs. Lipsey and Cullen (2007) and Henggeler (2011) reported a significant decrease (16%–46%) in criminal behavior in MST groups, and the benefit was estimated at \$17,694 per participant (Drake et al., 2009). Also, self-reported changes in symptoms and behavior problems change significantly in groups that receive MST (Borduin, Schaeffer, & Heiblum, 2009). To investigate the sustainability of therapy effectiveness, Sawyer and Borduin (2011) followed a group of juvenile offenders who took part in an RCT evaluating an MST intervention for more than 20 years and observed long-lasting effects in terms of recidivism reduction. Since this type of intervention with juvenile offenders is that promising, it has also been implemented in various facilities outside of the U.S. (e.g. Germany, Switzerland).

2.4.4 Intervention programs with no or potentially harmful effects

Not every prevention program for juveniles has effects as promising as those of MFTC, MST, and FFT. Some programs implemented in the criminal justice system (especially in the US) that aimed to reduce antisocial or problematic behaviors have been shown to be ineffective or to even to increase delinquency (Lipsey & Cullen, 2007; Lipsey, Howell, & Kelly, 2010).

Correctional boot camps and „Scared Straight“ programs can serve as examples for the latter whereas intensive probation or parole supervision by court, intensive probation as an alternative to incarceration or guided group interaction seem not to have an effect at all on future criminal behavior (Drake et al., 2009).

Boot camps

Boot camps are military-style correctional facilities in which juvenile offenders or antisocial youth are supposed to learn how to (re-)adapt to societal norms and to change their problem behaviors. The intervention focuses on educating moral concepts, reducing substance abuse and reinforcing responsible behavior. Rules in these camps are very strict following military standards like thrill, discipline and pressure on the offenders to follow the rules. Accordingly, noncompliance is seriously punished and also

obedience to authority, and acquisition of appropriate social skills is constantly emphasized (Lilienfeld, 2007). Nevertheless there are boot camp facilities that do offer rehabilitative treatment services such as substance abuse counseling or group therapy setting. However, as there is no standardized approach, these services differ in quality and availability of interventions.

Boot camps were widely promoted as a cost-effective approach for antisocial adolescents. Nevertheless, a large number of studies revealed no significant effects of boot camps on antisocial behavior. Meade and Steiner (2010) summarized that boot camps would not affect participants' odds of recidivism at all, and also found studies that even reported an increase of recidivism after attending a boot camp. The latter are characterized by the fact that in the investigated boot camps there were no rehabilitative services offered for juveniles. Cost- and benefit analyses report recidivism rates being on average about 10 percent higher for boot camp participants compared to juvenile offenders who went through regular juvenile institutional facilities (Aos et al., 2001). An average boot camp program costs about \$ 8.325 per participant (Drake et al., 2009).

„Scared Straight” programs

Another controversial juvenile correction program is the „Scared Straight” program, which was introduced in the late 1970s in the US. Based on the theory of criminal deterrence, the program aims to scare adolescents or children by showing them their potential future as a delinquent (Lilienfeld, 2007). Juveniles were taken on tours to prisons and were connected with prisoners to get scared off by their lives and history. At the same time, prisoners are often instructed to present themselves and their histories in a brutal and exaggerated way. The program, which became popular after a 1979 television documentary received substantial attention by the media, was implemented in more than 30 states in the US. Petrosino, Turpin-Petrosino, Hollis-Peel, and Lavenberg (2013) performed a systematic review of randomized and quasi-randomized studies evaluating the effectiveness of „Scared Straight” programs, which consisted of organized visits of prisons by juvenile offenders or pre-delinquents. The results after a follow-up of 3 to 24 months showed that arrests, convictions, police contacts, and self-reported offenses were substantially more prevalent in the experimental groups compared to the (randomized) controls (odds ratio [OR]=1.68). „Scared Straight” programs seem not to

have a reducing effect on recidivism. Instead, data suggests a harmful effect of those programs on juvenile offenders. Meta-analytic results report a 6.1% increase in criminal behavior after attending a „Scared Straight” program, with an estimated negative cost-benefit ratio of -\$17,470 per participant (Drake et al., 2009).

Surely there are many more correctional programs around. The aforementioned are found to be the most common and most often evaluated ones. Other programs such as dialectical behavior therapy for juvenile offenders, drug courts, job programs, therapeutic communities need more evaluation studies and/or conceptual framing in the next years.

2.5 Seewald K, Rossegger A, Endrass J (manuscript ready for submission). Examining the effectiveness of an offense-focused treatment program in violent and sexual offenders using propensity score estimation procedures and cox regression model

2.5.1 Abstract

Background: Relapse prevention is an important goal in correctional settings. Whereas there is strong evidence for certain specific treatment programs in misbehaving juveniles and young offenders, programs for adult violent and sexual offenders lack evidence of good methodological quality. The Department of Mental Health Services based in the Office of Corrections in the Canton of Zurich, Switzerland provides interventions with violent and sexual offenders following an offense-focused treatment approach. These interventions were evaluated in their effectiveness regarding relapse-prevention in the present study.

Methods: Since it is rather challenging to implement an RCT-design in forensic settings an observational design had to be used to explore effectiveness of therapeutic interventions. N=255 treated violent and sexual offenders were compared with n=285 untreated violent and sexual offenders, both groups administered by the Office of Corrections and the treated group receiving interventions by the Department of Mental Health Services. To control for selection bias, propensity score (PS) was estimated based on conditional logistic regression model for group assignment and average treatment effects were calculated using PS-matching, inverse probability weighted (IPW-) PS and a double-robust estimation using IPW and regression adjustment (RA). Additionally multivariate cox regression was used to model time-to-event-data and to explore time course of recidivism.

Results: Offenders were observed for 9.0 years (min: 0; max: 29.0 SD=5.6) and had an overall base rate of recidivism with severe violence and/or sexual offenses of 10.1% (n=57). Rates of untreated offenders were significantly higher (13.7%; n=39) than those of treated offenders (7.1%; n=18; $p<0.05$), resulting in OR=0.48 (95% CI: 0.27-0.87, $p<0.05$). With controlling for biases, average treatment effect using PS-matching resulted in a 7% decrease of recidivism in treatment group ($p<0.05$), and in a treatment effect of 8% decrease with PS-IPW ($p<0.05$). The doubly robust estimator (IPW-RA) was slightly not significant for a treatment effect of 6% with $p=0.058$. Hazard ratios for treatment-

only model indicating a decreased risk of recidivism in treatment group of 38-40% were not significant ($p=0.10$, $p=0.15$). However, cox regression models of recidivism in either PS-matched and also in IPW-PS groups fitted data when considering characteristics such as nationality other than Swiss, criminal record with violent or sexual offense, age at begin observational period, and VRAG-sumscore highly significant ($p<0.001$). Survival functions crossed after about 11 years of observation, but post-hoc analysis showed no significant difference in recidivism rates at this point in time.

Discussion: Results show that the effectiveness of interventions with offenders following the offense-focused treatment approach had a significant relapse-preventative effect, even when strictly controlling for potential biases. However, there seems to be evidence that effects of treatment decrease after a little more than ten years of observations. This indicates an urgent need of further research on lasting of treatment effects in larger offender samples. Moreover, potentially moderating characteristics such as prison climate, motivational aspects and reasons for drop-outs should be further investigated.

2.5.2 Introduction

Relapse Prevention by Correctional Intervention Programs

Public attention on violent and sexual offenders' legal probation regularly rises in the light of severe cases of recidivism that catch societal awareness. Hence, reducing recidivism rates of violent and sexual offenders is constantly a major concern in policy makers, juridical and mental health professionals. What kind of correctional supervision or intervention is the most effective and efficient to be implemented is frequently questioned and discussed in the field. Extensive meta-analyses have revealed that correctional interventions incorporating at least some form of rehabilitation are associated with lower recidivism rates (Lipsey & Cullen, 2007). However, compared with programs for misbehaving youth or juvenile offenders, less empirical evidence is available regarding the efficacy of rehabilitative interventions that aim to reduce recidivism rates of adult offenders and reduction rates from 8% to 32% are smaller than those for juveniles (Lipsey & Cullen, 2007). Evaluation of crime prevention strategies in juveniles has focused on specific intervention programs such as Multisystemic Therapy (MST; Henggeler, 2011) or Functional Family Therapy (FFT; Sexton & Alexander, 2003), whereas primary studies of adult offenders have concentrated on the effect of unspecific

interventions in specific types of offenders. Specifically, the evaluation of programs for sexual offenders has been emphasized in the last few years because of increasing societal awareness (Hanson et al., 2009; Jung & Gulayets, 2011; Lipsey & Cullen, 2007; Mandeville-Norden, Beech, & Hayes, 2008; Marques et al., 2005; Woessner & Schwedler, 2014; Woodrow & Bright, 2011) and an overall decreasing effect of recidivism in sexual offenders was identified by Lösel and Schmucker (2005) with 37%.

Not focusing on specific offender types but regarding general treatment approaches in psychotherapy and their effectiveness in the overall offender population, the meta-analysis of Landenberger and Lipsey (2005) showed that participants of cognitive-behavioral interventions were less likely to reoffend (overall OR=1.53), representing a 25% overall decrease in recidivism compared to controls. Furthermore, the most effective programs included in this meta-analysis reported a decrease in recidivism of more than 50%. Lösel and Schmucker (2005) summed up that cognitive-behavioral and classic behavioral therapy significantly decrease sexual recidivism (OR=1.45; OR=2.19), and Beech et al. (2007) reported a significant reduction of recidivism of 40% for sexual offenders who underwent cognitive-behavioral therapy. However, regardless of certain therapy techniques, intervention programs with offenders following the risk-needs-responsivity (RNR)-principle by (Andrews & Bonta, 2010a) are evident to be a widely effective approach: Risk management primarily depends on level of risk and the offender's criminogenic needs that determined the offense. Effective risk management is therefore individually formed and additionally considers the offender's cognitive abilities, willingness and motivation to change and responsiveness.

Offense-focused Therapy

The offense-focused approach used in the treatment delivered by the Department of Mental Health Services in the Office of Corrections, Canton of Zurich is based on this RNR-principle. The department holds the largest forensic mental health service in Switzerland with its main tasks being the delivery of psychiatric service in all prison facilities located in the Canton with about 12'000 psychiatric consultations per year and the implementation of court-mandated therapeutic treatments for violent and sex offenders. These interventions focus on the psychological mechanism that had lead to the offense, based on a comprehensive case formulation. Forensic case formulation generally follows the idea of RNR-principle and integrates various information of the

offender that can be used to determine the etiology of the criminal behavior and to anchor treatment foci (Hart et al., 2011). The practical implementation of case formulation in the Department of Mental Health Services is based on consensus good practice guidelines and aims to identify certain offense related „problem areas“ and criminogenic needs such as personality traits, situational circumstances, psychiatric symptoms which in their combination triggered the criminal behavior and lead to committing the index offense (Urbaniok, 2007). In group and single settings problem areas are targeted by certified forensic psychologists and therapists to enable a behavioral change and to encourage the learning of coping strategies. The latter should support the offender not to relapse with offense-related behavioral pattern and to recidivate. A special circumstance in the Swiss Correctional system is the fact that treatment is delivered in an in- and outpatient setting. No matter if the offenders are imprisoned, on (conditional) release or in freedom, offense-focused intervention is implemented on a regular, once a week basis, mostly starting during imprisonment and continuing through all subsequent settings from open prison facilities to final (conditional) release.

Evaluating Correctional Interventions

While being implemented since 1997 the offense-focused approach of forensic treatments in the Canton of Zurich lacks empirical validation until now. An outcome evaluation of forensic interventions however faces quite a few challenges regarding study design and definition of outcome criteria.

For more than 20 years experts around the globe have passionately debated the best way to evaluate treatment effects (Hollin, 2008; Ioannidis et al., 2001; Marshall & Marshall, 2007, 2008; Persons & Silberschatz, 1998; Seto et al., 2008; Sherman et al., 1998; Slade & Priebe, 2001). The discussion focuses primarily on the trade-off between internal validity (i.e., study design free from biases, allowing causal conclusions) and external validity (i.e., study design allowing generalization of results). Because of their high level of internal validity, randomized controlled trials (RCTs) are widely considered the “gold standard” for the evaluation of treatment effects in general. Especially in the forensic setting this consideration as the gold standard is widely discussed (Farrington & Welsh, 2005; Hollin, 2008; Marshall & Marshall, 2007, 2008; Seto et al., 2008; Slade & Priebe, 2001).

Attempting to avoid all possible forms of biases – what RCT-designs do - leads to „sterile“ study designs that – transferred to forensic settings - do not at all represent correctional reality and can hardly represent the common situation of mandatory interventions by court. In this case trying to eliminate substantial selection bias is only possible by excluding rare categories (e.g., minorities), subjects who are likely to drop out (e.g., delinquents with mental health issues), or subjects with long sentences, what all results in a homogeneous study population of relatively low-risk, mentally healthy offenders of Western European origin.

Alternatively, quasi-experimental studies can account for special circumstances in forensic settings and offer solutions for problems regarding randomization of group assignment, adequate control groups, and ethical dilemmas. For example, treatment responders can be compared with non-responders to evaluate the effect of therapy on recidivism rates (Beech et al., 2012), and a control group of untreated offenders can be matched to an experimental group of treated offenders by propensity score matching (Duwe & Goldman, 2009; Pascual-Leone et al., 2011) or by case-to-case matching if the sample size is small (Polaschek et al., 2005). The most important difference of quasi-experimental, observational designs with RCT-designs is the absence of random assignment of individuals to experimental and control groups, reducing internal validity. In systematic reviews and meta-analyses that included both quasi-experimental studies and RCTs, study design quality was however shown to have little or no influence on treatment effect sizes (Babcock et al., 2004; Lipsey & Cullen, 2007; Lösel & Schmucker, 2005). Nevertheless, knowing the downsides of quasi-experimental studies it needs to be questioned of which quality study outcomes are and how valid interpretations resulting from them can even be. Ioannidis et al. (2001) showed in a large meta-analysis that results from RCT studies are highly correlated with results of quasi-experimental studies. Although quasi-experimental studies tended to show a higher treatment effect, the absolute difference in effect sizes were significant but clinically negligible. In all comparisons, sample size, publication bias, and time lag regarding historical findings also influenced effect size. And theoretically comparing research designs regarding their positive predictive values, Ioannidis (2005) showed that a well-performed but underpowered RCT-design does not significantly outperform an adequately powered exploratory epidemiological study (PPV=0.23 vs. PPV=0.20). Thus, Ioannidis (2005) stated that the pure gold standard of research designs (i.e., unbiased with a PPV=0.85) is

unattainable. In systematic reviews and meta-analyses concerning specifically forensic interventions and including both quasi-experimental studies and RCTs, study design quality had little or no influence on treatment effect sizes (Babcock et al., 2004; Lipsey & Cullen, 2007; Lösel & Schmucker, 2005). In their meta-analysis Landenberger and Lipsey (2005) found rather moderator variables that were responsible for larger effects in some studies compared with others. A high general risk level of the offenders participating in the studies was significantly associated with larger effect sizes, as was the quality of treatment implementation.

Thus, the choice of quasi-experimental designs and observational studies in evaluating the effectiveness of offender therapies appears to be reasonable as long as statistical methods are chosen to address possible selection biases including general risk level and treatment integrity is monitored following standardized guidelines.

Current Study

The current study looks at the effectiveness of interventions that follow the RNR-based offense-focused therapy approach using a retrospective quasi-experimental design in a cohort of violent and sexual offenders treated in the Department of Mental Health Services of the Office of Corrections in the Canton of Zurich. An untreated offender cohort also administered by the Office of Corrections was considered as control group. Since all interventions of the Department are court-mandated it is obvious that individual characteristics influence treatment selection. Hence, treatment and control group are expected to differ significantly in characteristics that typically lead to mandating treatment and may also be influencing the outcome. The methodological approach in this study accounts for this selection bias by using propensity score estimation and calculates average treatment effects based on this score under different assumptions. A positive effect of offense-focused intervention, specifically in form of relapse-prevention is expected in the treatment group compared to the untreated control group.

2.5.3 Methods

Participants

Intervention group consisted of male offenders who had committed a serious violent and sexual offense and received an offense-focused intervention by forensic therapists

in the Department of Mental Health Services in the Canton of Zurich. Interventions in the cohort observed in this study started between 1997 and 2009. All interventions were mandated by court and had to take at least twelve months to be included in the evaluation (n=255).

Control group consisted of a cohort of male offenders convicted for a serious violent or sexual offense that were administered in the Office of Corrections in August 2000. Inclusion criteria were being fully responsible for criminal actions, having a minimum sentence length of 10 months, not receiving any form of court-mandated intervention nor receiving an intervention by choice (n=285).

Measures

For all participants an extensive list of variables was retrospectively collected by file information. These variables included demographic variables, information about the index offense, criminal records, prior contact with psychiatric services, course of therapy, static and dynamic risk assessments (PCL-R, VRAG; Hare, 2003; Quinsey et al., 2006); as well as circumstances of end of therapy and subsequent settings.

Treatment Integrity

Treatment integrity in the intervention group was operationalized by training and continuous further education of therapists and staff in case formulation and offense focusing therapeutic techniques. Case formulation followed the offense-orientated approach by Urbaniok (2007) and was always done by a group of certified therapists to ensure a comprehensive explanation of the specific case and a profound understanding of problem areas to focus on in therapy. Progress of interventions and the emphasis on specific offense-related topics was constantly monitored and there was a strict quality management with standardized progress and interim reports.

Taking these measures into account, most criteria of Andrews and Dowden (2005)'s treatment integrity concept by were met.

Observational Period and Post-Release Follow-up

To avoid underestimation of recidivism rates by including re-offenses that were committed during prison leaves and in open prison facilities observational period and post-release follow-up are considered separately (see figure 9). Observational period for offenders in the intervention group started 12 months after begins of therapy. The 12-

months-criterion was set as such along with therapists in the Department to ensure proper treatment implementation after triage period as well as program integrity and establishment of therapeutic relationship. To elevate comparability offenders in the control group were therefore observed starting 12 months after date of conviction for index offense. The date was used as the start point of observational period without regard to whether offenders were already released or in an open prison setting for both groups.

However, post-release follow-up started with date of conditional release in both groups. In case observational period had not started by then, begin of follow-up was set as begin of observational period. Post-release follow-up and observational period both ended for all participants with date of criminal record (retrieved in September 2013), except for those who died, were deported out the country or recidivated. In the latter cases observational period and post-release-follow-up ended with these specific dates.

The separation of observational period and post-release follow-up permits that the study sample includes offenders who were never actually free because they committed an offense that is classified as recidivism while in custody or in an open prison facility.

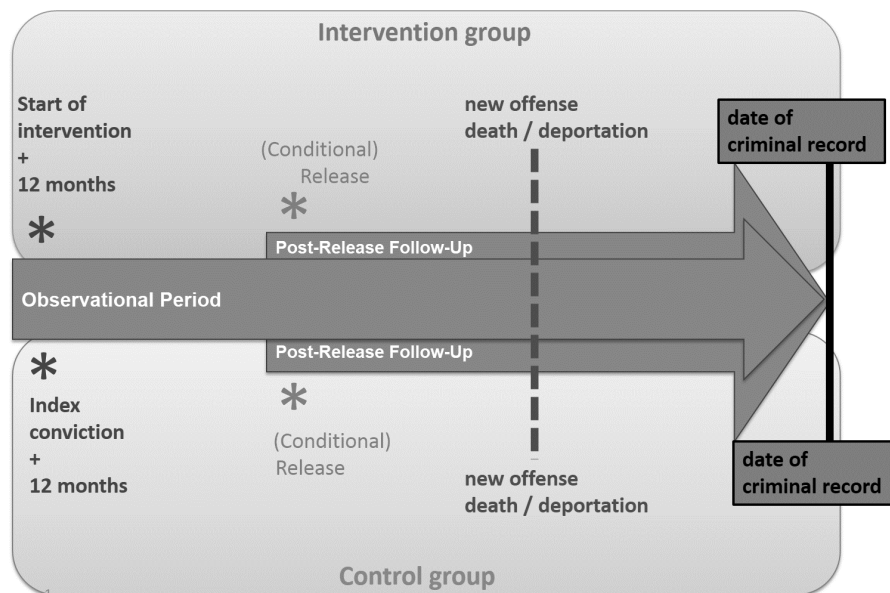


Figure 9: Observational period and post-release follow-up

Outcome Criterion

Positive treatment effect was defined as less recidivism rates in intervention group compared to control group. Recidivism rates included new convictions or current investigations for severe violent and/or sexual offenses committed after begin of observational period. Thus the outcome was not coded as such if the conviction happened after observational period had started but could be determined as a result of an offense committed prior to index offense. Offenses included in the definition of violent/sexual recidivism were homicide, manslaughter, robbery, assault (incl. violence against public servants), coercion, kidnapping, hostage-taking, putting others in life-threatening situations, rape, sexual abuse of children, incest, sexual coercion, sexual molesting. Recidivism was extracted out of criminal records retrieved in September 2013.

Statistical Measures

Analyses of Group Differences

As inherent in the evaluation of treatment in forensic settings and already mentioned in this manuscript, an RCT-study design could not be applied in the current study. Within the quasi-experimental nature of this study, the selection bias is expected to result in a priori differences between intervention and control group. Certain characteristics that are expected to differ between groups also may be related to the outcome what might strongly influence estimated treatment effects. An expert panel of the Department of Mental Health Services created a list of such clinically relevant characteristics (table 4). Descriptive statistics (t-tests and χ^2 tests according to nature of variables) were calculated to examine significant differences between control and intervention group.

Table 4: Differences in prematched groups on key characteristics identified by expert panel

Key Characteristics	Treated (n=255)	Controls (n=285)	Difference	% Missing Values
Nationality other than Swiss	32.2%	57.5%	**	0
Criminal record with violent or sexual offense	49.2%	34.2%	**	1.3
Failure on conditional release prior to index offense	33.5%	25%	*	11.1
Mean age at index conviction	36.3 (min: 19.2; max: 71.1; SD=11.0)	34.4 (min: 17.2; max: 71.0, SD=11.3)	*	0.3
Type of index offense (violent vs. sexual)	Violent: 56.5% Sexual: 43.5%	Violent: 60% Sexual: 40%	-	0
Mean age at begin observational period	38.4 (min: 19.3; max: 71.3, SD=11.0)	35.5 (min: 18.2; max: 72.0, SD=11.3)	**	0
PCL-R-sumscore (mean)	13.8	15.5	*	40.7
VRAG-sumscore (mean)	2.5	3.7	-	17.4
Psychiatric diagnoses	88.2%	65.6%	**	23.3
Influence of alcohol and/or drugs during index offense	49.6%	39.5%	*	4.2

* $p < .05$; ** $p < .01$

Estimation of Propensity Score

Statistical procedures had to be used in order to consider the aforementioned selection bias. The methods of choice in observational studies to assess treatment effects and to control for selection bias are based on propensity score estimation (Guo & Fraser, 2010; Rosenbaum & Rubin, 1983).

Propensity Score (PS) estimation and subsequent procedures provide a commonly used approach producing results most closely to those that could be expected with a randomized design (Yue, 2007). PS is estimated post-hoc and as such also provides the benefit of including a large number of individuals in both groups and hence increasing statistical power. PS represents the predicted probability of being in the intervention group. Stepwise binomial logistic regression was used to develop the conditioning model of group selection (see table 5). To create the most parsimonious regression model only those characteristics identified by the expert panel that differed significantly between the prematched groups and had less than 20% of missing values were included in the primary model. VRAG-sumscore was included in the model due to clinical relevance (Hosmer, Lemeshow, & Sturdivant, 2013) and age at begin observational period was excluded, as it could not have determined group assignment beforehand. Characteristics were stepwise excluded on a p-level of 0.05 (Hosmer et al., 2013). The final binomial logistic regression model was used to predict the PS for each offender in the study.

Table 5: Stepwise binary logistic regression modeling of group assignment to estimate propensity score

Key Characteristics	Primary Model		Model 2		Model 3		Model 4/Final Model	
	OR	p> z	OR	p> z	OR	p> z	OR	p> z
Nationality: Swiss/Non-Swiss	0.33	0.000	0.35	0.000	0.32	0.000	0.29	0.000
Criminal record with violent or sexual offense	2.22	0.002	2.19	0.002	2.02	0.002	2.08	0.001
VRAG-Summscore	0.96	0.002	0.96	0.001	0.97	0.011	0.97	0.003
Influence of alcohol and/or drugs during index offense	1.93	0.004	1.91	0.005	2.01	0.002	1.91	0.003
Type of index offense (Violent vs. sexual)	0.57	0.023	0.60	0.031	0.66	0.075		
Failure on conditional release prior to index offense	1.36	0.243	1.33	0.276				
Age at index	0.99	0.587						

Using Propensity Score: Matching and Inverse Probability Weighting

For the use of PS in time-to-event data (post-estimation survival analysis), Austin (2014) suggested procedures of either propensity score matching or inverse probability weighting using the propensity score. While matching procedures seem to be the more transparent method than complex weighting procedures, the latter additionally account for potential misestimating of the PS (Austin, 2014).

In the present study analysis of average treatment effects and post-estimation survival analysis are examined with both procedures of using propensity score. Advantages, disadvantages and potential effects of using either procedure on the results will be considered in the discussion section. For the matching procedure nearest neighbor matching within a caliper distance of 0.2 was chosen as this matching method is especially suggested for studies with post-estimation multivariate analysis and a dichotomous dependent outcome criterion (Guo & Fraser, 2010) and the specific width of 0.2 as it is recommended for models with at least one continuous covariate (Austin, 2011). Differences in key characteristics between post-matched groups are examined by descriptive statistics such as t-tests and χ^2 tests according to nature of characteristics.

Using Propensity Score: Double Robust Estimator

In addition, a double robust estimator of average treatment effect was calculated using inverse probability weighting (IPW) and regression adjustment (RA) in one model. This approach controls for misestimating PS and applies very strict criteria to the data set (Emsley, Lunt, Pickles, & Dunn, 2008). The regression adjustment followed again a stepwise logistic regression modeling the outcome criterion (severe violent and/or sexual recidivism) with characteristics identified by the expert panel. In this model, age at begin of observational period was included and along with the previous modeling of group assignment characteristics were stepwise excluded on a p-level of .05 (Hosmer et al., 2013). All average treatment effects were calculated using the *teffects* package in Stata 13 (StataCorp, 2013).

Post-Estimation Survival Analysis

Following propensity score estimation a survival analysis was conducted to explore time-dependent course of individuals in both groups controlling for varying follow-up times and right censored data: for some individuals the follow-up ended before the

outcome of interest is observed (Hosmer, Lemeshow, & May, 2008). Semi-parametric proportional hazards model (“Cox regression model”) was used to model the effect of treatment selection on time to event. Ties were handled with Breslow modification as suggested by (Hosmer et al., 2008). Cox regression model was separately run for (1) control and intervention group matched by PS and (2) with inverse probability weights predicted based on PS.

Post-estimation survival analysis was done with Stata 13.1 and a significance level of $p=0.05$ (StataCorp, 2013). Figure 10 gives an overview of all methodological steps.

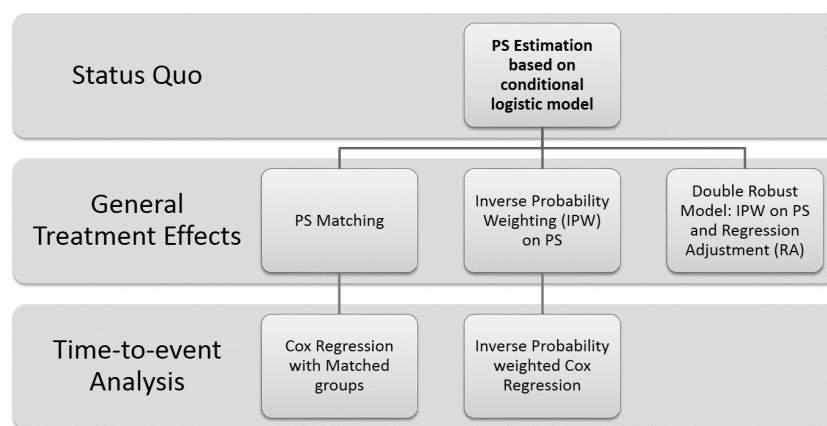


Figure 10: Methodological steps following propensity score estimation.

2.5.4 Results

Sample Characteristics

When observational period started, mean age in the overall sample was 36.8 years (min: 18.2; max: 72.0, SD=11.2) with 35.5 years (min: 18.2; max: 72.0, SD=11.3) in the control group and 38.4 (min: 19.3; max: 71.3, SD=11.0) in the treated group (table 3). Treated offenders were significantly more often diagnosed with personality disorder (53.5% vs. 26.3%, $p<0.01$), but there was no significant difference in diagnosis of schizophrenia (5.6 % vs. 4.7%). The amount of missing variables on diagnosis should be noted at that point (23.3%, see table 4).

Observational Period and Post-Release Follow-up

Mean observational period was 9.0 years (min: 0; max: 29.0 SD=5.6) in the overall sample. Control group was observed for an average time of 10.0 years (min: 0; max: 29.0 SD=6.3). Intervention group had a significantly shorter observational period ($t=4.56$,

$p < 0.01$) for an average of 7.8 years (min: 0; max: 15.7; SD=4.4). Mean post-release follow-up was 6.9 years (min: 0; max: 28.6, SD=5.4) in the overall sample. Control group had a post-release follow-up time of 7.6 years (min: 0; max: 28.6; SD=5.8). Intervention group had again a significantly shorter post-release follow-up time ($t=3.49$, $p < 0.01$) for an average of 5.9 years (min: 0; max: 15.4; SD=4.6; table 6).

Intervention Characteristics

In the treatment group, mean duration of intervention was 4.7 years (min: 1; max: 16.5; SD=3.1; table 6). About a quarter of interventions (25.8%) ended irregularly due to reasons such as dropouts due to acute crisis, dropouts due to external circumstances, etc).

Table 6: Descriptive study characteristics by group

Descriptive Study Characteristics	Treated (n=255)	Controls (n=285)	Difference
Observational period in years	7.8 (min: 0; max: 15.7; SD=4.4)	10.0 (min: 0; max: 29.0 SD=6.3)	**
Post-release follow-up period in years	5.9 (min: 0; max: 15.4; SD=4.6)	7.6 (min: 0; max: 28.6; SD=5.8)	**
Diagnoses of Personality Disorder	53.5%	26.3%	**
Diagnoses of Schizophrenia	4.7%	5.6%	n. s.
Mean duration of intervention in years	4.7 years (min: 1; max: 16.5; SD=3.1)		
Treatment attrition	25.9%		
Severe violent and/or sexual recidivism during observational period	7.1%	13.7%	*

* $p < .05$; ** $p < .01$, n.s.: no significant difference

Recidivism

Base rate of recidivism with severe violent and/or sexual offense was 10.1% (n=57) in the overall sample. 13.7% (n=39) of offenders in the control group recidivated during the observational period and 7.1% (n=18) of treated offenders in the control group did so. Rates were significantly different with $P\text{-}\chi^2=6.26$, $p<0.05$ (table 3), resulting in odds ratio of 0.48 (95% CI: 0.27-0.87, $p<0.05$).

Table 7 shows stepwise binary logistic regression results of modeling the outcome criterion in the overall sample. The final model predicting severe violent and/or sexual recidivism consisted of nationality, criminal record with violent and/or sexual offense, age at begin observational period and VRAG-sumscore with LR- $\chi^2=49.44$, $p<0.01$.

Propensity Score and Average Treatment Effects

Table 5 presents steps in model building of group assignment. The final model predicted PS based on nationality, criminal record with violent or sexual offense, VRAG-sumscore and influence of alcohol and/or drugs during the index offense with LR- $\chi^2=54.16$, $p<0.01$.

PS-Matching

There was a significant treatment effect of $\beta=-0.07$ ($p=0.042$), when matching control and intervention group on PS (table 8). Table 9 shows distribution of characteristics predicting PS in postmatched groups sample (control group: n=209, intervention group: n=221).

IPW

When using inverse probability weighted average treatment effect estimation based on prediction of PS, there was a significant treatment effect of $\beta=-0.08$ ($p=0.012$, table 8).

IPW-RA

When using the double robust estimator with inverse probability weights and regression adjustment the average treatment effect estimation was slightly not significant of $\beta=-0.06$ ($p=0.058$, table 8).

See figure 11 for an overview of different effects and significance levels

Table 8: Treatment effects assessed with different procedures based on estimation of propensity score

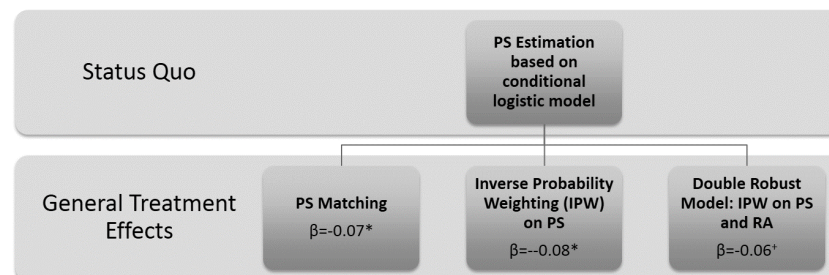
Procedure	Treatment Effects				
	β	Std.-Err.	z	p> z	95% CI
PSM					
ATE	-0.07	0.036	-2.03	0.042	-0.14; -0.00
IPW					
ATE	-0.08	0.03	-2.52	0.012	-0.14; -0.18
Double Robust: IPW-RA					
ATE	-0.06	0.03	-1.90	0.058	-0.12; 0.00

ATE= Average Treatment Effect; PSM=Propensity Score Matching, IPW=Inverse Probability Weighting; IPW-RA=Inverse Probability Weighting and Regression Adjustment

Table 9: Differences in characteristics predicting PS in postmatched groups

Characteristics	Treated (n=221)	Controls (n=209)	%bias	t	p> z
Nationality other than swiss	0.32	0.29	5.7	0.62	0.537
Criminal record with violent or sexual offense	0.52	0.48	6.4	0.66	0.507
VRAG-sumscore	2.38	3.50	-10.6	-1.06	0.288
Influence of alcohol and/or drugs during index offense	0.54	0.54	-7.3	-0.76	0.447

* $p < .05$; ** $p < .01$



* $p < .10$; * $p < .05$

Figure 10: Overview of effects of offense-focused treatment depending on methodological step

Time-to-event Analysis

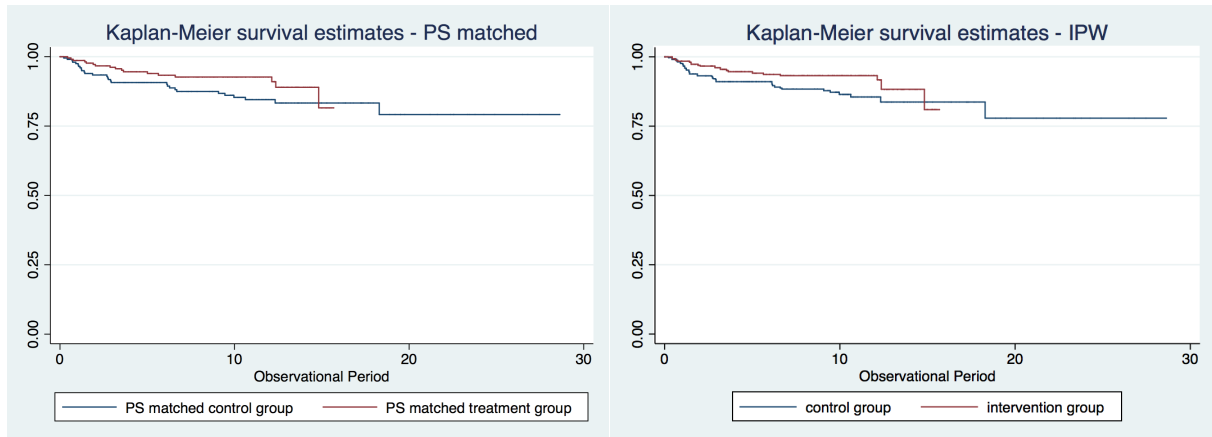
Descriptive Results

Kaplan-Meier estimates of the survival function by treatment for both cox models (PS and IPW) are shown in figures 12a and 12b. Log-rank test for the quality of Survival Functions showed a slightly non-significant difference of survival functions per treatment group in PS-matched groups ($p=0.09$) and a more clearly non-significant difference in IPW-groups ($p=0.15$). Exploring the survival functions figures 13a and 13b show the kernel-smoothed hazard function describing the underlying survival process per group and figures 6a and 6b the Schoenfeld Residuals to test for proportionality of hazards.

Even though hazard functions cross at a time point of 11 years, and Schoenfeld Residuals cross the base line of 0, hazards rates were assessed proportional for both cox models ($p=0.28$; $p=0.14$) and a cox model with characteristics predicting the outcome (see table 7) was fitted to first the PS matched data and second with inverse probability weights (table 10). Post-hoc χ^2 -Test revealed no significant difference in recidivism rates after the crossing point: 3.2% ($n=3$) in treatment group only consisting of $N=93$ offenders and 2.4% ($n=4$) in control group with remaining $N=165$ offenders.

Cox Regression with PS-Matched Groups and Inverse Probability Weights based on PS

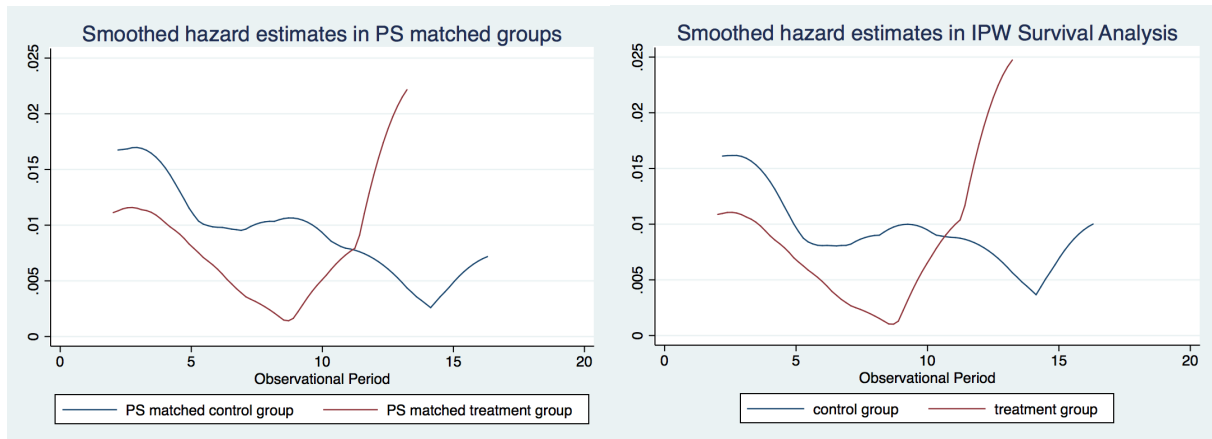
In the treatment-only model treatment group had a lower hazard (risk) of recidivism than control group both in the PS matched group analysis ($HR=0.60$) and also in the inverse probability weighted analysis ($HR=0.62$). Significance level of $p=0.05$ could though not be achieved in both treatment-only regression models (table 10). Multivariable cox regression was based on results of logistic modeling of recidivism data (see table 7) and revealed significant overall Wald-Statistics with $p<0.01$ in both analyses and gave evidence for significant influence of criminal record with violent and/or sexual offense ($HR=2.36/p=0.008$; $HR=2.25/p=0.015$), age at begin of observational period ($HR=0.93/p=0.001$; $HR=0.92/p=0.001$) and VRAG-sumscore ($HR=1.04/p=0.039$; $HR=1.04/p=0.005$) on recidivism with severe violent and/or sexual offenses (table 10).



Log-Rank test for equality of survivor functions:
 $\chi^2=2.74, p=0.09$

Log-Rank test for equality of survivor functions:
 $\chi^2=2.12, p=0.15$

Figure 12: Post-estimation Survival Analysis: Kaplan-Meier survival function in (a) PS matched groups and (b) IPW Survival Analysis



Global Test of Proportionality of Hazards:
 $\chi^2=6.31, p=0.28$

Global Test of Proportionality of Hazards:
 $\chi^2=8.23, p=0.14$

Figure 12: Post-estimation Survival Analysis: Kernel-smoothed hazard function per group in (a) PS matched groups and (b) IPW Survival Analysis

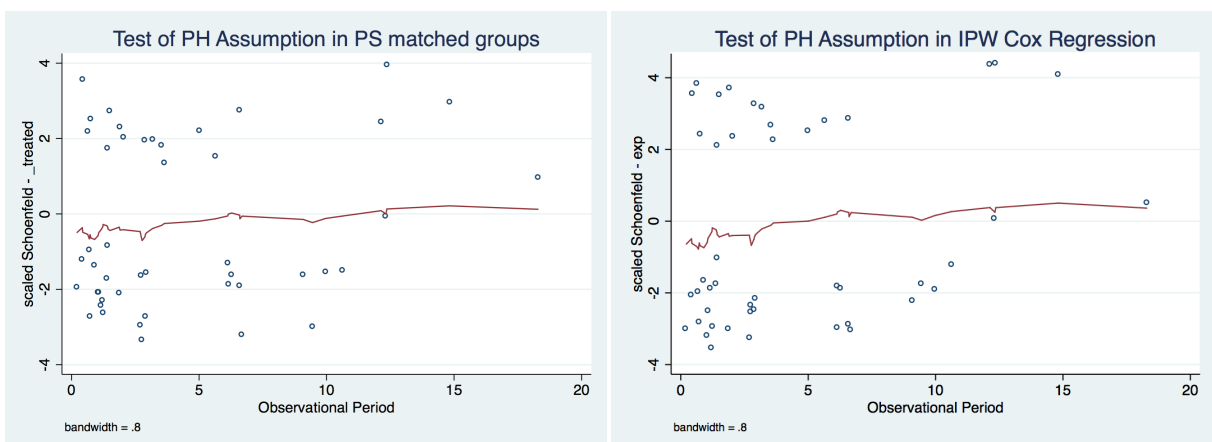


Figure 13: Post-estimation Survival Analysis: Schoenfeld residuals in (a) PS matched groups and (b) IPW Cox regression

Table 10: Post-estimation Survival Analysis: Cox Regression Results with PS-matched groups and IPW

Procedure	Cox Regression					
	HR	Std.-Err.	z	p> z	95% CI	
Treatment-only model	0.60	0.19	-1.64	0.101	0.32;1.10	
Wald $\chi^2=2.77$; $p=0.0961$						
Treatment yes/no	0.75	0.24	-0.89	0.374	0.39;1.42	
Nationality other than Swiss	0.62	0.20	-1.43	0.154	0.33;1.19	
Multi-variable model	2.36	0.77	2.64	0.008	1.25;4.47	
Age at begin observational period	0.93	0.02	-3.48	0.001	0.89;0.97	
VRAG-Sumscore	1.04	0.02	2.07	0.039	1.00;1.07	
Wald $\chi^2=46.21$; $p<0.01$						
IPW	HR	Std.-Err.	z	p> z 	95% CI	
Treatment-only model	0.62	0.21	-1.46	0.145	0.32;1.18	
Wald $\chi^2=2.12$; $p=0.1452$						
Treatment yes/no	0.74	0.26	-0.87	0.385	0.38;1.46	
Nationality other than Swiss	0.65	0.21	-1.32	0.188	0.34;1.23	
Multi-variable model	2.25	0.75	2.43	0.015	1.17;4.34	
Age at begin observational period	0.92	0.02	-3.41	0.001	0.88;0.97	
VRAG-Sumscore	1.04	0.02	2.83	0.005	1.01;1.07	
Wald $\chi^2=35.11$; $p<0.01$						

PSM: Propensity Score Matched groups; IPW: Inverse Probability weights on cox regression

2.5.5 Discussion

Main goal of the present study was to examine the effectiveness of RNR-based offense-focused therapy in a cohort of violent and sexual offenders using a retrospective observational study design. Average treatment effects were calculated controlling for possible biases with PS matching, IPW based on PS and doubly robust regression adjustment. Time-to-event analysis was used to control for right-censored data and varying follow-up times: Two cox models were fitted with either PS matched groups or IPW groups.

It needs to be stated at first, that the forensic cohort of violent and sexual offenders in the Canton of Zurich has a base rate of recidivism for severe violent and/or sexual offenses of 10.1% and that is substantially lower than base rates reported in other offender samples (Helmus, Hanson, Thornton, Babchishin, & Harris, 2012; Quinsey et al., 2006; Rossegger, Endrass, Gerth, & Singh, 2014). This aggravates, of course, the significant reduction by an effective intervention.

However, treated offenders had a significantly lower recidivism rate (7.1%) than controls (13.7%). Concerning the „pure“ rates, indicating a reduction of 6.6 percentage points or 48.2% less recidivism in treated individuals than in controls, these results are in line with previous results of offender treatment effectiveness (Lipsey & Cullen, 2007; Lösel & Schmucker, 2005). The very strict statistical controlling that was done in this study, for biases of group selection on the one hand and also for influences on recidivism over time on the other hand, revealed a substantially lower but significant treatment effect of 7%-8% using and PS matching and PS inverse probability weights. When using a doubly robust estimator that additionally controlled for errors in estimation of PS, treatment effect of 6% was slightly not significant. Yet this should not be interpreted as an alleviation of the overall effect as $p=0.058$ is still in an acceptable range concerning the strings doubly robust estimators put on the data.

Treatment-only cox regression model was not significant neither in PS matched nor in IPW groups for a risk reduction of offense-focused therapy of 38%-40% (HR: 0.62, $p=0.10$; HR: 0.60, $p=0.15$). As shown in the multivariable model characteristics like age, criminal record or risk level (based on VRAG) seem to have more influence on recidivism over time than treatment only. The crossing hazards functions at about 11 years of observational period indicate that the probability of recidivism in treated

offenders at this point in time exceeds the chances of recidivism in untreated offenders. However, post-hoc analyses showed that differences in recidivism rates were not significant, presumably due to small sample size remaining in both groups.

While preliminary results suggest that offense focused therapy is a promising approach (Urbaniok, Endrass, & Rossegger, 2008), the present study is a further evidence for the effectiveness of RNR-based treatment approaches in general and especially of the offense-focused approach (Urbaniok, 2007) used in the Department of Mental Health Services in Zurich. However, the comparatively low reduction rates when controlling for all possible biases ask for further research as also do the not significant results of cox regression. Moreover, future research is needed to explore long lasting effects of this therapy approach, especially considering the crossing hazards after ten years of observational period, presumably requiring larger sample.

Limitations

A first limitation is based on the longitudinal design of the study. To create a reasonable sample size, treated offenders were followed up starting with their therapy in 1997. Effectively, conceptualization of offense-focused therapy in the Department of Mental Health Services in Zurich has changed over the years due to new developments in the field of offender therapy. Hence, there may not be an actual bias in individual quality of treatment delivery but certainly in quality management, theoretical considerations and understanding of risk factors and criminogenic needs.

Furthermore, whereas the use of PS was inevitable to control for selection bias and to increase the chance of reporting true effect sizes of offender treatment in this study, the PS itself limits the reported results: PS can only consider observable, identified variables that may have had an influence on group assignment but misses variables and characteristics that may not be observed or identified by the authors or the expert panel. Accounting for all possible advantages of PS compared to RCT-designs in forensic settings, a randomized group assignment could also control for potential differences in variables between groups that cannot be identified by a research team (Biondi-Zoccai et al., 2011; Yue, 2007).

A third limitation of this analysis may stimulate further interpretation of the present data as it applies to treatment setting and dropout. In the present analysis an „intent-to-treat“-approach was used. All offenders for which the intervention had started between

1997 and 2009 were included with no regard to the practice of ending the intervention, whether this was regularly, or a dropout and to which circumstances. Additionally, some offenders in the treatment group have changed the treatment setting during observational period. According to Lösel and Schmucker (2005) there may be larger effect size to be expected in an ambulatory program than in an institutional setting. For further analysis a “treatment-as-delivered”-analysis, stratified by reasons for dropout and with consideration of changes in treatment setting as a time-varying covariate should be focused.

Practical Implications and Future Research

The major implication for practical work in forensic settings is that interventions based on an individual case formulation following a standardized formulation approach is of practical utility as it facilitates communication between various professional disciplines working in the juridical and correctional system. And not only that juridical and correctional staff supports the standardized use and implementation of case formulation in forensic practice, the identification of criminogenic needs relevant for treatment also seems to result in positive, relapse-preventative effects. Forensic psychiatrists and therapists should continue in the use and constant further development of RNR-based case formulation in treatment delivery.

Regarding future research directions, one needs to take into account that the present study was part of a longitudinal project of the Office of Corrections in the Kanton of Zurich, Switzerland. Data were extensively collected over years and there are numerous research questions that could be investigated in this offender cohort regarding epidemiological aspects, risk assessment accuracy, treatment attrition and individual progress. Effects of dynamic factors such as prison climate and motivational aspects on therapeutic change and relapse-prevention should be further investigated. Primarily major research is needed concerning the specific approach of offense-focused interventions including the accuracy, reliability and acceptability of case formulation and its specific implementation in the juridical and correctional system in Switzerland. Also, issues of specific efficacy of the approach for certain offender groups or in preventing certain re-offenses should guide further analyses of the data.

In addition, especially in governmental institutions such as the Swiss Office of Corrections, economic considerations such as cost-benefit analyses are becoming more

and more important and also demanded by civil society. Particularly in the context of sex and violent offender treatment, policy makers point out the limited acceptance of the general population regarding allocation of substantial financial resources for rehabilitative interventions. Assuming that reservations about this spending exist, cost-benefit analyses can help policy makers and forensic mental health professionals to make the case for implementing these rather costly but - as empirically shown - relapse preventing correctional programs.

EIGENLEISTUNG

Forensische Epidemiologie psychischer Störungen

1. Fazel S, Seewald K (2012) Severe mental illness in 33,588 prisoners worldwide: systematic review and meta-regression analysis. *The British Journal of Psychiatry* 200(5), 364-373.

Eigener Beitrag:

Systematische Literaturlaufarbeitung, Datenkollektion- und -extraktion, sowie Konzeption und Durchführung der Datenanalyse, Erstellung der Manuskriptinhalte in Methoden- und Resultateteil, Mitwirkung bei der Erstellung der übrigen Kapitel

Risikobeurteilungen

2. Seewald K, Rossegger A, Endrass J (2015). Risikoeinschätzungen bei Sexualstraftätern im institutionellen Bereich. In JM Fegert & M Wolff (Eds.) Kompendium „Sexueller Missbrauch in Institutionen“: Entstehungsbedingungen, Prävention und Intervention (pp 561-573). Weinheim: Verlagsgruppe Beltz.

Eigener Beitrag:

Inhaltliche Konzeption, Erstellung des Manuskripts

3. Seewald K, Rossegger A, Urbaniok F, Endrass J (under review) Assessing the risk of intimate partner violence: Expert evaluations versus the Ontario Domestic Assault Risk Assessment. *International Journal of Offender Therapy & Comparative Criminology*

Eigener Beitrag:

Datenbereinigung, Datenanalyse, Erstellung des Manuskripts

Risikoorientierte Interventionen

4. Seewald K, Yundina A, Endrass J (in press). Juvenile Corrections. In Kerley KR, Copes H, De Li S, Lane J, Sharp SF (Eds.) *The Encyclopedia of Corrections*. Wiley Blackwell

Eigener Beitrag:

Inhaltliche Konzeption, Erstellung des Manuskripts

5. Seewald K, Rossegger A, Endrass J (manuscript ready for submission) Examining the effectiveness of an offense-focused treatment program in violent and sexual offenders using propensity score estimation procedures and cox regression model

Eigener Beitrag:

Mitentwicklung des Studiendesigns, Datenerhebung und -bereinigung, Auswahl der Datenanalysemethoden, Erstellung des Manuskripts

LITERATURVERZEICHNIS

- Abdalla-Filho, E., De Souza, P. A., Tramontina, J. F., & Taborda, J. G. V. (2010). Mental disorders in prisons. *Current Opinion in Psychiatry*, 23(5), 463-466. doi: 10.1097/YCO.0b013e32833bb32f
- Ægisdottír, S., White, M., Spengler, P., & Maugherman, A (2006). The Meta-Analysis of Clinical Judgment Project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist*, 34(3), 341-382. doi: 10.1177/0011000006286696
- Agbahowe, S. A., Ohaeri, J. U., Ogunlesi, A. O., & Osahon, R. (1998). Prevalence of psychiatric morbidity among convicted inmates in a nigerian prison community. *East African Medical Journal*, 75, 19-26.
- Ajiboye, P. O., Yussuf, A. D., & Issa, B. A. (2009). Current and lifetime prevalence of mental disorders in a juvenile borstal institution in Nigeria. *International Journal of Research in Medical Sciences*, 3(1), 26-30.
- Alevizopoulos, G., & Igomenou, A. (2011). Mental health problems in male prisoners in Greece and their relation to the criminal history. *International Journal of Law and Psychiatry*(in press).
- Andersen, H. S., Sestoft, D., Lillebaek, T., Gabrielsen, G., & et al. (1996). Prevalence of ICD-10 psychiatric morbidity in random samples of prisoners on remand. *International Journal of Law and Psychiatry*, 19(1), 61-74. doi: 10.1016/0160-2527(95)00025-9
- Andrews, D. A., & Bonta, J. (2010a). *The psychology of criminal conduct*: Routledge.
- Andrews, D. A., & Bonta, J. (2010b). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy, and Law*, 16(1), 39-55. doi: 10.1037/a0018362
- Andrews, D. A., Bonta, J., & Wormith, J. S. (2004). *The Level of Service/ Case Management Inventory (LS/CMI)*. Toronto: Multi Health Systems.
- Andrews, D. A., Bonta, J., & Wormith, J. S. (2011). The risk-needs-responsivity model: Does adding the Good Lives Model contribute to effective crime prevention? *Criminal Justice and Behavior*, 38(7), 735-755. doi: 10.1177/0093854811406356

- Andrews, D. A., & Dowden, C. (2005). Managing correctional treatment for reduced recidivism: A meta-analytic review of programme integrity. *Legal and Criminological Psychology, 10*(2), 173-187. doi: 10.1348/135532505X36723
- Andrews, D. A., & Dowden, C. (2009). Keeping a respectful eye on crime prevention: A response to Birgden. *Canadian Journal of Criminology and Criminal Justice, 51*(1), 119-122. doi: 10.3138/cjccj.51.1.119
- Andrews, D. A., Zinger, I., Hoge, R. D., Bonta, J., Gendreau, P., & Cullen, F. T. (1990). Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology, 28*(3), 369-404. doi: 10.1111/j.1745-9125.1990.tb01330
- Aos, S., Phipps, P., Barnoski, R., & Lieb, R. (2001). The comparative costs and benefits of programs to reduce crime. Version 4.0.
- Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment, 87*(1), 84-94. doi: 10.1207/s15327752jpa8701_07
- Arkes, H. R. (1981). Impediments to accurate clinical judgment and possible ways to minimize their impact. *Journal of Consulting and Clinical Psychology, 49*(3), 323-330. doi: 10.1037/0022-006x.49.3.323
- Arnab, B., Prativa, S., & Ray, T. K. (2009). Persons with major psychiatric illness in prisons - a three years study. *Journal of the Indian Medical Association, 107*(1), 14.
- Assadi, S. M., Noroozian, M., Pakravannejad, M., Yahyazadeh, O., Aghayan, S., Shariat, S. V., & Fazel, S. (2006). Psychiatric morbidity among sentenced prisoners: prevalence study in Iran. *The British Journal of Psychiatry, 188*(2), 159-164. doi: 10.1192/bjp.188.2.159
- Austin, P. C. (2011). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics, 10*(2), 150-161. doi: 10.1002/pst.433
- Austin, P. C. (2014). The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments. *Statistics in Medicine, 33*(7), 1242-1258. doi: 10.1002/sim.5984
- Babcock, J. C., Green, C. E., & Robie, C. (2004). Does batterers' treatment work? A meta-analytic review of domestic violence treatment. *Clinical Psychology Review, 23*(8), 1023-1053. doi: 10.1016/j.cpr.2002.07.001

- Baillargeon, J., Binswanger, I. A., Penn, J. V., Williams, B. A., & Murray, O. J. (2009). Psychiatric disorders and repeat incarcerations: The revolving prison door. *American Journal of Psychiatry*, *166*(1), 103-109. doi: 10.1176/appi.ajp.2008.08030416
- Baillargeon, J., Penn, J. V., Knight, K., Harzke, A. J., Baillargeon, G., & Becker, E. A. (2010). Risk of reincarceration among prisoners with co-occurring severe mental illness and substance use disorders. *Administration and Policy in Mental Health and Mental Health Services Research*, *37*(4), 367-374. doi: 10.1007/s10488-009-0252-9
- Baillargeon, J., Penn, J. V., Thomas, C. R., Temple, J. R., Baillargeon, G., & Murray, O. J. (2009). Psychiatric disorders and suicide in the nation's largest State prison system. *Journal of the Academy of Psychiatry and the Law*, *37*(2), 188-193.
- Barnett, G. D., Manderville-Norden, R., & Rakestrow, J. (2014). The Good Lives Model or Relapse Prevention What Works Better in Facilitating Change? *Sexual Abuse: A journal of research and treatment*, *26*(1), 3-33. doi: 10.1177/1079063212474473
- Bartholomew, A. A., Brain, L. A., Douglas, A. S., & Reynolds, W. S. (1967). A medico-psychiatric diagnostic review of remanded (without a request for a psychiatric report) male minor offenders. *Medical Journal of Australia*, *1*(6), 267-269.
- Beck, A., Berzofsky, M., Caspar, R., & Krebs, C. P. (2013). Sexual Victimization in Prisons and Jails Reported by Inmates, 2011-12: Bureau of Justice Statistics, US Dept of Justice, Office of Justice Programs.
- Beech, A., Bourgon, G., Hanson, R. K., Harris, A. J. R., Langton, C., Marques, J., . . . Yates, P. M. (2007). Sexual Offender Treatment Outcome Research: CODC Guidelines for Evaluation Part 1: Introduction and Overview *Public Safety Canada*. Ottawa, Ontario.
- Beech, A., Friendship, C., Erikson, M., & Hanson, R. K. (2002). The relationship between static and dynamic risk factors and reconviction in a sample of U.K. child abusers. *Sexual Abuse: A journal of research and treatment*, *14*(2), 155-167; discussion 195-157. doi: 10.1177/107906320201400206

- Beech, A., Mandeville-Norden, R., & Goodwill, A. (2012). Comparing recidivism rates of treatment responders/nonresponders in a sample of 413 child molesters who had completed community-based sex offender treatment in the United Kingdom. *International Journal of Offender Therapy and Comparative Criminology*, 56(1), 29-49. doi: 10.1177/0306624X10387811
- Belfrage, H., & Strand, S. (2012). Measuring the outcome of structured spousal violence risk assessments using the B-SAFER: risk in relation to recidivism and intervention. *Behavioral Sciences & the Law*, 30(4), 420-430. doi: 10.1002/bsl.2019
- Bengtson, S., & Langstrom, N. (2007). Unguided clinical and actuarial assessment of re-offending risk: a direct comparison with sex offenders in Denmark. *Sexual Abuse: A journal of research and treatment*, 19(2), 135-153. doi: 10.1177/107906320701900205
- Bermúdez, C. E., Romero Mendoza, M. P., Rodríguez Ruiz, E. M., Durand-Smith, A. L., & Saldívar Hernández, G. J. (2007). Female depression and substance dependence in the Mexico City penitentiary system. *Salud Mental*, 30(6), 53-61.
- BFS. (2013). Bundesamt für Statistik, Schweiz. <http://www.bfs.admin.ch/>
- Binswanger, I. A., Stern, M. F., Deyo, R. A., Heagerty, P. J., Cheadle, A., Elmore, J. G., & Koepsell, T. D. (2007). Release from prison - A high risk of death for former inmates. *The New England Journal of Medicine*, 356(2), 157-165. doi: 10.1056/NEJMsa064115
- Biondi-Zoccai, G., Romagnoli, E., Agostoni, P., Capodanno, D., Castagno, D., D'Ascenzo, F., . . . Modena, M. G. (2011). Are propensity scores really superior to standard multivariable analysis? *Contemporary Clinical Trials*, 32(5), 731-740. doi: 10.1016/j.cct.2011.05.006
- Bird, S. M. (2008). Changes in male suicides in Scottish prisons: 10-year study. *The British Journal of Psychiatry*, 192(6), 446-449. doi: 10.1192/bjp.bp.107.038679
- Birgden, A. (2009). Crime-prevention jurisprudence? A response to Andrews and Dowden. *Canadian Journal of Criminology and Criminal Justice*, 51(1), 93-117. doi: 10.3138/cjccj.51.1.93
- Birmingham, L., Mason, D., & Grubin, D. (1996). Prevalence of mental disorder in remand prisoners: Consecutive case study. *British Medical Journal*, 313(7071), 1521-1524. doi: 10.1136/bmj.313.7071.1521

- Black, M. C., Basile, K. C., Breiding, M. J., Smith, S. G., Walters, M. L., Merrick, M. T., & Stevens, M. R. (2011). National intimate partner and sexual violence survey. Atlanta, GA: Centers for Disease Control and Prevention.
- Blackburn, A. G., Mullings, J. L., & Marquart, J. W. (2008). Sexual assault in prison and beyond toward an understanding of lifetime sexual assault among incarcerated women. *The Prison Journal*, *88*(3), 351-377. doi: 10.1177/0032885508322443
- Bland, R. C., Newman, S. C., Dyck, R. J., & Orn, H. (1990). Prevalence of psychiatric disorders and suicide attempts in a prison population. *Canadian Journal of Psychiatry*, *35*(5), 407-413.
- Bluglass, R. (1966). *A psychiatric study of Scottish Convicted prisoners*. MD Thesis.
- Bonta, J., & Andrews, D. A. (2007). Risk-need-responsivity model for offender assessment and rehabilitation. *Rehabilitation*, *6*, 1-22.
- Bonta, J., Law, M., & Hanson, K. (1998). The prediction of criminal and violent recidivism among mentally disordered offenders: a meta-analysis. *Psychological Bulletin*, *123*(2), 123-142. doi: 10.1037/0033-2909.123.2.123
- Borduin, C. M., Schaeffer, C. M., & Heiblum, N. (2009). A randomized clinical trial of multisystemic therapy with juvenile sexual offenders: effects on youth social ecology and criminal activity. *Journal of Consulting and Clinical Psychology*, *77*(1), 26-37. doi: 10.1037/a0013035
- Borum, R., Otto, R. K., & Golding, S. (1993). Improving clinical judgment and decision making in forensic evaluation. *Journal of Psychiatry & Law*, *21*(1), 35-76.
- Bradley-Engen, M. S., Cuddeback, G. S., Gayman, M. D., Morrissey, J. P., & Mancuso, D. (2010). Trends in State Prison admission of offenders with serious mental illness. *Psychiatric Services*, *61*(12), 1263-1265. doi: 10.1176/appi.ps.61.12.1263
- Brinded, P. M. J., Stevens, I., Mulder, R. T., Fairley, N., Malcolm, F., & Wells, J. E. (1999). Christchurch Prisons Psychiatric Epidemiology Study: Methodology and prevalence rates for psychiatric disorder. *Criminal Behaviour and Mental Health*, *9*(2), 131-143.
- Brooke, D., Taylor, C., Gunn, J., & Maden, A. (1996). Point prevalence of mental disorder in unconvicted male prisoners in England and Wales. *British Medical Journal*, *313*(7071), 1524-1527.

- Bulten, B. (1998). *Gevangen tussen strafen zorg [Caught between punishment and care]*. (Doctoral dissertation), Vrije Universiteit, Kluwer Academic Publishers. Amsterdam.
- Bulten, E., Nijman, H., & van der Staak, C. (2009). Psychiatric disorders and personality characteristics of prisoners at regular prison wards. *International Journal of Law and Psychiatry*, 32(2), 115-119. doi: 10.1016/j.ijlp.2009.01.007
- Butler, T., & Allnutt, S. (2003). Mental illness among New South Wales prisoners *State Health Publication No: (CHS) 030147*: NSW Correction Health Service.
- Campbell, J. C. (1995). Prediction of homicide of and by battered women. In J. C. Campbell (Ed.), *Assessing dangerousness: Violence by sexual offenders, batterers, and child abusers*. (pp. 96-113). Thousand Oaks, CA US: Sage Publications, Inc.
- Cheniaux, E., Landeira-Fernandez, J., & Versiani, M. (2009). The diagnoses of schizophrenia, schizoaffective disorder, bipolar disorder and unipolar depression: Interrater reliability and congruence between DSM-IV and ICD-10. *Psychopathology*, 42(5), 293-298. doi: 10.1159/000228838
- Cooke, D. J. (1994). Psychological disturbance in the scottish prison system: Prevalence, precipitants, and policy.
- Cooke, D. J., & Michie, C. (2010). Limitations of diagnostic precision and predictive utility in the individual case: A challenge for forensic practice. *Law and Human Behavior*, 34(4), 259-274. doi: 10.2307/40785221
- Curtin, K., Monks, S., Wright, B., Duffy, D. M., Linehan, S. A., & Kennedy, H. G. (2009). Psychiatric morbidity in male remanded and sentenced committals to Irish prisons. *Irish Journal of Psychological Medicine*, 26(4), 169-173. doi: 10.1017/S079096670000063X
- D'Alton, P., Guilfoyle, M., & Randall, P. (2013). Roman Catholic Clergy who have sexually abused children: Their perceptions of their developmental experience. *Child Abuse & Neglect*, 37(9), 698-702. doi: 10.1016/j.chiabu.2012.12.001
- Daniel, A. E., Robins, A. J., Reid, J. C., & Wilfley, D. E. (1988). Lifetime and six-month prevalence of psychiatric disorders among sentenced female offenders. *The Bulletin of the American Academy of Psychiatry and the Law*, 16(4), 333-342.
- Davidson, M., Humphreys, M. S., Johnstone, E. C., & Owens, D. G. C. (1995). Prevalence of psychiatric morbidity among remand prisoners in Scotland. *British Medical Journal*, 167(4), 545-548.

- Denton, B. (1995). Psychiatric morbidity and substance dependence among women prisoners: an Australian study. *Psychiatry Psychology and Law*, 2(2), 173-177. doi: 10.1080/13218719509524865
- DiCataldo, F., Greer, A., & Profit, W. E. (1995). Screening prison inmates for mental disorder: an examination of the relationship between mental disorder and prison adjustment. *The Bulletin of the American Academy of Psychiatry and the Law*, 23(4), 573-585.
- Douglas, K. S., Cox, D. N., & Webster, C. D. (1999). Violence risk assessment: Science and practice. *Legal and Criminological Psychology*, 4(2), 149-184. doi: 10.1348/135532599167824
- Douglas, K. S., & Ogloff, J. R. P. (2003). Multiple facets of risk for violence: The impact of judgmental specificity on structured decisions about violence risk. *International Journal of Forensic Mental Health*, 2(1), 19-34. doi: 10.1080/14999013.2003.10471176
- Doyle, M., & Dolan, M. (2002). Violence risk assessment: combining actuarial and clinical information to structure clinical judgements for the formulation and management of risk. *Journal of Psychiatric and Mental Health Nursing*, 9(6), 649-657. doi: 10.1046/j.1365-2850.2002.00535.x
- Drake, E. K., Aos, S., & Miller, M. G. (2009). Evidence-based public policy options to reduce crime and criminal justice costs: Implications in Washington State. *Victims and Offenders*, 4(2), 170-196. doi: 10.1080/15564880802612615
- Dressing, H., Kief, C., & Salize, H.-J. (2009). Prisoners with mental disorders in Europe. *The British Journal of Psychiatry*, 194(1), 88. doi: 10.1192/bjp.194.1.88
- Ducro, C., & Pham, T. (2006). Evaluation of the SORAG and the Static-99 on Belgian sex offenders committed to a forensic facility. *Sexual Abuse: A journal of research and treatment*, 18(1), 15-26. doi: 10.1007/s11194-006-9003-6
- Dudeck, M., Kopp, D., & Kuwert, P. (2009). Prevalence of psychiatric disorders in prisoners with a short imprisonment: results from a prison in north Germany. *Psychiatrische Praxis*, 36(5), 219. doi: 10.1055/s-0028-1090217
- Duffy, D., Linehan, S., & Kennedy, H. G. (2006). Psychiatric morbidity in the male sentenced Irish prisons population. *Irish Journal of Psychological Medicine*, 23(2), 54-62. doi: 10.1017/S0790966700009587

- Duwe, G., & Goldman, R. A. (2009). The impact of prison-based treatment on sex offender recidivism: evidence from Minnesota. *Sexual Abuse: A journal of research and treatment*, 21(3), 279-307. doi: 10.1177/1079063209338490
- Eddy, J. M., Whaley, R. B., & Chamberlain, P. (2004). The prevention of violent behavior by chronic and serious male juvenile offenders: A 2-year follow-up of a randomized clinical trial. *Journal of Emotional and Behavioral Disorders*, 12(1), 2-8.
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315, 629-634. doi: 10.1136/bmj.315.7109.629
- Eher, R., Matthes, A., & Rettenberger, M. (2012). Der STABLE-2007 - ein Instrument zur Erfassung des stabil-dynamischen Rückfallrisikos bei Sexualstraftätern. In J. Endrass, A. Rossegger, F. Urbaniok, & B. Borchard (Eds.), *Interventionen bei Gewalt- und Sexualstraftätern: Risk-Management, Methoden und Konzepte der forensischen Therapie* (pp. 123-132). Berlin: Medizinisch wissenschaftliche Verlagsgesellschaft.
- Eher, R., & Rettenberger, M. (2013). Der Static-99 zur Erfassung des statistischen Risikos bei Sexualstraftätern. In M. Rettenberger & F. von Franqué (Eds.), *Handbuch kriminalprognostischer Verfahren*. Göttingen: Hogrefe Verlag GmbH.
- Eher, R., Rettenberger, M., Schilling, F., & Pfaefflin, F. (2008). The prediction of recidivism with Static-99 and SORAG. A prospective study with 275 sex offenders. *Recht & Psychiatrie*, 26(2), 79-88.
- Elbogen, E. B., & Johnson, S. C. (2009). The intricate link between violence and mental disorder: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Archives of General Psychiatry*, 66(2), 152-161. doi: 10.1001/archgenpsychiatry.2008.537
- Emsley, R., Lunt, M., Pickles, A., & Dunn, G. (2008). Implementing double-robust estimators of causal effects. *Stata Journal*, 8(3), 334-353.
- Endrass, J., & Rossegger, A. (2012). *Forensisches Operationalisiertes Therapie-Risiko-Evaluations-System 2.0 (FOTRES 2.0)*. Zürich.

- Endrass, J., Rossegger, A., Frischknecht, A., Noll, T., & Urbaniok, F. (2008). Using the Violence Risk Appraisal Guide (VRAG) to predict in-prison aggressive behavior in a Swiss offender population. *International Journal of Offender Therapy and Comparative Criminology*, 52(1), 81-89. doi: 10.1177/0306624X07301643
- Endrass, J., Rossegger, A., & Kuhn, B. (2012). Kosten-Nutzen-Effizienz von Therapie. In J. Endrass, A. Rossegger, F. Urbaniok, & B. Borchard (Eds.), *Interventionen bei Gewalt- und Sexualstraftätern: Risk-Management, Methoden und Konzepte der forensischen Therapie* (pp. 77-88). Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft.
- Eyestone, L. L. (1994). An epidemiological study of attention-deficit hyperactivity disorder and depression in a male prison population. *The Bulletin of the American Academy of Psychiatry and the Law*, 22(2), 181-193.
- Falissard, B., Loze, J.-Y., Gasquet, I., Duburc, A., & de Beaurepaire, C. (2006). Prevalence of mental disorders in French prisons for men. *BMC Psychiatry*, 6, 33. doi: 10.1186/1471-244X-6-33
- Farrington, D. P., & Welsh, B. C. (2005). Randomized experiments in criminology: What have we learned in the last two decades? *Journal of Experimental Criminology*, 1(1), 9-38. doi: 10.1007/s11292-004-6460-0
- Faulk, M. (1976). A psychiatric study of men serving a sentence in Winchester prison. *Medicine, Science and the Law*, 16, 244.
- Fazel, S., & Baillargeon, J. (2011). The health of prisoners. *The Lancet*, 377(9656), 956-965. doi: 10.1016/S0140-6736(10)61053-7
- Fazel, S., Bains, P., & Doll, H. (2006). Substance abuse and dependence in prisoners: a systematic review. *Addiction*, 101(2), 181-191. doi: 10.1111/j.1360-0443.2006.01316.x
- Fazel, S., Cartwright, J., Norman-Nott, A., & Hawton, K. (2008). Suicide in prisoners: a systematic review of risk factors. *Journal of Clinical Psychiatry*, 69(11), 1721-1731.
- Fazel, S., & Danesh, J. (2002). Serious mental disorder in 23000 prisoners: a systematic review of 62 surveys. *The Lancet*, 359(9306), 545-550. doi: 10.1016/S0140-6736(02)07740-1

- Fazel, S., & Grann, M. (2004). Psychiatric morbidity among homicide offenders: A Swedish population study. *American Journal of Psychiatry*, *161*, 2129-2131. doi: 10.1176/appi.ajp.161.11.2129
- Fazel, S., Grann, M., Kling, B., & Hawton, K. (2010). Prison suicide in 12 countries: an ecological study of 861 suicides during 2003-2007. *Social Psychiatry and Psychiatric Epidemiology*, *46*(3), 191-195. doi: 10.1007/s00127-010-0184-4
- Fazel, S., Gulati, G., Linsell, L., Geddes, J. R., & Grann, M. (2009). Schizophrenia and violence: Systematic review and meta-analysis. *PLoS Medicine*, *6*(8), e1000120. doi: 10.1371/journal.pmed.1000120
- Fazel, S., Hope, T., O'Donnell, I., Piper, M., & Jacoby, R. (2001). Health of elderly male prisoners: worse than the general population, worse than younger prisoners. *Age Ageing*, *30*(5), 403-407. doi: 10.1093/ageing/30.5.403
- Fazel, S., Långström, N., Hjern, A., Grann, M., & Lichtenstein, P. (2009). Schizophrenia, substance abuse, and violent crime. *JAMA*, *301*(19), 2016-2023. doi: 10.1001/jama.2009.675
- Fazel, S., Lichtenstein, P., Grann, M., Goodwin, G. M., & Langstrom, N. (2010). Bipolar disorder and violent crime: new evidence from population-based longitudinal studies and systematic review. *Archives of General Psychiatry*, *67*(9), 931-938. doi: 10.1001/archgenpsychiatry.2010.97
- Fazel, S., & Seewald, K. (2012). Severe mental illness in 33,588 prisoners worldwide: systematic review and meta-regression analysis. *The British Journal of Psychiatry*, *200*(5), 364-373. doi: 10.1192/bjp.bp.111.096370
- Fazel, S., Singh, J. P., Doll, H., & Grann, M. (2012). Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24 827 people: Systematic review and meta-analysis. *British Medical Journal*, *24*(345), e4692. doi: 10.1136/bmj.e4692
- Fazel, S., Sjostedt, G., Langstrom, N., & Grann, M. (2007). Severe mental illness and risk of sexual offending in men: A case-control study based on Swedish national registers. *Journal of Clinical Psychiatry*, *68*(4), 588-596.
- Fazel, S., Wolf, A., Chang, Z., Larsson, H., Goodwin, G. M., & Lichtenstein, P. (2015). Depression and violence: a Swedish population study. *The Lancet*, *2*(3), 224-232. doi: 10.1016/S2215-0366(14)00128-X

- Fazel, S., & Yu, R. (2009). Psychotic disorders and repeat offending: systematic review and meta-analysis. *Schizophrenia Bulletin*, 37(4), 800-810. doi: 10.1093/schbul/sbp135
- Felson, R. B., Cundiff, P., & Painter-Davis, N. (2012). Age and sexual assault in correctional facilities: A blocked opportunity approach. *Criminology*, 50(4), 887-911. doi: 10.1111/j.1745-9125.2012.00282.x
- Fido, A. A., & al-Jabally, M. (1993). Presence of psychiatric morbidity in prison population in Kuwait. *Annals of Clinical Psychiatry*, 5(2), 107.
- Fisher, P. A., & Chamberlain, P. (2000). Multidimensional Treatment Foster Care: A program for intensive parenting, family support, and skill building. *Journal of Emotional and Behavioral Disorders*, 8(3), 155-164. doi: 10.1177/106342660000800303
- Fotiadou, M., Livaditis, M., Manou, I., Kaniotou, E., & Xenitidis, K. (2006). Prevalence of mental disorders and deliberate self-harm in Greek male prisoners. *International Journal of Law and Psychiatry*, 29(1), 68-73. doi: 10.1016/j.ijlp.2004.06.009
- Fraser, A., Møller, L., & van den Bergh, B. (2011). The health of prisoners. *The Lancet*, 377(9782), 2002. doi: 10.1016/S0140-6736(11)60859-3
- Gannon, T. A., King, T., Miles, H., Lockerbie, L., & Willis, G. M. (2011). Good Lives sexual offender treatment for mentally disordered offenders. *The British Journal of Forensic Practice*, 13(3), 153-168. doi: doi:10.1108/14636641111157805
- Ghubash, R., & El-Rufaie, O. (1997). Psychiatric morbidity among sentenced male prisoners in Dubai: transcultural perspectives. *Journal of Forensic Psychiatry & Psychology*, 8(2), 440-446. doi: 10.1080/09585189708412023
- Gibson, L. E., Holt, J. C., Fondacaro, K. M., Tang, T. S., Powell, T. A., & Turbitt, E. L. (1999). An examination of antecedent traumas and psychiatric comorbidity among male inmates with PTSD. *Journal of Traumatic Stress*, 12(3), 473-484.
- Gmür, C., Gerth, J., & Graber, C. (2012). Behandlungsprogramme für jugendliche Straftäter. In J. Endrass, A. Rossegger, F. Urbaniok, & B. Borchard (Eds.), *Interventionen bei Gewalt- und Sexualstraftätern: Risk-Management, Methoden und Konzepte der forensischen Therapie* (pp. 70-76): Medizinisch Wissenschaftliche Verlagsgesellschaft.

- Goff, A., Rose, S., & Purves, D. (2007). Does PTSD occur in sentenced prison populations? A systematic literature review. *Criminal Behaviour and Mental Health*, 17(3), 152-162. doi: 10.1002/cbm.653
- Gondolf, E. W. (2001). Limitations of experimental evaluation of batterer programs. *Trauma, Violence, & Abuse*, 2(1), 79-88. doi: 10.1177/1524838001002001005
- Gonsalves, V. M., Walsh, K., & Scalora, M. J. (2012). Staff perceptions of risk for prison rape perpetration and victimization. *The Prison Journal*, 92(2), 253-273. doi: 10.1177/0032885512439014
- Grady, M. D., Edwards, D., Pettus-Davis, C., & Abramson, J. (2012). Does volunteering for sex offender treatment matter? Using propensity score analysis to understand the effects of volunteerism and treatment on recidivism. *Sexual Abuse: A journal of research and treatment*, 25(4), 391-346. doi: 10.1177/1079063212459085
- Grann, M., Danesh, J., & Fazel, S. (2008). The association between psychiatric diagnosis and violent re-offending in adult offenders in the community. *BMC Psychiatry*, 8(92), 8-92. doi: 10.1186/1471-244X-8-92
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, 12(1), 19-30. doi: 10.1037/1040-3590.12.1.19
- Gunn, J., Maden, A., & Swinton, M. (1991). Treatment needs of prisoners with psychiatric disorders. *British Medical Journal*, 303(6798), 338-341. doi: 10.1136/bmj.303.6798.338
- Gunn, J., Robertson, G., Dell, S., & Way, C. (1978). Psychiatric aspects of imprisonment. In S. D. Academic Press (Ed.).
- Gunter, T. D., Arndt, S., Wenman, G., Allen, J., Loveless, P., Sieleni, B., & Black, D. W. (2008). Frequency of mental and addictive disorders among 320 men and women entering the Iowa prison system: Use of the MINI-Plus. *Journal of the American Academy of Psychiatry and the Law*, 36(1), 27-34.
- Guo, S., & Fraser, M. W. (2010). *Propensity score analysis: Statistical methods and applications* (Vol. 11). Thousand Oaks, US: Sage Publications, Inc.
- Guy, E., Platt, J. J., Zwerling, I., & Bullock, S. (1985). Mental health status of prisoners in an urban jail. *Criminal Justice and Behavior*, 12(1), 29-53. doi: 10.1177/0093854885012001004

- Gwet, K. (2001). *Handbook of inter-rater reliability*. Gaithersburg, MD: STATAXIS Publishing Company.
- Haapasalo, J. (2000). Vankien lapsuuden kaltoinkohtelu kaytosonogelmat ja aikuisian psyykkiset hairiot truama [Childhood maltreatment, behavioral problems and mental disorders among adult prisoners]. *Psykologia*, 35, 45-57.
- Habermeyer, E., Wolff, R., Gillner, M., Strohm, R., & Kutscher, S. (2010). Patienten mit schizophrenen Störungen im psychiatrischen Maßregelvollzug. *Der Nervenarzt*, 81(9), 1117-1124. doi: 10.1007/s00115-009-2913-z
- Hanson, R. K., Bourgon, G., Helmus, L., & Hodgson, S. (2009). The principles of effective correctional treatment also apply to sexual offenders a meta-analysis. *Criminal Justice and Behavior*, 36(9), 865-891. doi: 10.1177/0093854809338545
- Hanson, R. K., Broom, I., & Stephenson, M. (2004). Evaluating community sex offender treatment programs: A 12-Year follow-up of 724 offenders. *Canadian Journal of Behavioural Science*, 36(2), 87. doi: 10.1037/h0087220
- Hanson, R. K., Harris, A. J. R., Scott, T. L., & Helmus, L. (2007). Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project (Vol. 5). Ottawa: Public Safety Canada.
- Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: a meta-analysis of 118 prediction studies. *Psychological Assessment*, 21(1), 1-21. doi: 10.1037/a0014421
- Hanson, R. K., & Thornton, D. (1999). Static-99: Improving actuarial risk assessments for sex offenders. from Department of the Solicitor General of Canada, Ottawa
- Hare, R. D. (2003). *The Hare Psychopathy Checklist-Revised*. Toronto: Multi-Health Systems.
- Harkins, L., Flak, V. E., Beech, A. R., & Woodhams, J. (2012). Evaluation of a community-based sex offender treatment program using a Good Lives Model approach. *Sexual Abuse: A journal of research and treatment*, 24(6), 519-543. doi: 10.1177/1079063211429469
- Harper, D., & Barry, D. (1979). Estimated prevalence of psychiatric disorder in a prison population. *Abstracts on Criminology and Penology*, 19.
- Harris, A., Phenix, A., Hanson, K., & Thornton, D. (2003). Static-99 Coding rules revised.

- Harris, G. T., & Rice, M. E. (2015). Progress in violence risk assessment and communication: Hypothesis versus evidence. *Behavioral Sciences & the Law*, 33(1), 128-145. doi: 10.1002/bsl.2157
- Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumière, M. L., Boer, D. P., & Lang, C. (2003). A multisite comparison of actuarial risk instruments for sex offenders. *Psychological Assessment*, 15(3), 413-425. doi: 10.1037/1040-3590.15.3.413
- Hart, S. D., & Logan, C. (2011). Formulation of violence risk using evidence-based assessments: The structured professional judgment approach. In P. Sturmey, M. McMurrin, S. D. Hart, & C. Logan (Eds.), *Forensic Case Formulation* (pp. 83-106). Chichester, UK.: John Wiley & Sons, Ltd.
- Hart, S. D., Michie, C., & Cooke, D. J. (2007). Precision of actuarial risk assessment instruments: Evaluating the 'margins of error' of group v. individual predictions of violence. *The British Journal of Psychiatry*, 190(49), 60-65. doi: 10.1192/bjp.190.5.s60
- Hart, S. D., Sturmey, P., Logan, C., & McMurrin, M. (2011). Forensic case formulation. *International Journal of Forensic Mental Health*, 10(2), 118-126. doi: 10.1080/14999013.2011.577137
- Hassan, L., Birmingham, L., Harty, M. A., Jarrett, M., Jones, P., King, C., . . . Shaw, J. (2011). Prospective cohort study of mental health during imprisonment. *The British Journal of Psychiatry*, 198(1), 37-42. doi: 10.1192/bjp.bp.110.080333
- Helmus, L., Hanson, R. K., Thornton, D., Babchishin, K. M., & Harris, A. J. R. (2012). Absolute recidivism rates predicted by Static-99R and Static-2002R sex offender risk assessment tools vary across samples: A meta-analysis. *Criminal Justice and Behavior*, 39(9), 1148-1171. doi: 10.1177/0093854812443648
- Henggeler, S. W. (2011). Efficacy studies to large-scale transport: The development and validation of multisystemic therapy programs. *Annual review of Clinical Psychology*, 7, 351-381. doi: 10.1146/annurev-clinpsy-032210-104615
- Hensley, C., Koscheski, M., & Tewksbury, R. (2005). Examining the characteristics of male sexual assault targets in a southern maximum-security prison. *Journal of Interpersonal Violence*, 20(6), 667-679. doi: 10.1177/0886260505276069
- Hensley, C., & Tewksbury, R. (2005). Wardens' perceptions of prison sex. *The Prison Journal*, 85(2), 186-197.

- Herrman, H., McGorry, P., Mills, J., & Singh, B. (1991). Hidden severe psychiatric morbidity in sentenced prisoners - an Australian study. *American Journal of Psychiatry*, *148*(2), 236-239.
- Higgins, J. P. T. (2008). Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. *International Journal of Epidemiology*, *37*(5), 1158-1160. doi: 10.1093/ije/dyn204
- Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539-1558. doi: 10.1002/sim.1186
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*(7414), 557-560. doi: 10.1136/bmj.327.7414.557
- Hilton, N. Z., Carter, A. M., Harris, G. T., & Sharpe, A. J. B. (2008). Does using nonnumerical terms to describe risk aid violence risk communication? Clinician agreement and decision making. *Journal of Interpersonal Violence*, *23*(2), 171-188. doi: 10.1177/0886260507309337
- Hilton, N. Z., & Harris, G. T. (2009). How nonrecidivism affects predictive accuracy: evidence from a cross-validation of the Ontario Domestic Assault Risk Assessment (ODARA). *Journal of Interpersonal Violence*, *24*(2), 326-337. doi: 10.1177/0886260508316478
- Hilton, N. Z., Harris, G. T., & Rice, M. E. (2001). Predicting violence by serious wife assaulters. *Journal of Interpersonal Violence*, *16*(5), 408-423. doi: 10.1177/088626001016005002
- Hilton, N. Z., Harris, G. T., & Rice, M. E. (2010). *Risk Assessment for Domestically Violent Men*. Washington, DC: American Psychological Association.
- Hilton, N. Z., Harris, G. T., Rice, M. E., Houghton, R. E., & Eke, A. W. (2008). An in-depth actuarial assessment for wife assault recidivism: The Domestic violence risk appraisal guide. *Law and Human Behavior*, *32*(2), 150-163. doi: 10.1007/s10979-007-9088-6
- Hilton, N. Z., Harris, G. T., Rice, M. E., Lang, C., Cormier, C. A., & Lines, K. J. (2004). A brief actuarial assessment for the prediction of wife assault recidivism: the Ontario domestic assault risk assessment. *Psychological Assessment*, *16*(3), 267-275. doi: 10.1037/1040-3590.16.3.267

- Höffler, K., & Stadtland, C. (2012). Mad or bad? Der Begriff »psychische Störung« des ThUG im Lichte der Rechtsprechung des BVerfG und des EGMR1. *StV*, 4, 239.
- Hollin, C. R. (2008). Evaluating offending behaviour programmes Does only randomization glister? *Criminology and Criminal Justice*, 8(1), 89-106. doi: 10.1177/1748895807085871
- Home Office. (2004). Probation statistics England and Wales. London: Home Office Research, Development and Statistics Directorate.
- Hosmer, D. W., Lemeshow, S., & May, S. (2008). *Applied Survival Analysis: Regression Modeling of Time to Event Data* (2nd Edition ed.): Wiley-Interscience.
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression* (3 ed.). New York: Wiley.
- Huang, Y., Kotov, R., de Girolamo, G., Preti, A., Angermeyer, M., Benjet, C., . . . Kessler, R. C. (2009). DSM-IV personality disorders in the WHO World Mental Health Surveys. *The British Journal of Psychiatry*, 195(1), 46-53. doi: 10.1192/bjp.bp.108.058552
- Hurley, W., & Dunne, M. P. (1991). Psychological distress and psychiatric morbidity in women prisoners. *Australian & New Zealand Journal of Psychiatry*, 25(4), 461-470.
- Hurwitz, S., & Christiansen, K. (1983). *Criminology* (2nd ed.). London: George Allen & Unwin.
- Hyde, P., & Seiter, R. (1987). The prevalence of mental illness among inmates in the Ohio prison system. Ohio Department of Mental Health and the Ohio Department of Rehabilitation and Correction.
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), e124. doi: 10.1371/journal.pmed.0020124
- Ioannidis, J. P. A., Haidich, A., Pappa, M., Pantazis, N., Kokori, S. I., Tektonidou, M. G., . . . Lau, J. (2001). Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA*, 286(7), 821-830. doi: 10.1001/jama.286.7.821.
- Ioannidis, J. P. A., Patsopoulos, N. A., & Evangelou, E. (2007). Uncertainty in heterogeneity estimates in meta-analyses. *British Medical Journal*, 335(7626), 914-916. doi: 10.1136/bmj.39343.408449.80
- Ioannidis, J. P. A., Patsopoulos, N. A., & Rothstein, H. R. (2008). Reasons or excuses for avoiding meta-analysis in forest plots. *British Medical Journal*, 336(7658), 1413-1415. doi: 10.1136/bmj.a117

- James, J. F., Gregory, D., Jones, R. K., & Rundell, O. H. (1980). Psychiatric morbidity in prisons. *Hospital and Community Psychiatry, 31*, 674.
- Jones, T. R., & Pratt, T. C. (2008). The prevalence of sexual violence in prison: The state of the knowledge base and implications for evidence-based correctional policy making. *International Journal of Offender Therapy and Comparative Criminology, 52*(3), 280-295. doi: 10.1177/0306624X07307631
- Jordan, B. K., Schlenger, W. E., Fairbank, J. A., & Caddell, J. M. (1996). Prevalence of psychiatric disorders among incarcerated women: II. Convicted felons entering prison. *Archives of General Psychiatry, 53*(6), 513-519. doi: 10.1001/archpsyc.1996.01830060057008
- Joukamaa, M. (1993). Mental health of Finnish prisoners - Results of a survey. *Journal of Forensic Psychiatry & Psychology, 4*(2), 261-271. doi: 10.1080/09585189308407975
- Joukamaa, M. (1995). Psychiatric morbidity among finnish prisoners with special reference to sociodemographic factors - results of the health survey of finnish prisoners (WATTU project). *Forensic Science International, 73*(2), 85-91. doi: 10.1016/0379-0738(95)01713-S
- Jung, S., & Gulayets, M. (2011). Using clinical variables to evaluate treatment effectiveness in programmes for sexual offenders. *Journal of Sexual Aggression, 17*(2), 166-180. doi: 10.1080/13552601003802238
- Kanyanya, I. M., Othieno, C. J., & Ndeti, D. M. (2007). Psychiatric morbidity among convicted male sex offenders at Kamiti Prison, Kenya. *East African Medical Journal, 84*(4), 151-155.
- Kariminia, A., Law, M., Butler, T., Corben, S., Levy, M., Kaldor, J., & Grant, L. (2007). Factors associated with mortality in a cohort of Australian prisoners. *European Journal of Epidemiology, 22*(7), 417-428. doi: 10.1007/s10654-007-9134-1
- Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., . . . Wang, P. S. (2003). The epidemiology of major depressive disorder. *JAMA, 289*(23), 3095-3105. doi: 10.1001/jama.289.23.3095
- Kessler, R. C., & Bromet, E. J. (2013). The epidemiology of depression across cultures. *Annual review of public health, 34*, 119.

- Kilvinger, F., Rossegger, A., Urbaniok, F., & Endrass, J. (2012). Risikokalkulation bei Häuslicher Gewalt. *Fortschritte der Neurologie und Psychiatrie*, *80*(6), 312-319. doi: 10.1055/s-0031-1273200
- Kingsnorth, R. (2006). Intimate partner violence: predictors of recidivism in a sample of arrestees. *Violence Against Women*, *12*(10), 917-935. doi: 10.1177/1077801206293081
- Kingston, D. A., Olver, M. E., Harris, M., Wong, S. C. P., & Bradford, J. M. (2015). The relationship between mental disorder and recidivism in sexual offenders. *International Journal of Forensic Mental Health*, *14*(1), 10-22. doi: 10.1080/14999013.2014.974088
- Klein, A. R., & Tobin, T. (2008). A longitudinal study of arrested batterers, 1995-2005: Career criminals. *Violence Against Women*, *14*(2), 136-157. doi: 10.1177/1077801207312396
- Klepfisz, G., O'Brien, K., & Daffern, M. (2014). Violent offenders' within-treatment change in anger, criminal attitudes, and violence risk: associations with violent recidivism. *International Journal of Forensic Mental Health*, *13*(4), 348-362. doi: 10.1080/14999013.2014.951107
- Koehler, J. A., Lösel, F., Akoensi, T. D., & Humphreys, D. K. (2013). A systematic review and meta-analysis on the effects of young offender treatment programs in Europe. *Journal of Experimental Criminology*, *9*(1), 19-43. doi: 10.1007/s11292-012-9159-7
- Kreffft, K. M., & Brittain, T. H. (1983). A prisoner assessment survey: Screenings of a municipal prison population. *International Journal of Law and Psychiatry*, *6*(1), 113-124.
- Kropp, P. R. (2008). Intimate partner violence risk assessment and management. *Violence and Victims*, *23*(2), 202-220. doi: 10.1891/0886-6708.23.2.202
- Kropp, P. R., & Hart, S. D. (2000). The Spousal Assault Risk Assessment (SARA) Guide: Reliability and validity in adult male offenders. *Law and Human Behavior*, *24*(1), 101-118. doi: 10.1023/a:1005430904495
- Kugu, N., Akyuz, G., & Dogan, O. (2008). Psychiatric morbidity in murder and attempted murder crime convicts: A Turkey study. *Forensic Science International*, *175*(2-3), 107-112. doi: 10.1016/j.forsciint.2007.05.016

- Lamb, H. R., Weinberger, E. L., Marsh, S. J., & Gross, H. B. (2007). Treatment prospects for persons with severe mental illness in an urban county jail. *Psychiatric Services*, *58*(6), 782. doi: 10.1176/ps.2007.58.6.782
- Landenberger, N. A., & Lipsey, M. W. (2005). The positive effects of cognitive-behavioral programs for offenders: A meta-analysis of factors associated with effective treatment. *Journal of Experimental Criminology*, *1*(4), 451-476. doi: 10.1007/s11292-005-3541-7
- Langan, P., & Lewin, D. Recidivism of prisoners released in 1994. Bureau of Justice Statistics Special Report Retrieved March 2011, from <http://bjs.ojp.usdoj.gov/index.cfm?ty=pbdetail&iid=1134>
- Langevin, R., Curnoe, S., & Bain, J. (2000). A study of clerics who commit sexual offenses: Are they different from other sex offenders? *Child Abuse & Neglect*, *24*(4), 535-545. doi: 10.1016/S0145-2134(00)00113-7
- Lasher, M. P., McGrath, R. J., Wilson, D., & Cumming, G. F. (2015). Collaborative treatment planning using the Sex Offender Treatment Intervention and Progress Scale (SOTIPS): Concordance of therapist evaluation and client self-evaluation. *International Journal of Forensic Mental Health*, *14*(1), 1-9. doi: 10.1080/14999013.2014.974087
- Laub, J. H., & Sampson, R. J. (2001). Understanding desistance from crime. *Crime and Justice*, *28*, 1-69.
- Laubacher, A., Rossegger, A., Endrass, J., Angst, J., Urbaniok, F., & Vetter, S. (2014). Adolescent delinquency and antisocial tendencies as precursors to adult violent offending: a prospective study of a representative sample of swiss men. *International Journal of Offender Therapy and Comparative Criminology*, *58*(5), 537-549. doi: 10.1177/0306624X13479935
- Levander, S., Svalenius, H., & Jensen, J. (1997). Psykiska skador vanliga bland interner [Common psychiatric disorders among inmates]. *Lakartidningen*, *94*, 46-50.
- Levy, M. (2011). Health services for prisoners. *British Medical Journal*, *342*(342), d351. doi: 10.1136/bmj.d351
- Lilienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science*, *2*(1), 53-70. doi: 10.1111/j.1745-6916.2007.00029.x

- Linehan, S. A., Duffy, D. M., Wright, B., Curtin, K., Monks, S., & Kennedy, H. G. (2005). Psychiatric morbidity in a cross-sectional sample of male remanded prisoners. *Irish Journal of Psychological Medicine*, 22(4), 128-132. doi: 10.1017/S079096670000923X
- Lipsey, M. W., & Cullen, F. T. (2007). The effectiveness of correctional rehabilitation: A review of systematic reviews. *Annual Review of Law and Social Science*, 3(1), 297-320. doi: 10.1146/annurev.lawsocsci.3.081806.112833
- Lipsey, M. W., Howell, J. C., & Kelly, M. R. (2010). Improving the effectiveness of juvenile justice programs. Washington DC: Center for Juvenile Justice Reform at Georgetown University.
- Logan, C., & Johnstone, L. (2010). Personality disorder and violence: making the link through risk formulation. *Journal of Personality Disorders*, 24(5), 610-633. doi: 10.1521/pedi.2010.24.5.610
- Lohner, J., & Konrad, N. (2007). Risk factors for self-injurious behaviour in custody: problems of definition and prediction. *International Journal of Prisoner Health*, 3(4), 135-161. doi: 10.1080/17449200701321654
- Lösel, F., & Schmucker, M. (2005). The effectiveness of treatment for sexual offenders: A comprehensive meta-analysis. *Journal of Experimental Criminology*, 1(1), 117-146. doi: 10.1007/s11292-004-6466-7
- Lussier, P., & Gress, C. L. Z. (2014). Community re-entry and the path toward desistance: A quasi-experimental longitudinal study of dynamic factors and community risk management of adult sex offenders. *Journal of Criminal Justice*, 42(2), 111-122. doi: 10.1016/j.jcrimjus.2013.09.006
- Maden, A., Taylor, C., Brooke, D., & Gunn, J. (1996). Mental disorder in remand prisoners: London: Home Office Research and Planning Unit.
- Maden, T., Swinton, M., & Gunn, J. (1994). Psychiatric disorder in women serving a prison sentence. *The British Journal of Psychiatry*, 164(1), 44-54. doi: 10.1192/bjp.164.1.44
- Mandeville-Norden, R., Beech, A., & Hayes, E. (2008). Examining the effectiveness of a UK community-based sexual offender treatment programme for child abusers. *Psychology, Crime & Law*, 14(6), 493-512. doi: 10.1080/10683160801948907

- Marques, J. K., Wiederanders, M., Day, D. M., Nelson, C., & Van Ommeren, A. (2005). Effects of a relapse prevention program on sexual recidivism: Final results from California's Sex Offender Treatment and Evaluation Project (SOTEP). *Sexual Abuse: A journal of research and treatment*, *17*(1), 79-107. doi: 10.1177/107906320501700108
- Marshall, W. L., & Marshall, L. E. (2007). The utility of the random controlled trial for evaluating sexual offender treatment: The gold standard or an inappropriate strategy? *Sexual Abuse: A journal of research and treatment*, *19*(2), 175-191. doi: 10.1177/107906320701900207
- Marshall, W. L., & Marshall, L. E. (2008). Good clinical practice and the evaluation of treatment: A response to Seto et al. *Sexual Abuse: A journal of research and treatment*, *20*(3), 256-260. doi: 10.1177/1079063208323839
- McGrath, J., Saha, S., Welham, J., El Saadi, O., MacCauley, C., & Chant, D. (2004). A systematic review of the incidence of schizophrenia: the distribution of rates and the influence of sex, urbanicity, migrant status and methodology. *BMC Medicine*, *2*, 13. doi: 10.1186/1741-7015-2-13
- McMurrin, M., & Ward, T. (2010). Treatment readiness, treatment engagement and behaviour change. *Criminal Behaviour and Mental Health*, *20*(2), 75-85. doi: 10.1002/cbm.762
- Meade, B., & Steiner, B. (2010). The total effects of boot camps that house juveniles: A systematic review of the evidence. *Journal of Criminal Justice*, *38*(5), 841-853. doi: 10.1016/j.jcrimjus.2010.06.007
- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis, MN, US: University of Minnesota Press.
- Merikangas, K. R., Jin, R., He, J., & et al. (2011). Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. *Archives of General Psychiatry*, *68*(3), 241-251. doi: 10.1001/archgenpsychiatry.2011.12
- Messing, J. T., & Thaller, J. (2013). The average predictive validity of intimate partner violence risk assessment instruments. *Journal of Interpersonal Violence*, *28*(7), 1537-1558. doi: 10.1177/0886260512468250

- Mihalic, S. F., & Irwin, K. (2003). Blueprints for violence prevention: From research to real world settings – Factors influencing the successful replication of model programs. *Youth Violence and Juvenile Justice*, 1(4), 307-329. doi: 10.1177/1541204003255841
- Mohan, D., Scully, P., Collins, C., & Smith, C. (1997). Psychiatric disorder in an Irish female prison. *Criminal Behaviour and Mental Health*, 7(3), 229-235. doi: 10.1002/cbm.175/pdf
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. *PLoS Medicine*, 6(7), e1000097. doi: 10.1371/journal.pmed.1000097
- Moher, D., & Tsertsvadze, A. (2006). Systematic reviews: when is an update an update? *The Lancet*, 367(9514), 881-883. doi: 10.1016/S0140-6736(06)68358-X
- Moore, E., & Drennan, G. (2013). Complex forensic case formulation in recovery-oriented services: Some implications for routine practice. *Criminal Behaviour and Mental Health*, 23(4), 230-240. doi: 10.1002/cbm.1885
- Morgan, R. D., Fisher, W. H., Duan, N. H., Mandracchia, J. T., & Murray, D. (2010). Prevalence of criminal thinking among state prison inmates with serious mental illness. *Law and Human Behavior*, 34(4), 324-336. doi: 10.1007/s10979-009-9182-z
- Morrissey, J. P., Fagan, J. A., & Cocozza, J. J. (2009). New models of collaboration between criminal justice and mental health systems. *American Journal of Psychiatry*, 166(11), 1211-1214. doi: 10.1176/appi.ajp.2009.09050670
- Mossman, D. (2013). Evaluating risk assessments using receiver operating characteristic analysis: Rationale, advantages, insights, and limitations. *Behavioral Sciences & the Law*, 31(1), 23-39. doi: 10.1002/bsl.2050
- Motiuk, L., & Porporino, F. (1992). The prevalence, nature and severity of mental health problems among federal male inmates in Canadian penitentiaries: Ottawa: Research and Statistics Branch, Correctional Service.
- Murray, J., & Thomson, M. E. (2010). Clinical judgement in violence risk assessment. *Europe's Journal of Psychology*, 6(1), 128-149. doi: 10.5964/ejop.v6i1.175
- Nedopil, N., & Müller, J. L. (2012). *Forensische Psychiatrie: Klinik, Begutachtung und Behandlung zwischen Psychiatrie und Recht*: Georg Thieme Verlag.

- Neighbors, H., Williams, D., Gunnings, T., Lipscomb, W., Broman, C., & Lepkowski, J. (1987). The prevalence of mental disorder in Michigan prison: Detroit: Michigan Department of Corrections.
- O'Keefe, M. L., & Schnell, M. J. (2007). Offenders with mental illness in the correctional system. *Journal of Offender Rehabilitation, 45*(1-2), 81-104. doi: 10.1300/J076v45n01_08
- Olver, M. E., Nicholaichuk, T. P., Gu, D., & Wong, S. C. P. (2013). Sex offender treatment outcome, actuarial risk, and the aging sex offender in Canadian corrections: a long-term follow-up. *Sexual Abuse: A journal of research and treatment, 25*(4), 396-422. doi: 10.1177/1079063212464399
- Otto, R. K. (2013). *Improving clinical decision making in forensic evaluation*. Paper presented at the Florida Psychological Association, West Palm Beach
- Parker, G. (2000). Classifying depression: Should paradigms lost be regained? *American Journal of Psychiatry, 157*(8), 1195-1203. doi: 10.1176/appi.ajp.157.8.1195
- Parsons, S., Walker, L., & Grubin, D. (2001). Prevalence of mental disorder in female remand prisons. *Journal of Forensic Psychiatry & Psychology, 12*(1), 194-202. doi: 10.1080/09585180122050
- Pascual-Leone, A., Bierman, R., Arnold, R., & Stasiak, E. (2011). Emotion-focused therapy for incarcerated offenders of intimate partner violence: A 3-year outcome using a new whole-sample matching method. *Psychotherapy Research, 21*(3), 331-347. doi: 10.1080/10503307.2011.572092
- Patsopoulos, N. A., Evangelou, E., & Ioannidis, J. P. (2008). Sensitivity of between-study heterogeneity in meta-analysis: proposed metrics and empirical evaluation. *International Journal of Epidemiology, 37*(5), 1148-1157. doi: 10.1093/ije/dyn065
- Perillo, A. D., Mercado, C. C., & Terry, K. J. (2008). Repeat offending, victim gender, and extent of victim relationship in Catholic Church sexual abusers: Implications for risk assessment. *Criminal Justice and Behavior, 35*(5), 600-614. doi: 10.1177/0093854808314368
- Persons, J. B., & Silberschatz, G. (1998). Are results of randomized controlled trials useful to psychotherapists? *Journal of Consulting and Clinical Psychology, 66*(1), 126.

- Petrosino, A., Turpin-Petrosino, C., Hollis-Peel, M. E., & Lavenberg, J. G. (2013). 'Scared Straight' and other juvenile awareness programs for preventing juvenile delinquency. *Cochrane Database of Systematic Reviews*, 9(5). doi: 10.1002/14651858.CD002796.pub2
- Piselli, M., Elisei, S., Murgia, N., Quartesan, R., & Abram, K. M. (2009). Co-occurring psychiatric and substance use disorders among male detainees in Italy. *International Journal of Law and Psychiatry*, 32(2), 101-107. doi: 10.1016/j.ijlp.2009.01.006
- PKS. (2013). Polizeiliche Kriminalstatistik Bundesrepublik Deutschland, Berichtsjahr 2012. In Bundeskriminalamt (Ed.). Wiesbaden.
- Polaschek, D. L., & Ross, E. C. (2010). Do early therapeutic alliance, motivation, and stages of change predict therapy change for high - risk, psychopathic violent prisoners? *Criminal Behaviour and Mental Health*, 20(2), 100-111. doi: 10.1002/cbm.759
- Polaschek, D. L., Wilson, N. J., Townsend, M. R., & Daly, L. R. (2005). Cognitive-behavioral rehabilitation for high-risk violent offenders an outcome evaluation of the violence prevention unit. *Journal of Interpersonal Violence*, 20(12), 1611-1627. doi: 10.1177/0886260505280507
- Pollähne, H. (2013). StGB §63 Unterbringung in einem psychiatrischen Krankenhaus *Kindhäuser/Neumann/Paeffgen, Strafgesetzbuch, Rn. 56-79.*
- Ponde, P. P., Cruz Freire, A. C., & Santos Mendonca, M. S. (2011). The prevalence of mental disorders in detainees in the city of Salvador, Bahia, Brazil. *Journal of Forensic Sciences*, 56(3), 679-682. doi: 10.1111/j.1556-4029.2010.01691
- Powell, T. A., Holt, J. C., & Fondacaro, K. M. (1997). Prevalence of mental illness among inmates in a rural state. *Law and Human Behavior*, 21(4), 427-438.
- Poythress, N., Hoge, S., Bonnie, R., Monahan, J., Eisenberg, M., & Feucht-Haviar, T. (1998). The competence-related abilities of women criminal defendants. *Journal of the American Academy of Psychiatry and the Law*, 26(2), 215-222.
- Pratt, D., Appleby, L., Piper, M., Webb, R., & Shaw, J. (2010). Suicide in recently released prisoners: a case-control study. *Psychological Medicine*, 40(5), 827-835. doi: 10.1017/S0033291709991048

- Priebe, S., Frottier, P., Gaddini, A., Kilian, R., Lauber, C., Martinez-Leal, R., . . . Wright, D. (2008). Mental health care institutions in nine European countries, 2002 to 2006. *Psychiatric Services, 59*(5), 570-573. doi: 10.1176/appi.ps.59.5.570
- Prince, M., Patel, V., Saxena, S., Maj, M., Maselko, J., Phillips, M. R., & Rahman, A. (2007). No health without mental health. *The Lancet, 370*(9590), 859-877. doi: 10.1016/S0140-6736(07)61238-0
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk (2nd ed.)*. Washington, DC US: American Psychological Association.
- Raimer, B., & Stobo, J. (2004). Health care delivery in the texas prison system: the role of academic medicine. *JAMA, 292*(4), 485-489. doi: 10.1001/jama.292.4.485.
- Randall, P., Carr, A., Dooley, B., & Rooney, B. (2011). Psychological characteristics of Irish clerical sexual offenders. *The Irish Journal of Psychology, 32*(1-2), 4-13. doi: 10.1080/03033910.2011.610191
- Rasch, W., & Konrad, N. (1986). *Forensische Psychiatrie*. Stuttgart: Kohlhammer.
- Rasmussen, K., Storsaeter, O., & Levander, S. (1998). Psychiatric disorders in a Norwegian prison population. *Nordisk Psykiatrisk Supplement, 22*(1), 91-97.
- Rettenberger, M., & Eher, R. (2013). Actuarial risk assessment in sexually motivated intimate-partner violence. *Law and Human Behavior, 37*(2), 75-86. doi: 10.1037/b0000001
- Rettenberger, M., Matthes, A., Boer, D. P., & Eher, R. (2009). Prospective actuarial risk assessment: A comparison of five risk assessment instruments in different sexual offender subtypes. *International Journal of Offender Therapy and Comparative Criminology, 54*(2), 169-186. doi: 10.1177/0306624x08328755
- Rice, M. E., Harris, G. T., & Lang, C. (2013). Validation of and revision to the VRAG and SORAG: the Violence Risk Appraisal Guide-Revised (VRAG-R). *Psychol Assess, 25*(3), 951-965. doi: 10.1037/a0032878
- Robertson, G. (1990). Correlates of crime among women offenders. *Medicine, Science and the Law, 30*(2), 165-174. doi: 10.1177/002580249003000216
- Robins, R., & Reiger, D. (1991). *Psychiatric disorders in America: The Epidemiologic Catchment Area Study*. New York: The Free Press.
- Roesch, R. (1995). Mental health interventions in pretrial jails. *Psychology and Law: Advances in Research. Berlin: De Greuter, 520-531*.

- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41-55. doi: 10.1093/biomet/70.1.41
- Rosenfeld, B., & Penrod, S. D. (2011). *Research methods in forensic psychology*: John Wiley & Sons.
- Rossegger, A., Endrass, J., & Gerth, J. (2012a). Einführung ins Risk-Assessment. In J. Endrass, A. Rossegger, F. Urbaniok, & B. Borchard (Eds.), *Interventionen bei Gewalt- und Sexualstraftätern: Risk-Management, Methoden und Konzepte der forensischen Therapie* (pp. 91-97). Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft.
- Rossegger, A., Endrass, J., & Gerth, J. (2012b). Mechanische Risk-Assessment Instrumente. In J. Endrass, A. Rossegger, F. Urbaniok, & B. Borchard (Eds.), *Interventionen bei Gewalt- und Sexualstraftätern: Risk-Management, Methoden und Konzepte der forensischen Therapie* (pp. 98-122). Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft.
- Rossegger, A., Endrass, J., Gerth, J., & Singh, J. P. (2014). Replicating the violence risk appraisal guide: a total forensic cohort study. *PLoS ONE*, *9*(3), e91845. doi: 10.1371/journal.pone.0091845
- Rossegger, A., Endrass, J., Urbaniok, F., & Borchard, B. (2012). Thesen deliktpräventiver Therapien. In J. Endrass, A. Rossegger, F. Urbaniok, & B. Borchard (Eds.), *Interventionen bei Gewalt- und Sexualstraftätern: Risk-Management, Methoden und Konzepte der forensischen Therapie* (pp. 135-141). Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft.
- Rossegger, A., Gerth, J., Seewald, K., Urbaniok, F., Singh, J. P., & Endrass, J. (2013). Current obstacles in replicating risk assessment findings: a systematic review of commonly used actuarial instruments. *Behavioral Sciences & the Law*, *13*(10), 154-164. doi: 10.1002/bsl.2044
- Rossegger, A., Gerth, J., Singh, J. P., & Endrass, J. (2013). Examining the Predictive Validity of the SORAG in Switzerland. *Sexual Offender Treatment*, *8*(2).
- Rossegger, A., Gerth, J., Urbaniok, F., Laubacher, A., & Endrass, J. (2010). The Sex Offender Risk Appraisal Guide (SORAG) - Validität und autorisierte deutsche Übersetzung. *Fortschritte der Neurologie und Psychiatrie*, *78*(11), 658-667. doi: 10.1055/s-0029-1245688

- Rossegger, A., Laubacher, A., Moskvitin, K., Villmar, T., Palermo, G. B., & Endrass, J. (2011). Risk assessment instruments in repeat offending: The usefulness of FOTRES. *International Journal of Offender Therapy and Comparative Criminology*, 55(5), 716-731. doi: 10.1177/0306624X09360662
- Rossegger, A., Urbaniok, F., Danielsson, C., & Endrass, J. (2009). The Violence Risk Appraisal Guide (VRAG) - Ein Instrument zur Kriminalprognose bei Gewaltstraftätern: Übersichtsarbeit und autorisierte deutsche Übersetzung. *Fortschritte der Neurologie und Psychiatrie*, 77(10), 577-584. doi: 10.1055/s-0028-1109705
- Saß, H. (1985). Ein psychopathologisches Referenzsystem für die Beurteilung der Schuldfähigkeit. *Forensia*, 6, 33-43.
- Saß, H., Wittchen, H.-U., Zaudig, M., & Houben, I. (2003). *Diagnostisches und Statistisches Manual Psychischer Störungen – Textrevision – DSM-IV-TR* Göttingen: Hogrefe.
- Sawyer, A. M., & Borduin, C. M. (2011). Effects of multisystemic therapy through midlife: a 21.9-year follow-up to a randomized clinical trial with serious and violent juvenile offenders. *Journal of Consulting and Clinical Psychology*, 79(5), 643. doi: 10.1037/a0024862
- Saxena, S., Thornicroft, G., Knapp, M., & Whiteford, H. (2007). Resources for mental health: scarcity, inequity, and inefficiency. *The Lancet*, 370(9590), 878-889. doi: 10.1016/S0140-6736(07)61239-2
- Schauer, C. (2006). *Aufforderung zum Spiel: Foucault und das Recht*: Böhlau Verlag Köln Weimar.
- Schenk, A. M., & Fremouw, W. J. (2012). Individual characteristics related to prison violence: A critical review of the literature. *Aggression and Violent Behavior*, 17(5), 430-442. doi: 10.1016/j.avb.2012.05.005
- Schoemaker, D., & VanZessen, G. (1997). *Psychische stoornissen bij gedetineerden [Psychiatric disorders in prisoners]*. Utrecht:Trimbos-Instituut.
- Schuckit, M. A., Hermann, G., & Schuckit, J. J. (1977). Importance of psychiatric illness in newly arrested prisoners. *The Journal of Nervous and Mental Disease*, 165(2), 118-125.
- Segagni Lusignani, G., Giacobone, C., Pozzi, F., Dal Canton, F., Alecci, P., & Carra, G. (2006). Disturbi mentali in una casa circondariale: uno studio di prevalenza [Mental disorders in a local prison: a prevalence study]. *NÓOS*, 1, 23-34.

- Seto, M. C., Marques, J. K., Harris, G. T., Chaffin, M., Lalumière, M. L., Miner, M. H., . . . Quinsey, V. L. (2008). Good science and progress in sex offender treatment are intertwined: A response to Marshall and Marshall (2007). *Sexual Abuse: A journal of research and treatment*, 20(3), 247-255. doi: 10.1177/1079063208317733
- Sexton, T. L., & Alexander, J. F. (2003). Functional family therapy: A mature clinical model for working with at-risk adolescents and their families. In T. L. Sexton, G. R. Weeks, & M. S. Robbins (Eds.), *Handbook of family therapy: The science and practice of working with families and couples* (pp. 323-348). New York, NY, US: Brunner-Routledge.
- Sharma, A., Nijhawan, M., Sharma, D. K., & Sushil, C. S. (2010). Psychosocial and psychiatric aspect of criminal behavior. *Indian Journal of Psychiatry*, 52, S37.
- Sherman, L. W., Gottfredson, D. C., MacKenzie, D. L., Eck, J., Reuter, P., & Bushway, S. D. (1998). What Works, what Doesn't, what's promising: A report to the United States Congress, National Institute of Justice.
- Shorey, R. C., Febres, J., Brasfield, H., & Stuart, G. L. (2012). The prevalence of mental health problems in men arrested for domestic violence. *Journal of Family Violence*, 27(8), 741-748. doi: 10.1007/s10896-012-9463-z
- Simpson A, Brinded P, Laidlaw T, Fairley N, & Malcolm F. (1999). The national study of psychiatric morbidity in New Zealand prisons. Auckland:: Department of Corrections, .
- Singh, D. (2006). Publication bias-A reason for the decreased research output in developing countries. *South African Psychiatry Review*, 9(3), 153-155. doi: 10.4314/ajpsy.v9i3.30216
- Singh, J. P. (2013). Predictive validity performance indicators in violence risk assessment: A methodological primer. *Behavioral Sciences & the Law*, 31(1), 8-22. doi: 10.1002/bsl.2052
- Singh, J. P., Desmarais, S. L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., de Ruiter, C., . . . Otto, R. K. (2014). International perspectives on the practical application of violence risk assessment: A global survey of 44 countries. *International Journal of Forensic Mental Health*, 13(3), 193-206. doi: 10.1080/14999013.2014.922141
- Skeem, J. L., Manchak, S., & Peterson, J. (2011). Correctional policy for offenders with mental illness: creating a new paradigm for recidivism reduction. *Law and Human Behavior*, 35(2), 110-126. doi: 10.1007/s10979-010-9223-7

- Skeem, J. L., & Monahan, J. (2011). Current directions in violence risk assessment. *Current Directions in Psychological Science, 20*(1), 38-42. doi: 10.1177/0963721410397271
- Slade, M., & Priebe, S. (2001). Are randomised controlled trials the only gold that glitters? *The British Journal of Psychiatry, 179*(4), 286-287. doi: 10.1192/bjp.179.4.286
- Smid, W. J., Kamphuis, J. H., Wever, E. C., & Van Beek, D. (2014). A quasi-experimental evaluation of high-intensity inpatient sex offender treatment in the Netherlands. *Sexual Abuse: A journal of research and treatment*(Epub ahead of print). doi: 1079063214535817
- Smith, C., O'Neill, H., Tobin, J., Walshe, D., & Dooley, E. (1996). Mental disorders detected in an Irish prison sample. *Criminal Behaviour and Mental Health, 6*(2), 177-183. doi: 10.1002/cbm.85
- Smith, P., Gendreau, P., & Swartz, K. (2009). Validating the principles of effective intervention: A systematic review of the contributions of meta-analysis in the field of corrections. *Victims and Offenders, 4*(2), 148-169. doi: 10.1080/15564880802612581
- Sreenivasan, S., Kirkish, P., Garrick, T., Weinberger, L. E., & Phenix, A. (2000). Actuarial risk assessment models: A review of critical issues related to violence and sex-offender recidivism assessments. *Journal of the Academy of Psychiatry and the Law, 28*(4), 438-448.
- Stadtland, C., & Nedopil, N. (2007). Prognosebegutachtung bei Sexualstraftaetern - Der Static-99. Prediction of recidivism in sex offenders - The Static-99. *Forensische Psychiatrie und Psychotherapie, 14*(2), 53-64.
- Staggs, S. L., & Schewe, P. A. (2011). Primary prevention of domestic violence. In M. P. Koss, J. W. White, & A. E. Kazdin (Eds.), *Violence Against women and Children, Vol 2: Navigating Solutions*. (pp. 237-257). Washington, DC, US: American Psychological Association.
- StataCorp. (2013). *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP. : StataCorp.
- Steadman, H. J., Osher, F. C., Robbins, P. C., Case, B., & Samuels, S. (2009). Prevalence of serious mental illness among jail inmates. *Psychiatric Services, 60*(6), 761-765. doi: 10.1176/ps.2009.60.6.761

- Stinson, J. D., Becker, J. V., & McVay, L. A. (2015). Treatment progress and behavior following 2 years of inpatient sex offender treatment: A pilot investigation of Safe Offender Strategies. *Sexual Abuse: A journal of research and treatment*(Epub ahead of print). doi: 10.1177/1079063211432475
- Stompe, T., Brandstätter, N., Ebner, N., & Fischer-Danzinger, D. (2010). Psychiatric disorders in prison inmates (Psychiatrische Störungen bei Haftinsassen). *Journal für Neurologie, Neurochirurgie, Psychiatrie, 11*, 20-23.
- Storey, J. E., Kropp, P. R., Hart, S. D., Belfrage, H., & Strand, S. (2014). Assessment and management of risk for intimate partner violence by police officers using the Brief Spousal Assault Form for the evaluation of risk. *Criminal Justice and Behavior, 41*(2), 256-271. doi: 10.1177/0093854813503960
- Swank, G. E., & Winer, D. (1976). Occurrence of psychiatric disorder in a county jail population. *American Journal of Psychiatry, 133*(11), 1331-1333. doi: 10.1176/ajp.133.11.1331
- Sweeting, M. J., Sutton, A. J., & Lambert, P. C. (2004). What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Statistics in Medicine, 23*(9), 1351-1375. doi: 10.1002/sim.1761
- Szmukler, G., Everitt, B., & Leese, M. (2012). Risk assessment and receiver operating characteristic curves. *Psychological Medicine, 42*(5), 895-898. doi: 10.1017/S003329171100208X
- Teplin, L. A. (1994). Psychiatric and substance abuse disorders among male urban jail detainees. *American Journal of Public Health, 84*(2), 290-293. doi: 10.2105/AJPH.84.2.290
- Teplin, L. A., Abram, K. M., & McClelland, G. M. (1996). Prevalence of psychiatric disorders among incarcerated women: I. Pretrial jail detainees. *Archives of General Psychiatry, 53*(6), 505-512. doi: 10.1001/archpsyc.1996.01830060047007
- The World Bank. (2010). How we classify countries: Country and lending groups. The World Bank Website. Retrieved January, 2011, from <http://data.worldbank.org/about/country-classifications/country-and-lending-groups>

- Torrey, E., Kennard, A., Eslinger, D., Lamb, R., & Pavle, J. (2010). More mentally ill persons are in jails and prisons than hospitals: A survey of the States. Arlington, VA: Treatment Advocacy Center.
- Tort, V., Vicens E. (2009). *Prevalence of mental disorders in Spanish prisons: PRECA study*. Paper presented at the 9th Annual IAFMHS Conference, Edinburgh, Scotland.
- Travers, R., Wakeling, H. C., Mann, R. E., & Hollin, C. R. (2013). Reconviction following a cognitive skills intervention: An alternative quasi-experimental methodology. *Legal and Criminological Psychology, 18*(1), 48-65. doi: 10.1111/j.2044-8333.2011.02026.x
- Trestman, R. L., Ford, J., Zhang, W., & Wiesbrock, V. (2007). Current and lifetime psychiatric illness among inmates not identified as acutely mentally ill at intake in Connecticut's jails. *Journal of the American Academy of Psychiatry and the Law, 35*(4), 490-500.
- Urbaniok, F. (2003). *Was sind das für Menschen–was können wir tun. Nachdenken über Straftäter*. Bern: Zytglogge Verlag.
- Urbaniok, F. (2007). *FOTRES - Forensisches Operationalisiertes Therapie-Risiko-Evaluations-System* (2. Aufl. ed.). Schoren: Zytglogge
- Urbaniok, F., Endrass, J., & Rossegger, A. (2008). Das Zürcher Modell und Ergebnisse der Zürcher Forensik-Studie unter besonderer Berücksichtigung von Präventionsdefiziten in Deutschland. from BIOS Baden-Württemberg <http://www.bios-bw.de/images/stories/pdfs/urbaniok-rueckfalluntersuchung-bios-11-11-2008-karlsruhe.pdf>
- Urbaniok, F., Rinne, T., Held, L., Rossegger, A., & Endrass, J. (2008). Forensische Risikokalkulationen: Grundlegende methodische Aspekte zur Beurteilung der Anwendbarkeit und Validität verschiedener Verfahren. *Fortschritte der Neurologie und Psychiatrie, 76*(8), 470-477. doi: 10.1055/s-2008-1038228
- Van Dorn, R. A., Volavka, J., & Johnson, N. (2012). Mental disorder and violence: is there a relationship beyond substance use? *Social Psychiatry and Psychiatric Epidemiology, 47*(3), 487-503. doi: 10.1007/s00127-011-0356-x
- Van Gemmeren, G. (2012). StGB §63 Unterbringung in einem psychiatrischen Krankenhaus *Münchener Kommentar: StGB, Rn. 56-59*.
- van Panhuis, P. (1997). *De psychotische patient in de TBS. [The psychotic patient in the TBS]*. Deventer: Gouda Quint.

- Violence against women: An EU-wide survey. (2014): European Union Agency for Fundamental Rights (FRA).
- Volavka, J., & Swanson, J. (2010). Violent behavior in mental illness: the role of substance abuse. *JAMA*, *304*(5), 563-564. doi: 10.1001/jama.2010.1097
- von Schoenfeld, C. E., Schneider, F., Schroder, T., Widmann, B., Botthof, U., & Driessen, M. (2006). Prevalence of psychiatric disorders, psychopathology, and the need for treatment in female and male prisoners. *Der Nervenarzt*, *77*(7), 830.
- Walmsley, R. (2009). World prison population list (eighth edition). London: King's College London International Centre for Prison Studies.
- Walter, M., Wiesbeck, G. A., Dittmann, V., & Graf, M. (2011). Criminal recidivism in offenders with personality disorders and substance use disorders over 8 years of time at risk. *Psychiatry Research*, *186*(2-3), 443-445. doi: 10.1016/j.psychres.2010.08.009
- Ward, T., & Brown, M. (2004). The Good Lives Model and conceptual issues in offender rehabilitation. *Psychology, Crime & Law*, *10*(3), 243-257. doi: 10.1080/10683160410001662744
- Ward, T., Yates, P. M., & Willis, G. M. (2012). The Good Lives Model and the Risk Need Responsivity Model: A critical response to Andrews, Bonta, and Wormith (2011). *Criminal Justice and Behavior*, *39*(1), 94-110. doi: 10.1177/0093854811426085
- Washington, P., & Diamond, R. J. (1985). Prevalence of mental illness among women incarcerated in five California county jails. *Research in Community & Mental Health*, *5*(6), 33-41. doi: 10.1001/archpsyc.1996.01830060047007
- Watt, F., Tomison, A., & D, T. (1993). The prevalence of psychiatric disorder in a male remand population: a pilot study. *Journal of Forensic Psychiatry & Psychology*, *4*(1), 75-83. doi: 10.1080/09585189308407793
- Watzke, S., Ullrich, S., & Marneros, A. (2006). Gender- and violence-related prevalence of mental disorders in prisoners. *European Archives of Psychiatry and Clinical Neuroscience*, *256*(7), 414-421. doi: 10.1007/s00406-006-0656-4
- Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20: Assessing Risk for Violence (Version 2)*. Burnaby, Canada: Simon Fraser University.

- Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2004). The empirical status of empirically supported psychotherapies: assumptions, findings, and reporting in controlled clinical trials. *Psychological Bulletin*, *130*(4), 631-663. doi: 10.1037/0033-2909.130.4.631
- Whitaker, D. J., Morrison, S., Lindquist, C., Hawkins, S. R., O'Neil, J. A., Nesius, A. M., . . . Reese, L. R. (2006). A critical review of interventions for the primary prevention of perpetration of partner violence. *Aggression and Violent Behavior*, *11*(2), 151-166. doi: 10.1016/j.avb.2005.07.007
- Wilkins, J., & Coid, J. (1997). Self-mutilation in female remanded prisoners: I. An indicator of severe psychopathology. *Criminal Behaviour and Mental Health*, *1*(3), 247-267.
- Willis, G. M., Ward, T., & Levenson, J. S. (2014). The good lives model (GLM): An evaluation of GLM operationalization in North American treatment programs. *Sexual Abuse: A journal of research and treatment*, *26*(1), 58-81. doi: 10.1177/1079063213478202
- Wilper, A. P., Woolhandler, S., Boyd, J. W., Lasser, K. E., McCormick, D., Bor, D. H., & Himmelstein, D. U. (2009). The health and health care of US prisoners: Results of a nationwide survey. *American Journal of Public Health*, *99*(4), 666-672. doi: 10.2105/AJPH.2008.144279
- Wilson, D. B., Bouffard, L. A., & Mackenzie, D. L. (2005). A quantitative review of structured, group-oriented, cognitive-behavioral programs for offenders. *Criminal Justice and Behavior*, *32*(2), 172-204. doi: 10.1177/0093854804272889
- Woessner, G., & Schwedler, A. (2014). Correctional treatment of sexual and violent offenders: Therapeutic change, prison climate, and recidivism. *Criminal Justice and Behavior*, *41*(7), 862-879. doi: 10.1177/0093854813520544
- Woodrow, A. C., & Bright, D. A. (2011). Effectiveness of a sex offender treatment programme: A risk band analysis. *International Journal of Offender Therapy and Comparative Criminology*, *55*(1), 43-55. doi: 10.1177/0306624X09352162
- World Health Organization. (2009). World Health Organisation, Women's health in prison. Copenhagen: WHO Regional Office for Europe.
- World Health Organization. (2010). *The ICD-10 Classification of Mental and Behavioural Disorders*. Geneva: WHO.

- World Health Organization. (2013). Global and regional estimates of violence against women: Prevalence and health effects of intimate partner violence and non-partner sexual violence. Geneva.
- Wright, B., Duffy, D., Curtin, K., Linehan, S., Monks, S., & Kennedy, H. G. (2006). Psychiatric morbidity among women prisoners newly committed and amongst remanded and sentenced women in the Irish prison system. *Irish Journal of Psychological Medicine*, 23(2), 47-53. doi: 10.1017/S0790966700009575
- Yu, R., Geddes, J. R., & Fazel, S. (2012). Personality disorders, violence, and antisocial behavior: a systematic review and meta-regression analysis. *Journal of Personality Disorders*, 26(5), 775-792. doi: 10.1521/pedi.2012.26.5.775
- Yue, L. Q. (2007). Statistical and regulatory issues with the application of propensity score analysis to nonrandomized medical device clinical studies. *Journal of Biopharmaceutical Statistics*, 17(1), 1-13. doi: 10.1080/10543400601044691
- Zahari, M. M., Hwan Bae, W., Zainal, N. Z., Habil, H., Kamarulzaman, A., & Altice, F. L. (2010). Psychiatric and substance abuse comorbidity among HIV seropositive and HIV seronegative prisoners in Malaysia. *American Journal of Drug and Alcohol Abuse*, 36(1), 31-38. doi: 10.3109/00952990903544828
- Zoccali, R., Muscatello, M. R., Bruno, A., Cambria, R., & Cavallaro, L. (2008). Mental disorders and request for psychiatric intervention in an Italian local jail. *International Journal of Law and Psychiatry*, 31(5), 447-450. doi: 10.1016/j.ijlp.2008.08.004