

Evaluation der risikoorientierten Bewährungshilfe in Hessen: Die Prognose- und Interventionspraxis im Sicherheitsmanagement II

Breiling, Lisanne; Reese, Viktoria; Rettenberger, Martin

Veröffentlichungsversion / Published Version

Forschungsbericht / research report

Empfohlene Zitierung / Suggested Citation:

Breiling, L., Reese, V., & Rettenberger, M. (2021). *Evaluation der risikoorientierten Bewährungshilfe in Hessen: Die Prognose- und Interventionspraxis im Sicherheitsmanagement II*. (Berichte und Materialien (BM-Online), 27). Wiesbaden: Eigenverlag Kriminologische Zentralstelle. <https://nbn-resolving.org/urn:nbn:de:hebis:2378-opus-2174>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

KRIMZ

KRIMINOLOGISCHE ZENTRALSTELLE E.V.

Evaluation der risikoorientierten Bewährungshilfe in Hessen

**Die Prognose- und Interventionspraxis im
Sicherheitsmanagement II**

Lisanne Breiling

Viktoria Reese

Martin Rettenberger

BM-Online
Elektronische Schriftenreihe der KrimZ

Band 27

Berichte und Materialien (BM-Online)
Elektronische Schriftenreihe der Kriminologischen Zentralstelle e.V. (KrimZ)
Band 27

Evaluation der risikoorientierten Bewährungshilfe in Hessen

Die Prognose- und Interventionspraxis im Sicherheitsmanagement II

Lisanne Breiling

Viktoria Reese

Martin Rettenberger

Wiesbaden 2021

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie. Detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Diese Publikation wird gefördert aus Mitteln des Hessischen Ministeriums der Justiz.

© **Eigenverlag** Kriminologische Zentralstelle e.V.
Alle Rechte vorbehalten

KRIMZ

KRIMINOLOGISCHE ZENTRALSTELLE E.V.

Luisenstraße 7, 65185 Wiesbaden
<http://www.krimz.de/publikationen>
ISSN 2199-4188
ISBN 978-3-945037-39-3

Vorwort

Die vorliegende Publikation stellt den vorläufigen Abschluss eines mehr als 2-jährigen Forschungsprojekts dar, dessen Ergebnisse aus unserer Sicht von hoher kriminalpolitischer und kriminologischer Bedeutung sind, da sie eindrücklich veranschaulichen, wie mittels vergleichsweise einfacher Mittel signifikante Verbesserungen der Betreuungs- und Behandlungspraxis von straffällig gewordenen Menschen erreicht werden können. Wir möchten uns an dieser Stelle sehr herzlich bei allen Personen und Institutionen bedanken, die die Durchführung der vorliegenden Studie überhaupt erst ermöglicht bzw. unterstützt haben. Zuallererst möchten wir dem Hessischen Ministerium der Justiz (HMdJ) für das entgegengebrachte Vertrauen bedanken, für die Finanzierung der Studie und die durchgehend aktive und engagierte Unterstützung. Die Mitarbeiter/-innen, die im HMdJ für die vorliegende Studie zuständig waren, standen uns stets unterstützend zur Seite, so dass jede Herausforderung schnell bewältigt und jede offene Frage unmittelbar beantwortet werden konnte.

Des Weiteren möchten wir vor allem den SIMA II-Fachbereichsleiter/-innen für den regen, immer ausgesprochen freundlichen und konstruktiven Austausch danken, die einen nicht unerheblichen Aufwand betrieben haben, uns die notwendigen Daten zur Verfügung zu stellen, und die uns mit ihrer offenen und wertschätzenden Art diese Evaluation erst ermöglicht haben. Besonders erwähnt werden sollen die Interviewpartner/-innen, denen wir für ihre Zeit, ihre Offenheit und ihre erkenntnisreichen Perspektiven danken möchten. Darüber hinaus möchten wir allen im SIMA II arbeitenden Bewährungshelfer/-innen, die zum Gelingen des Projektes beigetragen haben, herzlich danken. Neben der Bewährungshilfe haben auch andere Fachbereiche unverzichtbare Beiträge geleistet, so zum Beispiel die für uns bzw. das Projekt zuständigen Mitarbeiter/-innen der IT-Stelle der hessischen Justiz, denen unser besonderer Dank gilt. Abschließend möchten wir uns bei Wanja Hemmerich, Olivia Rosebrock und Anna Monschau sowie bei unseren Kollegen/-innen an der Kriminologischen Zentralstelle bedanken, die das Projekt in unterschiedlicher Form unterstützt haben.

Das vorliegende Projekt zeigt eindrucksvoll, wie wichtig und sinnvoll es sein kann, wenn Praxisveränderungen wissenschaftlich eingeleitet und begleitet werden. Gleichzeitig möchten wir unserer Hoffnung Ausdruck verleihen, dass die vorliegende Studie dazu beiträgt, weitere und vergleichbare Maßnahmen im Umgang

mit straffällig gewordenen Personen anzustoßen: Das Risikoprinzip wirkt und das bedeutet weniger Taten, weniger Betroffene und weniger Täter/-innen.

Wiesbaden, im Januar 2021

Prof. Dr. Martin Rettenberger
Direktor der Kriminologischen
Zentralstelle (KrimZ)

Lisanne Breiling

Viktoria Reese

Inhaltsverzeichnis

Vorwort	5
Inhaltsverzeichnis	7
Tabellenverzeichnis.....	9
Abbildungsverzeichnis	12
0 Zusammenfassung	13
1 Einleitung.....	16
2 Fragestellungen	19
2.1 Projektteil 1: Güte der Risikoeinschätzung	19
2.2 Projektteil 2: Kriminalpräventive Wirksamkeit der Betreuung	22
2.3 Projektteil 3: Qualitative Interviews mit Bewährungshelfer/-innen.....	22
3 Methodisches Vorgehen	24
3.1 Datenzugänge	24
3.1.1 Probandendaten der Bewährungshilfe.....	24
3.1.2 Bundeszentralregister (BZR)	24
3.1.3 Mehrländer-Staatsanwaltschafts-Automation (MESTA).....	26
3.2 Stichprobenbildung	27
3.3 Projektteil 1: Güte der Risikoeinschätzung	31
3.3.1 Reliabilität	31
3.3.2 Validität	32
3.4 Projektteil 2: Kriminalpräventive Wirksamkeit der Betreuung	34
3.4.1 Bildung der Kontrollgruppe mittels Matching.....	35
3.4.2 Auswertung von Rückfallereignissen	40
3.5 Projektteil 3: Qualitative Interviews mit Bewährungshelfer/-innen.....	40
4 Ergebnisse	43
4.1 Deskriptive Statistiken.....	43
4.1.1 Stichprobenbeschreibung	43
4.1.2 Rückfallereignisse	46
4.2 Projektteil 1: Güte der Risikoeinschätzung	56
4.2.1 Güte des SVG-5	56
4.2.2 Güte des Erhebungsbogens.....	63
4.2.3 Diskussion	71
4.3 Projektteil 2: Kriminalpräventive Wirksamkeit der Betreuung	78
4.3.1 Vergleich der Rückfallraten.....	79

4.3.2	Diskussion	85
4.4	Projektteil 3: Qualitative Interviews mit Bewährungshelfer/-innen.....	90
4.5.1	Auswertungsergebnisse der Interviews.....	91
4.5.2	Diskussion	95
5	Fazit	98
	Literaturverzeichnis	102
	Anhang.....	108

Tabellenverzeichnis

Tabelle 1:	Betreuungseinstufungen auf Basis des SVG-5.....	20
Tabelle 2:	Betreuungseinstufungen auf Basis des Erhebungsbogens	21
Tabelle 3:	Erläuterungen zu den analysierten Rückfallereignissen.....	40
Tabelle 4:	Unterstellungsgründe der SIMA II-Probanden/-innen für die Bewährungshilfe (n = 607)	44
Tabelle 5:	Rechtliche Gründe für den Eintritt der Führungsaufsicht bei den SIMA II-Probanden/-innen (n = 857).....	44
Tabelle 6:	Indexdelikte der SIMA II-Probanden/-innen (n = 1 339)	45
Tabelle 7:	Neuerliche rechtskräftige Verurteilungen (BZR-Einträge) der SIMA II-Probanden/-innen (n = 1 339).....	48
Tabelle 8:	Neuerliche rechtskräftige Verurteilungen (BZR-Einträge) der SIMA II-Probanden/-innen (n = 1 339), getrennt nach Anlassdelikt (Gewaltdelikte)	48
Tabelle 9:	Neuerliche rechtskräftige Verurteilungen (BZR-Einträge) der SIMA II-Probanden/-innen (n = 1 339), getrennt nach Anlassdelikt (BtMG-Verstoß)	49
Tabelle 10:	Neuerliche rechtskräftige Verurteilungen (BZR-Einträge) der SIMA II-Probanden/-innen (n = 1 339), getrennt nach Anlassdelikt (Tötungsdelikte)	49
Tabelle 11:	Neuerliche rechtskräftige Verurteilungen (BZR-Einträge) der SIMA II-Probanden/-innen (n = 1 339), getrennt nach Unterstellungsgrund (Führungsaufsicht).....	50
Tabelle 12:	Neuerliche rechtskräftige Verurteilungen (BZR-Einträge) der SIMA II-Probanden/-innen (n = 1 339), getrennt nach Unterstellungsgrund (Vollverbüßung).....	50
Tabelle 13:	Neuerliche rechtskräftige Verurteilungen (BZR-Einträge) der SIMA II-Probanden/-innen (n = 1 339), getrennt nach Geschlecht	51
Tabelle 14:	Neuerliche rechtskräftige Verurteilungen (BZR-Einträge) der SIMA II-Probanden/-innen (n = 1 339), getrennt nach Betreuungsstufe.....	51
Tabelle 15:	Neuerliche Strafanzeigen (MESTA-Einträge) gegen die SIMA II-Probanden/-innen (n = 1 339).....	52

Tabelle 16: Neuerliche Strafanzeigen (MESTA-Einträge) gegen die SIMA II-Probanden/-innen ($n = 1\,339$), getrennt nach Anlassdelikt (Gewaltdelikt).....	53
Tabelle 17: Neuerliche Strafanzeigen (MESTA-Einträge) gegen die SIMA II-Probanden/-innen ($n = 1\,339$), getrennt nach Anlassdelikt (BtMG-Verstoß)	53
Tabelle 18: Neuerliche Strafanzeigen (MESTA-Einträge) gegen die SIMA II-Probanden/-innen ($n = 1\,339$), getrennt nach Anlassdelikt (Tötungsdelikte)	54
Tabelle 19: Neuerliche Strafanzeigen (MESTA-Einträge) gegen die SIMA II-Probanden/-innen ($n = 1\,339$), getrennt nach Unterstellungsgrund (Führungsaufsicht).....	54
Tabelle 20: Neuerliche Strafanzeigen (MESTA-Einträge) gegen die SIMA II-Probanden/-innen ($n = 1\,339$), getrennt nach Unterstellungsgrund (Vollverbüßung).....	55
Tabelle 21: Neuerliche Strafanzeigen (MESTA-Einträge) gegen die SIMA II-Probanden/-innen ($n = 1\,339$), getrennt nach Geschlecht	55
Tabelle 22: Neuerliche Strafanzeigen (MESTA-Einträge) gegen die SIMA II-Probanden/-innen ($n = 1\,339$), getrennt nach Betreuungsstufe.....	56
Tabelle 23: Cohen's Kappa für die einzelnen Items des SVG-5	57
Tabelle 24: SVG-5-Summenwerte.....	58
Tabelle 25: Deskriptive Statistiken der einzelnen Items des SVG-5 für die Gesamtstichprobe ($n = 1\,689$)	59
Tabelle 26: Interpretationsrichtlinien für <i>Area Under the Curve</i> (AUC)-Werte	.60
Tabelle 27: Vorhersage neuerlicher Strafanzeigen (MESTA) durch den SVG-5	61
Tabelle 28: Vorhersage neuerlicher rechtskräftiger Verurteilungen (BZR) durch den SVG-5	61
Tabelle 29: Vorhersage neuerlicher rechtskräftiger Verurteilungen (BZR) für verschiedene Subgruppen durch SVG-5-Summenwert	62
Tabelle 30: Vergleich der Vorhersageleistung von SVG-5 und OGRS 3 in Bezug auf neuerliche rechtskräftige Verurteilungen (BZR).....	63
Tabelle 31: Vergleich der Vorhersageleistung von SVG-5 und OGRS 3 in Bezug auf neuerliche Strafanzeigen (MESTA)	63

Tabelle 32: Grad der Beurteilerübereinstimmung für die 10 Items des Erhebungsbogens	64
Tabelle 33: Summenwerte des Erhebungsbogens	65
Tabelle 34: Deskriptive Statistiken der einzelnen Items des Erhebungsbogens für die gesamte Stichprobe ($n = 410$)	65
Tabelle 35: Vorhersage neuerlicher Strafanzeigen (MESTA) durch den Erhebungsbogen	68
Tabelle 36: Vorhersage neuerlicher rechtskräftiger Verurteilungen (BZR) durch den Erhebungsbogen.....	69
Tabelle 37: Vorhersage neuerlicher strafrechtlicher Verurteilungen (BZR) für verschiedene Subgruppen durch Summenwert des Erhebungsbogens	69
Tabelle 38: Vergleich der Vorhersageleistung von Erhebungsbogen und OGRS 3 in Bezug auf neuerliche rechtskräftige Verurteilungen (BZR)	70
Tabelle 39: Vergleich der Vorhersageleistung von Erhebungsbogen und OGRS 3 in Bezug auf neuerliche Strafanzeigen (MESTA).....	71
Tabelle 40: <i>Offender Group Reconviction Scale</i> (OGRS 3)-Merkmale der SIMA II-Probanden/-innen und der Kontrollgruppe	79
Tabelle 41: Vergleich der Rückfälligkeit in Form rechtskräftiger Verurteilungen (BZR) zwischen SIMA II-Probanden/-innen und der Kontrollgruppe	80
Tabelle 42: Vergleich der Rückfälligkeit in Form rechtskräftiger Verurteilungen (BZR) zwischen SIMA II-Probanden/-innen und der Kontrollgruppe, getrennt nach OGRS 3-Risikogruppen.....	82
Tabelle 43: Rückfallvergleiche der Vollverbüßer/-innen.....	83
Tabelle 44: Rückfallvergleiche von Probanden/-innen mit bzw. ohne Gewaltanlassdelikt.....	85

Abbildungsverzeichnis

Abbildung 1: Schematische Darstellung der Stichprobenbildung	30
Abbildung 2: Verteilung der Gesamtstichprobe auf die 20 OGRS 3- Risikokategorien ($N = 2\,614$)	78

0 Zusammenfassung

Der folgende Bericht stellt die Ergebnisse eines 2-jährigen Projektes vor, in dessen Rahmen die Kriminologische Zentralstelle (KrimZ) die Prognose- und Interventionspraxis eines auf Gewaltdelikte spezialisierten Fachbereichs der Bewährungshilfe in Hessen evaluierte: des 2017 implementierten Sicherheitsmanagements (SIMA) II. Das SIMA II verfolgt mit seiner Konzeption einerseits das Ziel, die Bewährungshilfe stärker als zuvor am aktuellen Stand wissenschaftlicher Erkenntnisse zu orientieren. Erreicht werden sollte dieses Ziel durch die systematische Berücksichtigung des etablierten und wissenschaftlich gut untersuchten *Risk-Need-Responsivity* (RNR)-Modells (Andrews/Bonta 1990; 2010). Andererseits sollte eine Spezialisierung der Betreuung auf einzelne Gruppen von Probanden/-innen, die mit spezifischen Bedarfslagen in die Bewährungshilfe kommen, zuvor festgestellte Betreuungsdefizite kompensieren und fachliche, zielgruppenspezifische Vertiefungen ermöglichen.

Im SIMA II werden Probanden/-innen betreut, die zuvor aufgrund mindestens eines Gewaltdelikts verurteilt wurden und gleichzeitig ein erhöhtes Rückfallrisiko aufweisen. Darüber hinaus ist das SIMA II für Probanden/-innen zuständig, die unter Führungsaufsicht stehen und zwar keine aktuelle Gewaltstraftat begangen haben, gleichzeitig aber eine negative Sozialprognose aufweisen. Die Einstufung und damit auch die Festlegung der Betreuungsintensität erfolgt im SIMA II mithilfe zweier Prognoseinstrumente, welche der Einschätzung des individuellen Risikos krimineller Rückfälle dienen. Diese Risikoeinschätzung soll eine risikoorientierte und daher effektive und effiziente Ressourcensteuerung gewährleisten. Die kriminalprognostische und betreuerische Qualität dieses risikoorientierten und spezialisierten Betreuungskonzeptes sollte im Rahmen der hier vorgestellten Evaluation überprüft werden. Hierzu wurden das Projekt anhand dreier unterschiedlicher Fragestellungen unterteilt: Wie reliabel und valide sind die im SIMA II vorgenommenen Risikoeinstufungen (*Projektteil 1*)? Wie wirksam ist die Betreuung im SIMA II in Bezug auf die Legalbewährung bzw. ist sie wirksamer als die Betreuung vor der Implementierung (*Projektteil 2*)? Wie gut ließen sich die Neuerungen in der betreuerischen Praxis aus Sicht der im Fachbereich tätigen Bewährungshelfer/-innen umsetzen (*Projektteil 3*)?

Zur Umsetzung von Projektteil 1 wurden Risikoeinschätzungen von Probanden/-innen des SIMA II ausgewertet. Die Reliabilität der beiden Instrumente wurde auf Basis der Beurteilerübereinstimmung bestimmt. Das *Screeninginstrument zur Vorhersage des Gewaltrisikos* (SVG-5; Eher u.a. 2012), das bei allen aufgrund von

Gewaltdelikt verurteilten Probanden/-innen eingesetzt wird, erzielte bei den $n = 58$ doppelt eingeschätzten Probanden/-innen einen als sehr gut einzustufenden *Intra Class Correlation*-Koeffizienten (ICC) von .89 (siehe Abschnitt 4.2.1). Der von der Bewährungshilfe selbst konstruierte Erhebungsbogen, der bei allen sonstigen Probanden/-innen Anwendung findet, erzielte bei den $n = 50$ doppelt eingeschätzten Probanden/-innen einen ebenfalls sehr guten Übereinstimmungswert von $ICC = .85$ (siehe Abschnitt 4.2.2). Auf Ebene der einzelnen Items zeigten sich beim Erhebungsbogen allerdings Probleme in Form niedriger Übereinstimmungswerte und hoher Anteile fehlender Werte.

Zur Überprüfung der Validität des SVG-5 und des Erhebungsbogens wurden wiederum Einschätzungen von insgesamt mehr als 2 000 Probanden/-innen des Fachbereichs ausgewertet. Als Kriterium für die Validität wurden Rückfalldaten des Bundeszentralregisters (BZR) und der Mehrländer-Staatsanwaltschafts-Automation (Mesta) herangezogen. Der SVG-5-Summenwert sagte alle betrachteten Rückfallereignisse signifikant vorher. Die Effektstärken der Zusammenhänge mit dem jeweiligen Rückfallereignis lagen dabei im kleinen bis moderaten Bereich. Auch der Summenwert des Erhebungsbogens konnte den Großteil der betrachteten Rückfallereignisse signifikant vorhersagen. Die Effektstärken der Zusammenhänge mit dem jeweiligen Rückfallereignis variierten ebenfalls von klein bis moderat. Die Vorhersageleistungen beider Instrumente wurden mit jenen eines alternativen Instrumentes verglichen: der in Großbritannien etablierten 3. Auflage der *Offender Group Reconviction Scale* (OGRS 3; Francis/Soothill/Humphreys 2007; Howard u.a. 2009), die das allgemeine kriminelle Rückfallrisiko während eines 1-Jahreszeitraumes (bzw. 2-Jahreszeitraumes) erfasst. Insbesondere rechtskräftig verurteilte Rückfälle (BZR) sagte die OGRS 3 zum Teil besser vorher als der SVG-5. Im Vergleich zum Erhebungsbogen erreichte die OGRS 3 durchweg bessere Prognoseleistungen, obgleich der Unterschied nicht in jedem Fall statistisch signifikant war.

Im Rahmen von Projektteil 2, welcher sich der Untersuchung der kriminalpräventiven Wirksamkeit der Betreuung im SIMA II widmete, wurden $n = 1\,307$ Probanden/-innen des SIMA II mit einer geeigneten Kontrollgruppe verglichen. Diese Kontrollgruppe umfasste $n = 1\,307$ risikoäquivalente Probanden/-innen, die vor der Implementierung des SIMA II von der hessischen Bewährungshilfe betreut worden waren, jedoch die Eingangskriterien des SIMA II erfüllten. Risikoäquivalenz wurde mithilfe eines Matching-Verfahrens gewährleistet, welches auf den retrospektiv erhobenen Risikowerten der OGRS 3 basierte (siehe Abschnitt 3.4.1). Durch das Matching-Verfahren wurde sichergestellt, dass potentiell unterschiedliche Rückfallraten der beiden Gruppe auf die Wirkung

der im Fachbereich erfolgten Intervention zurückgeführt werden konnten (vgl. z.B. Duggan 2014; Keßler/Rettenberger 2017; Schmucker 2004). Die im SIMA II betreuten Probanden/-innen hatten hinsichtlich nahezu aller betrachteter Rückfallkriterien geringere Rückfallraten als die Kontrollgruppe (siehe Abschnitt 4.3.1). Die Rückfallquote reduzierte sich im Vergleich zur Kontrollgruppe – je nach betrachteter Subgruppe bzw. betrachtetem Rückfallparameter – zwischen 8 % und 74 %. Obgleich bei der Interpretation dieser Ergebnisse potentiell limitierende Faktoren berücksichtigt werden müssen (siehe Abschnitt 4.3.2), wiesen die Ergebnisse in der Gesamtschau auf eine rückfallpräventive Wirkung des im SIMA II realisierten Betreuungskonzepts hin.

Projektteil 3 ergänzte die quantitative Methodik von Projektteil 1 und 2 durch einen qualitativen Ansatz der Datenerhebung: Zur Überprüfung der praktischen Umsetzung der Neuerungen und zur Identifikation von Optimierungsmöglichkeiten wurden semi-strukturierte Experteninterviews mit $n = 9$ Mitarbeiter/-innen des SIMA II geführt (siehe Abschnitt 3.5). Die Gespräche ergaben, dass der Großteil der Befragten die grundsätzliche Ausrichtung des neuen Betreuungskonzeptes begrüßte, die in einer stärkeren wissenschaftlichen Fundierung und einem stärker standardisierten und damit besser vergleichbaren Arbeiten resultierte (siehe Abschnitt 4.4). Die Befragten identifizierten jedoch auch Schwachstellen, die aus ihrer Sicht zu Ressourcenengpässen führten und mitunter ihren Handlungsspielraum zu stark einschränkten. Die SIMA II-Klientel wurde als recht heterogen eingeschätzt, was bei einigen Befragten Zweifel am Erfolg des Spezialisierungsgedankens erzeugte. Auch Optimierungsbedarf in Bezug auf den Erhebungsbogen wurde kommuniziert.

Insgesamt zeigten die im Rahmen der hier vorgestellten wissenschaftlichen Evaluation gewonnenen Erkenntnisse diverse Möglichkeiten auf, die Risikoeinschätzung und die Betreuung im SIMA II noch effektiver und effizienter zu gestalten sowie noch stärker in Einklang mit den Bedürfnissen und Wünschen der Mitarbeiter/-innen zu bringen. Trotz dieser möglichen Ansatzpunkte zeigten die vorliegenden Ergebnisse jedoch eindrücklich die Wirksamkeit der Prognose- und Interventionspraxis des SIMA II. Die vorliegenden Ergebnisse verdeutlichen damit mit Nachdruck die praktische Relevanz wissenschaftlicher Erkenntnisse der Evaluationsforschung für die Implementierung und Durchführung kriminalpräventiver Maßnahmen.

1 Einleitung

Am 1. Februar 2017 wurde in der Bewährungshilfe Hessen ein neuer Fachbereich implementiert, das *Sicherheitsmanagement II* (SIMA II; Hessisches Ministerium der Justiz 2017). Jeder der neun hessischen Landgerichtsbezirke (Darmstadt, Frankfurt, Fulda, Gießen, Hanau, Kassel, Limburg, Marburg und Wiesbaden) wurde um diesen Fachbereich erweitert. Aufgabe des SIMA II ist die Betreuung von Probanden/-innen, die zuvor aufgrund mindestens eines Gewaltdelikts verurteilt wurden und gleichzeitig ein hohes Rückfallrisiko aufweisen. Des Weiteren werden Probanden/-innen betreut, die unter Führungsaufsicht stehen und zwar keine aktuelle Gewaltstraftat begangen haben, gleichzeitig aber eine negative Sozialprognose aufweisen.

Mit der Implementierung des SIMA II wird unter Berücksichtigung des *Risk-Need-Responsivity* (RNR)-Modells (Andrews/Bonta 2010) das Ziel einer risikoorientierten und somit möglichst effektiven und effizienten Betreuungspraxis verfolgt (Müller/Nixdorff/Pirner 2018). Das Risikoprinzip (*Risk*) besagt, dass Personen mit hohem Rückfallrisiko eine intensivere Betreuung erhalten sollten als Personen mit niedrigerem Risiko; dem Bedürfnisprinzip (*Need*) wird Genüge geleistet, wenn die Betreuung die individuellen kriminogenen Faktoren (empirisch belegte dynamisch-veränderbare Risikofaktoren) einer jeweiligen Person adressiert; das Ansprechbarkeitsprinzip (*Responsivity*) besagt schließlich, dass Betreuungen an den individuellen Fähigkeiten und Voraussetzungen einer Person sowie ihrem biographischen Hintergrund orientiert sein sollten (Andrews/Bonta 1994; 2010). Mehrere Metaanalysen konnten mittlerweile belegen, dass Interventionsprogramme, die sich an den RNR-Prinzipien orientieren, kriminelle Rückfälle effektiver reduzieren können als Programme, die diese Prinzipien nicht berücksichtigen (z. B. Bourgon/Armstrong 2005; Hanson u.a. 2009; Koehler u.a. 2013; Lowenkamp/Latessa/Holsinger 2006; Prendergast u.a. 2013). Das SIMA II orientiert sich mit seinem Prognose- und Interventionskonzept damit an den aktuellen wissenschaftlichen Erkenntnissen bezüglich der Wirksamkeit von Betreuungsmaßnahmen.

Das SIMA II wurde vor dem Hintergrund der fortschreitenden Spezialisierung innerhalb der hessischen Bewährungshilfe eingeführt. Das gesamte Aufgabenspektrum der Bewährungshilfe wurde als zu komplex betrachtet, um von einzelnen Mitarbeiter/-innen vollständig abgedeckt werden zu können. Angesichts der umfangreichen und heterogenen Anforderungen wurde angenommen, dass nicht jede einzelne Aufgabe mit gleicher Qualität ausgeführt werden könne

(siehe Müller/Nixdorff/Pirner 2018). Diese Entwicklungen waren mitunter von Befürchtungen und Widerständen begleitet. Es bestand beispielsweise die Sorge, man würde mit der Spezialisierung das in der Bewährungshilfe grundlegende Prinzip der Ganzheitlichkeit einbüßen und die Probanden/-innen auf ihre Probleme und Risiken reduzieren (vgl. Mayer 2018; Müller/Nixdorff/Pirner 2018). In der Überzeugung, Spezialisierung stünde in keinem grundsätzlichen Widerspruch zu Ganzheitlichkeit, sondern sei vielmehr als Ergänzung zu betrachten, und überdies eine notwendige Reaktion auf die komplexer werdenden sozialarbeiterischen Anforderungen, hielt man in Hessen jedoch an den Zielsetzungen fest. Ein Beispiel der Realisierung dieser Zielsetzungen stellt das bereits 2008 implementierte *Sicherheitsmanagements I* (SIMA I) dar, mit dem ein Konzept zur risikoorientierten Betreuung von Personen eingeführt wurde, die aufgrund eines Sexualdelikts verurteilt worden waren (Müller/Nixdorff/Pirner 2018).

Das SIMA I wurde in einem vorangegangenen Evaluationsprojekt untersucht, wobei die Ergebnisse auf eine rückfallpräventive Wirkung der Betreuung hinwiesen (Keßler/Rettenberger 2017), wobei gleichzeitig Maßnahmen zur Qualitätssicherung identifiziert werden konnten (Bockshammer/Rettenberger 2015). Bei der Konzeption des SIMA II orientierte man sich maßgeblich an den Erfahrungen, die im SIMA I gemacht wurden, wobei sich das SIMA II analog zu diesem Fachbereich in die bereits vorhandene Organisationsstruktur einfügen sollte (Müller/Nixdorff/Pirner 2018). Auch in Bezug auf die Evaluation sollte das SIMA II dem Beispiel des SIMA I folgen. Insofern konnte man bei der Evaluation des SIMA II von den Erfahrungen, die im Zuge der Evaluation des SIMA I gemacht worden waren, profitieren und eine methodisch vergleichbare Strategie verfolgen.

Während im SIMA I lediglich wegen Sexualdelikten verurteilte Personen betreut werden, richtet sich die Spezialisierung im SIMA II an eine größere und hinsichtlich ihrer Eintrittsgründe heterogenere Klientel. Mit der Aufnahme von Führungsaufsichtsprobanden/-innen in die Zielgruppe des SIMA II sollten die Befunde aus Untersuchungen der Betreuungssituation dieser Klientel berücksichtigt werden. Im Zuge ihrer bundesweiten Evaluation der reformierten Führungsaufsicht stellten Baur und Kinzig (2015) ein Betreuungsdefizit bei Führungsaufsichtsprobanden/-innen mit negativer Sozialprognose fest. Darüber hinaus sah man bei besonders rückfallgefährdeten und wegen Gewaltdelikten verurteilten Personen strukturell ähnliche Problem- und Bedarfslagen (vgl. Müller/Nixdorff/Pirner 2018). Diese Abwägungen führten zu der nun bestehenden Zusammensetzung der SIMA II-Zielklientel.

Die im vorliegenden Bericht vorgestellten Ergebnisse sollen dem Auftraggeber Erkenntnisse bzgl. der Effektivität (Wirksamkeit) und Effizienz (Kosten-Nutzen-Abwägungen) von Maßnahmen, die im SIMA II umgesetzt wurden, bereitstellen. Dabei wurden insbesondere die prognostische sowie die darauf aufbauende Interventionspraxis evaluiert. Dies soll eine wissenschaftlich fundierte Entscheidungsgrundlage darüber zur Verfügung stellen, ob die im SIMA II implementierten Maßnahmen zur Erreichung der ursprünglich intendierten Zielsetzungen beitragen und wo gegebenenfalls noch Optimierungsmöglichkeiten bestehen könnten.

Die nachfolgenden Kapitel des Berichts sind jeweils auf Basis der drei im Projekt behandelten Fragestellungen in entsprechende Abschnitte unterteilt. Projektteil 1 widmet sich der Evaluation der Risikoeinschätzung bzw. Rückfallprognostik, in Projektteil 2 wird die Wirksamkeit der SIMA II-Interventionen untersucht und Projektteil 3 ergänzt diese quantitativen Ansätze um Interviews mit SIMA II-Mitarbeiter/-innen. Zusätzlich zu diesen projektteilspezifischen Abschnitten enthält der Bericht übergeordnete Abschnitte, die nicht einem bestimmten Projektteil zuzuordnen sind, wie etwa allgemeine Beschreibungen zu den Datenzugängen, deskriptive Merkmale der Gesamtstichprobe und ein übergreifendes und zusammenfassendes Fazit am Ende des Berichts.

2 Fragestellungen

Im Rahmen eines 2-jährigen durch das Hessische Ministerium der Justiz (HMdJ) beauftragten und finanziell geförderten Projekts wurden unterschiedliche Aspekte des SIMA II evaluiert. Die Evaluation umfasste drei Themenbereiche bzw. Fragestellungen, die im Folgenden näher ausgeführt werden.

2.1 Projektteil 1: Güte der Risikoeinschätzung

Um eine risikoorientierte Ressourcenverteilung gemäß dem RNR-Modell (Andrews/Bonta 1994; 2010) zu gewährleisten, setzt der Fachbereich SIMA II routinemäßig zwei Prognoseinstrumente ein. Diese dienen der Zuteilung der Probanden/-innen zu einer von drei Betreuungsstufen zu Beginn der Betreuung. Die Einstufungen gehen wiederum mit unterschiedlichen Betreuungs- und Berichtsintensitäten und -standards einher. Hoch rückfallgefährdete Probanden/-innen erhalten eine intensivere Betreuung als Probanden/-innen mit niedrigem Rückfallrisiko, was sich u.a. in einer höheren Kontakt- sowie Berichtsfrequenz niederschlägt (Hessisches Ministerium der Justiz 2017).

Um eine wissenschaftlich fundierte Einstufung zu gewährleisten, wurde für männliche Probanden, die aufgrund eines Gewaltdelikts verurteilt wurden (für eine genauere Beschreibung dieser Personengruppe siehe Abschnitt 3.2), ein auf empirischem Wege konstruiertes Prognoseinstrument implementiert. Dieses Instrument ist das aus fünf Items bestehende *Screeninginstrument zur Vorhersage des Gewaltrisikos* (SVG-5; Eher u.a. 2012). Die Items des SVG-5 beziehen sich auf das Alter, die Art und Ausprägung möglicher strafrechtlicher Vorbelastungen sowie das Vorliegen psychischer Auffälligkeiten (siehe Anhang A). Die aufsummierten Gesamtwerte liegen zwischen -22 und +16, wobei -22 das niedrigste und +16 das höchstmögliche Risiko für gewalttätige Rückfälle indiziert.

Probanden mit Summenwerten ab +2 werden dem Fachbereich SIMA II zugewiesen, Probanden mit niedrigeren Werten werden von der allgemeinen Bewährungshilfe betreut. Wegen Gewaltdelinquenz verurteilte Probanden/-innen, die unter Führungsaufsicht stehen, werden in jedem Fall im SIMA II betreut. In Tabelle 1 sind die Einstufungen in Abhängigkeit von den SVG-5-Summenwerten dargestellt. Niedrigere Betreuungsstufen (Spalte 3) indizieren hierbei ein höheres Risiko.

Tabelle 1: Betreuungseinstufungen auf Basis des SVG-5

SVG-5-Summenwert	Risikoeinschätzung		Zuständiger Fachbereich	
	Risikokategorie(n) SVG-5	Betreuungsstufe	Probanden/-innen mit Führungsaufsicht	Probanden/-innen ohne Führungsaufsicht
< 2	niedrig/niedrig-moderat/moderat hoch	Stufe 3	SIMA II	Allgemeine Bewährungshilfe
2-8	hoch	Stufe 2	SIMA II	SIMA II
> 8	hoch/sehr hoch	Stufe 1	SIMA II	SIMA II

Der SVG-5 wurde bereits an einer Stichprobe von $n = 307$ aufgrund von Gewaltdelinquenz verurteilten Männern validiert, die zwischen 2001 und 2003 aus der Haft entlassen worden waren. Dabei erzielte das Instrument *Area Under the Curve* (AUC)-Werte von $AUC = .79$ für die Vorhersage des gewalttätigen Rückfalls und $AUC = .82$ für die Vorhersage des gewalttätigen Rückfalls, der zu einer Inhaftierung führte (Eher u.a. 2012). Gemäß internationaler Interpretationsrichtlinien für AUC-Werte können diese Ergebnisse als gute Vorhersageleistungen interpretiert werden (Rice/Harris 2005; generell werden AUC-Werte über .71 als gute, Werte über .63 als moderate und darunterliegende Werte, die das Signifikanzniveau noch erreichen, als geringe Effektstärken eingestuft). Der AUC-Wert bildet die Wahrscheinlichkeit ab, mit der eine zufällig aus der Stichprobe gezogene rückfällige Person einen höheren mit dem interessierenden Instrument festgestellten Risikowert aufweist als eine zufällig aus der Stichprobe gezogene nichtrückfällige Person. Insofern sprechen alle Werte über .50 für eine Klassifikationsleistung des Instruments, die besser ist als eine auf dem Zufall basierende Klassifikationsleistung. Eine angemessene Anwendung des Instruments in der Praxis wurde und wird durch adäquate Schulungsmaßnahmen von Seiten des HMdJ gewährleistet.

Bei allen übrigen Probanden/-innen, also Personen, die wegen anderer (d.h. nichtgewalttätiger) Delikte verurteilt und gleichzeitig der Führungsaufsicht unterstellt wurden, wird die Risikoeinschätzung mittels eines von der Bewährungshilfe entwickelten standardisierten *Erhebungsbogens* vorgenommen. Dieser Erhebungsbogen besteht aus insgesamt zehn Items, die diverse biographische Merkmale (bspw. Schulbildung und Herkunftsfamilie betreffend), ggf. strafrechtliche Vorbelastungen sowie Aspekte der aktuellen Lebenssituation

(bspw. protektives Freizeitverhalten und Tagesstruktur) erfassen (siehe Anhang B). Zur Feststellung des Betreuungsbedarfs werden die Werte der einzelnen Items zu einem Summenwert aufaddiert, der zwischen -1 und 20 liegen kann, wobei 20 den höchsten Betreuungsbedarf indiziert. Die psychometrischen Eigenschaften (Reliabilität, Validität) dieses Erhebungsbogens wurden bislang noch nicht überprüft (Müller/Nixdorff/Pirner 2018). In Tabelle 2 sind die Betreuungseinstufungen in Abhängigkeit von den Summenwerten des Erhebungsbogens dargestellt.

Tabelle 2: Betreuungseinstufungen auf Basis des Erhebungsbogens

Summenwert Erhebungsbogen	Betreuungsstufe	Zuständiger Fachbereich
< 6	Stufe 3	SIMA II
6-8	Stufe 2	SIMA II
> 8	Stufe 1	SIMA II

Im Rahmen des vorliegenden Evaluationsprojekts wurden die Zuverlässigkeit (Reliabilität) und die Aussagekraft bzw. Gültigkeit (Validität) der Risikoeinschätzungen, die im SIMA II mittels SVG-5 bzw. Erhebungsbogen vorgenommen werden, untersucht.

Zusätzlich sollten nach Möglichkeit weitere als relevant erachtete Risikofaktoren dahingehend geprüft werden, ob sie einen zusätzlichen prognostischen Erklärungswert besitzen, der über die bestehende Praxis hinausgeht. Auf diese Weise sollten ggf. Verbesserungsoptionen für den zukünftigen Einsatz der Instrumente empirisch geprüft werden. Als mögliche ergänzende Risikofaktoren kamen kriminologische Variablen (z.B. weitere Altersvariablen oder Merkmale der strafrechtlichen Vorbelastung) oder alternative, aber ähnlich strukturierte Prognoseinstrumente (z.B. die *Offender Group Reconviction Scale* – OGRS 3; Francis/Soothill/Humphreys 2007; Howard u.a. 2009) infrage. Im Hinblick auf den bislang noch nicht evaluierten Erhebungsbogen sollten Möglichkeiten hinsichtlich der Optimierung von Operationalisierungen und/oder des Verzichts auf vorhersageschwache Items identifiziert werden. Zusätzlich sollten bei Bedarf Vorschläge für die Neu- oder Umformulierung von (Ersatz-)Prädiktoren erarbeitet werden.

2.2 Projektteil 2: Kriminalpräventive Wirksamkeit der Betreuung

Ein zentrales Ziel von Bewährungshilfe besteht in der Verhinderung neuerlicher Straftaten. Eine in diesem Sinne wirksame Weiterentwicklung der Betreuungspraxis sollte folglich mit einer Reduktion krimineller Rückfälle der betreffenden Probanden/-innen einhergehen. Aus diesem Grund war die Analyse unterschiedlicher Rückfallparameter der SIMA II-Probanden/-innen zentraler Bestandteil des Projektes. Die konkrete Frage, die es zu beantworten galt, lautete: Werden Personen, die im SIMA II betreut wurden, seltener kriminell rückfällig als vergleichbare Personen, die nicht im SIMA II betreut wurden? Oder mit anderen Worten formuliert: Führte die Implementierung des SIMA II zu einer Verringerung der einschlägigen Rückfälligkeit bei Personen, die bereits zuvor aufgrund von Gewaltdelikten verurteilt wurden, und überdies zu einer Verringerung der allgemein kriminellen Rückfälligkeit bei Führungsaufsichtsprobanden/-innen mit negativer Sozialprognose, die wegen anderer als gewalttätiger Delikte verurteilt wurden?

Ein weiterer zu berücksichtigender Aspekt betraf die Erfassung möglicher differentieller Effekte, insbesondere im Hinblick auf sogenannte (mutmaßliche) Hoch-Risiko-Probanden/-innen, wie z.B. Personen mit vergleichsweise hohem Ausgangsrisiko oder auch hinsichtlich sog. „Vollverbüßer/-innen“ (§ 68f Abs. 1 StGB; diese letztgenannte Gruppe wurde in vorangegangenen Studien nachweislich vergleichsweise häufig rückfällig; vgl. Jehle u.a. 2016). Bei diesen Gruppen sollte die Wirksamkeit der Interventionsmaßnahmen nach Möglichkeit separat untersucht werden, sodass über diese Subgruppen spezifische Aussagen getroffen werden können.

2.3 Projektteil 3: Qualitative Interviews mit Bewährungshelfer/-innen

Letztlich sollte die Evaluation durch qualitativ gewonnene Erkenntnisse erweitert werden. Umgesetzt wurde dies in Form von Befragungen von Bewährungshelfer/-innen, die zum Zeitpunkt der Evaluation im SIMA II tätig waren. Hierdurch sollte sowohl der Stand der Umsetzung der neuen Richtlinien zum Zeitpunkt der Evaluation aus praktischer Perspektive als auch die Akzeptanz der Neuregelungen erfasst und abgebildet werden. Dabei standen folgende Fragen im Zentrum: Wie gut lassen sich die Vorgaben des SIMA II-Erlasses in der

Praxis umsetzen? Welche Erfahrungen haben die im Fachbereich beschäftigten Bewährungshelfer/-innen mit den Prognoseinstrumenten gemacht? Bzgl. welcher Vorgaben sehen sie Optimierungsbedarf bzw. -möglichkeiten? Hat sich durch die Umstrukturierung die reale Betreuungsarbeit maßgeblich verändert?

3 Methodisches Vorgehen

3.1 Datenzugänge

Die Gewinnung der Probandendaten erfolgte über drei unterschiedliche Datenquellen: Über Probandendateien der Bewährungshilfe, das Bundeszentralregister (BZR) und die hessische Mehrländer-Staatsanwaltschafts-Automation (MESTA).

3.1.1 Probandendaten der Bewährungshilfe

Den Ausgangspunkt und die Grundlage für alle weiteren Untersuchungen bildeten Probandendaten der hessischen Bewährungshilfe. Diese wurden entsprechend der Studienanforderungen durch die IT-Stelle der hessischen Justiz aus SoPart¹, der digitalen Dokumentationsanwendung der sozialen Dienste, extrahiert. Neben personenbezogenen Angaben (Alter, Geschlecht, Geburtsdatum, Geburtsort etc.) enthielten diese Probandendaten Informationen zum Anlassdelikt, das die Betreuung im Fachbereich begründete, zu den (straf-)rechtlichen Unterstellungsgründen sowie zu relevanten Betreuungsmerkmalen (z.B. Dauer der Betreuung oder die Fachbereichszuständigkeit). Die Datenextraktion wurde durch das HMdJ autorisiert und koordiniert.

3.1.2 Bundeszentralregister (BZR)

Um das Kriterium der kriminellen Rückfälligkeit abzubilden, wurden Daten aus dem BZR angefordert. Das BZR enthält bundesweite Informationen über sämtliche rechtskräftige Verurteilungen. Die BZR-Datei jeder Person ist in nummerierte Einträge unterteilt, von denen i.d.R. jeder eine rechtskräftige Verurteilung behandelt. Pro Eintrag sind verschiedene Informationen gelistet. Neben dem Datum der Verurteilung sind ggf. das Datum der (letzten) mit dem Urteil assoziierten Tat und ggf. das Datum, zu dem das Urteil rechtskräftig wurde, enthalten. Darüber hinaus enthält jeder Eintrag eine Tatbezeichnung, also eine kurze Erläuterung der begangenen Straftat(en), wobei das im Rahmen dieser Erläuterung erstgenannte Delikt das jeweilige Hauptdelikt darstellt. Des Weiteren enthält jeder Eintrag alle im Rahmen der Verurteilung angewandten gesetzlichen Vorschriften. Dabei kann es sich zum einen um Strafvorschriften

¹ SoPart ist die gemeinsame Plattform für die sozialen Dienste innerhalb und außerhalb des Vollzugs und wird seit 2007 standardmäßig in der Bewährungshilfe in Hessen eingesetzt. SoPart dient der Erfassung, der Pflege, dem Zugriff sowie der Übergabe von Falldaten und unterstützt die Zusammenarbeit zwischen den verschiedenen an stationären und ambulanten Maßnahmen beteiligten Instanzen bzw. Institutionen.

handeln, welche die Art des begangenen Delikts analog zur Tatbezeichnung kennzeichnen, zum anderen werden auch Vorschriften gelistet, die Auskunft über Verfahrenseigenheiten und Nebenfolgen geben, bspw. Vorschriften des Jugendgerichtsgesetzes (JGG) oder der Strafprozessordnung (StPO), die also nicht direkt mit der begangenen Tathandlung in Verbindung stehen müssen.

Der Antrag auf Erteilung von Auskünften aus dem BZR zu wissenschaftlichen Zwecken gem. § 42a BZRG wurde am 29. April 2019 beim Bundesamt für Justiz (BfJ) gestellt. Dem Antrag, der durch das BfJ am 13. Januar 2020 abschließend bewilligt wurde, lag ein – wie für derartige Abfragen üblich – ausführliches Datenschutzkonzept sowie eine Begründung des wissenschaftlichen Interesses bei. Der Export der Auszüge erfolgte am 27. März 2020, sodass dieses Datum als Stichtag (Ende des Nachbeobachtungszeitraums) festgesetzt werden konnte. Aufgrund von zum Teil fehlenden Angaben in der Ausgangspersonendatei (bspw. fehlender Geburtsort) musste eine Bereinigung vorgenommen werden, sodass schließlich für $n = 5\,412$ Personen eine Anfrage gestellt werden konnte. Es wurden Auszüge zu $n = 5\,169$ Personen übermittelt. Für eine Reihe der angefragten Personen konnten aufgrund fehlerhafter oder unvollständiger Personendaten keine Auszüge extrahiert werden. Zu 62 Personen wurde außerdem mitgeteilt, dass diese verstorben seien und insofern keine Auskunft erteilt werden dürfe bzw. könne.

Für die vorliegende Evaluationsstudie wurden pro BZR-Eintrag nur diejenigen gelisteten Vorschriften extrahiert, bei denen es sich um Strafvorschriften der verschiedenen Gesetzesbücher handelte, da die restlichen Vorschriften keinen Aufschluss über die begangene Straftat beinhalteten. Für jeden BZR-Eintrag wurde zunächst das Hauptdelikt identifiziert. Da diese Auswertung aufgrund der Datenmenge automatisiert erfolgen musste, konnten die Hauptdelikte nicht aus den (sehr individuellen und nicht ausreichend standardisierten) Tatbezeichnungen extrahiert werden, sondern mussten basierend auf einer Kategorisierung der gelisteten Strafvorschriften identifiziert werden. Dabei wurde das jeweils schwerste Delikt als das Hauptdelikt klassifiziert. An die Schwere des Delikts wurde sich mithilfe eines Indexes angenähert. Dieser Index wurde basierend auf dem mit dem Delikt assoziierten Strafraumen berechnet. Es wurde eine Liste erstellt, in der für jede in den BZR-Auszügen gelistete Strafvorschrift der Strafraumen bzw. die minimale und die maximale Freiheitsstrafe (in Jahren) dokumentiert wurde, die laut Gesetzestext für das jeweilige Delikt vergeben werden kann. Geldstrafen wurden mit dem Wert 0 kodiert. Lebenslange Freiheitsstrafen (d.h. Freiheitsstrafen, die mind. 15 Jahre andauern und

anschließend die Möglichkeit zur Aussetzung zur Bewährung besteht) wurden mit dem Wert 15.1 kodiert, um eine Abgrenzung zum Höchstmaß von 15 Jahren zu gewährleisten. Für die Berechnung des Indexes wurden das jeweilige Mindest- und Höchstmaß (Freiheitsstrafe in Jahren) addiert². Diese Summe indiziert folglich die Schwere des Delikts, wobei höhere Indexwerte schwerere Delikte kennzeichnen. Basierend auf diesem Index wurde pro Eintrag dasjenige Delikt mit dem höchsten Indexwert als Hauptdelikt klassifiziert. Waren in einem Eintrag mehrere Delikte mit demselben Indexwert gelistet, wurde der in der Auflistung erstgenannte Tatbestand als Hauptdelikt verwendet. Die Plausibilität dieser Strategie wurde geprüft, indem abgeglichen wurde, ob auf diese Art dasjenige Delikt als Hauptdelikt identifiziert wurde, welches auch in der Tatbezeichnung als Hauptdelikt geführt wurde. Bei einer Überprüfung von 25 Fällen wurden auf diese Weise 23 korrekte Identifizierungen des Hauptdelikts (Trefferquote: 92 %) erzielt.

3.1.3 Mehrländer-Staatsanwaltschafts-Automation (MESTA)

Darüber hinaus wurden nicht nur rechtskräftige Verurteilungen, sondern auch gegen die jeweiligen Probanden/-innen eingeleitete Strafverfahren als Maß des kriminellen Rückfalls herangezogen. Zu diesem Zweck wurden bei der Staatsanwaltschaft Frankfurt am Main Daten aus der MESTA angefordert. Die MESTA ist eine von den hessischen Amtsanwaltschaften, Staatsanwaltschaften und der Generalstaatsanwaltschaft genutzte digitale Anwendung, die der automatisierten Verfahrensverwaltung dient. Sie enthält Informationen darüber, ob innerhalb von Hessen ein Ermittlungsverfahren gegen eine Person eingeleitet wurde und um welchen Straftatbestand es sich dabei handelte, und zwar unabhängig davon, ob das Verfahren später eingestellt oder die Person tatsächlich rechtskräftig verurteilt wurde. Im Vergleich zu Verurteilungsraten ist dieses Kriterium krimineller Rückfälligkeit als sensitiver anzusehen (Craig/Beech/Browne 2008; Schmucker 2004). Für die Untersuchungsstichprobe wurden Auszüge aus der MESTA ab dem 01.01.2014 (d.h. zum Beginn des Untersuchungszeitraums der Vergleichsgruppe, siehe dazu Abschnitt 3.2) angefordert. Autorisiert wurde die Extraktion durch die Staatsanwaltschaft Frankfurt am Main, ausgeführt wurde sie von der IT-Stelle der hessischen Justiz. Der Antrag über Erteilung von Auskünften aus der MESTA zu wissenschaftlichen Zwecken wurde am 29. April 2019 gestellt. Es wurden Auskünfte zu $n = 7\,435$ Personen angefordert. Der Antrag wurde am 08.05.2019 bewilligt, die Extraktion der Daten erfolgte am

² Exemplarische Berechnung des Index für Raubdelikte (§ 249 StGB): Mindeststrafe = 0.5 Jahre, Höchststrafe = 15 Jahre → Deliktschwereindex für die vorliegende Studie = 15.5

17.06.2019, der folglich den Stichtag darstellte. Alle Ermittlungsverfahren vor 2014, nach dem Stichtag oder außerhalb Hessens wurden nicht erfasst, da diesbezüglich in MESTA keine Datenspeicherung stattfindet.

Die extrahierten MESTA-Daten enthielten Eintragungen für insgesamt $n = 3\,845$ Probanden/-innen (61.2 %) der Untersuchungsstichprobe (zur Unterscheidung zwischen Behandlungs- und Vergleichsstichprobe siehe Abschnitt 3.2). Wenn für eine Person innerhalb des abgefragten Zeitraums in Hessen keinerlei Strafanzeige eingegangen war, hatte dies zur Folge, dass für die entsprechende Person keine Daten geliefert wurden. Dementsprechend wurden Personen, für die keine Daten übermittelt wurden, als nicht rückfällig eingestuft. Einschränkend zu erwähnen ist jedoch, dass auf Basis der MESTA-Daten nicht verlässlich unterschieden werden kann, ob es sich in solchen Fällen tatsächlich um das Ausbleiben von strafbarem Handeln oder lediglich um fehlende Werte, bspw. aufgrund von Problemen bei der Datenabfrage, handelte. Das Fehlen von MESTA-Daten kann weiterhin darauf zurückzuführen sein, dass eine Person in ein anderes Bundesland verzogen oder verstorben ist.

In den folgenden Analysen werden MESTA- und BZR-Daten als getrennte Rückfallindikatoren behandelt und die jeweiligen Ergebnisse getrennt voneinander dargestellt.

3.2 Stichprobenbildung

Aus der Extraktion aus SoPart resultierte zunächst eine Stichprobe von $N = 9\,099$ verurteilten und von der hessischen Bewährungshilfe betreuten Probanden/-innen. Dabei handelte es sich zum einem um Personen, die aufgrund eines Gewaltdelikts³ verurteilt und deren Strafen gemäß § 56 StGB oder § 57

³ Folgende Indexdelikte werden als Gewaltdelikte im Sinne des SIMA II-Eingangskriteriums definiert (Hessisches Ministerium der Justiz 2017): § 211 StGB (Mord), § 212 StGB (Totschlag), § 223 StGB (Körperverletzung), § 224 StGB (Gefährliche Körperverletzung), § 225 StGB (Misshandlung von Schutzbefohlenen), § 226 StGB (Schwere Körperverletzung), § 227 StGB (Körperverletzung mit Todesfolge), § 231 StGB (Beteiligung an einer Schlägerei), § 239 StGB (Freiheitsberaubung), § 239a StGB (Erpresserischer Menschenraub), § 239b StGB (Geiselnahme), § 240 StGB Abs. 4 (Nötigung im besonders schweren Fall), § 249 StGB (Raub), § 250 StGB (Schwerer Raub), § 251 StGB (Raub mit Todesfolge), § 252 StGB (Räuberischer Diebstahl), § 255 StGB (Räuberische Erpressung), § 306a StGB (Schwere Brandstiftung), § 306b StGB (Besonders schwere Brandstiftung) und § 306c StGB (Brandstiftung mit Todesfolge).

StGB zur Bewährung (rest)ausgesetzt worden waren. Zum anderen handelte es sich um aus der Haft oder dem Maßregelvollzug entlassene Personen, die kraft richterlicher Anordnung (§ 68 Abs. 1 StGB) oder kraft Gesetzes (§§ 67d Abs. 2, 67d Abs. 4-6, 68f Abs. 1 StGB) der Führungsaufsicht unterstellt worden waren⁴. Diese Gruppen wiesen eine gemeinsame Schnittmenge auf, da der Führungsaufsicht unterstellte Personen auch Verurteilungen aufgrund von Indexdelikten aus der Kategorie der Gewaltdelikte aufweisen konnten.

Die im Folgenden als *Behandlungsgruppe* bezeichnete Gruppe umfasste Probanden/-innen, welche *nach* der Implementierung von SIMA II von der hessischen Bewährungshilfe betreut wurden, bei denen also das neue Prognose- und Interventionskonzept bereits angewandt wurde. Berücksichtigt wurden Personen, welche die o.g. Einschlusskriterien aufwiesen und zwischen dem 01.07.2017 und dem 30.06.2018 in Betreuung gewesen waren⁵. Alle Führungsaufsichtsprobanden/-innen der Behandlungsgruppe waren zumindest zeitweise im SIMA II in Betreuung. Wegen Gewaltdelinquenz verurteilte Probanden/-innen (ohne Führungsaufsicht) waren – abhängig von ihrem ermittelten Rückfallrisiko – entweder (zumindest zeitweise) vom SIMA II oder der allgemeinen Bewährungshilfe betreut worden. Insofern enthält die Behandlungsgruppe nicht nur SIMA II-Probanden/-innen, da für Teile der Evaluation auch an die allgemeine Bewährungshilfe abgegebene Probanden/-innen relevant waren.

Die im Folgenden als *Vergleichsgruppe* bezeichnete Gruppe bildeten Probanden/-innen, welche *vor* der Implementierung von SIMA II von der hessischen Bewährungshilfe betreut wurden. Für den Vergleichsprobandenpool wurden alle Personen berücksichtigt, welche die o.g. Einschlusskriterien (Unterstellungsgründe bzw. Gewaltindexdelikt) aufwiesen und zwischen dem 01.01.2014 und dem 31.12.2014 in Betreuung gewesen waren.

Eine Reihe von Ausschlusskriterien führte zu einer Reduktion der anfänglichen Stichprobe. Die Ausgangsdaten wurden zunächst um Personen bereinigt, die

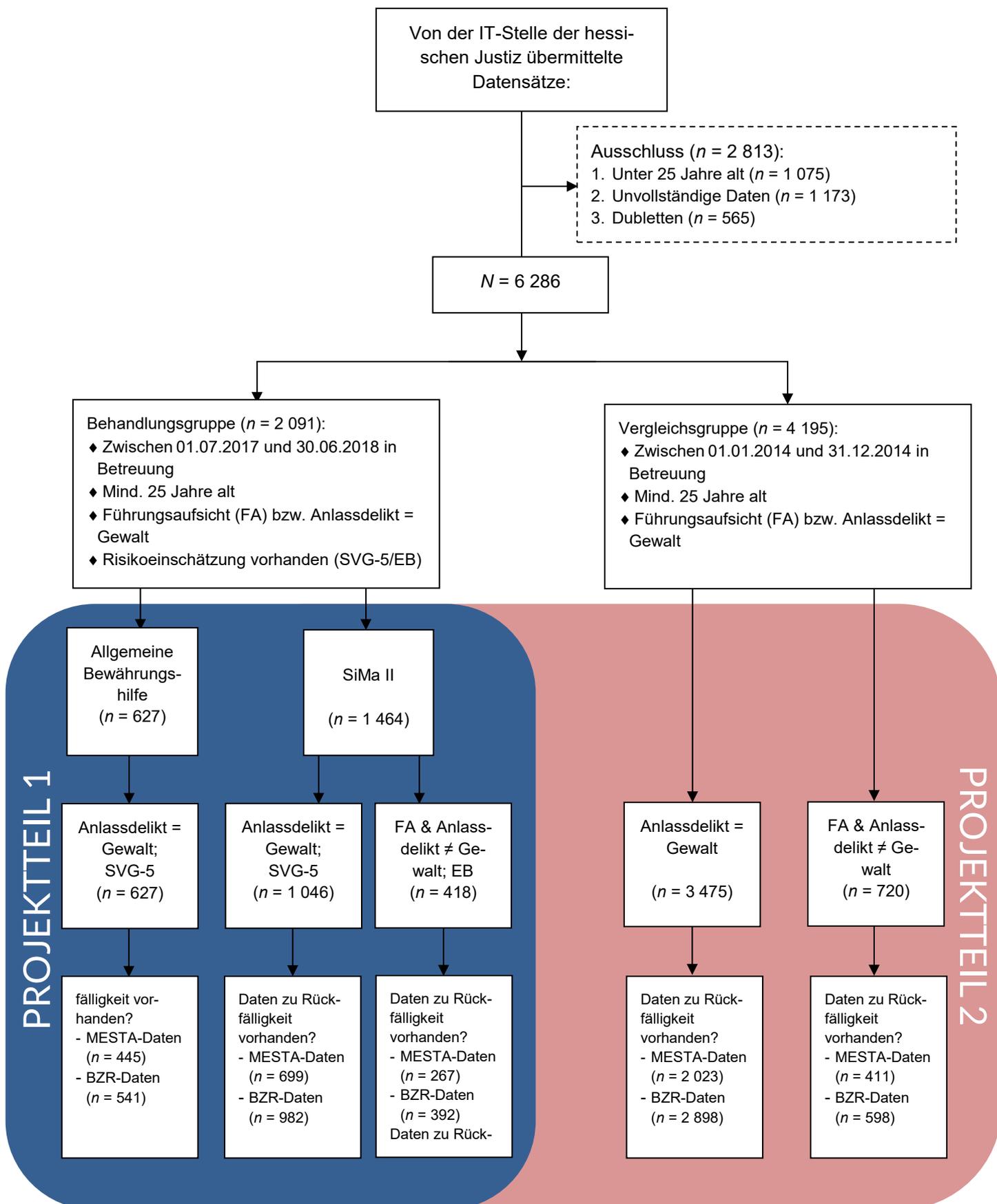
⁴ Davon ausgeschlossen waren Personen, die wegen (mindestens) eines Sexualdelikts verurteilt wurden, da diese vom Fachbereich SIMA I betreut werden.

⁵ Der Erlass zum SIMA II (Hessisches Ministerium der Justiz 2017) sah vor, dass Probanden, die zum Zeitpunkt der Implementierung bereits in Betreuung waren, an den neuen Fachbereich abgegeben wurden, sofern sie zur Zielgruppe (Indexdelikt und hohes Rückfallrisiko und/oder Führungsaufsicht) gehörten. Eine Ausnahme für diese Regelung bildeten Personen, deren restliche Betreuungszeit zum Zeitpunkt der Implementierung weniger als ein Jahr betrug. Aus diesem Grund wurden Probanden, deren Betreuung vor dem 01.02.2018 endete, aus unseren Untersuchungen ausgeschlossen.

zum Beginn ihres Nachbeobachtungszeitraums unter 25 Jahre alt gewesen waren ($n = 1\,173$), da dies das Mindestalter der im SIMA II betreuten Personen darstellt. Anschließend wurden Personen aus der Behandlungsgruppe entfernt, für die keine Risikoeinschätzung vorlag ($n = 1\,075$; siehe auch Abschnitt 3.3.2 bzw. 4.1.2). Es wurden darüber hinaus Dubletten im Datensatz identifiziert, d.h. Personen, die sowohl in der Behandlungs- als auch in der Vergleichsgruppe aufgeführt waren, weil sie in beiden definierten Untersuchungszeiträumen von der hessischen Bewährungshilfe betreut worden waren ($n = 565$). Diese Dopplungen wurden beseitigt; um Verzerrungen zu vermeiden, wurde ein Teil dieser Personen (erste Hälfte der Gruppe, $n = 283$, geordnet nach Nachnamen) der Behandlungs- und der Rest (zweite Hälfte der Gruppe, $n = 282$, geordnet nach Nachnamen) der Vergleichsgruppe zugeordnet. Die verbleibenden $n = 6\,286$ Probanden/-innen bildeten die Behandlungs- und die Vergleichsgruppe.

Abbildung 1 zeigt schematisch die Stichprobenbildung für die vorliegende Studie. Die blau unterlegten Felder markieren die Teile der Stichprobe, welche für die Untersuchung der Risikoeinschätzungen (Projektteil 1) analysiert wurden. Die rot unterlegten Felder markieren die Teile der Stichprobe, die bei der Untersuchung der Effektivität der SIMA II-Interventionspraxis (Projektteil 2) analysiert wurden. Für Projektteil 2 wurde folglich ein Teil der Behandlungsgruppe ausgeschlossen: Dabei handelte es sich um wegen Gewaltdelinquenz verurteilte Personen, die aufgrund eines niedrigen SVG-5-Ratings nicht vom SIMA II, sondern durch die allgemeine Bewährungshilfe betreut wurden.

Abbildung 1: Schematische Darstellung der Stichprobenbildung



3.3 Projektteil 1: Güte der Risikoeinschätzung

Die Qualität von Risikoeinschätzungen gilt dann als gegeben, wenn die verwendeten Instrumente – in diesem Fall der SVG-5 und der Erhebungsbogen – zuverlässige und treffsichere Voraussagen machen, d.h. wenn die Risikoeinschätzungen mit den tatsächlichen (Nicht-)Rückfällen überzufällig häufig korrespondieren. Zuverlässigkeit (psychometrisch: Reliabilität) liegt dann vor, wenn das Instrument – unter Voraussetzung derselben Ausgangssituation – bei wiederholter Messung identische oder zumindest vergleichbare Ergebnisse liefert. Dies ist eine notwendige, aber keine hinreichende Voraussetzung für die Gültigkeit (psychometrisch: Validität) der Instrumente. Diese ist erst dann gegeben, wenn die durch das Instrument ermittelten Werte (in diesem Fall: Risikoeinschätzungen) in hohem Maße mit objektiven und praktisch relevanten Kriterien des gemessenen Konstrukts (in diesem Fall: Rückfälligkeit) zusammenhängen.

Um Reliabilität und Validität der beiden vom SIMA II eingesetzten Prognoseinstrumente untersuchen zu können, wurden SVG-5- und Erhebungsbogendaten bei den Mitarbeiter/-innen der Bewährungshilfe Hessen angefordert (siehe Anhang C). Als Koordinationsschnittstelle dieser Abfragen dienten die SIMA II-Fachbereichsleiter/-innen (in Ausnahmefällen deren Vertretung). Da Risikoeinschätzungen erst seit der Implementierung von SIMA II vorgenommen wurden, war für diesen Teil des Projektes nur die Behandlungs-, nicht aber die Vergleichsgruppe relevant.

3.3.1 Reliabilität

Eine Möglichkeit des Nachweises der Reliabilität von (Prognose-)Instrumenten besteht in der Untersuchung der Beurteilerübereinstimmung (sog. Interrater-Reliabilität). Diese Reliabilitätsform ist im Bereich der Prognoseforschung die am häufigsten untersuchte. Ein Instrument gilt als reliabel, wenn verschiedene Anwender/-innen des Instruments bei gleichen Voraussetzungen (Proband/-in mit spezifischen Merkmalsausprägungen) zu möglichst gleichen oder zumindest vergleichbaren Einschätzungen kommen.

Um diese Maßgabe zu überprüfen, wurde an jeden Fachbereich des SIMA II die Bitte adressiert, doppelte Risikoeinschätzungen vorzunehmen. Das bedeutet, dass sich jeweils (mindestens) zwei Mitarbeiter/-innen jedes Fachbereichs bereit erklärten, eine Reihe von Probanden/-innen unabhängig voneinander einzuschätzen (siehe Anhang C). Pro Instrument (SVG-5 und Erhebungsbogen) sollten in jedem Fachbereich mindestens sechs Probanden/-innen auf diese Weise dop-

pelt bewertet werden. Es wurden keine Vorgaben hinsichtlich der Auswahl der Probanden/-innen gemacht, sondern lediglich darum gebeten, durch die Einschätzungen eine möglichst große Bandbreite möglicher Fälle abzudecken. Sofern möglich, sollten also weder ausschließlich Personen mit mutmaßlich besonders hohem, noch ausschließlich Personen mit mutmaßlich besonders niedrigem Rückfallrisiko berücksichtigt werden. Die Mitarbeiter/-innen wurden gleichzeitig darüber informiert, dass die Auswertungsergebnisse nicht einzelnen Personen oder Dienststellen zugeordnet werden würden, sondern ausschließlich in aggregierter Form für gruppenstatistische Auswertungen von Interesse seien.

Für die einzelnen Items wurde als statistisches Maß der Beurteilerübereinstimmung *Cohen's Kappa* (K) bestimmt. Im Falle von Items mit mehr als zweistufigem Antwortformat wurde das *gewichtete Cohen's Kappa* (K_w) bestimmt (vgl. Wirtz/Kutschmann 2007). Cohen's Kappa eignet sich zur Berechnung der Übereinstimmung bei kategorialen Variablen und erfasst das relative Ausmaß übereinstimmender Beurteilungen unter Berücksichtigung der Wahrscheinlichkeit für zufällig zustande gekommene Übereinstimmungen. Für die Summenwerte der Instrumente, für die Intervallskalenniveau angenommen werden konnte, wurden Intraklassen-Korrelationskoeffizienten (ICC) berechnet (vgl. Landers 2015). Diese geben Auskunft darüber, wie stark die Gesamturteile der Beurteiler/-innen zusammenhängen.

3.3.2 Validität

Die Aussagekraft bzw. Gültigkeit (prädiktive oder auch prognostische Validität) der Instrumente kann dann als gegeben eingestuft werden, wenn die mittels SVG-5 und Erhebungsbogen festgestellten Rückfallrisiken und die daraus resultierenden Einstufungen einen signifikanten Zusammenhang mit der tatsächlichen kriminellen Rückfälligkeit aufweisen. Deshalb wurden für jenen Teil der Gesamtstichprobe, der nach Implementierung von SIMA II in Betreuung war ($n = 4\,361$), die entsprechenden Risikoeinschätzungen von den jeweils (zuletzt) zuständigen Dienststellen der hessischen Bewährungshilfe angefordert.

Für $n = 2\,091$ Personen der Gesamtstichprobe (exklusive Dopplungen oder nicht zuzuordnenden Daten) wurden Einschätzungen übermittelt, die vom SIMA II mithilfe des SVG-5 bzw. des Erhebungsbogens vorgenommen worden waren. Dieser Teil der Stichprobe bildete im weiteren Verlauf die Behandlungsgruppe (siehe **Fehler! Verweisquelle konnte nicht gefunden werden.**). Die Überprüfung der Validität der Instrumente konnte für solche Personen vorgenommen wer-

den, für die einerseits eine Risikoeinschätzung vorlag und für die andererseits Rückfalldaten vorhanden waren.

Um den Zusammenhang zwischen den Einschätzungen und der tatsächlichen Rückfälligkeit zu untersuchen, wurden Auszüge aus dem Bundeszentralregister (BZR) und der Mehrländer-Staatsanwaltschafts-Automation (MESTA) herangezogen (siehe dazu Abschnitt 3.1.2 und 3.1.3). Die Anzahl der Personen, für die diese Validitätsanalysen vorgenommen werden konnten, unterschied sich in Abhängigkeit von der Rückfallquelle, da aus der BZR-Abfrage andere Daten und Fallzahlen resultierten als aus der MESTA-Abfrage. Für die Personengruppe, anhand derer der Zusammenhang zwischen den Vorhersagen der Instrumente und neuerlichen strafrechtlichen Verurteilungen (BZR-Einträge) untersucht wurde ($n = 1\,845$), betrug die Nachbeobachtungszeit (Zeit ab Beginn der Nachbeobachtung bis zum Zeitpunkt des Datenexports) $M = 2.47$ Jahre ($SD = 0.79$ Jahre). Für diese Personengruppe ($n = 2\,295$), anhand derer der Zusammenhang zwischen den Vorhersagen der Instrumente und neuerlichen Strafanzeigen (MESTA-Einträge) untersucht wurde, betrug die Nachbeobachtungszeit (Zeit ab Beginn der Nachbeobachtung bis zum Zeitpunkt des Datenexports) $M = 1.68$ Jahre ($SD = 0.81$ Jahre). Wenn für eine Person, deren MESTA-Daten abgefragt worden waren, keinerlei Anzeige seit dem 01.01.2014 dokumentiert war, wurde dieser Fall als nichtrückfällig eingestuft.

Auswertung von Rückfallereignissen

Um potentielle Rückfälle einer Person festzustellen, mussten alle neuerlichen Anzeigen (MESTA) bzw. rechtskräftigen Verurteilungen (BZR) ab Beginn des Nachbeobachtungszeitraumes berücksichtigt werden. Der Beginn des Nachbeobachtungszeitraumes orientierte sich bei Projektteil 1 (der Untersuchung der Validität der Risikoeinschätzungen) an dem Datum, an dem die Risikoeinschätzung vom Fachbereich SIMA II vorgenommen worden war. In den Fällen, in denen zwar Informationen zur Risikoeinschätzung, jedoch nicht zum Zeitpunkt der Einschätzung vorlagen, wurde stattdessen das Datum des (frühesten) Beginns der Betreuung im SIMA II als Beginn des Nachbeobachtungszeitraums festgelegt.

Vergleich mit alternativen Prädiktoren

Um Optimierungspotentiale hinsichtlich der Risikoeinschätzung im SIMA II zu identifizieren, wurden die Vorhersageleistungen von SVG-5 und Erhebungsbogen mit der Vorhersageleistung eines anderen Prognoseinstrumentes zur Vorhersage krimineller Rückfälle verglichen; dabei kam die im britischen Justiz-

vollzug etablierte *Offender Group Reconviction Scale, Version 3* (OGRS 3; Francis/Soothill/Humphreys 2007; Howard u.a. 2009) zum Einsatz, deren Aufbau und psychometrische Eigenschaften an anderer Stelle des Berichts (Abschnitt 3.4.1) genauer beschrieben werden.

3.4 Projektteil 2: Kriminalpräventive Wirksamkeit der Betreuung

Die zentrale Frage, die es im Zuge der Evaluationsstudie zu beantworten galt, war, inwieweit die stärkere Risikoorientierung und Spezialisierung innerhalb der Bewährungshilfe, die durch die Implementierung des SIMA II realisiert wurde, zu einer Steigerung der Effektivität der Betreuung führte. Wenn das neue Konzept sich besser als das vorherige Vorgehen eignet, ehemals straffälligen Personen zu einem straffreien Leben zu verhelfen, sollte Folgendes gegeben sein: Personen, die im SIMA II betreut wurden, werden signifikant seltener rückfällig als vergleichbare Personen (mit ähnlicher Risikokonstellation), die nicht im SIMA II betreut wurden.

Um dieser Frage nachzugehen, mussten SIMA II-Probanden/-innen mit Personen verglichen werden, die nicht durch das SIMA II betreut wurden, sich ansonsten jedoch nicht von den SIMA II-Probanden/-innen unterschieden. Die bevorzugte Methode, um kausale Zusammenhänge zwischen einer Intervention (SIMA II-Betreuung) und einem Kriterium (kriminelle Rückfälligkeit) nachzuweisen, ist ein randomisiert-kontrolliertes Studiendesign, bei dem Probanden/-innen der interessierenden Klientel randomisiert (d.h. zufällig) auf zwei Gruppen aufgeteilt werden (z.B. Farrington u.a. 2002): Dabei erhält Gruppe 1, die Experimentalgruppe (auch: Behandlungsgruppe) genannt wird, die Intervention, deren Wirkung untersucht werden soll; Gruppe 2, als Kontrollgruppe bezeichnet, erhält sie hingegen nicht (oder lediglich eine Placebo-Intervention). Unterschiede zwischen den Gruppen bzgl. des interessierenden Kriteriums können dann kausal bei ausreichend großer Stichprobe auf den Interventionseffekt zurückgeführt werden. Aus ethischen, resourcentechnischen und rechtlichen Gründen war ein solches Vorgehen im vorliegenden Fall nicht realisierbar. Ein alternatives Vorgehen, welches ebenfalls die Aufteilung in Behandlungs- und Kontrollgruppe vorsieht, jedoch die genannten Probleme randomisiert-kontrollierter Studiendesigns umgeht, besteht in der Verwendung eines sog. *Matching*-Verfahrens.

3.4.1 Bildung der Kontrollgruppe mittels Matching

Im Gegensatz zum randomisiert-kontrollierten Studiendesign wird beim Matching die Aufteilung in Behandlungs- und Kontrollgruppe nicht zufällig, sondern systematisch und nachträglich (*ex-post-facto*) vorgenommen. Bei *ex-post-facto*-Designs besteht allerdings das Problem potentieller Störvariablen, die das Kriterium beeinflussen könnten. Dieses Problem kann umgangen werden, wenn jeder Person der Behandlungsgruppe ein bzgl. relevanter Variablen äquivalenter Zwilling aus einem Pool potentieller Vergleichsprobanden/-innen zugewiesen wird (vgl. z.B. Bacher 2002; Gangl 2010).

Im vorliegenden Fall dienten Personen, die vor der Implementierung von SIMA II von der hessischen Bewährungshilfe betreut worden waren, jedoch ansonsten die Charakteristika der SIMA II-Klientel aufwiesen, als Vergleichsgruppe für die SIMA II-Probanden/-innen. Für die statistische Umsetzung des Matchings wurde auf die Methode des sog. *Propensity Score Matchings* zurückgegriffen. Beim Propensity Score Matching wird unter Verwendung regressionsanalytischer Methoden für alle Personen der Stichprobe ein Wahrscheinlichkeitswert, der sog. Propensity Score, gebildet. Im Rahmen des Verfahrens wird zunächst basierend auf den Matching-Variablen eine logistische Regression durchgeführt, bei der die Gruppenzugehörigkeit zur Behandlungs- bzw. Vergleichsgruppe die abhängige Variable darstellt. Das Ergebnis dieser Regressionsanalyse ist der Propensity Score, der demnach alle rückfallrelevanten Variablen zu einem Wert integriert, welcher die Wahrscheinlichkeit angibt, mit der eine Person der Behandlungsgruppe angehört (Austin 2011; Thoemmes 2012). Basierend auf diesem Propensity Score wird für jede Person ein Abgleich mit den Personen der Vergleichsgruppe durchgeführt und ein passender Zwilling mit möglichst ähnlichem Propensity Score ausgewählt und zugeordnet. Die Einstufung als (noch) ausreichend ähnlich, wird durch die Abgleichstoleranz festgesetzt. Diese setzt die zu tolerierende Differenz zwischen den Propensity Scores einer Person der Behandlungsgruppe und einer Person der Vergleichsgruppe (dem potentiellen Matching-Zwilling), also die Unterschiedlichkeit in Bezug auf die betrachteten Matching-Variablen, fest. Um Informationsverluste zu vermeiden, wurde angestrebt, möglichst jeder Person aus der Behandlungsgruppe eine Person aus dem Vergleichsprobanden-Pool zuzuteilen. Personen, für die kein geeigneter Zwilling gefunden werden kann, müssen aus den Analysen ausgeschlossen werden. Dieses Ziel sollte allerdings nicht auf Kosten der Vergleichbarkeit der beiden Gruppen erreicht werden, weswegen die Abgleichstoleranz in jedem Fall möglichst gering ausfallen sollte. Für das Matching-Prozedere standen für die

vorliegende Studie 1 339 SIMA II-Probanden/-innen 3 223 mögliche Matching-Zwillinge in der Vergleichsgruppe gegenüber.

Auswahl und Erhebung der Matching-Variablen

Das Matching musste auf Basis von Variablen erfolgen, die nachweislich das Rückfallrisiko beeinflussen. Dazu zählen Alters- und Geschlechtsvariablen sowie Angaben zur kriminellen Vorgeschichte und dem Anlassdelikt: Eine Vielzahl von Studien konnte belegen, dass Männer im Vergleich zu Frauen häufiger (allgemein und gewalttätig) rückfällig werden (vgl. z.B. Fazel u.a. 2016; Jehle u.a. 2016). In Bezug auf das Alter zeigte sich in der Vergangenheit mehrfach, dass die Wahrscheinlichkeit eines allgemeinen (Fazel u.a. 2016; Jehle u.a. 2016) und eines gewalttätigen Rückfalls (Fazel u.a. 2016; Harris/Rice 2007) mit zunehmendem Alter abnimmt. Das Alter bei der ersten Verurteilung hängt ebenfalls mit dem Rückfallrisiko zusammen, d.h. je jünger Personen zum Zeitpunkt ihrer ersten Verurteilung waren, desto mehr sind sie gefährdet, erneut Straftaten zu begehen (Harris/Rice 2007; Långström/Grann 2002). Vergleichbare Zusammenhänge zeigten sich auch bei aufgrund von Gewaltdelikten verurteilten Personen (Piquero u.a. 2015).

Darüber hinaus belegten Studien, dass die Art des Indexdeliktes mit der Rückfallwahrscheinlichkeit in Zusammenhang steht (vgl. z.B. Cottle/Lee/Heilbrun 2001). So beeinflusst es das Rückfallrisiko einer Person, ob bzw. wie viele Gewaltstraftaten bereits zuvor von ihr begangen wurden: Personen mit Gewaltindexdelikten haben ein erhöhtes Rückfallrisiko im Vergleich zu Personen mit anderen Indexdelikten, zusätzlich wirkten sich Vorstrafen aufgrund von Gewaltdelinquenz risikosteigernd aus (Fazel u.a. 2016). Demgegenüber haben bspw. Personen, die wegen Sexualstraftaten verurteilt wurden, vergleichsweise niedrige Rückfallquoten (Rehn u.a. 2001; Rettenberger u.a. 2015; Spicer/Glicksman 2004). In einer metaanalytischen Studie von Cottle, Lee und Heilbrun (2001) erwies sich die kriminelle Vorgeschichte, u.a. die Anzahl früherer Verurteilungen, als stärkster Prädiktor neuerlicher Straftaten.

Systematisch Berücksichtigung finden die genannten Faktoren in der *Offender Group Reconviction Scale - Version 3* (OGRS 3; Francis/Soothill/Humphreys 2007; Howard u.a. 2009; National Offender Management Service 2009), der revidierten Version der OGRS (Copas/Marshall 1998). Bei der OGRS 3 handelt es sich um ein Prognoseinstrument, das in dieser revidierten Ausführung seit 2008 in England und Wales routinemäßig in der Bewährungshilfe eingesetzt wird. Dabei wird das Risiko straffällig gewordener Personen für allgemeine kriminelle Rückfälle während eines 1- bzw. 2-jährigen Nachbeobachtungszeitraums nach der

Erledigung freiheitsentziehender Strafen oder dem In-Kraft-Treten ambulanter Maßnahmen (bspw. Bewährung mit oder ohne Führungsaufsicht) ermittelt. Zur Feststellung des Risikos integriert die OGRS 3 die folgenden sechs Risikofaktoren: *Geschlecht, Alter bei der letzten Verurteilung, Alter zu Beginn des Prognosezeitraum, Anzahl früherer rechtskräftiger Verurteilungen, Alter bei der ersten Verurteilung* sowie das *aktuelle Hauptdelikt*. Als Hauptdelikt ist jenes Delikt zu werten, für das im Verfahren das höchste Strafmaß festgesetzt wurde. Zur Klassifikation des aktuellen Hauptdelikts werden zwanzig Deliktkategorien vorgegeben⁶. Der individuelle Risikowert wird ermittelt, indem die sechs Risikofaktoren in einem ordinal-logistischen Regressionsmodell mit variierender Gewichtung verrechnet werden. Die Gewichtung erfolgt mithilfe vorgegebener Kategorien, für die – basierend auf einer umfangreichen Normierungsstichprobe von über 79 000 Personen aus dem Vereinigten Königreich (Francis/Soothill/Humphreys 2007) – unterschiedliche Regressionsgewichte zur Verfügung gestellt werden. Der resultierende OGRS 3-Wert ist als Wahrscheinlichkeit zu interpretieren, mit der die Person innerhalb von einem bzw. zwei Jahren erneut rechtskräftig verurteilt wird.

Obwohl die OGRS 3 lediglich auf statische Faktoren zurückgreift, wurden in vergangenen empirischen Studien hohe Vorhersageleistungen belegt. Howard u.a. (2009) erzielten bei ihrer Normierungsstichprobe von über 79 000 Personen einen AUC-Wert von .80 für die allgemeine Rückfälligkeit, was als starker Effekt interpretiert werden kann (ab einem AUC-Wert von .71 kann von einer hohen Vorhersageleistung ausgegangen werden; Rice/Harris 2005). Auch die Vorgängerversionen der OGRS 3 erwiesen sich bereits als valide im Sinne einer treffsicheren Vorhersage neuerlicher Delinquenz. So erzielte die OGRS 2 bei 1 400 Männern, die wegen Gewalt- oder Sexualdelinquenz verurteilt worden waren, einen AUC-Wert von .79 für die Vorhersage der allgemeinen Rückfälligkeit. Darüber hinaus war sie anderen ebenfalls untersuchten Prognoseinstrumenten hinsichtlich der Vorhersageleistung überlegen (Coid u.a. 2007).

Damit stellt die OGRS 3 ein ökonomisch anwendbares und gleichzeitig sehr vorhersagestarkes Instrument dar. Für ihre Anwendung bedarf es – ähnlich wie im

⁶ 1. Gewalt, 2. Raub, 3. Straftaten gegen die öffentliche Ordnung, 4. Sexualdelikte (Opfer ist ein Kind), 5. Sexualdelikte (Opfer ist kein Kind), 6. Zuhälterei/Prostitution, 7. Hausfriedensbruch, 8. Sonstige Einbrüche, 9. Diebstahl, 10. Hehlerei, 11. Betrug/Fälschung, 12. Flucht/Weisungsverstöße, 13. KFZ-Diebstahl, 14. KFZ-Aufbruch, 15. Verkehrsdelikte, 16. Trunkenheit am Steuer, 17. Sachbeschädigung, 18. Drogenhandel/-produktion, 19. Drogenbesitz, 20. Sonstige Delikte.

Fall des SVG-5 – lediglich der Informationen aus dem BZR. Zudem hat die OGRS 3 den Vorteil, nicht auf eine spezielle Klientel (bspw. nur wegen Gewalt- oder Sexualdelinquenz verurteilte oder nur männliche Personen) beschränkt zu sein und lässt sich damit bei der vergleichsweise heterogenen SIMA II-Klientel gut anwenden. Aus diesem Grund wurde die OGRS 3 ausgewählt, um risikoäquivalente Zwillingspaare in der Behandlungs- und der Vergleichsgruppe zu identifizieren und auf diese Weise eine geeignete Kontroll- bzw. Vergleichsgruppe zu bilden.

Für alle Probanden/-innen der Stichprobe, für die BZR-Daten vorlagen, wurden deshalb retrospektiv OGRS 3-Werte ermittelt. Um die *Anzahl früherer Verurteilungen* festzustellen, wurden alle dem aktuellen Hauptdelikt vorausgegangenen BZR-Einträge gezählt, die eine Straftat erfassten⁷. Als das *aktuelle Hauptdelikt* (bzw. Index- oder Anlassdelikt) kamen nur BZR-Einträge vor dem Beginn des Nachbeobachtungszeitraumes des/der jeweiligen Probanden/-in infrage⁸. Bei Projektteil 2 (der Untersuchung der Wirksamkeit SIMA II-Interventionen) wurde der Beginn des Nachbeobachtungszeitraumes der Behandlungsgruppe durch den Beginn der Betreuung im SIMA II markiert. Im Fall der Vergleichsgruppe orientierte sich der Beginn des Nachbeobachtungszeitraumes hingegen am (spätesten) Beginn der Betreuung in der Bewährungshilfe, welcher vor oder innerhalb des Auswertungszeitraumes (bei der Vergleichsgruppe: 01.01.2014 bis 31.12.2014) lag. Wenn ein/e Proband/-in der Vergleichsgruppe also bspw. am 01.01.2010 und darüber hinaus am 01.01.2013 eine Betreuungsleistung durch die Bewährungshilfe in Anspruch genommen hatte, wurde das letztgenannte Datum als Beginn der Nachbeobachtung festgesetzt. War ein/e Proband/-in der Vergleichsgruppe bspw. am 01.01.2010, am 01.01.2013 und darüber hinaus am 01.06.2014 der Bewährungshilfe zugewiesen worden, wurde der 01.06.2014 als Beginn der Nachbeobachtung festgesetzt. Das aktuelle Hauptdelikt wurde in der Folge aus demjenigen BZR-Eintrag extrahiert, der den

⁷ Manche Einträge beinhalteten bspw. die Information zur Bildung einer Gesamtstrafe für Straftaten, die in vorangegangenen Einträgen dokumentiert waren. Solche Einträge wurden von der Zählung der Vorstrafen ausgeschlossen.

⁸ Das Indexdelikt, welches in den SoPart-Daten hinterlegt war, war das Delikt, das die Betreuung im SIMA II begründete. Dieses konnte jedoch – insbesondere bei Personen, die schon relativ lange in Betreuung waren, weiter in der Vergangenheit liegen und daher auch längere Zeit vor dem Beginn der Betreuung im SIMA II begangen worden sein und insofern das Risiko zum Beginn des Nachbeobachtungszeitraums nur unzureichend abbilden. Aus diesem Grund wurde das Indexdelikt auf die oben beschriebene alternative Art aus den BZR-Daten extrahiert.

geringsten zeitlichen Abstand zum Beginn des Nachbeobachtungszeitraumes aufwies. Maßgeblich dabei war das Datum der Verurteilung. Dieses Datum war dem ebenfalls im BZR-Eintrag dokumentierten Datum der (letzten) Tat vorzuziehen, da der Urteilspruch der frühestmögliche Zeitpunkt war, ab dem die Intervention (die Betreuung im SIMA II) beginnen konnte. Das Urteilsdatum war ebenso dem Datum der Rechtskraft des Urteils vorzuziehen, da das letztgenannte Datum nicht konsistent dokumentiert war. Die Informationen wurden gemäß des OGRS 3-Berechnungsmodells integriert und für jede Person der Stichprobe wurden die beiden OGRS 3-Risikowerte (1-Jahrezeitraum und 2-Jahrezeitraum) bestimmt. Da die Untersuchungsstichprobe auch Personen mit Nachbeobachtungszeiträumen unter zwei Jahren enthielt, wurde der daher aussagekräftigere Risikowert für den 1-Jahrezeitraum als Grundlage für das Matching herangezogen. Diese Risikowerte wurden wiederum in 20 Risikostufen kategorisiert, d.h. in 5 %-Schritten gesetzt.

Neben dem OGRS 3-Risikowert wurde beim Matching berücksichtigt, ob der rechtliche Grund des Eintritts der Betreuung eine vollverbüßte Freiheitsstrafe (§ 68f StGB) war. Diese zusätzliche Matching-Variable wurde auf Basis der großangelegten bundesweiten Rückfalluntersuchung von Jehle u.a. (2016) ausgewählt, bei denen die sog. Vollverbüßer/-innen vergleichsweise hohe Rückfallraten aufwiesen. Insbesondere erneute unbedingt ausgesprochene Freiheitsstrafen waren bei den von Jehle u.a. (2016) untersuchten Vollverbüßer/-innen häufiger als bei den Personen mit Straf(rest)aussetzung. Durch die Berücksichtigung dieses Merkmals als Matching-Variable sollte sich zusätzlich die Möglichkeit eröffnen, für diese Gruppe differentielle Wirksamkeitsanalysen durchführen zu können. Für eine solche Überprüfung bedarf es annähernd gleicher Fallzahlen in Behandlungs- und Vergleichsgruppe.

Das Matching erfolgte schließlich auf Basis der 20-stufigen OGRS 3-Risikovvariable und der nominalen Variable in Bezug auf den Status als Vollverbüßer/-in: Um einander zugeordnet werden zu können, mussten zwei Personen sowohl in Bezug auf die OGRS 3-Risikokategorie, als auch in Bezug auf das Vorliegen des Unterstellungsgrundes nach § 68f StGB (0 = nein, 1 = ja) vergleichbar sein. Durch die Einteilung der Risikovvariable in 20 Kategorien konnte die Abgleichstoleranz schließlich auf null gesetzt werden; Matching-Zwillinge mussten folglich exakt identische Propensity Scores aufweisen, um einander zugeordnet werden zu können. Auf diese Weise sollte Risikoäquivalenz gewährleistet werden. Die so identifizierten Zwillinge innerhalb des Vergleichsprobanden-Pools bildeten schließlich die Kontrollgruppe.

3.4.2 Auswertung von Rückfallereignissen

Als Rückfallkriterium, hinsichtlich dessen die beiden Gruppen verglichen wurden, wurden für diesen Projektteil nur die Daten aus dem BZR, nicht aber die MESTA-Daten berücksichtigt, da die letztgenannte Datenquelle für die Vergleichsgruppe nicht in ausreichender Qualität und Vollständigkeit vorlag. Um potentielle Rückfälle einer Person festzustellen, wurden alle neuerlichen rechtskräftigen Verurteilungen nach Beginn des Nachbeobachtungszeitraumes berücksichtigt, die in Form von Eintragungen im BZR dokumentiert waren. Tabelle 3 gibt eine Übersicht über die bei den Analysen berücksichtigten möglichen Rückfallereignisse.

Tabelle 3: Erläuterungen zu den analysierten Rückfallereignissen

Rückfallereignis	Erläuterung
Allgemeiner Rückfall	Jede neue rechtskräftige Verurteilung aufgrund eines Verstoßes gegen Strafvorschriften nach AO, AufenthG, BtMG, FreizügG/EU, MarkenG, PflVG, StAG, StGB, StVG oder WaffG
StGB-Verstoß	Verurteilung aufgrund eines StGB-Verstoßes
Nichtsexuelles Gewaltdelikt	Verurteilung aufgrund eines nichtsexuellen Gewaltdelikts
Sexualdelikt	Verurteilung aufgrund eines Sexualdelikts (Hand-On oder Hands-Off)
Freiheitsstrafe	Rückfall inklusive Sanktionierung in Form einer Freiheitsstrafe

3.5 Projektteil 3: Qualitative Interviews mit Bewährungshelfer/-innen

Schließlich wurden unter Verwendung einer qualitativen Forschungsmethodik Experteninterviews (vgl. Gläser/Laudel 2010) mit Bewährungshelfer/-innen geführt, die zum Zeitpunkt der Evaluation im SIMA II tätig waren. Im Rahmen dieser Interviews sollte insbesondere herausgearbeitet werden, wie aus Sicht der SIMA II-Mitarbeiter/-innen die praktische Umsetzung der vorgegebenen Richtlinien und Zielsetzungen zum Zeitpunkt der Evaluation zu bewerten war. Ein weiterer Schwerpunkt lag auf der Frage, wo (noch) Probleme und Defizite zu verorten waren und wie diese ggf. behoben werden könnten. Insbesondere soll-

te auch die Frage nach einem möglichst effektiven Einsatz der vorhandenen Ressourcen gestellt werden (werden die „richtigen“ Probanden/-innen mit ausreichenden Ressourcen erreicht?). Da die Klientel der Führungsaufsichtsprobanden/-innen als vergleichsweise heterogen eingeschätzt werden musste, erschienen hier spezifische Nachfragen jedenfalls angezeigt.

Der qualitativen Methodik folgend stand in diesem Projektteil nicht die Erreichung einer möglichst hohen Anzahl an Interviews im Mittelpunkt, sondern eine möglichst repräsentative Auswahl von Bewährungshelfer/-innen, die individuelle und möglichst detaillierte Einblicke in ihre Arbeitsweise geben konnten. Zur Umsetzung dieses Vorhabens wurde ein Leitfaden entwickelt, welcher der Methodik semi-strukturierter Interviews folgte (Döring/Bortz 2016; Gläser/Laudel 2010; siehe Anhang D). Semi- oder auch halbstrukturierte Interviews basieren nach Döring/Bortz (2016) auf einem „*Katalog offener Fragen, zu denen sich die Befragungspersonen in eigenen Worten äußern sollen. Der Leitfaden gibt die Fragen und ihre Reihenfolge vor, erlaubt aber den Interviewenden je nach Interviewsituation in gewissen Grenzen individuelle Anpassungen. So können Fragen vorgezogen, übersprungen oder vertieft werden, um den Interviewfluss am Laufen zu halten. Häufig ist von ‚teilstandardisierten‘ Interviewleitfäden die Rede, damit ist gemeint, dass die Fragen oder Themenblöcke (grob) vorgegeben sind, jedoch keine Antwortalternativen angeboten werden, sondern sich die Befragten in eigenen Worten äußern*“ (Ebd.: 358).

Das semi-strukturierte Interview lässt demnach Raum für individuelle Abweichungen im Gesprächsverlauf. Das hat den Vorteil, dass auch Aspekte des Forschungsgegenstands Berücksichtigung finden können, die zuvor nicht antizipiert wurden, welche der/dem jeweiligen Interviewpartner/-in im jeweiligen Kontext jedoch wichtig sind (Döring/Bortz 2016).

In Kooperation mit dem HMdJ wurde eine Liste möglicher Interviewpartner/-innen erstellt ($n = 9$), wobei aus jedem Landgerichtsbezirk ein/-e SIMA II-Mitarbeiter/-in vorgeschlagen wurde. Dabei wurde auf eine möglichst ausgeglichene Verteilung hinsichtlich Alter und Geschlecht geachtet. Die Personen wurden via E-Mail kontaktiert, über Kontext und Zweck der Befragung informiert und eingeladen, daran teilzunehmen. Für die Gespräche, die bevorzugt in Form eines persönlichen vis-à-vis-Kontaktes erfolgten, war ein zeitlicher Rahmen von etwa 45 Minuten (Netto-Interviewzeit, exkl. Begrüßung, Aufklärung, Verabschiedung etc.) vorgesehen. In Ausnahmefällen wurden die Interviews am Telefon geführt. Von allen Gesprächen wurden Audioaufzeichnungen erstellt. Alle Teilnehmer/-innen wurden ausführlich über die

Freiwilligkeit und den Zweck ihrer Teilnahme informiert und nur unter der Voraussetzung ihres schriftlichen Einverständnisses befragt (siehe Anhang E).

Im Anschluss an die Interviews wurden die Tonaufnahmen transkribiert. Die Analyse der Daten wurde durch das Programm MAXQDA unterstützt, welches dabei half, das Datenmaterial inhaltlich aufzubereiten. Die Auswertung der Interviews erfolgte in Form einer qualitativen Inhaltsanalyse nach Mayring (2010). Hierbei wurden die Daten zuerst zusammengefasst und strukturiert. Im nächsten Schritt wurden die Antworten der Bewährungshelfer/-innen anhand des erstellten Leitfadens thematischen Kategorien zugeordnet und bei Bedarf neue Kategorien gebildet. Auf dieser Basis wurden inhaltliche Vergleiche zwischen den Aussagen der befragten Personen und die Identifikation thematischer Schwerpunkte ermöglicht. Diese Schwerpunkte basierten größtenteils auf Aussagen, die von mindestens einem Drittel der Bewährungshelfer/-innen getätigt wurden. Des Weiteren bildeten sie die Grundlage für weitere Interpretationen und für die Beantwortung der Fragen, inwiefern die Vorgaben des SIMA II zum Abfragezeitpunkt praktisch umgesetzt werden konnten, welche positiven und negativen Aspekte mit der Einführung einhergegangen sind und welche Verbesserungsansätze möglich und denkbar wären.

4 Ergebnisse

4.1 Deskriptive Statistiken

4.1.1 Stichprobenbeschreibung

Im Folgenden werden zunächst die deskriptiven Merkmale der SIMA II-Probanden/-innen ($n = 1\,339$) dargestellt, die im Rahmen von Projektteil 2 (der Untersuchung der Wirksamkeit der SIMA II-Interventionen) berücksichtigt wurden.

Von den 1 339 SIMA II-Probanden/-innen waren 96.6 % männlich, 66.2 % waren in Deutschland geboren. Die nächsthäufigsten Geburtsländer waren die Türkei (4.7 %), Kasachstan (4.6 %) und Polen (2.5 %). Das mittlere Alter zum Stichtag (01.07.2017) lag bei $M = 38.42$ Jahren ($SD = 10.55$, Minimum = 25, Maximum = 76). Die durchschnittliche Betreuungszeit im SIMA II betrug zum Zeitpunkt der Evaluation $M = 1.63$ Jahre ($SD = 0.61$). Die überwiegende Mehrheit (93.0 %, $n = 1\,358$) wurde mindestens sechs Monate durch den Fachbereich betreut. Es wiesen 71.2 % ($n = 959$) Indexdelikte der Kategorie Gewalt auf.

In Tabelle 4 ist aufgeführt, welche rechtlichen Voraussetzungen die Betreuung durch die Bewährungshilfe begründeten. Die größten Gruppen bildeten Personen mit primärer Strafaussetzung (§ 56 StGB), gefolgt von Personen mit Aussetzung des Restes der Unterbringung (§ 57 StGB). Tabelle 5 listet die rechtlichen Grundlagen des Eintritts der Führungsaufsicht für den Teil der Probanden/-innen auf, deren Betreuung an eine Führungsaufsichtsunterstellung geknüpft war. Annähernd zwei Drittel der SIMA II-Probanden/-innen (64.0 %, $n = 857$) war der Führungsaufsicht unterstellt. Die größte Gruppe innerhalb der Personen mit Führungsaufsicht bildeten die sog. Vollverbüßer/-innen, bei denen die (restliche) Dauer der Unterbringung nicht ausgesetzt wurde, die also ihre Freiheitsstrafe vollständig verbüßten.

Tabelle 4: Unterstellungsgründe der SIMA II-Probanden/-innen für die Bewährungshilfe ($n = 607$)

Rechtlicher Unterstellungsgrund	Häufigkeit	%
§ 56 StGB	365	60.1
§ 57 Abs. 1 StGB	179	29.5
§ 57 Abs. 2 StGB	17	2.8
§ 57a StGB	4	0.7
§§ 35, 36 BtMG	36	5.9
§ 21 JGG	5	0.8
§ 88 JGG	16	2.6
Erllass des HMdJ	7	1.2
Strafaussetzung: im Wege der Gnade	5	0.8
Sonstige Gründe	11	1.8

Tabelle 5: Rechtliche Gründe für den Eintritt der Führungsaufsicht bei den SIMA II-Probanden/-innen ($n = 857$)

Rechtlicher Unterstellungsgrund	Häufigkeit	%
§ 67b Abs. 2 StGB (Aussetzung der Unterbringung mit Anordnung)	19	2.2
§ 67c Abs. 1 StGB (Aussetzung der Unterbringung nach Vollzug d. Strafe)	5	0.6
§ 67d Abs. 2 StGB (Aussetzung des Restes der Unterbringung)	104	12.1
§ 67d Abs. 3 StGB (Entlassung aus der Sicherungsverwahrung)	2	0.2
§ 67d Abs. 4 StGB (Erledigung nach Ablauf der Höchstfrist)	25	2.9
§ 67d Abs. 5 StGB (Erledigung nach Nichterreichen des Zweckes)	77	9.0
§ 67d Abs. 6 StGB (Erledigung nach Wegfall der Voraussetzungen)	13	1.5
§ 68 Abs. 1 StGB (Anordnung neben Strafe)	15	1.8
§ 68f Abs. 1 StGB (Vollverbüßung der Strafe)	578	67.4

In Tabelle 6 sind die Indexdelikte der SIMA II-Probanden/-innen gelistet, d.h. jene Delikte, welche den Verurteilungen zugrunde lagen, und an die sich die Betreuung in der Bewährungshilfe anschloss. Die Tabelle berücksichtigt alle Deliktarten, die bei mindestens zehn Personen das Indexdelikt bzw. einen Teil des Indexdelikts darstellten. Die häufigsten Indexdelikte stellten einfache, gefolgt von gefährlichen Körperverletzungsdelikten dar.

Tabelle 6: Indexdelikte der SIMA II-Probanden/-innen (n = 1 339)

Indexdelikt ^a	Häufigkeit	%
§ 211 StGB (Mord)	20	1.5
§ 212 StGB (Totschlag)	50	3.7
§ 223 StGB (Körperverletzung)	648	48.4
§ 224 StGB (Gefährliche Körperverletzung)	447	33.4
§ 239 StGB (Freiheitsberaubung)	30	2.2
§ 239a StGB (Erpresserischer Menschenraub)	10	0.7
§ 242 StGB (Diebstahl)	16	1.2
§ 243 StGB (Besonders schwerer Diebstahl)	76	5.7
§ 244 StGB (Diebstahl mit Waffen; Bandendiebstahl; Wohnungseinbruchdiebstahl)	31	2.3
§ 249 StGB (Raub)	248	18.5
§ 250 StGB (Schwerer Raub)	228	17.0
§ 252 StGB (Räuberischer Diebstahl)	51	3.8
§ 255 StGB (Räuberische Erpressung)	135	10.1
§ 263 StGB (Betrug)	67	5.0

§ 29 BtMG (Drogenhandel, -herstellung, -einfuhr, -veräußerung, -erwerb, -besitz)	35	1.7
§ 267 StGB (Urkundenfälschung)	12	0.9
§ 306a StGB (Schwere Brandstiftung)	19	1.4
§ 29 BtMG (Unerlaubtes Handeltreiben, Anbauen, Herstellen, Ein- oder Ausführen, Veräußern, Abgeben, Sonst-in-Verkehr-Bringen, Erwerben sowie Besitz)	27	2.0
§ 29a BtMG (Betäubungsmittelhandel, -herstellung, -abgabe, -besitz)	60	4.5
§ 30 BtMG (Drogenhandel, -herstellung als Mitglied einer Bande)	32	2.4
§ 30a BtMG (Gewerbsmäßige Abgabe an Jugendliche)	10	0.7

^a Mehrfachnennungen möglich, wenn das Indexdelikt mehrere Gesetzesverstöße inkludierte, die der Kategorie Gewalt zuzuordnen waren.

4.1.2 Rückfallereignisse

Neuerliche rechtskräftige Verurteilungen (BZR)

Die BZR-Daten enthielten Informationen zu allen rechtskräftigen Verurteilungen bis zum 27.03.2020, dem Datum der Abfrage durch das BfJ. Von den $n = 5\,169$ erteilten Auskünften waren 36.1 % ($n = 1\,867$) der Behandlungsgruppe und 63.9 % ($n = 3\,302$) der Vergleichsgruppe zuzuordnen. Tabelle 7 enthält Informationen zur kriminellen Rückfälligkeit der 1 339 SIMA II-Probanden/-innen, für die BZR-Daten vorlagen, getrennt nach verschiedenen Rückfallereignissen. Bei dieser Personengruppe betrug der mittlere Nachbeobachtungszeitraum in Bezug auf die BZR-Daten $M = 2.75$ Jahre ($SD = 0.46$, Minimum = 0.98 Jahre, Maximum = 3.21 Jahre). Beim Großteil der allgemeinen Rückfälle der SIMA II-Probanden/-innen handelte es sich um StGB-Verstöße (67.4 %), gefolgt von BtMG-Verstößen (21.5 %), StVG-Verstößen (8.4 %) und WaffG-Verstößen (2.1 %). Diebstahldelikte stellten die häufigste Kategorie – angelehnt an die Deliktkategorien der OGRS 3 (siehe Abschnitt 3.4.1) – im Bereich der allgemeinen Rückfälligkeit (30.4 %) dar, gefolgt von Drogen- (20.8 %) und Gewaltdelikten (16.7 %). Im Mittel ereigneten sich die Rückfälle nach 337 Tagen ($SD = 231$ Tage).

In Tabelle 8 bis Tabelle 14 sind die Rückfallraten getrennt nach verschiedenen Subgruppen der SIMA II-Klientel dargestellt. Es zeigte sich, dass Probanden/-innen, die aufgrund von Gewaltdelikten in die Betreuung kamen, im Vergleich zu Personen mit anderen Indexdelikten häufiger aufgrund erneuter Gewaltdelikte verurteilt wurden, sich hinsichtlich der anderen Rückfallkategorien jedoch nicht signifikant von den übrigen Probanden/-innen unterscheiden (siehe Tabelle 8). Probanden/-innen, die aufgrund von BtMG-Verstößen in die Betreuung kamen, hatten hinsichtlich der meisten betrachteten Rückfallkategorien geringere Rückfallquoten als Probanden/-innen mit anderen Indexdelikten (siehe Tabelle 9). Gleiches gilt für Personen, die aufgrund von Tötungsdelikten verurteilt wurden (siehe Tabelle 10). Führungsaufsichtsprobanden/-innen unterschieden sich hinsichtlich der Rückfälligkeit kaum von den restlichen Probanden/-innen ohne entsprechende Weisung, wurden jedoch seltener gewalttätig und sexuell rückfällig (siehe Tabelle 11). Personen, die nach Vollverbüßung einer Freiheitsstrafe in die Betreuung kamen, hatten eine im Vergleich zur restlichen Stichprobe jedoch erhöhte allgemeine und auf StGB-Verstößen basierende Rückfallquote und wurden darüber hinaus häufiger erneut mit Freiheitsstrafen sanktioniert (siehe Tabelle 12). Gewalttätige Rückfälle waren in der Gruppe der Vollverbüßer/-innen jedoch seltener als in der übrigen Stichprobe. Auch zeigte sich, dass mit höherem Rückfallrisiko und damit einer intensiver werdenden Betreuungsstufe die Rückfallquoten stiegen (siehe Tabelle 14). Personen, die aufgrund eines als hoch eingeschätzten Rückfallrisikos einer intensiveren Betreuungsstufe zugewiesen worden waren, wurden also – entsprechend des festgestellten Risikos – häufiger rückfällig (mehr dazu in Abschnitt 4.2.2). Generell sei angemerkt, dass alle Daten in Bezug auf neuerliche Sexualdelikte aufgrund der insgesamt sehr niedrigen Quote entsprechender Rückfallereignisse zurückhaltend interpretiert werden sollten.

Tabelle 7: Neuerliche rechtskräftige Verurteilungen (BZR-Einträge) der SIMA II-Probanden/-innen ($n = 1\,339$)

Rückfallereignis	Verurteilungsquote in % (n)
Allgemeiner Rückfall	32.7 (438)
StGB-Verstoß	26.6 (356)
Nichtsexuelle Gewalt	7.9 (106)
Sexualdelikt	0.4 (5)
Freiheitsstrafe	16.9 (226)

Tabelle 8: Neuerliche rechtskräftige Verurteilungen (BZR-Einträge) der SIMA II-Probanden/-innen ($n = 1\,339$), getrennt nach Anlassdelikt (Gewaltdelikte)

Rückfallereignis	Verurteilungsquote in % (n)		Testwerte	
	Anlassdelikt = Gewalt ($n = 959$)	Anlassdelikt ≠ Gewalt ($n = 380$)	χ^2	p
Allgemeiner Rückfall	32.5 (312)	33.2 (126)	0.05	.826
StGB-Verstoß	26.6 (255)	26.6 (101)	0.00	.997
Nichtsexuelle Gewalt	10.0 (96)	2.6 (10)	20.33	.000
Sexualdelikt	0.5 (5)	0.0 (0)	1.99	.158
Freiheitsstrafe	16.3 (156)	18.4 (70)	0.90	.343

Signifikante Testwerte sind fett hervorgehoben

Tabelle 9: Neuerliche rechtskräftige Verurteilungen (BZR-Einträge) der SIMA II-Probanden/-innen ($n = 1\,339$), getrennt nach Anlassdelikt (BtMG-Verstoß)

Rückfallereignis	Verurteilungsquote in % (n)		Testwerte	
	Anlassdelikt = Anlassdelikt ≠		χ^2	p
	BtMG-Verstoß (n = 117)	BtMG-Verstoß (n = 1 222)		
Allgemeiner Rückfall	23.9 (28)	33.6 (410)	4.49	.034
StGB-Verstoß	17.1 (20)	27.5 (336)	5.92	.015
Nichtsexuelle Gewalt	2.6 (3)	8.4 (103)	5.04	.025
Sexualdelikt	0.0 (0)	0.4 (5)	0.48	.488
Freiheitsstrafe	10.3 (12)	17.5 (214)	0.01	.045

Signifikante Testwerte sind fett hervorgehoben

Tabelle 10: Neuerliche rechtskräftige Verurteilungen (BZR-Einträge) der SIMA II-Probanden/-innen ($n = 1\,339$), getrennt nach Anlassdelikt (Tötungsdelikte)

Rückfallereignis	Verurteilungsquote in % (n)		Testwerte	
	Anlassdelikt = Anlassdelikt ≠		χ^2	p
	Tötung (n = 70)	Tötung (n = 1 269)		
Allgemeiner Rückfall	15.7 (11)	33.6 (427)	9.69	.002
StGB-Verstoß	8.6 (6)	27.6 (350)	12.28	.000
Nichtsexuelle Gewalt	1.4 (1)	8.3 (105)	4.27	.039
Sexualdelikt	0.0 (0)	0.4 (5)	0.28	.599
Freiheitsstrafe	4.3 (3)	17.6 (223)	8.35	.004

Signifikante Testwerte sind fett hervorgehoben

Tabelle 11: Neuerliche rechtskräftige Verurteilungen (BZR-Einträge) der SI-MA II-Probanden/-innen ($n = 1339$), getrennt nach Unterstellungsgrund (Führungsaufsicht)

Rückfallereignis	Verurteilungsquote in % (n)		Testwerte	
	Mit Führungs- aufsicht ($n = 857$)	Ohne Führungs- aufsicht ($n = 482$)	χ^2	p
Allgemeiner Rückfall	33.0 (283)	32.2 (155)	0.11	.746
StGB-Verstoß	26.5 (227)	26.8 (129)	0.01	.913
Nichtsexuelle Gewalt	6.1 (52)	11.2 (54)	11.16	.001
Sexualdelikt	0.1 (1)	0.8 (4)	4.22	.040
Freiheitsstrafe	17.2 (147)	16.4 (79)	0.13	.721

Signifikante Testwerte sind fett hervorgehoben

Tabelle 12: Neuerliche rechtskräftige Verurteilungen (BZR-Einträge) der SI-MA II-Probanden/-innen ($n = 1339$), getrennt nach Unterstellungsgrund (Vollverbüßung)

Rückfallereignis	Verurteilungsquote in % (n)		Testwerte	
	Vollverbüßung ($n = 581$)	Keine Vollver- büßung ($n = 758$)	χ^2	p
Allgemeiner Rückfall	36.3 (211)	29.9 (227)	6.06	.014
StGB-Verstoß	29.4 (171)	24.4 (185)	4.26	.039
Nichtsexuelle Gewalt	6.0 (35)	9.4 (71)	5.04	.025
Sexualdelikt	0.2 (1)	0.5 (4)	1.12	.290
Freiheitsstrafe	19.3 (112)	15.0 (114)	4.21	.040

Signifikante Testwerte sind fett hervorgehoben

Tabelle 13: Neuerliche rechtskräftige Verurteilungen (BZR-Einträge) der SIMA II-Probanden/-innen ($n = 1\,339$), getrennt nach Geschlecht

Rückfallereignis	Verurteilungsquote in % (n)		Testwerte	
	Männer ($n = 1294$)	Frauen ($n = 45$)	χ^2	p
Allgemeiner Rückfall	33.0 (427)	24.4 (11)	1.45	.229
StGB-Verstoß	26.7 (345)	24.4 (11)	0.11	.741
Nichtsexuelle Gewalt	8.2 (106)	0.0 (0)	4.00	.045
Sexualdelikt	0.4 (5)	0.0 (0)	0.18	.676
Freiheitsstrafe	17.0 (220)	13.3 (6)	0.42	.518

Signifikante Testwerte sind fett hervorgehoben

Tabelle 14: Neuerliche rechtskräftige Verurteilungen (BZR-Einträge) der SIMA II-Probanden/-innen ($n = 1\,339$), getrennt nach Betreuungsstufe

Rückfallereignis	Verurteilungsquote in % (n)			Testwerte	
	Stufe 1 ($n = 393$)	Stufe 2 ($n = 514$)	Stufe 3 ($n = 432$)	χ^2	p
Allgemeiner Rückfall	39.2 (154)	35.4 (182)	23.6 (102)	25.44	.000
StGB-Verstoß	32.3 (127)	29.0 (149)	18.5 (80)	22.54	.000
Nichtsexuelle Gewalt	10.2 (40)	9.7 (50)	3.7 (16)	15.59	.000
Sexualdelikt	0.8 (3)	0.4 (2)	0.0 (0)	3.23	.199
Freiheitsstrafe	21.1 (83)	18.3 (94)	11.3 (49)	15.20	.000

Signifikante Testwerte sind fett hervorgehoben

Neuerliche Strafanzeigen (MESTA)

Die extrahierten MESTA-Daten enthielten Eintragungen für 67.5 % ($n = 1\,411$) der Behandlungsgruppe und 58.0 % ($n = 2\,434$) der Vergleichsgruppe. In Bezug auf die MESTA-Daten lag der mittlere Nachbeobachtungszeitraum der SIMA II-Probanden/-innen ($n = 1\,339$) bei $M = 1.98$ Jahren ($SD = 0.46$, Minimum = 0.21 Jahre, Maximum = 2.43 Jahre). Die Nachbeobachtung in Bezug

auf neuerliche Anzeigen war demnach kürzer als im Fall der BZR-Daten, da der Datenexport aus dem BZR aufgrund des längeren Prüfprozesses beinahe ein Jahr später erfolgte als der Export der MESTA-Daten. Dieser Unterschied muss bei der Interpretation und dem Vergleich der Rückfallzahlen berücksichtigt werden. Rückfalldelikte wurden im Mittel nach 203 Tagen registriert ($SD = 190$ Tage). Tabelle 15 enthält Informationen zu neuerlichen Strafanzeigen gegen die 1 339 SIMA II-Probanden/-innen, getrennt nach verschiedenen Rückfallereignissen. Die Unterschiede zwischen den Subgruppen in Bezug auf neuerliche Anzeigen sind in Tabelle 16 bis Tabelle 22 dargestellt und entsprechen weitgehend den auf den BZR-Daten basierenden und im vorherigen Abschnitt dargestellten Rückfallraten, weswegen an dieser Stelle nicht gesondert darauf eingegangen werden soll.

Tabelle 15: Neuerliche Strafanzeigen (MESTA-Einträge) gegen die SIMA II-Probanden/-innen ($n = 1\,339$)

Rückfallereignis	Verurteilungsquote in % (<i>n</i>)
Allgemeiner Rückfall	45.4 (608)
StGB-Verstoß	41.4 (554)
Nichtsexuelle Gewalt	18.8 (252)
Sexualdelikt	1.3 (17)
BtMG-Verstoß	17.0 (227)

Tabelle 16: Neuerliche Strafanzeigen (MESTA-Einträge) gegen die SI-MA II-Probanden/-innen ($n = 1\,339$), getrennt nach Anlassdelikt (Gewaltdelikt)

Rückfallereignis	Verurteilungsquote in % (n)		Testwerte	
	Anlassdelikt =	Anlassdelikt ≠	χ^2	p
	Gewalt (n = 959)	Gewalt (n = 380)		
Allgemeiner Rückfall	46.1 (442)	43.7 (166)	0.64	.425
StGB-Verstoß	41.7 (400)	40.5 (154)	0.16	.692
Nichtsexuelle Gewalt	22.1 (212)	10.5 (40)	23.89	.000
Sexualdelikt	1.6 (15)	0.5 (2)	2.34	.126
BtMG-Verstoß	17.4 (167)	15.8 (60)	0.51	.475

Signifikante Testwerte sind fett hervorgehoben

Tabelle 17: Neuerliche Strafanzeigen (MESTA-Einträge) gegen die SI-MA II-Probanden/-innen ($n = 1\,339$), getrennt nach Anlassdelikt (BtMG-Verstoß)

Rückfallereignis	Verurteilungsquote in % (n)		Testwerte	
	Anlassdelikt =	Anlassdelikt ≠	χ^2	p
	BtMG- Verstoß (n = 117)	BtMG- Verstoß (n = 1 222)		
Allgemeiner Rückfall	33.3 (39)	46.6 (569)	7.54	.006
StGB-Verstoß	26.5 (31)	42.8 (523)	11.70	.001
Nichtsexuelle Gewalt	10.3 (12)	19.6 (240)	6.15	.013
Sexualdelikt	0.9 (1)	1.3 (16)	0.18	.675
BtMG-Verstoß	17.1 (20)	16.9 (207)	0.00	.966

Signifikante Testwerte sind fett hervorgehoben

Tabelle 18: Neuerliche Strafanzeigen (MESTA-Einträge) gegen die SI-MA II-Probanden/-innen ($n = 1\,339$), getrennt nach Anlassdelikt (Tötungsdelikte)

Rückfallereignis	Verurteilungsquote in % (n)		Testwerte	
	Anlassdelikt = Tötung (n = 70)	Anlassdelikt ≠ Tötung (n = 1 269)	χ^2	p
Allgemeiner Rückfall	31.4 (22)	46.2 (586)	5.82	.016
StGB-Verstoß	27.1 (19)	42.2 (535)	6.17	.013
Nichtsexuelle Gewalt	10.0 (7)	19.3 (245)	3.76	.052
Sexualdelikt	1.4 (1)	1.3 (16)	0.02	.903
BtMG-Verstoß	10.0 (7)	17.3 (220)	2.54	.111

Signifikante Testwerte sind fett hervorgehoben

Tabelle 19: Neuerliche Strafanzeigen (MESTA-Einträge) gegen die SI-MA II-Probanden/-innen ($n = 1\,339$), getrennt nach Unterstellungsgrund (Führungsaufsicht)

Rückfallereignis	Verurteilungsquote in % (n)		Testwerte	
	Mit Führungs- aufsicht (n = 857)	Ohne Führungs- aufsicht (n = 482)	χ^2	p
Allgemeiner Rückfall	44.7 (383)	46.7 (225)	0.49	.483
StGB-Verstoß	41.3 (354)	41.5 (200)	0.00	.947
Nichtsexuelle Gewalt	16.8 (144)	22.4 (108)	6.34	.012
Sexualdelikt	0.9 (8)	1.9 (9)	2.15	.143
BtMG-Verstoß	18.1 (155)	14.9 (72)	2.17	.141

Signifikante Testwerte sind fett hervorgehoben

Tabelle 20: Neuerliche Strafanzeigen (MESTA-Einträge) gegen die SIMA II-Probanden/-innen ($n = 1339$), getrennt nach Unterstellungsgrund (Vollverbüßung)

Rückfallereignis	Verurteilungsquote in % (n)		Testwerte	
	Vollverbüßung ($n = 581$)	Keine Vollver- büßung ($n = 758$)	χ^2	p
Allgemeiner Rückfall	49.9 (290)	42.0 (318)	8.41	.004
StGB-Verstoß	46.6 (271)	37.3 (283)	11.75	.001
Nichtsexuelle Gewalt	17.9 (104)	19.5 (148)	0.57	.451
Sexualdelikt	1.2 (7)	1.3 (10)	0.03	.853
BtMG-Verstoß	19.8 (115)	14.8 (112)	5.88	.015

Signifikante Testwerte sind fett hervorgehoben

Tabelle 21: Neuerliche Strafanzeigen (MESTA-Einträge) gegen die SIMA II-Probanden/-innen ($n = 1339$), getrennt nach Geschlecht

Rückfallereignis	Verurteilungsquote in % (n)		Testwerte	
	Männer ($n = 1294$)	Frauen ($n = 45$)	χ^2	p
Allgemeiner Rückfall	45.4 (587)	46.7 (21)	0.03	.863
StGB-Verstoß	41.3 (543)	44.4 (20)	0.18	.671
Nichtsexuelle Gewalt	19.0 (246)	13.3 (6)	0.92	.338
Sexualdelikt	1.3 (17)	0.0 (0)	0.60	.439
BtMG-Verstoß	17.1 (221)	13.3 (6)	0.43	.510

Tabelle 22: Neuerliche Strafanzeigen (MESTA-Einträge) gegen die SI-MA II-Probanden/-innen ($n = 1\,339$), getrennt nach Betreuungsstufe

Rückfallereignis	Verurteilungsquote in % (n)			Testwerte	
	Stufe 1 (n = 393)	Stufe 2 (n = 514)	Stufe 3 (n = 432)	χ^2	p
Allgemeiner Rückfall	51.1 (201)	48.6 (250)	36.3 (157)	21.70	.000
StGB-Verstoß	47.6 (187)	42.4 (218)	34.5 (149)	14.91	.001
Nichtsexuelle Gewalt	23.4 (92)	20.6 (106)	12.5 (54)	17.81	.000
Sexualdelikt	1.8 (7)	1.4 (7)	0.7 (3)	2.00	.369
BtMG-Verstoß	22.1 (87)	19.5 (100)	9.3 (40)	27.95	.000

Signifikante Testwerte sind fett hervorgehoben

4.2 Projektteil 1: Güte der Risikoeinschätzung

4.2.1 Güte des SVG-5

Reliabilität des SVG-5

Für den SVG-5 belief sich der Rücklauf von (doppelten) Einschätzungen auf 58 Probanden/-innen mit einem mittleren Summenwert von $M = 5.45$; dies entspricht im Mittel der Betreuungsstufe 2. Der Übereinstimmungsgrad der Summenwerte lag bei $ICC = .89$ ($p < .001$); dieser Wert spricht für eine hohe Übereinstimmung, weshalb die Reliabilität des Instruments als gegeben ein- gestuft werden kann (Wirtz/Caspar 2002). Tabelle 23 gibt Auskunft über die Übereinstimmungsgüte der einzelnen Items des SVG-5. Die Bewertung wurde entsprechend der von Döring und Bortz (2016) vorgeschlagenen Richtwerte vorgenommen. Demzufolge gelten Werte von $K/K_w > .75$ als sehr gut, Werte von $.60$ bis $.75$ als gut und Werte von $.40$ bis $.60$ als mittelmäßig bzw. als gerade noch ausreichend. Vier der fünf SVG-5-Items wiesen sehr gute Übereinstimmungswerte auf, während Item 3, welches nach ggf. begangenen Tötungsdelikten fragt, mit einem lediglich mittelmäßigen Übereinstimmungswert auffiel. Bei diesen doppelten SVG-5-Einschätzungen waren keinerlei fehlende Werte zu verzeichnen.

Tabelle 23: Cohen's Kappa für die einzelnen Items des SVG-5

Item	Cohen's Kappa	<i>p</i>	<i>n</i>	Bewertung ^a
1. Frühere Gewaltdelikte	$K_{\omega} = .81$	< .001	58	sehr gut
2. Deliktfrequenz	$K = .81$	< .001	58	sehr gut
3. Jemals Tötung	$K = .55$	< .001	58	mittelmäßig
4. Psychische Auffälligkeiten	$K = .85$	< .001	58	sehr gut
5. Alter bei erstem Gewaltdelikt	$K_{\omega} = .85$	< .001	58	sehr gut

^a Einordnung nach Döring und Bortz (2016)

Validität des SVG-5

Insgesamt lagen $n = 1\,689$ Risikoeinschätzungen mittels SVG-5 vor. Tabelle 24 zeigt die Verteilungsmerkmale der Summenwerte des SVG-5 für die gesamte Behandlungsgruppe (SIMA II-Probanden/-innen + Probanden/-innen der allgemeinen Bewährungshilfe mit Gewaltindexdelikten) sowie getrennt nach einzelnen Subgruppen. Tabelle 25 enthält die deskriptiven Verteilungsmerkmale der einzelnen SVG-5-Items für die gesamte Behandlungsgruppe.

Tabelle 24: SVG-5-Summenwerte

(Sub-)Gruppe	<i>n</i>	<i>M</i>	<i>SD</i>	Minimum	Maximum
Probanden/-innen mit Führungsaufsicht	595	1.13	9.40	-22	16
Vollverbüßer/-innen (§ 68f Abs. 1 StGB)	331	3.84	8.15	-22	16
Maßregelgruppe (§§ 67d Abs. 2, 67d Abs. 4-6 StGB)	208	-1.72	9.64	-22	16
Richterliche Anordnung (§ 68 Abs. 1 StGB)	13	-4.00	11.64	-22	14
Geschlecht:					
Männer	1 584	1.62	8.20	-22	16
Frauen ^a	105	-2.91	6.87	-22	14
Bewährungshilfe eingetreten aufgrund:					
§ 56 StGB	810	1.39	7.48	-22	16
§ 57 StGB ^b	354	0.89	8.14	-16	16
Zuständiger Fachbereich:					
SIMA II	1 070	4.13	8.19	-22	16
Allg. Bewährungshilfe	619	-3.50	5.53	-22	16
Gesamt	1 689	1.34	8.20	-22	17

^a Der SVG-5 ist zwar lediglich für die Anwendung bei Männern entwickelt worden, kann jedoch grundsätzlich auch bei Frauen angewendet werden

^b umfasst §§ 57a, 57 Abs. 1-2 StGB

Tabelle 25: Deskriptive Statistiken der einzelnen Items des SVG-5 für die Gesamtstichprobe ($n = 1\,689$)

Item	<i>n</i>	%
Item 1 (Anzahl früherer Gewaltdelikte)	1 689	100.0
0	613	36.3
1 oder 2	631	37.4
3 bis 5	377	22.3
6 und mehr	68	4.0
fehlend	0	0.0
Item 2 (Deliktfrequenz)	1 689	100.0
ja	823	48.7
nein	866	51.3
fehlend	0	0.0
Item 3 (jemals Tötung eines Opfers)	1 685	100.0
ja	100	5.9
nein	1 585	94.1
fehlend	4	0.2
Item 4 (psychische Auffälligkeiten)	1 685	100.0
ja	170	10.1
nein	1 515	89.9
fehlend	4	0.2
Item 5 (Alter beim ersten Gewaltdelikt)	1 685	100.0
jünger als 20 Jahre	415	24.6
20 bis 40 Jahre	1 062	63.0
älter als 40 Jahre	208	12.3
fehlend	4	0.2

Nachfolgend wird die Vorhersageleistung des SVG-5 sowie seiner einzelnen Items in Bezug auf unterschiedliche Rückfallkriterien dargestellt. Die Vorhersageleistung wird dabei in Form von AUC-Werten angegeben, dem psychometrischen Goldstandard der Prognoseforschung (z.B. Mossman 1994; Rice/Harris 1995). Ob es sich bei einem AUC-Wert um ein statistisch signifikantes Ergebnis handelt, kann anhand der angegebenen Konfidenzintervalle nachvollzogen werden, die – im Falle einer signifikanten Vorhersageleistung – den Wert .50 nicht beinhalten sollten⁹. Tabelle 26 stellt Interpretationsrichtlinien nach Rice und Harris (2005) zur Verfügung, mithilfe derer die Effektstärken der AUC-Werte bestimmt werden können.

⁹ Teilweise sind Intervalle, deren Untergrenze mit .50 angegeben ist, nicht signifikant. Dies ist u.a. auf Aufrundungen von Werten < .50 zurückzuführen.

Tabelle 26: Interpretationsrichtlinien für *Area Under the Curve* (AUC)-Werte

AUC-Wert	Bewertung der Effektstärke
≥ .556	Klein
≥ .639	Mittel
≥ .714	Groß

In Anlehnung an Rice und Harris (2005)

In Tabelle 27 sind die AUC-Werte des SVG-5 in Bezug auf die Vorhersage neuerlicher Strafanzeigen innerhalb von Hessen in Form von Eintragungen in MESTA und in Tabelle 28 die AUC-Werte in Bezug auf die Vorhersage neuerlicher rechtskräftiger Verurteilungen in Form von Eintragungen im BZR aufgeführt. Statistisch signifikante Vorhersagewerte sind mit fetter Schrift hervorgehoben. Der SVG-5-Summenwert sagte alle betrachteten Rückfallereignisse signifikant vorher. Die Effektstärken der Zusammenhänge mit dem jeweiligen Rückfallereignis lagen dabei im kleinen bis moderaten Bereich. Die auf dem SVG-5-Summenwert basierende Betreuungsstufe stand in etwas geringerem Ausmaß mit den Rückfallkriterien im Zusammenhang. Auf der Ebene der einzelnen Items fanden sich hauptsächlich für die Items 1, 2 und 5 signifikante Vorhersagewerte, jedoch mitunter nicht für alle Rückfallereignisse. In Bezug auf die Vorhersage von Sexualdelikten ist zu berücksichtigen, dass in der Gesamtstichprobe sehr wenige Rückfälle aufgetreten sind und die Aussagekraft der diesbezüglichen Analysen entsprechend beschränkt ist. In Tabelle 29 werden verschiedene Subgruppen, bei denen der SVG-5 angewendet wurde, hinsichtlich der Vorhersageleistung der Risikobewertungen gegenübergestellt. Es fällt auf, dass der SVG-5 bei Frauen keine signifikanten Rückfallvorhersagen leisten konnte. Auch bei der Klientel, die schließlich nicht durch das SIMA II, sondern von der allgemeinen Bewährungshilfe betreut wurde, erreichten die Vorhersagewerte nicht die statistische Signifikanzgrenze.

Tabelle 27: Vorhersage neuerlicher Strafanzeigen (MESTA) durch den SVG-5

Prädiktor	n	AUC [95%-CI]				
		Allgemeiner Verstoß ^a	StGB-Verstoß	Gewaltdelikt (nicht sexuell)	BtMG-Verstoß	Sexualdelikt
Item 1	1 894	.62 [.59-.64]	.60 [.57-.63]	.63 [.60-.66]	.63 [.60-.67]	.63 [.54-.73]
Item 2	1 894	.60 [.58-.63]	.59 [.56-.61]	.61 [.57-.64]	.62 [.59-.66]	.58 [.48-.67]
Item 3	1 890	.52 [.49-.54]	.51 [.49-.54]	.51 [.48-.55]	.53 [.49-.56]	.51 [.41-.61]
Item 4	1 889	.53 [.50-.55]	.53 [.50-.55]	.52 [.48-.55]	.53 [.49-.57]	.49 [.38-.59]
Item 5	1 889	.58 [.56-.61]	.57 [.55-.60]	.58 [.55-.62]	.61 [.58-.65]	.56 [.47-.66]
Summenwert	1 894	.65 [.62-.67]	.62 [.60-.65]	.65 [.62-.68]	.68 [.65-.72]	.62 [.52-.71]
Betreuungsstufe	1 894	.61 [.59-.64]	.59 [.57-.62]	.63 [.59-.66]	.65 [.61-.68]	.59 [.49-.70]

^a Verstöße gegen BtMG, KrWaffKontrG, SprengG, StGB und WaffG; signifikante Vorhersagewerte sind fett hervorgehoben.

Tabelle 28: Vorhersage neuerlicher rechtskräftiger Verurteilungen (BZR) durch den SVG-5

Prädiktor	n	AUC [95%-CI]				
		Allgemeiner Verstoß	StGB-Verstoß	Gewaltdelikt (nicht sexuell)	Sexualdelikt	Freiheitsstrafe
Item 1	1 498	.56 [.53-.59]	.55 [.52-.58]	.55 [.50-.60]	.79 [.68-.90]	.55 [.51-.59]
Item 2	1 498	.58 [.55-.61]	.58 [.54-.61]	.53 [.49-.58]	.58 [.36-.81]	.56 [.53-.60]
Item 3	1 494	.52 [.49-.55]	.52 [.49-.55]	.52 [.48-.56]	.53 [.31-.75]	.53 [.49-.57]
Item 4	1 495	.52 [.49-.56]	.52 [.49-.55]	.51 [.47-.56]	.38 [.12-.64]	.53 [.49-.57]
Item 5	1 495	.55 [.52-.58]	.54 [.51-.57]	.53 [.48-.58]	.72 [.53-.92]	.52 [.49-.57]
Summenwert	1 498	.60 [.57-.63]	.58 [.55-.62]	.56 [.51-.60]	.70 [.46-.94]	.58 [.55-.62]
Betreuungsstufe	1 498	.57 [.54-.60]	.56 [.53-.60]	.54 [.49-.58]	.72 [.52-.92]	.56 [.52-.60]

Signifikante Vorhersagewerte sind fett hervorgehoben

Tabelle 29: Vorhersage neuerlicher rechtskräftiger Verurteilungen (BZR) für verschiedene Subgruppen durch SVG-5-Summenwert

Subgruppe	n	AUC [95%-CI]			
		Allgemeiner Verstoß	StGB-Verstoß	Gewaltdelikt (nicht sexuell)	Freiheitsstrafe
Führungsaufsicht					
Ja	539	.61 [.56-.66]	.60 [.54-.65]	.61 [.53-.70]	.64 [.58-.69]
Nein	959	.59 [.55-.63]	.58 [.54-.62]	.53 [.48-.59]	.55 [.51-.60]
Geschlecht					
Männer	1 412	.60 [.57-.63]	.59 [.56-.62]	.56 [.52-.61]	.59 [.55-.62]
Frauen	86	.55 [.42-.68]	.55 [.42-.68]	.39 [.22-.55]	.53 [.39-.67]
SIMA II-Proband/-in					
Ja	978	.59 [.55-.63]	.58 [.54-.62]	.63 [.58-.68]	.59 [.55-.64]
Nein	520	.56 [.50-.61]	.54 [.49-.60]	.56 [.49-.63]	.58 [.51-.64]

Signifikante Vorhersagewerte sind fett hervorgehoben

Vergleich von SVG-5 und OGRS 3

Bei der Personengruppe, für die SVG-5-Einschätzungen vorlagen und ausgewertet werden konnten, wurde auch die Vorhersageleistung der OGRS 3 (siehe Abschnitt 3.4.1) untersucht, um zu prüfen, inwiefern sich ggf. Verbesserungen der Prognoseleistung erzielen lassen könnten. Dabei zeigten sich Unterschiede in Bezug auf die Vorhersage verschiedener Rückfallereignisse, die anschließend auf ihre statistische Signifikanz untersucht wurden. Diese Ergebnisse sind in Tabelle 30 (BZR) und Tabelle 31 (MESTA) dargestellt. Es zeigte sich, dass die OGRS 3 einen Teil der Rückfallereignisse signifikant besser vorhersagen konnte als der SVG-5; bzgl. der Vorhersage rechtskräftiger Verurteilungen war die OGRS 3 dem SVG-5 überwiegend überlegen, während sich in Bezug auf die Vorhersage von Strafanzeigen mittels MESTA-Daten keine signifikanten Unterschiede ergaben. Vergleiche der Vorhersagen sexueller Rückfälle wurden nicht berechnet, da keines der Instrumente diesbezüglich signifikante Prognosen tätigte – dies war mit hoher Wahrscheinlichkeit auf die generell sehr niedrige Anzahl an sexuell motivierten Rückfalldelikten zurückzuführen.

Tabelle 30: Vergleich der Vorhersageleistung von SVG-5 und OGRS 3 in Bezug auf neuerliche rechtskräftige Verurteilungen (BZR)

Rückfallereignis	AUC [95%-CI]		Testwerte	
	SVG-5 (n = 1 498)	OGRS 3 (n = 1 437)	z	p
Allgemeiner Rückfall	.60 [.57-.63]	.64 [.61-.67]	1.73	.083
StGB-Verstoß	.58 [.55-.62]	.65 [.62-.68]	2.52	.012
Nichtsexuelle Gewalt	.56 [.51-.60]	.64 [.59-.69]	2.27	.023
Freiheitsstrafe	.58 [.55-.62]	.68 [.64-.72]	3.17	.002

Signifikante Unterschiede sind fett hervorgehoben

Tabelle 31: Vergleich der Vorhersageleistung von SVG-5 und OGRS 3 in Bezug auf neuerliche Strafanzeigen (MESTA)

Rückfallereignis	AUC [95%-CI]		Testwerte	
	SVG-5 (n = 1 498)	OGRS 3 (n = 1 437)	z	p
Allgemeiner Rückfall	.63 [.60-.66]	.64 [.61-.66]	0.09	.925
StGB-Verstoß	.61 [.58-.64]	.62 [.59-.65]	0.48	.632
Nichtsexuelle Gewalt	.63 [.59-.67]	.63 [.60-.66]	0.00	1.00
BtMG-Verstoß	.69 [.64-.72]	.68 [.64-.72]	0.06	.951

4.2.2 Güte des Erhebungsbogens

Reliabilität des Erhebungsbogens

Insgesamt konnten $n = 50$ doppelte, mit dem Erhebungsbogen vorgenommene Einschätzungen ausgewertet werden, um die Messgenauigkeit (Zuverlässigkeit bzw. Reliabilität) zu prüfen. Der mittlere Summenwert dieser Einschätzungen lag bei $M = 8.31$ (das entspricht im Mittel der Betreuungsstufe 1, d.h. der intensivsten Betreuungsstufe). Der Übereinstimmungswert der Summenwerte lag bei $ICC = .85$ ($p < .001$); auch dieser Wert spricht für einen hohen Übereinstimmungsgrad (Wirtz/Caspar 2002). Bei den Übereinstimmungswerten der einzelnen Items (siehe Tabelle 32) zeigte sich hingegen ein heterogeneres Bild im Vergleich zum SVG-5. Hinsichtlich eines Items wurde eine perfekte Überein-

stimmung erzielt, und jeweils drei Items hatten sehr gute, gute bzw. moderate Übereinstimmungswerte. Die Einschätzungen waren allerdings zum Teil unvollständig: Insbesondere die Items 2, 5 und 8 wiesen mit 24 %, 32 % und 26 % relativ hohe Anteile fehlender Werte auf. Die Übereinstimmungswerte wurden in diesen Fällen nur basierend auf den vorhandenen Antworten berechnet, es muss jedoch berücksichtigt werden, dass auch der Anteil fehlender Werte ein (negativer) Indikator für die psychometrische Güte eines Instruments darstellt.

Tabelle 32: Grad der Beurteilerübereinstimmung für die 10 Items des Erhebungsbogens

Item	Cohen's Kappa	<i>p</i>	<i>n</i>	Bewertung ^a
1. Alter	1.00	< .001	50	perfekt
2. Familiäre Gewalt	.52	.001	38	mittelmäßig
3. Schulbildung	.86	< .001	48	sehr gut
4. Tagesstruktur	.75	< .001	41	gut
5. Freizeitverhalten	.53	.002	34	mittelmäßig
6. Vorstrafen	.74	< .001	50	gut
7. Vorstrafen Gewalt	.96	< .001	49	sehr gut
8. Erfolgreiche Bewährungen	.42	.006	37	mittelmäßig
9. Dissozialität	.82	< .001	50	sehr gut
10. Substanzmissbrauch	.66	< .001	50	gut

^a Nach Döring und Bortz (2016: 346)

Validität des Erhebungsbogens

Für die Auswertung der prädiktiven Validität (d.h. der tatsächlichen Vorhersageleistung) lagen für den Erhebungsbogen insgesamt 410 Einschätzungen vor. Tabelle 33 zeigt die Verteilungsmerkmale der Summenwerte des Erhebungsbogens für die gesamte Behandlungsgruppe sowie getrennt für einzelne Subgruppen.

Tabelle 34 enthält die deskriptiven Verteilungsmerkmale der einzelnen Items des Erhebungsbogens für die gesamte Behandlungsgruppe. Mit Anteilen von

13.4 %, 19.5 % und 17.3 % wiesen Item 2 (Gewalt in Herkunftsfamilie), Item 5 (Protektives Freizeitverhalten) und Item 8 (Erfolgreich durchlaufene Bewährung) des Erhebungsbogens vergleichsweise hohe Raten fehlender Werte auf.

Tabelle 33: Summenwerte des Erhebungsbogens

(Sub-)Gruppe	<i>n</i>	<i>M</i>	<i>SD</i>	Minimum	Maximum
Rechtliche Gründe für Führungsaufsicht:					
Vollverbüßer/-innen (§ 68f Abs. 1 StGB)	312	7.40	3.34	1	17
Maßregelgruppe (§§ 67d Abs. 2, 67d Abs. 4-6 StGB)	78	7.96	3.86	-1	18
Richterliche Anordnung (§ 68 Abs. 1 StGB)	7	6.71	2.29	4	11
Geschlecht:					
Männer	397	7.46	3.42	-1	18
Frauen	13	6.70	2.69	3	12
Gesamt	410	7.44	3.40	-1	18

Tabelle 34: Deskriptive Statistiken der einzelnen Items des Erhebungsbogens für die gesamte Stichprobe ($n = 410$)

Item	<i>n</i>	%
Item 1 (Alter: Bis 30 Jahre alt?)	410	100.0
ja	42	10.2
nein	368	89.8
fehlend	0	0.0
Item 2 (Gewalt in Herkunftsfamilie)	355	100.0
ja	97	27.3
nein	258	72.7
fehlend	55	13.4
Item 3 (Schulbildung)	401	100.0
Hauptschulabschluss oder höher	314	78.3
kein Hauptschulabschluss oder höher	87	21.7
fehlend	9	2.2
Item 4 (Tagesstruktur)	381	100.0
ja	196	51.4
nein	185	48.6

fehlend	29	7.1
Item 5 (Protektives Freizeitverhalten)	330	100.0
ja	154	46.7
nein	176	53.3
fehlend	80	19.5
Item 6 (Strafrechtliche Vorbelastung)	405	100.0
keine oder eine Vorstrafe	46	11.4
2 Vorstrafen	29	7.2
3 Vorstrafen und mehr	330	81.5
fehlend	5	1.2
Item 7 (Vorstrafe wegen Gewalt oder Raub)	405	100.0
ja	141	34.8
nein	264	65.2
fehlend	5	1.2
Item 8 (Erfolgreiche Bewährung)	339	100.0
ja	139	41.0
nein	200	59.0
fehlend	71	17.3
Item 9 (Dissozialität)	393	100.0
ja	68	17.3
Dissozialität laut Gutachten	31	7.9
nein	325	82.7
fehlend	17	4.1
Item 10 (Substanzmissbrauch)	397	100.0
kein Substanzmissbrauch	125	31.6
Substanzmissbrauch ohne Straftateneinfluss	32	8.1
Substanzmissbrauch bei Straftaten	228	57.7
abgeschl. Therapie u. mind. 6 Monate abstinent UND regelmäßige Suchtberatung u. abstinent/Methadonprogramm	20	5.0
abgeschl. Therapie u. mind. 6 Monate abstinent ODER regelmäßige Suchtberatung u. abstinent/Methadonprogramm	105	26.4
fehlend	13	3.2

Nachfolgend wird die Vorhersageleistung des Erhebungsbogens sowie seiner einzelnen Items in Bezug auf die unterschiedlichen Rückfallkriterien dargestellt. Die Vorhersageleistung wird dabei abermals in Form von AUC-Werten¹⁰ angegeben. In Tabelle 35 sind die AUC-Werte des Erhebungsbogens in Bezug auf die Vorhersage neuerlicher Strafanzeigen (MESTA) und in Tabelle 36 die AUC-Werte in Bezug auf die Vorhersage neuerlicher rechtskräftiger Verurteilungen (BZR) abgetragen.

¹⁰ Zur Interpretation der AUC-Werte wird auf Tabelle 26 verwiesen.

Der Summenwert des Erhebungsbogens konnte den Großteil der betrachteten Rückfallereignisse signifikant vorhersagen. Die Effektstärken der Zusammenhänge mit dem jeweiligen Rückfallereignis variierten von klein bis moderat. Die auf dem Erhebungsbogen basierende Betreuungsstufe hing abermals in etwas geringerem Ausmaß – ebenso wie beim SVG-5 – mit den Rückfallkriterien zusammen. Auf der Ebene der einzelnen Items fanden sich relativ betrachtet weniger signifikante Vorhersagewerte als beim SVG-5. Bei der Vorhersage von BtMG-Verstoßen waren mehr Zusammenhänge auf Item-Ebene zu verzeichnen als bei den restlichen Rückfallereignissen (siehe Tabelle 35). Bei der Vorhersage neuerlicher rechtskräftiger Verurteilungen (siehe Tabelle 36) konnten für Item 10 diverse signifikante Zusammenhänge mit den betrachteten Rückfallereignissen identifiziert werden. In Bezug auf die Vorhersage von Sexualdelikten ist erneut einschränkend zu berücksichtigen, dass es bei lediglich vier der mit dem Erhebungsbogen eingeschätzten Probanden/-innen im Nachbeobachtungszeitraum zu Anzeigen und lediglich bei einem Probanden zu einer Verurteilung aufgrund eines Sexualdelikts kam. Deswegen konnte die diesbezügliche Klassifikationsleistung des Erhebungsbogens nur in sehr eingeschränktem Maße überprüft werden. In Tabelle 37 werden verschiedene Subgruppen, bei denen der Erhebungsbogen angewendet wurde, hinsichtlich der Vorhersageleistung der vorgenommenen Risikobewertungen gegenübergestellt.

Tabelle 35: Vorhersage neuerlicher Strafanzeigen (MESTA) durch den Erhebungsbogen

Prädiktor	n	AUC [95%-CI]				
		Allgemeiner Verstoß ^a	StGB-Verstoß	Gewaltdelikt (nicht sexuell)	BtMG-Verstoß	Sexualdelikt ^c
Item 1	490	.52 [.47-.57]	.51 [.46-.56]	.50 [.42-.57]	.52 [.45-.59]	.46 [.20-.71]
Item 2	427	.53 [.48-.59]	.52 [.46-.58]	.56 [.47-.64]	.57 [.50-.65]	.73 [.49-.98]
Item 3	480	.50 [.45-.55]	.48 [.43-.53]	.50 [.42-.57]	.54 [.47-.61]	.52 [.23-.80]
Item 4	457	.53 [.48-.58]	.52 [.47-.57]	.57 [.49-.64]	.58 [.51-.64]	.58 [.27-.90]
Item 5	398	.51 [.46-.57]	.52 [.46-.57]	.51 [.42-.59]	.59 [.53-.66]	.56 [.24-.88]
Item 6	485	.55 [.49-.60]	.53 [.48-.58]	.54 [.47-.62]	.51 [.44-.58]	.59 [.35-.82]
Item 7	484	.56 [.51-.61]	.54 [.49-.60]	.62 [.54-.70]	.57 [.50-.63]	.57 [.29-.86]
Item 8	411	.52 [.46-.57]	.53 [.48-.59]	.54 [.46-.63]	.52 [.45-.59]	.58 [.31-.84]
Item 9	465	.51 [.46-.57]	.52 [.47-.58]	.52 [.44-.60]	.48 [.41-.55]	.59 [.23-.96]
Item 10	475	.54 [.49-.59]	.52 [.47-.58]	.57 [.50-.64]	.66 [.61-.72]	.71 [.39-1.0]
Summenwert	490	.60 [.55-.65]	.56 [.51-.62]	.63 [.55-.70]	.67 [.62-.73]	.67 [.39-.95]
Betreuungsstufe	490	.58 [.53-.63]	.55 [.50-.60]	.61 [.53-.69]	.66 [.60-.72]	.57 [.28-.86]

^a Verstöße gegen BtMG, KrWaffKontrG, SprengG, StGB und WaffG; ^c lediglich n = 4 Anzeigen aufgrund von Sexualdelinquenz in der Gruppe der Personen, für die eine Einschätzung mit dem Erhebungsbogen vorgenommen wurde; signifikante Vorhersagewerte sind fett hervorgehoben.

Tabelle 36: Vorhersage neuerlicher rechtskräftiger Verurteilungen (BZR) durch den Erhebungsbogen

Prädiktor	n	AUC [95%-CI]				
		Allgemeiner Verstoß ^a	StGB-Verstoß	Gewaltdelikt (nicht sexuell) ^b	Sexualdelikt ^a	Freiheitsstrafe
Item 1	377	.51 [.45-.57]	.50 [.43-.56]	.51 [.36-.67]	.45 [.00-.95]	.50 [.43-.58]
Item 2	326	.54 [.47-.61]	.53 [.46-.60]	.60 [.41-.75]	-	.52 [.44-.60]
Item 3	368	.49 [.43-.55]	.49 [.42-.55]	.46 [.31-.60]	.39 [.00-.83]	.50 [.43-.57]
Item 4	348	.56 [.49-.62]	.57 [.51-.64]	.49 [.33-.65]	.26 [.00-.55]	.54 [.46-.62]
Item 5	301	.56 [.50-.63]	.61 [.54-.68]	.51 [.33-.68]	.23 [.00-.50]	.55 [.47-.63]
Item 6	372	.54 [.48-.60]	.55 [.49-.62]	.49 [.34-.64]	.59 [.13-1.00]	.55 [.48-.62]
Item 7	372	.53 [.46-.59]	.52 [.45-.58]	.56 [.41-.71]	.32 [.00-.69]	.52 [.44-.59]
Item 8	314	.55 [.48-.62]	.55 [.49-.62]	.48 [.31-.64]	.21 [.00-.44]	.53 [.45-.61]
Item 9	361	.48 [.42-.54]	.47 [.41-.54]	.48 [.33-.62]	.42 [.00-.89]	.47 [.39-.54]
Item 10	363	.60 [.54-.66]	.61 [.54-.67]	.69 [.59-.79]	-	.60 [.53-.67]
Summenwert	377	.59 [.53-.65]	.60 [.54-.66]	.55 [.38-.72]	.09 [.04-.14]	.58 [.51-.65]
Betreuungsstufe	377	.58 [.52-.64]	.59 [.52-.65]	.53 [.38-.69]	.16 [.00-.35]	.56 [.49-.64]

^a Nur $n = 1$ Verurteilung aufgrund von Sexualdelinquenz in der Gruppe der Personen, für die eine Einschätzung mit dem Erhebungsbogen vorgenommen wurde; signifikante Vorhersagewerte sind fett hervorgehoben.

Tabelle 37: Vorhersage neuerlicher strafrechtlicher Verurteilungen (BZR) für verschiedene Subgruppen durch Summenwert des Erhebungsbogens

Subgruppe	n	AUC [95%-CI]			
		Allgemeiner Verstoß	StGB-Verstoß	Gewaltdelikt (nicht sexuell)	Freiheitsstrafe
Vollverbüßung:					
Ja	292	.63 [.57-.70]	.62 [.55-.69]	.64 [.43-.85]	.62 [.53-.69]
Nein	85	.45 [.32-.58]	.51 [.37-.65]	.42 [.17-.66]	.43 [.53-.69]
BTM-Indexdelikt:					
Ja	112	.53 [.41-.65]	.54 [.41-.67]	.65 [.35-.96]	.53 [.36-.69]
Nein	265	.61 [.54-.68]	.61 [.54-.68]	.52 [.32-.71]	.59 [.51-.67]
gesamt	377	.59 [.53-.65]	.60 [.54-.66]	.55 [.38-.72]	.58 [.51-.65]

Signifikante Vorhersagewerte sind fett hervorgehoben.

Vergleich von Erhebungsbogen und OGRS 3

Auch für die Personengruppe, für die Einschätzungen auf Basis des Erhebungsbogens vorlagen und ausgewertet werden konnten, wurde die Vorhersageleistung der OGRS 3 untersucht. Dabei ergaben sich Unterschiede in Bezug auf die Vorhersage verschiedener Rückfallereignisse, die daraufhin auf statistische Signifikanz geprüft wurden. Die Ergebnisse sind in Tabelle 38 (BZR) und Tabelle 39 (MESTA) dargestellt. Es zeigte sich, dass die OGRS 3 dem Erhebungsbogen bzgl. der Vorhersage neuerlicher rechtskräftiger Verurteilungen (BZR) überlegen war. Lediglich bei der Vorhersage von Verurteilungen aufgrund gewalttätiger Delikte erreichte der Unterschied in der Vorhersageleistung beider Instrumente nicht die Schwelle statistischer Signifikanz. In Bezug auf die Vorhersage neuerlicher Strafanzeigen (MESTA) zeigten sich hingegen keine signifikanten Unterschiede zwischen den Vorhersageleistungen der beiden Instrumente.

Tabelle 38: Vergleich der Vorhersageleistung von Erhebungsbogen und OGRS 3 in Bezug auf neuerliche rechtskräftige Verurteilungen (BZR)

Rückfallereignis	AUC [95%-CI]		Testwerte	
	Erhebungsbogen (n = 377)	OGRS 3 (n = 369)	z	p
Allgemeiner Rückfall	.59 [.53-.65]	.71 [.65-.77]	2.71	.007
StGB-Verstoß	.60 [.54-.66]	.74 [.69-.80]	3.18	.001
Nichtsexuelle Gewalt	.55 [.38-.72]	.63 [.47-.78]	0.61	.539
Freiheitsstrafe	.58 [.51-.65]	.69 [.63-.76]	2.15	.031

Ein Vergleich der Vorhersagen sexueller Rückfälle wurde aufgrund der niedrigen Basisrate bei dieser Rückfallform nicht durchgeführt; signifikante Unterschiede sind fett hervorgehoben.

Tabelle 39: Vergleich der Vorhersageleistung von Erhebungsbogen und OGRS 3 in Bezug auf neuerliche Strafanzeigen (MESTA)

Rückfallereignis	AUC [95%-CI]		Testwerte	
	Erhebungsbogen (n = 377)	OGRS 3 (n = 369)	z	p
Allgemeiner Rückfall	.58 [.52-.64]	.64 [.59-.70]	1.59	.113
StGB-Verstoß	.54 [.48-.61]	.63 [.57-.69]	1.61	.108
Nichtsexuelle Gewalt	.58 [.47-.69]	.51 [.41-.62]	0.87	.384
BtMG-Verstoß	.65 [.58-.72]	.69 [.63-.76]	0.82	.415

Ein Vergleich der Vorhersagen sexueller Rückfälle wurde aufgrund der niedrigen Basisrate bei dieser Rückfallform nicht durchgeführt.

4.2.3 Diskussion

Der SVG-5 und der Erhebungsbogen wurden in Bezug auf ihre Zuverlässigkeit (Messgenauigkeit, Reliabilität) und ihre Vorhersageleistung (prädiktive Validität) überprüft. Beide Instrumente erzielten dabei gute Beurteilerübereinstimmungen (Interrater-Reliabilität) auf der Ebene der Summenwerte. Die Effekte in Bezug auf die Vorhersage neuerlicher Delinquenz waren klein bis moderat, wobei der SVG-5 tendenziell bessere Vorhersageleistungen zeigte.

Im Folgenden werden die Analyseergebnisse beider Instrumente getrennt voneinander diskutiert und im Anschluss allgemeine Limitationen erörtert, welche die Überprüfung beider Instrumente gleichermaßen betroffen haben könnten.

SVG-5

Der SVG-5 erzielte überwiegend sehr gute Übereinstimmungswerte, sowohl auf der Ebene des Summenwertes, als auch auf der Ebene der einzelnen Items. Lediglich ein Item bildete bzgl. dieses sehr guten Ergebnisses eine Ausnahme: Item 3 (*Jemals Tötung eines Opfers?*) musste als lediglich moderat reliabel eingestuft werden. Da auch dieses Item eine relativ objektive und auf Grundlage des Vorstrafenregisters erhebbare Information erfasst, erscheint dieses Ergebnis zunächst überraschend. Bei näherer Betrachtung können jedoch hierfür unterschiedliche Erklärungen in Betracht gezogen werden. Zum einen mag es statistisch-methodische Gründe geben, weshalb der Übereinstimmungswert bei

diesem Item so gering ausfiel: Die Basiswahrscheinlichkeit dafür, dass jemand bereits Tötungsdelikte begangen hatte, war vergleichsweise gering, was auch an der Verteilung der Antworten bei diesem Item (siehe Tabelle 25) ersichtlich war. Das verwendete Maß *Cohen's Kappa* gilt jedoch als relativ sensitiv gegenüber ungleichen Wahrscheinlichkeiten einzelner Antwortoptionen und tendiert in solchen Fällen zur Angabe geringer Übereinstimmung bei abweichenden Antworten. Dass es überhaupt zu diskrepanten Urteilen kam, könnte außerdem mit der Polung des Items zusammenhängen: Für Personen, die bereits Tötungsdelikte begangen hatten, musste ein negativer und folglich risikoreduzierender Wert eingetragen werden, was von Anwender/-innen als kontraintuitiv empfunden werden konnte, selbst wenn dieser Zusammenhang in der kriminologischen und forensischen Prognoseforschung als gut abgesichert gelten kann (Quinsey u.a. 2006). Diesem Problem könnte durch eine digitalisierte bzw. teil-automatisierte Risikoerfassung entgegengewirkt werden, bei der Anwender/-innen bei den einzelnen Items lediglich zwischen den verbalen Antwortoptionen (z.B. „Ja“ oder „Nein“) wählen, der korrespondierende Wert automatisch ergänzt und der Summenwert somit nicht von Hand berechnet werden muss. Eine solche digitale Lösung könnte die Objektivität bzw. Reliabilität des Instrumentes steigern und könnte bspw. mithilfe von Microsoft Excel mit relativ geringem Aufwand realisiert werden. Alternativ oder ergänzend kann das angesprochene Problem im Rahmen von Schulungen aufgegriffen und die Anwender/-innen bzgl. mutmaßlich missverständlicher Items noch stärker sensibilisiert werden.

Bezüglich der Vorhersageleistungen des SVG-5-Summenwertes lässt sich feststellen, dass diese zwischen kleinen und moderaten Effekten ($AUC = .56$ bis $.68$) variierten. Da der SVG-5 primär für die Prognose gewalttätiger Rückfälle konstruiert wurde, war insbesondere die Vorhersage neuerlicher Gewaltdelikte von Interesse. In Bezug auf dieses Rückfallkriterium erzielte der SVG-5 einen kleinen (BZR) bzw. moderaten (MESTA) Effekt. Damit wies er etwas geringere Effektstärken auf im Vergleich zu früheren Studien (Eher u.a. 2012; Rettenberger u.a. 2010), bei denen er (AUC)-Werte von $.79$ für die Vorhersage gewalttätiger Rückfälle und $.82$ für die Vorhersage gewalttätiger Rückfälle, die zu einer Inhaftierung führten, erzielte (Eher u.a. 2012). Die in dieser genannten Studie beobachteten Probanden/-innen wiesen mit durchschnittlich 6.8 Jahren allerdings deutlich längere Nachbeobachtungszeiträume auf. Als aktuarisches Instrument zieht der SVG-5 primär statische und somit unveränderbare Informationen heran. Es erscheint plausibel, dass während laufender Betreuung auch dynamische Faktoren risikoreduzierend und damit prognoserelevant wirken können. Aufgrund der Betreuungssituation stehen der betreffenden Person an-

dere, jedoch zeitlich determinierte Ressourcen zur Verfügung, die sie bei einem straffreien Leben unterstützen sollen. Endet die Betreuung, verlieren diese mit ihr assoziierten dynamischen Faktoren möglicherweise wieder in gewissem Umfang an Bedeutung. Da sich Probanden/-innen nach beendeter Bewährung wieder eher ihrem ursprünglichen Lebensstil annähern könnten, wäre es denkbar, dass sich der Vorhersagewert der statischen Faktoren dann wieder erhöht. Damit im Einklang steht der Befund, dass sich Behandlungseffekte nach Abschluss der Behandlung mit zunehmender Beobachtungszeit relativieren können (Sauter/Voss/Dahle 2015) und sich somit Behandlungs- und Kontrollgruppe bzgl. ihrer Rückfälligkeit wieder annähern können. Es wäre also möglich, dass der SVG-5 bei längeren Nachbeobachtungszeiträumen, die auch Zeiten abseits der Betreuung einschließen, bessere Vorhersagen trifft. Darüber hinaus bestehen zwischen den früheren SVG-Studien und der vorliegenden Untersuchung weitere methodische Unterschiede, die zur Erklärung der voneinander abweichenden Ergebnismuster beitragen können: So liegen mehrere Unterschiede im Forschungsdesign vor (z.B. prospektive vs. retrospektive Datenerhebung; Datenerhebung durch Wissenschaftler/-innen ausschließlich anhand von Akteninformationen vs. Erhebung von Daten im Zuge der alltäglichen Routinepraxis der Bewährungshilfe), die üblicherweise unmittelbare Auswirkungen auf Reliabilitäts- und Validitätsberechnungen besitzen können.

Bei der weiblichen Subgruppe leistete der SVG-5 keine signifikanten Vorhersagen. Dies scheint jedoch nicht zuletzt auf die geringe Fallzahl zurückzuführen zu sein, welche die Detektion signifikanter Effekte deutlich erschwerte. Bei der Klientel, die schließlich nicht durch das SIMA II, sondern von der allgemeinen Bewährungshilfe betreut wurde, erreichten die Vorhersagewerte ebenfalls keine statistische Signifikanz. Diesbezüglich ist jedoch zu bedenken, dass die SVG-5-Werte dieser Gruppe niedriger und daher in ihrer Varianz systematisch eingeschränkt waren. Varianzeinschränkungen reduzieren ebenfalls die Möglichkeit, statistische Zusammenhänge zu untersuchen.

Zumindest zum Teil könnte die Validität auch durch die Wirkung der SIMA II-Interventionen (siehe auch Abschnitt 4.3) eingeschränkt worden sein. Im Sinne des Risikoprinzips erhalten in der hessischen Bewährungshilfe diejenigen Probanden/-innen mit besonders hohem SVG-5-Risikowert (Betreuungsstufe 1) die intensivste Betreuung. Jene Probanden/-innen mit dem niedrigsten Risiko (Betreuungsstufe 3) werden hingegen, so sie nicht unter Führungsaufsicht stehen, von der allgemeinen Bewährungshilfe übernommen und erhalten eine weniger intensive Betreuung, die sich bspw. in einer niedrigeren Kontaktfrequenz nie-

derschlägt. Studien belegen, dass die Behandlungseffekte in Hochrisikogruppen zumeist stärker ausgeprägt sind als in Niedrigrisikogruppen (siehe z.B. Koehler u.a. 2013). Dies bedeutet, dass das zu Beginn der Betreuung festgestellte Ausgangsrisiko bei denjenigen mit hohem Risiko eventuell stärker reduziert werden konnte als bei denjenigen mit geringerem Risiko. Eine solcher differentieller Wirkmechanismus würde wiederum den Zusammenhang zwischen dem eingangs festgestellten Risikowert und der späteren Rückfälligkeit systematisch reduzieren. Hinzu kommt, dass die Betreuung im SIMA II insbesondere auf die Reduktion einschlägiger, d.h. gewalttätiger Rückfälle abzielte. Der „wahre“ Zusammenhang zwischen Risikoeinschätzung und Rückfälligkeit ließe sich insofern nur dann feststellen, wenn die Probanden/-innen keine oder eine standardisierte bzw. nicht-risikoorientierte Behandlung erhalten hätten. Dennoch bleibt festzustellen, dass im Schnitt Personen der Hochrisikogruppe trotz intensiverer Betreuung signifikant häufiger rückfällig wurden als Personen mit geringerem Risiko (siehe Tabelle 30 bzw. Tabelle 31), weshalb diese Einflussgröße in ihrer Wirkung nicht überschätzt werden sollte.

Im Vergleich mit der OGRS 3 (Francis/Soothill/Humphreys 2007; Howard u.a. 2009), welche das allgemeine kriminelle Rückfallrisiko während eines 1-Jahreszeitraumes (bzw. 2-Jahreszeitraumes) erfasst, zeigten sich signifikante Unterschiede zugunsten der OGRS 3. Insbesondere neuerliche rechtskräftige Verurteilungen sagte die OGRS 3 zum Teil besser vorher als der SVG-5. Diese Überlegenheit zeigte sich auch in Bezug auf die gewalttätige Rückfälligkeit und somit das hauptsächlich interessierende Kriterium des SVG-5. Der SVG-5 ist im Gegensatz zur OGRS 3 nur für eine spezifische Gruppe delinquenter Personen entwickelt worden. Insofern berücksichtigt er in erster Linie Informationen, die sich auf Gewaltdelinquenz beziehen. Die OGRS 3 ist hingegen breiter anwendbar und berechnet das Rückfallrisiko u.a. unter Berücksichtigung des zuletzt verübten Deliktes (*Aktuelles Hauptdelikt*; siehe Abschnitt 3.4.1). In der vorliegenden Untersuchung wurde das Delikt, welches zuletzt vor der Risikoeinschätzung begangen wurde, zur Berechnung des OGRS 3-Risikowertes verwendet. Dieses Delikt muss – selbst bei Personen, die initial aufgrund von Gewaltdelikten ins SIMA II kamen – nicht immer ein Gewaltdelikt gewesen sein. Insofern differenziert die OGRS 3 hinsichtlich des Indexdeliktes anders als der SVG-5, wodurch es zu Unterschieden in der Anwendungspraxis der beiden Verfahren kam, die sich evtl. auch auf die Analyseergebnisse ausgewirkt haben könnten.

Erhebungsbogen

Auf der Ebene des Summenwertes indizierte die Reliabilitätsuntersuchung des Erhebungsbogens eine sehr gute Übereinstimmung. Auf Item-Ebene zeigten die Analysen jedoch auch Schwachstellen: Zunächst gab es bei einzelnen Items hohe Anteile fehlender Werte, so fehlten z.T. über 30 % der Antworten. Dieser Umstand sollte auch hinsichtlich der Interpretation des Übereinstimmungsgrades des Summenwertes berücksichtigt werden. Von hohen Anteilen fehlender Werte waren nämlich v.a. diejenigen Items betroffen, die keine objektiven bzw. aktenbasierten Informationen erfassten und schlechter hinsichtlich ihrer Übereinstimmung abschnitten. Dass folglich mehr aktenbasierte und daher objektive Informationen in den Summenwert mit eingingen als es der Erhebungsbogen eigentlich vorsah, könnte eine Überschätzung der tatsächlichen Reliabilität zur Folge gehabt haben. Diese Ergebnisse weisen auf einen Überarbeitungsbedarf der betroffenen Items hin. Da der Summenwert des Erhebungsbogens auch dann noch berechnet wurde, wenn Teile der Antworten fehlten, führten fehlende Werte zu einer Reduktion des Risikowertes und damit zu einer niedrigeren Einstufung der/des jeweiligen Proband/-in. Konkretere Operationalisierungen und klarere Vorgaben hinsichtlich des Umgangs mit fehlenden Werten könnten dementsprechend zu einer Verbesserung des Erhebungsbogens beitragen. Eine Steigerung von Objektivität und Reliabilität könnte außerdem durch Schulungen in der konsistenten Anwendung des Instrumentes gefördert werden.

Die Vorhersageleistungen des Summenwertes des Erhebungsbogens variierten – ebenso wie im Falle des SVG-5 – zwischen kleinen und moderaten Effekten (AUC = .56 bis .67). Ebenso wie im Zuge der Reliabilitätsuntersuchung offenbarten die analysierten Daten jedoch auf Item-Ebene die Anfälligkeit des Erhebungsbogens für Auslassungen bzw. fehlende Antworten. Hohe Raten fehlender Werte traten bei denselben drei Items auf wie im Kontext der Reliabilitätsprüfung, wenngleich sie etwas schwächer ausgeprägt waren. Es handelte sich wiederum um Item 2 und Item 5, die i.d.R. nicht aktenbasiert, sondern nur in Zusammenarbeit mit dem/der Probanden/-in bearbeitet werden können. Im Falle von Item 8, welches nach einer bereits erfolgreich durchlaufenen Bewährung fragt, könnte eine ungenaue Operationalisierung zu Auslassungen geführt haben. Die vorgegebenen Antwortmöglichkeiten berücksichtigten eine Teilmenge der einzustufenden Personen nicht angemessen, nämlich Probanden/-innen, die erstmalig eine Bewährung durchliefen. Diese erhielten hier eine möglicherweise unzutreffende Einschätzung; bei ihnen musste die Frage nach bereits erfolgreich durchlaufenen Bewährungen verneint werden, woraus gemäß der Logik des Instrumentes eine Erhöhung des Risikowertes

resultieren hätte können, was inhaltlich jedoch wenig plausibel erscheint. Tatsächlich spräche es eher für eine geringere strafrechtliche Vorbelastung und daher ein tendenziell geringeres Rückfallrisiko, wenn ein/-e Proband/-in bisher keine Bewährungszeit aufwies. Nach Kenntnis der Autor/-innen wurde das Item inzwischen überarbeitet, um das genannte Problem zu beseitigen; die der Evaluation zugrundeliegenden Daten waren jedoch mutmaßlich noch von diesem Problem betroffen, wodurch die Objektivität des Erhebungsbogens eingeschränkt gewesen sein könnte.

Die auf Item-Ebene identifizierten Probleme des Erhebungsbogens legen nahe, dass eine Überarbeitung der entsprechenden Items zu einer Steigerung der Objektivität und Reliabilität beitragen könnte. Auf diese Weise würden sich auch zusätzliche Potentiale hinsichtlich der Validität des Instruments ergeben, da Objektivität und Reliabilität Voraussetzungen für eine (hohe) Validität darstellen. Eine weitere Möglichkeit, die mit dem Erhebungsbogen verbundenen Probleme zu beseitigen, könnte darin bestehen, ihn teilweise oder vollständig durch ein anderes Instrument zu ersetzen, welches die identifizierten Schwächen nicht oder zumindest in einem geringeren Umfang aufweist. Die eingeschränkte Objektivität des Erhebungsbogens geht auch damit einher, dass es zur Beantwortung einzelner Items zum Teil mehrerer Sitzungen mit dem/der Proband/-in bedarf, woraus ein relativ ressourcenintensiver diagnostischer Aufwand resultiert, der wiederum mit Verzögerungen bei der Einstufung verbunden sein kann (siehe auch Abschnitt 4.5.1). Es könnte insofern zu einer Stärkung der Reliabilität und Validität sowie zu einer ökonomischeren Risikoeinschätzung beitragen, ein Instrument zu nutzen, das – ebenso wie der SVG-5 – auf aktenbasierte Informationen zurückgreift. Hierfür böte sich die bereits etablierte und erfolgreich validierte OGRS 3 an, welche auch im Rahmen der vorliegenden Studie zum Einsatz kam. Wie der Vergleich zwischen OGRS 3 und Erhebungsbogen zeigte, erreichte die OGRS 3 durchweg bessere Prognoseleistungen, obgleich der Unterschied nicht in jedem Fall statistisch signifikant war. Bisher wird die OGRS 3 vorwiegend in Großbritannien angewandt und ist hinsichtlich einiger Konstruktionsmerkmale an die dortigen strafrechtlichen Gegebenheiten angepasst. Zwar wurde das Instrument im Zuge der vorliegenden Evaluationsstudie an die hiesigen Bedingungen adaptiert, eine publizierte deutsche Übersetzung, inklusive eines ausführlichen Manuals, existiert bislang jedoch nicht. Nichtsdestotrotz geben die vorliegenden Ergebnisse einen vielversprechenden Ausblick auf die Verwendungsmöglichkeiten der OGRS 3. Gleichzeitig muss jedoch auch bedacht werden, dass ein Verzicht auf den Erhebungsbogen dazu führen würde, dass einzelne konzeptionelle Vorteile (wie z.B.

die inhaltlichen Stärken, die u.a. aufgrund der einschlägigen Erfahrungswerte der Bewährungshilfe nutzbar gemacht werden konnten, oder der dynamische Charakter einzelner Items) in Zukunft nicht mehr nutzbar wären. Insofern müssten vor konkreten Entscheidungen verschiedene Vor- und Nachteile der einzelnen Verfahren abgewogen werden.

Allgemeine Limitationen

Wenngleich gute Reliabilitätskoeffizienten und signifikante Zusammenhänge mit unterschiedlichen Rückfallkriterien identifiziert werden konnten, schränkten diverse Faktoren die Möglichkeiten aussagestarker Analysen ein. Im Zuge der Datenerhebung zeigte sich, dass die Risikoeinschätzungen, welche die Basis für die Validitätsuntersuchungen bilden sollten, mit zum Teil recht späten Datumsangaben versehen waren. Mitunter lagen die Einschätzungen erst kurze Zeit zurück und waren längere Zeit nach Beginn der Betreuung im SIMA II vorgenommen worden. Das hing mutmaßlich damit zusammen, dass im Laufe der Betreuung einer konkreten Person wiederholt Risikoeinschätzungen mit dem SVG-5 bzw. dem Erhebungsbogen vorgenommen wurden. Somit wurde in einigen Fällen wahrscheinlich nur die zuletzt getätigte Einschätzung übermittelt, jedoch nicht die zu Beginn der Betreuung vorgenommene. Dadurch, dass in der Folge für manche Probanden/-innen nur recht aktuelle Einschätzungen vorlagen, verkürzten sich für die betreffenden Personen die Nachbeobachtungszeiträume erheblich. Demnach hatten diese Personen einen entsprechend kürzeren Nachbeobachtungszeitraum.

Zudem konnten für etwa 30 % der Probanden/-innen, für die Risikoeinschätzungen bei der Bewährungshilfe angefragt wurden, keine Daten geliefert werden. Ein gewisser Anteil an fehlenden Daten ist zunächst für jedes praxisnahe Projekt unvermeidlich, relevant in diesem Zusammenhang ist die Frage, inwieweit es sich dabei um zufällige oder systematische Einschränkungen handelte, da ein systematisch reduzierter Rücklauf die Interpretierbarkeit und Generalisierbarkeit der Analyseergebnisse einschränken würde. Für die vorliegende Studie lagen uns keine Hinweise auf systematische Verzerrungen hinsichtlich fehlender Werte vor, es kann jedoch auch nicht gänzlich ausgeschlossen werden.

Insgesamt war bei einigen Rückfallereignissen die Basisrate sehr gering, zum einen in Bezug auf Sexualdelikte, aber auch – in geringerem Umfang – in Bezug auf gewalttätige Rückfälle, welche für die vorliegenden Untersuchungen von besonderer Bedeutung waren. Im Falle niedriger Rückfallraten reduziert sich aus statistischen Gründen die Möglichkeit der Überprüfung der Klassifikationsleistung, da bei geringen Fallzahlen die Generalisierbarkeit weniger

gewährleistet ist. Zwar ist die eingesetzte Methodik (AUC-Werte) relativ robust gegenüber geringen Basisraten, trotzdem steigt auch hier mit geringer werdender Basisrate die Wahrscheinlichkeit zufällig variierender Ergebnisse (Eher u.a. 2008; Hanczar u.a. 2010; Lobo/Jiménez-Valverde/Real 2008).

4.3 Projektteil 2: Kriminalpräventive Wirksamkeit der Betreuung

Im Zuge der Untersuchung der Wirksamkeit der Interventionspraxis der SI-MA II-Betreuung konnte auf Basis der gewählten Matching-Variablen für 97.6 % ($n = 1\,307$) der 1 339 SIMA II-Probanden/-innen ein risikoäquivalenter Zwilling innerhalb der Gruppe potentieller Vergleichsprobanden/-innen identifiziert werden. Bei diesen Vergleichsprobanden/-innen handelte es sich um Personen, die vor der Implementierung des SIMA II in der hessischen Bewährungshilfe betreut worden waren. Abbildung 2 zeigt die Verteilung der Stichprobe in Bezug auf die unterschiedlichen Risikokategorien der OGRS 3.

Abbildung 2: Verteilung der Gesamtstichprobe auf die 20 OGRS 3-Risikokategorien ($N = 2\,614$)

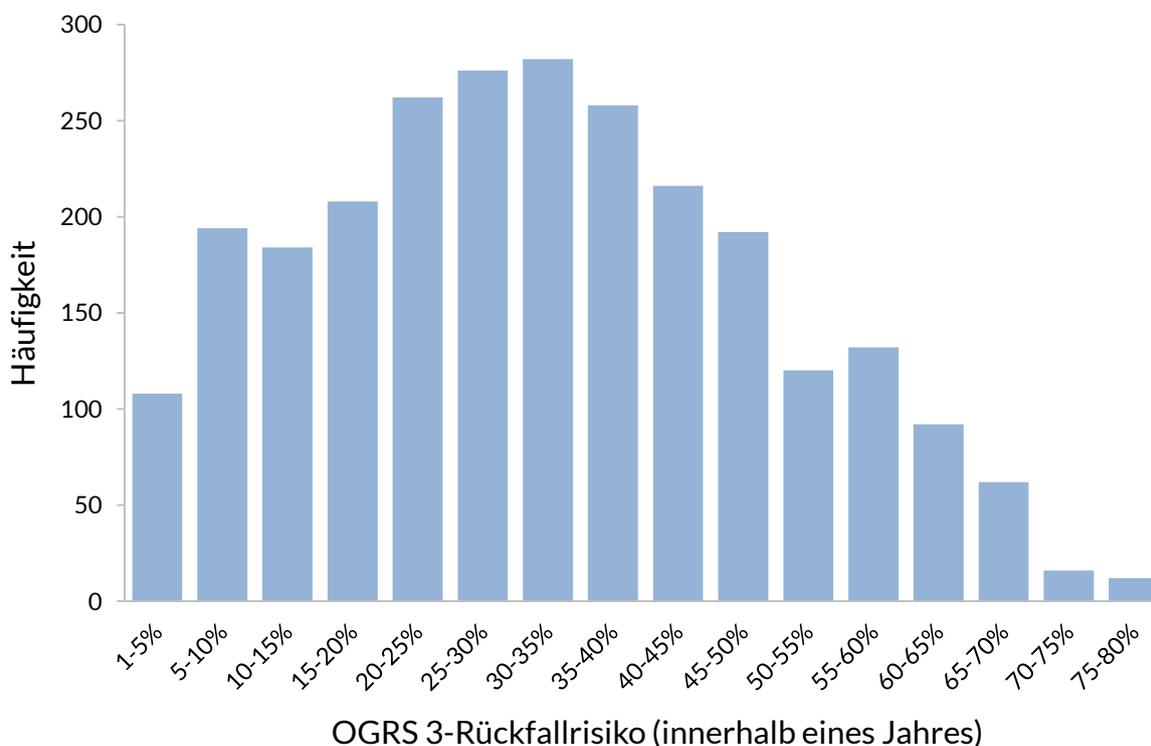


Tabelle 40 enthält weitere Merkmale der für Projektteil 2 untersuchten Stichprobe, bezogen auf die einzelnen Items der OGRS 3. Dabei wurden die SIMA II-Probanden/-innen der Kontrollgruppe gegenübergestellt. Die Kontrollgruppe war zum Zeitpunkt der anlassbezogenen Verurteilung und zum Beginn des Nachbeobachtungszeitraumes im Mittel etwas jünger als die SIMA II-Klientel. Die SIMA II-Probanden/-innen hatten im Mittel wiederum mehr Vorstrafen als die Probanden/-innen der Kontrollgruppe. Anhand des ebenfalls in der Tabelle dargestellten Effektmaßes Cohen's d ist ersichtlich, dass es sich bei diesen Unterschieden jedoch lediglich um kleine Effekte handelte (Cohen 1988; Rice/Harris 2005). In Bezug auf das Rückfallrisiko und damit die Risikoäquivalenz der beiden Gruppen können diese Unterschiede allerdings als irrelevant eingestuft werden, da sie durch die Anwendung der OGRS 3 als globaler Gesamtrisikoindex ausreichend berücksichtigt worden waren.

Tabelle 40: *Offender Group Reconviction Scale* (OGRS 3)-Merkmale der SIMA II-Probanden/-innen und der Kontrollgruppe

Merkmal	SIMA II- Probanden/- innen (<i>n</i> = 1 307)	Kontrollgruppe (<i>n</i> = 1 307)	Testwerte		
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>t</i>	<i>p</i>	<i>d</i>
Alter bei erster Verurteilung	21.87 (8.91)	21.81 (9.41)	0.16	.871	-
Alter bei anlassbezogener Verurteilung	35.47 (10.07)	33.09 (9.30)	6.25	< .001	0.25
Alter zu Beginn der Prognose	38.31 (10.44)	35.38 (9.63)	7.46	< .001	0.29
Anzahl der Vorstrafen	9.26 (6.55)	7.96 (6.15)	5.23	< .001	0.21

4.3.1 Vergleich der Rückfallraten

Unterschiede zwischen den SIMA II-Probanden/-innen und der Kontrollgruppe in Bezug auf verschiedene Rückfallereignisse sind in Tabelle 41 dargestellt. Gegenüber der Kontrollgruppe reduzierten sich allgemein kriminelle Rückfälle bei der SIMA II-Klientel um 22.8 %, StGB-Verstöße um 31.0 % und Gewaltdelikte um 37.4 %. Die Rate an Freiheitsstrafen während des Nachbeobachtungszeit-

raums reduzierte sich sogar um 44.7 %. Als zusätzliches, in Rückfallstudien häufig verwendetes Effektmaß (vgl. z.B. Koehler u.a. 2013) wurde die *Odds Ratio* (OR) berechnet, welche die Rückfallraten beider Gruppen miteinander ins Verhältnis setzt. Dabei drücken Werte > 1 ein erhöhtes Rückfallrisiko der Kontrollgruppe im Vergleich zur Behandlungsgruppe aus. Die OR wurde nur für signifikante Unterschiede berechnet. Im Vergleich zu der SIMA II-Klientel hatten Personen der Kontrollgruppe bspw. ein 1.73-fach erhöhtes Risiko, erneut wegen StGB-Verstößen verurteilt zu werden. Aufgrund der geringen Anzahl an sexuell motivierten Rückfällen – lediglich 13 Personen der Gesamtstichprobe wurden mit Sexualdelikten rückfällig – waren die Auswertungen in Bezug auf dieses Rückfallereignis nicht aussagekräftig.

Tabelle 41: Vergleich der Rückfälligkeit in Form rechtskräftiger Verurteilungen (BZR) zwischen SIMA II-Probanden/-innen und der Kontrollgruppe

Rückfallereignis	Rückfallquote in % (n)		Testwerte		
	SIMA II- Probanden/- innen (n = 1 307)	Kontroll- gruppe (n = 1 307)	χ^2	p	OR
Allgemeiner Rückfall	32.7 (428)	42.4 (554)	26.06	< .001	1.51
StGB-Verstoß	26.6 (348)	38.6 (504)	42.56	< .001	1.73
Nichtsexuelle Gewalt	8.0 (104)	12.7 (166)	15.93	< .001	1.68
Sexualdelikt	0.4 (5)	0.6 (8)	0.70	.404	-
Freiheitsstrafe	16.7 (218)	30.1 (394)	66.09	< .001	2.16

In Tabelle 42 wurden die Rückfallvergleiche getrennt nach unterschiedlichen Risikogruppen berechnet. Dafür wurde die Stichprobe anhand ihrer OGRS 3-Risikowerte (1-Jahreszeitraum) in drei Gruppen geteilt. Die Gruppe mit niedrigem Risiko enthielt alle Personen mit einem OGRS 3-Rückfallrisiko unter 25 %, die Gruppe mit moderatem Risiko enthielt Personen mit einem Rückfallrisiko von 25 % bis 50 % und Personen mit einem OGRS 3-Rückfallrisiko über 50 % bildeten die Hochrisikogruppe. Für die SIMA II-Probanden/-innen mit niedrigem Risiko reduzierten sich im Vergleich zur Kontrollgruppe allgemeine Rückfälle um

11.2 %, erneute StGB-Verstöße um 16.4 % und Freiheitsstrafen um 14.9 %. Innerhalb der Gruppe mit moderatem Risiko reduzierten sich allgemeine Rückfälle um 24.2 %, StGB-Verstöße um 32.7 % und Rückfälle in Form von Gewaltdelinquenz um 35.7 %. Erneute Freiheitsstrafen reduzierten sich bei Personen mit moderatem Risiko um 50.5 %. Innerhalb der Hochrisikogruppe reduzierten sich allgemeine Rückfälle um 20.3 %, erneute StGB-Verstöße um 26.2 % und erneute Gewaltstraftaten sogar um 47.6 %, Freiheitsstrafen gingen um 37.1 % zurück.

Tabelle 42: Vergleich der Rückfälligkeit in Form rechtskräftiger Verurteilungen (BZR) zwischen SIMA II-Probanden/-innen und der Kontrollgruppe, getrennt nach OGRS 3-Risikogruppen

Rückfallereignisse pro Risikogruppe ^a	Rückfallquote in % (n)		Testwerte		
	SIMA II-Probanden/-innen	Kontrollgruppe	χ^2	p	OR
Niedriges Risiko: <25 %	n = 478	n = 478			
Allgemeiner Rückfall	22.6 (108)	29.1 (139)	5.25	.022	1.40
StGB-Verstoß	16.7 (80)	24.9 (119)	9.65	.002	1.64
Nichtsexuelle Gewalt	5.4 (26)	7.7 (37)	2.06	.152	-
Sexualdelikt	0.0 (0)	0.4 (2)	2.00	.157	-
Freiheitsstrafe	10.0 (48)	16.5 (79)	8.73	.003	1.77
Moderates Risiko: 25-50 %	n = 612	n = 612			
Allgemeiner Rückfall	35.0 (214)	46.2 (282)	15.87	< .001	1.59
StGB-Verstoß	28.1 (172)	41.7 (255)	25.00	< .001	1.83
Nichtsexuelle Gewalt	9.2 (56)	14.2 (87)	7.67	.006	1.65
Sexualdelikt	0.5 (3)	0.8 (5)	0.50	.478	-
Freiheitsstrafe	17.0 (104)	34.3 (210)	48.13	< .001	2.55
Hohes Risiko: >50 %	n = 217	n = 217			
Allgemeiner Rückfall	48.8 (106)	61.3 (133)	6.79	.009	1.66
StGB-Verstoß	44.2 (96)	59.9 (130)	10.67	.001	1.88
Nichtsexuelle Gewalt	10.1 (22)	19.4 (42)	7.33	.007	2.13
Sexualdelikt	0.9 (2)	0.5 (1)	0.34	.562	-
Freiheitsstrafe	30.4 (66)	48.4 (105)	14.69	< .001	2.15

^a Die Gruppenbildung basierte auf den Rückfallprognosen der *Offender Group Reconviction Scale* (OGRS 3) für einen 1-Jahreszeitraum

Des Weiteren wurden Rückfallvergleiche speziell für diejenigen Probanden/-innen der Stichprobe berechnet, die nach vollverbüßter Freiheitsstrafe und dementsprechend unter Führungsaufsicht in die Betreuung kamen. Auch bei dieser Subgruppe gab es keinen signifikanten Unterschied zwischen der SIMA II-Klientel und der Kontrollgruppe hinsichtlich des Ausgangsrisikos (erneut quantifiziert anhand des OGRS 3-Risikowerts), insofern konnte auch bei diesem Vergleich von Risikoäquivalenz der Kontrollgruppe aufgrund des durchgeführten Matching-Verfahrens ausgegangen werden. Die Ergebnisse dieser Berechnungen sind in Tabelle 43 dargestellt und verdeutlichen, dass auch bei dieser speziellen Substichprobe der SIMA II-Klientel – abgesehen von den Sexualdelikten, die allerdings auch hier nur eine sehr geringe Zahl ausmachten und insofern kaum interpretierbar waren – alle Rückfallmaße signifikant niedriger waren im Vergleich zur Kontrollgruppe. Im Gegensatz zur Kontrollgruppe reduzierten sich die allgemeinen Rückfälle der im SIMA II behandelten Vollverbüßer/-innen um 26.1 %, StGB-Verstöße um 37.8 % und nichtsexuelle Gewaltdelikte um 64.5 %. Die im SIMA II betreuten Vollverbüßer/-innen hatten zudem eine um 53.2 % reduzierte Rate von Freiheitsstrafen im Vergleich zu den Vollverbüßer/-innen der Kontrollgruppe.

Tabelle 43: Rückfallvergleiche der Vollverbüßer/-innen

Rückfallereignis	Rückfallquote in % (n)		Testwerte		
	SIMA II- Probanden/- innen	Kontroll- gruppe	χ^2	p	OR
Vollverbüßer/-innen (§ 68f StGB)	n = 549	n = 549			
Allgemeiner Rückfall	36.6 (201)	49.5 (272)	18.72	< .001	1.70
StGB-Verstoß	29.7 (163)	47.7 (262)	37.63	< .001	2.16
Nichtsexuelle Gewalt	6.0 (33)	16.9 (93)	32.28	< .001	3.19
Sexualdelikt	0.2 (1)	0.2 (1)	0.00	1.000	-
Freiheitsstrafe	18.9 (104)	40.4 (222)	60.75	< .001	2.91

In Tabelle 44 werden Rückfallvergleiche für Probanden/-innen mit bzw. Probanden/-innen ohne Gewaltanlassdelikt dargestellt. Bei beiden Subgruppen gab es keine signifikanten Unterschiede zwischen der SIMA II-Klientel und der Kon-

trollgruppe hinsichtlich des Ausgangsrisikos (wiederum quantifiziert durch den OGRS 3-Risikowert), insofern kann auch bei diesen Berechnungen von Risikoäquivalenz der Experimental- und Kontrollgruppe ausgegangen werden. Die Berechnungen für beide Gruppen zeigten erneut signifikante Rückfallreduktionen auf: Bei den Probanden/-innen, die aufgrund von Gewaltdelinquenz in die Betreuung kamen, reduzierten sich allgemeine Rückfälle um 8.3 %, StGB-Rückfälle um 11.3 %, gewalttätige Rückfälle um 12.3 % und Freiheitsstrafen um 17.5 %. Bei Probanden/-innen, die wegen anderer als gewalttätiger Delikte in Betreuung kamen, reduzierten sich allgemeine Rückfälle hingegen um 36.4 %, StGB-Rückfälle um 47.5 %, gewalttätige Rückfälle um 73.7 % und Freiheitsstrafen um 60.5 %.

Tabelle 44: Rückfallvergleiche von Probanden/-innen mit bzw. ohne Gewaltanlassdelikt

Rückfallereignis	Rückfallquote in % (n)		Testwerte		
	SIMA II- Probanden/- innen	Kontroll- gruppe	χ^2	p	OR
Anlassdelikt = Gewalt	n = 945	n = 983			
Allgemeiner Rückfall	32.8 (310)	39.5 (388)	9.38	.002	1.34
StGB-Verstoß	26.8 (253)	34.8 (342)	14.64	< .001	1.46
Nichtsexuelle Gewalt	9.9 (94)	13.4 (132)	5.68	.017	1.41
Sexualdelikt	0.5 (5)	0.7 (7)	0.26	.610	-
Freiheitsstrafe	16.3 (154)	25.3 (249)	23.79	< .001	1.74
Anlassdelikt ≠ Gewalt	n = 362	n = 324			
Allgemeiner Rückfall	32.6 (118)	51.2 (166)	24.48	< .001	2.17
StGB-Verstoß	26.2 (95)	50.0 (162)	41.19	< .001	2.81
Nichtsexuelle Gewalt	2.8 (10)	10.5 (34)	17.03	< .001	4.13
Sexualdelikt	0.0 (0)	0.3 (1)	1.12	.290	-
Freiheitsstrafe	17.7 (64)	44.8 (145)	59.16	< .001	3.77

4.3.2 Diskussion

Projektteil 2 hatte zum Ziel, die Neuerungen, die mit der Implementierung des SIMA II einhergingen, hinsichtlich ihrer Auswirkungen zu evaluieren. Basierend auf einer Stichprobe von über 1 000 Probanden/-innen des SIMA II-Fachbereichs konnte die rückfallpräventive Wirkung der rückfallprognostischen und betreuerischen Maßnahmen analysiert werden. Verglichen wurden diese SIMA II-Probanden/-innen mit einer Kontrollgruppe, die Personen mit äquivalentem Ausgangsrisiko enthielt, die allerdings nicht durch das SIMA II betreut worden waren. Durch diese Risikoäquivalenz wurde sichergestellt, dass potentiell unterschiedliche Rückfallraten der beiden Gruppe

auf die Wirkung der im Fachbereich erfolgten Intervention zurückgeführt werden konnten (vgl. z.B. Duggan 2014; Keßler/Rettenberger 2017; Schmucker 2004). Im Zuge der Vergleichsanalysen zeigten sich zahlreiche Unterschiede in Bezug auf die erhobenen Rückfallparameter. In den Fällen, in denen sich signifikante Unterschiede zeigten, fielen diese zugunsten der SIMA II-Klientel aus: Bei den im SIMA II betreuten Probanden/-innen reduzierte sich die Rückfallquote im Vergleich zur Kontrollgruppe – je nach betrachteter Subgruppe bzw. betrachtetem Rückfallparameter – zwischen 8 % und 74 %. Damit zeigte sich die rückfallpräventive Wirkung des SIMA II in eindrucksvoller Form: Durch die Implementierung dieser neuen Form der Bewährungshilfe konnten signifikant mehr Rückfälle verhindert werden als mit der zuvor durchgeführten konventionellen Bewährungshilfe.

Bei differentieller Betrachtung der drei auf Basis des Ausgangsrisikos gebildeten Gruppen zeigten sich über alle Risikogruppen hinweg Unterschiede zur Kontrollgruppe. Dabei fielen die rückfallpräventiven Effekte bei den Gruppen mit moderatem und hohem Ausgangsrisiko etwas stärker aus als bei der Gruppe mit niedrigem Ausgangsrisiko. Dieses Teilergebnis der vorliegenden Evaluation steht dabei in Einklang mit früheren metaanalytischen Befunden, die stärkere Behandlungseffekte bei Hochrisikogruppen Straffälliger im Vergleich zu Personen mit niedrigerem Ausgangsrisiko nahelegten (Koehler u.a. 2013). Zudem kann dieses Ergebnis auch als Bestätigung des Risiko-Prinzips im Sinne des RNR-Modells (Andrews/Bonta 1994; 2010) gedeutet werden, welches ein grundlegendes Element des Betreuungskonzepts des SIMA II darstellt und dafür plädiert, Behandlungsressourcen zugunsten von Hochrisikogruppen zu verteilen.

Angesichts der Fokussierung des Fachbereichs auf Personen, die wegen Gewaltdelinquenz verurteilt wurden, waren Effekte in Bezug auf gewalttätige Rückfälle von besonderem Interesse. Ein Hauptziel bei der Implementierung von SIMA II bestand in der Verhinderung einschlägiger, d.h. neuerlicher gewalttätiger Rückfälle der betreuten Probanden/-innen, bei denen es sich größtenteils um wegen Gewaltdelikten verurteilte Personen handelte (Müller/Nixdorff/Pirner 2018). Hinsichtlich dieses Rückfallkriteriums zeigten sich nahezu durchweg signifikante Unterschiede zugunsten der SIMA II-Klientel: Sowohl in der Gesamtgruppe als auch in den betrachteten Subgruppen wiesen die SIMA II-Probanden/-innen im Vergleich zur Kontrollgruppe reduzierte Raten gewalttätiger Rückfälle auf. Lediglich in der Gruppe mit niedrigem Ausgangsrisiko erreichte dieser Unterschied nicht die Schwelle statistischer Signifikanz. Positiv hervorzuheben ist, dass auch in Gruppen, die wegen anderer als gewalt-

tätiger Anlassdelikte in Betreuung kamen, signifikante Reduktionen gewalttätiger Rückfälle zu verzeichnen waren.

Auch schwerwiegende Sanktionierungen in Form von Freiheitsstrafen traten bei SIMA II-Probanden/-innen signifikant seltener als in der Kontrollgruppe auf. Dieses Ergebnis spricht dafür, dass sich die kriminalpräventive Wirkung der Betreuung nicht nur auf minderschwere Delikte beschränkte, sondern auch die kriminalpolitisch und viktimologisch besonders relevanten Straftaten, die mit schwerwiegenderen Sanktionen verknüpft waren, reduzieren konnte.

Ebenfalls separat betrachtet wurden die Rückfallentwicklungen in der Gruppe der sog. Vollverbüßer/-innen, die nach vollständiger Erledigung einer Freiheitsstrafe kraft Gesetzes der Führungsaufsicht unterstellt (§ 68f StGB) wurden. Diese Klientel wurde seit 2017 im SIMA II betreut und stellte dort die mit Abstand größte Gruppe innerhalb der Probanden/-innen mit Führungsaufsichtsweisung dar. Nachdem Baur und Kinzig (2015) im Zuge ihrer Evaluation der Führungsaufsicht bei dieser besonders rückfallgefährdeten Klientel (siehe auch Jehle u.a. 2016) mit negativer Sozialprognose ein Betreuungsdefizit feststellten, sollte die Implementierung des SIMA II diesem Defizit entgegenwirken. Auch die Auswertungen im Rahmen der vorliegenden Arbeiten verdeutlichten ein besonderes Risikopotential für diese Gruppe: Im Vergleich zur restlichen SIMA II-Klientel wurden Personen, die nach Vollverbüßung ihrer Freiheitsstrafe in die Betreuung kamen, signifikant häufiger rückfällig (siehe Abschnitt 4.1.2). Wie die im vorherigen Abschnitt 4.3.1 dargestellten Analysen jedoch zeigten, wiesen die im SIMA II betreuten Probanden/-innen, die nach einer Vollverbüßung ihrer Strafe in Betreuung kamen, signifikant niedrigere Rückfallraten als vergleichbare Probanden/-innen ohne SIMA II-Betreuung auf. Damit sind solche Probanden/-innen gemeint, die ebenfalls nach einer Vollverbüßung, allerdings vor Implementierung des SIMA II von der hessischen Bewährungshilfe betreut wurden. Probanden/-innen dieser letztgenannten Vergleichsgruppe wurden mehr als dreimal so häufig mit erneuten Gewaltdelikten rückfällig wie die im SIMA II betreuten Vollverbüßer/-innen. Auch wenn von dieser Gruppe nach Implementierung des neuen Betreuungskonzepts vergleichsweise viele kriminelle Rückfälle ausgingen, sprechen die Ergebnisse doch für nachweislich bessere Betreuungseffekte durch das SIMA II im Vergleich zu vorher.

Einen Großteil der SIMA II-Klientel bildeten Personen, die nach Verurteilungen aufgrund von Gewaltdelikten in die Betreuung kamen. Im Vergleich zu dieser Gruppe ist die Führungsaufsichtsklientel, die aufgrund sehr unterschiedlicher

Delikte in die Betreuung gekommen war, deutlich heterogener. Das impliziert, dass Probanden/-innen sehr unterschiedliche kriminogene Faktoren und Bedürfnislagen aufweisen können, woraus besondere Herausforderungen für die Betreuungsarbeit resultieren. Aus diesem Grund sollten differentielle Rückfallvergleiche Aufschluss darüber geben, ob sich dieser Umstand – d.h., ob Personen vor ihrer Aufnahme im SIMA II wegen gewalttätiger oder anderer Delikte verurteilt wurden – auf die kriminalpräventive Effektivität der Betreuung auswirkte. Dabei wurde deutlich, dass auch Personen, die wegen anderer als gewalttätiger Delinquenz in Betreuung waren, von der Betreuung im SIMA II signifikant profitieren konnten. Im Vergleich zur Kontrollgruppe wurden die im SIMA II betreuten Probanden/-innen dieser Subgruppe (Anlassdelikt ≠ Gewalt) deutlich seltener rückfällig.

In der Gesamtschau wiesen die Ergebnisse auf eine rückfallpräventive Wirkung des im SIMA II realisierten Betreuungskonzepts hin. Wie stark diese Wirkung ist, lässt sich mithilfe von Vergleichen mit den Ergebnissen ähnlicher Untersuchungen zur Wirksamkeit der Behandlung von straffällig gewordenen Personen veranschaulichen. In seinem Übersichtsartikel fasste Lösel (2016) die mittleren Behandlungseffekte diverser Metaanalysen zusammen und berichtete von Effektstärken zwischen $d = 0.10$ und $d = 0.28$, bzw. – umgerechnet in *Odds Ratios* – zwischen $OR = 1.20$ und $OR = 1.67$. Die Analysen, die im Rahmen der vorliegenden Arbeit durchgeführt wurden, erzielten Effekte, welche überwiegend am oberen Rand dieser von Lösel (2016) berichteten Spannweite lagen. Die von Lösel (2016) ausgewerteten Untersuchungen umfassten allerdings ein sehr breites Spektrum unterschiedlicher Behandlungsprogramme verschiedener Ausrichtungen und es wurden keine separaten Effektmaße für Behandlungskonzepte berichtet, die auf dem RNR-Modell basierten. Koehler u.a. (2013) führten eine Metaanalyse zu Behandlungsprogrammen jugendlicher bzw. junger Straftäter/-innen durch und berichteten im Mittel über alle im Rahmen ihrer Untersuchung betrachteten Studien einen Behandlungseffekt von $OR = 1.34$. Separat berichteten sie Effekte für Programme, die am RNR-Modell orientiert waren. Diese erzielten wiederum einen mittleren Behandlungseffekt von $OR = 1.90$. Die Effekte der vorliegenden Evaluation übertrafen die erstgenannte, jedoch nicht in allen Fällen die letztgenannte von Koehler u.a. (2013) berichtete Effektstärke. Es lässt sich somit feststellen, dass die Implementierung des SIMA II die Rückfälligkeit der Probanden/-innen ähnlich stark reduzierte wie andere auf den RNR-Prinzipien basierende Programme.

Limitationen

Bei der Interpretation der vorgestellten Ergebnisse müssen unterschiedliche methodische Beschränkungen berücksichtigt werden. Zunächst kann auf Basis der durchgeführten Analysen nicht beurteilt werden, worin genau die rückfall-reduzierenden Mechanismen lagen. Gegenstand der vorliegenden Studie war die Systemwirksamkeit des Gesamtkonzepts, die spezifische Wirksamkeit einzelner Komponenten konnte anhand des vorliegenden Untersuchungsdesigns nicht überprüft werden. Der Wirksamkeitsnachweis einzelner Aspekte des Gesamtbetreuungskonzepts muss jedoch aus methodischer Sicht ohnehin zurückhaltend eingestuft werden. Um entsprechend differenzierte Aussagen treffen zu können, bedürfte es jedenfalls weiterer Evaluationen, bei der einzelne Betreuungsaspekte systematisch variiert werden müssten, um sie hinsichtlich ihrer Wirkung untersuchen zu können. Zudem könnten einzelne Behandlungselemente lediglich einen (kleinen) Teil der Varianz der kriminellen Rückfälligkeit erklären, was bedeutet, dass zusätzlich andere Faktoren entscheidend für die Legalbewährung einer Person waren und sind (vgl. Lösel 2012).

Obgleich durch das Matching Risikoäquivalenz auf der Basis eines bereits etablierten und mehrfach validierten Prognoseinstrumentes hergestellt werden konnte, war nicht auszuschließen, dass es andere systematische und risikorelevante Unterschiede zwischen den beiden Gruppen gab, die für die gefundenen Unterschiede mitverantwortlich sein könnten. Da es einen mehrjährigen Unterschied im Betreuungszeitpunkt der verglichenen Probanden/-innengruppen gab, könnten sich bspw. potentiell zeitabhängige Diskrepanzen in der Zusammensetzung der beiden Gruppen systematisch auf deren Legalbewährung ausgewirkt haben. Ein solcher Unterschied könnte auch die Zusammensetzung der Führungsaufsichtsklientel betreffen.

Seit der Reform der Führungsaufsicht im Jahre 2007 (vgl. Dessecker 2015) sind die Zahlen der Personen, die der Führungsaufsicht unterstellt wurden, laufend gestiegen (vgl. Kinzig 2015). Kinzig (2015) führte dies u.a. auf die Ausweitung der Maßregel nach § 68f Abs. 1 StGB zurück und vermutete überdies eine zunehmend restriktive Handhabung der Strafrestaussatzung nach § 57 StGB. Beides habe zur Folge, dass bei mehr Personen als zuvor die Führungsaufsicht nach Vollverbüßung einer Freiheitsstrafe eingesetzt habe. Ein damit übereinstimmender Trend deutete sich auch in der vorliegenden Stichprobe an: Der Anteil an Vollverbüßer/-innen war in der Behandlungsgruppe höher als in der Vergleichsgruppe, deren Auswertungszeitraum¹¹ 3.5 Jahre vor dem der Behand-

¹¹ Auswertungszeitraum der Vergleichsgruppe: 01.01.2014 – 31.12.2014.

lungsgruppe¹² angesetzt wurde. Dies könnte zur Folge gehabt haben, dass sich die Gruppen der Vollverbüßer/-innen zwischen den beiden Auswertungszeiträumen systematisch unterschieden, da sich die rechtlichen Selektionskriterien für diese Gruppe im Laufe der letzten Jahre aus den o.g. Gründen verändert hatten.

Anzumerken ist auch, dass sich die gefundenen Unterschiede in den Rückfallraten lediglich auf den analysierten Nachbeobachtungszeitraum bezogen. Angesichts des eher kurzen Nachbeobachtungszeitraums war es wahrscheinlich, dass sich einige der Probanden/-innen während des gesamten Nachbeobachtungszeitraumes noch in Betreuung befanden. Unklar ist folglich, auf welche Weise bzw. ob sich die vorgefundenen Unterschiede ändern würden, sobald die Bewährung und damit die mitunter engmaschige Betreuung endet. Aus anderen Rückfalluntersuchungen wurde der Effekt berichtet, dass sich Unterschiede zwischen Behandlungs- und Kontrollgruppe im Laufe der Nachbeobachtungszeit reduzierten, insbesondere nachdem die Betreuung bzw. Behandlung beendet war (Sauter/Voss/Dahle 2015). Vor diesem Hintergrund muss bedacht werden, dass anhand der vorliegenden Daten keine Aussagen zur Nachhaltigkeit des Betreuungseffekts getroffen werden konnte.

4.4 Projektteil 3: Qualitative Interviews mit Bewährungshelfer/-innen

Für diesen Projektteil wurden neun qualitative Interviews mit Mitarbeiter/-innen des SIMA II durchgeführt. Bei den Teilnehmenden handelte es sich sowohl um Personen von der Ausgangsliste geeigneter Personen ($n = 7$), als auch um Mitarbeiter/-innen, die sich eigeninitiativ um die Teilnahme an der Befragung bewarben ($n = 2$)¹³. Zwei Interviews wurden in den Räumlichkeiten der Kriminologischen Zentralstelle, fünf in Räumlichkeiten der Bewährungshilfe und zwei am Telefon geführt. Die Interviews dauerten zwischen 30 und 50 Minuten und im Mittel 42 Minuten (reine Tonaufnahme; exklusive Begrüßung, Aufklärung und Verabschiedung). Die folgenden inhaltlichen Auswertungen der Interviews wurden stellenweise mit wörtlichen Zitaten versehen, die das jeweilige Teilergebnis illustrieren sollten. Die Zitate wurden in Anführungszeichen gesetzt und durch kursive Schrift hervorgehoben. Um die Anonymität der Befragten zu ge-

¹² Auswertungszeitraum der Behandlungsgruppe: 01.07.2017 – 30.06.2018.

¹³ Aus Gründen der Wahrung der Anonymität wird im Folgenden keine genauere Auskunft bzgl. der Eigenschaften der Teilnehmer/-innen sowie der Teilnahmebedingungen gegeben.

währleisten, wurde auch auf eine pseudonymisierte Form der Zuordnung der Zitate verzichtet; eine solche Zuordnung könnte die Identifizierung einzelner Personen über die Kombination mehrerer Zitate ermöglichen.

4.5.1 Auswertungsergebnisse der Interviews

Vom Großteil der Befragten wurde nach der Einführung von SIMA II ein deutlicher Unterschied in der Arbeitsweise der Bewährungshilfe konstatiert. Früher habe man eher bedarfsorientiert gearbeitet, teilweise jedoch auch kriminogene und Risikofaktoren bei der Betreuungsarbeit berücksichtigt (*„Ich denke schon, dass man vorher auch auf Risiko geguckt hat, aber man hatte [...] einen unscharfen Begriff von Risiko.“*). Nach dem Übergang zum SIMA II habe das risikoorientierte Arbeiten durch eine stärkere Berücksichtigung der bisherigen Delikte sowie einer höheren Zahl an kriminogenen Faktoren jedoch noch einmal deutlich zugenommen. In den meisten Fällen wurde diese intensivere Fallarbeit aufgrund des Fokus auf individuelle Risikofaktoren und der höheren Kontaktfrequenz als positiv eingestuft. Weiterhin begrüßten viele Bewährungshelfer/-innen die Orientierung an wissenschaftlichen Kriterien. Man arbeite nicht mehr *„nur nach Bauchgefühl“* und eine *„valide, begründete Einschätzung“* mittels einer Risikoeinschätzung sei *„nachvollziehbar [...] und gut“*. Die Risikoeinschätzung trage zudem zur Strukturierung der Arbeit bei und habe in Kombination mit der genaueren Dokumentation *„eine Vergleichbarkeit geschaffen“*. Demgegenüber habe man vor der Einführung des SIMA II zuweilen das Gefühl gehabt, seine Arbeit als *„Einzelkämpfer“* zu verrichten, bei der jeder *„seine Sozialarbeit, seine Bewährungshilfe gemacht hat, wie er sie für richtig hielt“*.

Insbesondere für die Gruppe der Personen, die wegen Gewaltdelikten verurteilt wurden, wurde das SIMA II als positiv eingeschätzt. In dem Zusammenhang war ein häufig genanntes Problem, dass dem Fachbereich auch Personen zugewiesen wurden, deren Bedarfslagen nicht unbedingt mit der Spezialisierung des SIMA II übereinstimmten. Dies betreffe v.a. Personen mit Führungsaufsicht, die nicht aufgrund von Gewaltstraftaten verurteilt worden seien. Probleme bei der SIMA II-spezifischen Betreuung ergäben sich insbesondere bei Personen, die aufgrund von Straftaten gemäß BtMG verurteilt worden seien, aber auch bei Personen mit psychischen Störungen ohne Krankheitseinsicht sowie bei Personen, die aufgrund der Vollverbüßung ihrer Freiheitsstrafe in Betreuung seien (§ 68f StGB). Insbesondere bei letztgenannter Gruppe träten des Öfteren Diskrepanzen zwischen den richterlich angeordneten Kontaktfrequenzen und der Maßgabe auf Grundlage der Risikoeinschätzung auf. Dies führe zu Unsicherheiten und beeinträchtige eine positive Beziehungsgestaltung zwischen

Probanden/-innen und Bewährungshelfer/-innen. Diese Probanden/-innen müssten sich, trotz eines Gerichtsbeschlusses, der lediglich monatliche Meldungen bei der/dem zugeteilten Bewährungshelfer/-in fordere, nach der Risikoeinschätzung mitunter wöchentlich melden. Dies werde von den Probanden/-innen teilweise nur schwer akzeptiert. Eine Bewährungshelferin berichtete in diesem Zusammenhang von der Erfahrung, die Akzeptanz und Beziehungsgestaltung fördern zu können, indem man diese Problematiken aktiv im Gespräch mit den Probanden/-innen aufgreife und erkläre, warum dieses Vorgehen nun notwendig sei.

Zudem wurde sehr häufig die Klientel mit Suchtproblematik aus der Gruppe der Personen mit Führungsaufsicht erwähnt, deren Bedarfslagen man im Rahmen der SIMA II-Konzeption nicht angemessen berücksichtigen könne. Diese Probanden/-innen seien „sehr unstrukturiert“ und man könne sie nicht in „so eine Struktur zwingen“, die mit dem SIMA II einhergehe. Solche Personen bräuchten eher ein „klinisches Setting“, das im Rahmen des SIMA II jedoch nicht zu leisten sei. Dies träfe auch auf Personen mit psychischen Störungen ohne Krankheitseinsicht zu. Die eigentliche Spezialisierung im SIMA II auf Personen, die wegen Gewaltdelikten verurteilt wurden, verfehle aufgrund der heterogenen Zusammensetzung der Probanden/-innen damit teilweise ihren Zweck: „[...] Es gibt Klienten, die Ressourcen fressen, und die da [ins SIMA II] aus betreuender Sicht nicht hingehören“.

Ein weiterer Schwachpunkt bei der Umsetzung des Erlasses sei die unzureichende Berücksichtigung von Frauen, auf die weder die Risikoeinschätzung, noch die Betreuung zugeschnitten sei. Ein weiteres, häufig genanntes Defizit betreffe die zeitliche Umsetzbarkeit des Erlasses im Alltag. In diesem Aspekt hätten sich zuvor gehegte Befürchtungen bestätigt, dass die Vorgaben zu Hausbesuchen bei den Klienten/-innen hohen Fahrtaufwand und Mehrkosten erzeugen würden. Gleichzeitig sei der Dokumentationsaufwand stark gestiegen. Dies sei zwar, wie bereits erwähnt, ein positives Strukturierungsmerkmal, beanspruche jedoch auch viel Zeit und setze die Bewährungshelfer/-innen zu Lasten der praktischen Betreuungsarbeit unter Druck: „[...] Weil wir dauernd die Idee haben: du musst das dokumentieren, [...], also diese ganzen Vorgaben nehmen so viel Zeit in Anspruch, die eigentlich dem Probanden ein Stück weit fehlen.“ Hinsichtlich der Hausbesuche wurde zudem vereinzelt kritisiert, dass während dieser kein Zugriff auf die Daten der Probanden/-innen möglich sei, und nicht immer ein vertraulicher Gesprächsraum gewährleistet werden könne. Einige Bewährungshelfer/-innen sahen in den Hausbesuchen aufgrund der Eigenheiten der SIMA II-

Klientel auch ein Risiko für ihre eigene Sicherheit; so sollte man teilweise Hausbesuche durchführen, „wo selbst die Polizei nur zu zweit hingeht.“

Auch die Fallzahl pro Bewährungshelfer/-in würden einige Befragte als zu hoch empfinden, woraus mitunter Arbeitsüberlastungen resultieren könnten. Ein (teilweise) dadurch bedingter erhöhter Krankenstand führe wiederum zu einer Verschärfung der Situation für die restlichen Mitarbeiter/-innen.

Die Durchführung von Fallkonferenzen wurde ambivalent aufgenommen. Auf der einen Seite könnten bei diesen Anlässen Einzelfälle besprochen sowie Schwierigkeiten zur Einhaltung von Fristen kommuniziert werden (z.B. aufgrund der bereits erwähnten hohen zeitlichen Beanspruchung), gleichzeitig seien die Fallkonferenzen ein Bestandteil der Vorgaben, die viel Zeit in Anspruch nähmen. In diesem Zusammenhang wurde zudem ein fehlender Ermessensspielraum seitens der Bewährungshelfer/-innen kritisiert, beispielsweise für eine erste Abstufung der Klienten/-innen: „Das Überformale ist letztlich sehr, sehr einengend“.

Risikoeinschätzung mit SVG-5 und Erhebungsbogen

Der Zugriff auf die für die Risikoeinschätzung notwendigen Informationen sei zumeist gewährleistet, Informationen aus dem Vollzug hätten jedoch, v.a. in der Anfangsphase nach der Einführung, gefehlt. Das Informationsdefizit habe zu Verzögerungen bei der Betreuung bzw. zu unpassenden Einstufungen geführt, da der Erlass für Probanden/-innen die maximale Kontaktfrequenz vorschreibe, solange keine Risikoeinschätzung durchgeführt worden sei. Insgesamt habe sich jedoch der Zugriff auf die Daten über die Zeit hinweg erheblich verbessert.

Der SVG-5 erhielt von den Befragten eine durchschnittliche Gesamtnote von 2.1 (zu interpretieren wie Schulnoten). Dabei wurde vor allem die schnelle und einfache Anwendbarkeit des Fragebogens geschätzt, was zu einer hohen Akzeptanz sowohl bei den Bewährungshelfer/-innen als auch den Probanden/-innen führen würde. Vorteilhaft sei auch, dass der SVG-5 ohne persönlichen Kontakt zu den Probanden/-innen angewendet werden könne. Nach Ansicht der meisten Befragten bilde der Score das Risiko angemessen ab und man arbeite mit einem „validierten Instrument, das wissenschaftlich abgesichert“ sei. Kritisch anzumerken sei jedoch, dass der Fragebogen einen zu großen Zeitraum bzw. keine deliktfreien Episoden berücksichtige und Bagatelldelikte mit einbezogen werden müssten. Ein weiteres Defizit sei, dass die Haftentwicklung nicht berücksichtigt werde.

Der Erhebungsbogen erhielt von den Befragten hingegen die Durchschnittsnote 4.5: Von einzelnen Bewährungshelfer/-innen wurde positiv erwähnt, dass der Fragebogen Anhaltspunkte für die Gesprächsführung mit den Probanden/-innen liefern würde. Außerdem berücksichtige er Aspekte für ein straffreies Leben und die persönliche Sicht der Probanden/-innen, da das Instrument in Zusammenarbeit mit den Probanden/-innen bearbeitet werden müsse. Dieser letztgenannte Aspekt führe jedoch zuweilen auch zu Problemen, da diese Zusammenarbeit mit den Probanden/-innen zeitintensiv und das Instrument daher vergleichsweise unökonomisch sei (*„Erst wenn er zum Erstgespräch da war und ich den Erhebungsbogen machen kann, dann könnte ich sagen, ich kann ihn runterstufen“*).

Ein wichtiger Kritikpunkt waren zudem einzelne Items des Erhebungsbogens (z.B. Item 4 und 5), die aufgrund eines (zu) großen Ermessensspielraums bei der Beantwortung als wenig objektiv wahrgenommen werden würden. Des Weiteren hätten die Items zu Sucht (Item 10) bzw. zur strafrechtlichen Vorbelastung (Item 6 und 7) ein unpassendes Antwortformat bzw. eine unpassende Gewichtung. Allgemein sei *„der Erhebungsbogen [...] überhaupt nicht ausgereift.“* Es würden *„[...] Faktoren gemischt und miteinander in Relation gesetzt, die nichts miteinander zu tun haben“*. *„Am Ende fühlt es sich nicht wie Wissenschaft an, sondern wie ein Bauchgefühl, und das sollte ja eigentlich in SIMA II vermieden werden.“* Des Weiteren sei der Erhebungsbogen unpassend für Personen, die nach einer Verurteilung aufgrund von Betrugs- oder Betäubungsmitteldelikten in die Betreuung kämen. Der Fragebogen berücksichtige kaum Ressourcen und die Cut-off Werte seien zu niedrig, sodass sehr viele Personen einer hohen Betreuungsstufe zugeordnet werden müssten (*„[...] das bringt uns nicht weiter.“*). Insbesondere die Klientel mit Betäubungsmitteldelikten würde oftmals hohe Scorings und somit einen intensiven Fokus auf die risikoorientierte Rückfallprävention erhalten. Diese Personen hätten jedoch meistens, aufgrund der Suchtproblematik, mit ganz anderen Schwierigkeiten zu kämpfen wie bspw. der Organisation des Alltags oder der Integration in den Arbeitsmarkt: *„[Solch ein Klient] kriegt ja so schon sein Leben nicht hin.“* Insgesamt wurde der Fragebogen von fast allen Bewährungshelfer/-innen eher als defizitär eingestuft. Der Erhebungsbogen sei *„schlicht unpraktikabel, und zwar [...] nahezu komplett.“*

Empfehlungen und Ausblick

Auf Basis der genannten Kritikpunkte formulierten die Bewährungshelfer/-innen Ideen und Verbesserungsvorschläge. Die Schulungsmaßnahmen im Übergang zu SIMA II wurden positiv bewertet, man wünsche sich jedoch noch mehr

Fort- und Weiterbildung, insbesondere angesichts der Menge an Aufgaben und der heterogenen Klientel. Dies beziehe sich auch auf weitere Schulungen zur Anwendung der Prognoseinstrumente, sodass Unklarheiten bezüglich der theoretischen Hintergründe einzelner Items und unterschiedlicher Gefährlichkeitseinschätzungen trotz ähnlicher Ausgangslagen ausgeräumt werden könnten. In diesem Zusammenhang würde man einen intensiveren Austausch befürworten, auch zwischen den verschiedenen Landgerichtsbezirken, *„sodass ein standardisiertes Arbeiten ermöglicht und vorangetrieben werden könnte“*.

Um die Bewährungshelfer/-innen zeitlich zu entlasten sowie eine korrekte und umfassende Dokumentation zu gewährleisten, müsse die Fallzahl der Probanden/-innen pro Bewährungshelfer/-in gesenkt werden. Des Weiteren wurde der Wunsch nach mehr Entscheidungsfreiheit und Eigenverantwortung geäußert, um beispielsweise Kontaktfrequenzen, die auf einer unzureichenden bzw. nicht erfolgten Einstufung der Probanden/-innen durch die Risikoinstrumente beruhten, selbstständig und schneller korrigieren zu können.

Nach Einschätzung eines Großteils der Befragten könnte man einem großen Teil der Probleme entgegenwirken, indem der Schwerpunkt von SIMA II auf die wegen Gewaltdelikten verurteilten Probanden/-innen gelegt werde. An deren Bedürfnissen und Problemen habe sich die Konzeption des Fachbereichs ursprünglich orientiert – vergleichbar mit dem Fachbereich SIMA I, in dem ausschließlich Personen betreut würden, die aufgrund von Sexualstraftaten verurteilt worden waren.

Um Unklarheiten zwischen den Kontaktfrequenzvorgaben der Gerichte und der Vorgabe durch die Risikoeinschätzung entgegen zu wirken, bestehe der Wunsch nach einer eindeutigen Anweisung, wie in solchen Fällen zu verfahren sei. Hinsichtlich der Prognoseinstrumente selbst sei eine stärkere Berücksichtigung von dynamischen bzw. protektiven Faktoren wünschenswert.

4.5.2 Diskussion

Die Befragung von Mitarbeiter/-innen des SIMA II verfolgte das Ziel, eine praxisnahe Bestandaufnahme der Implementierung des neuen Betreuungskonzeptes vorzunehmen. Hierzu wurden neun Gespräche mit im Fachbereich arbeitenden Bewährungshelfer/-innen durchgeführt und qualitativ ausgewertet.

Die Auswertungen offenbarten ein gemischtes Bild hinsichtlich der Bewertung der neuen Richtlinien und deren Umsetzung in der Arbeitspraxis. Die grundsätzliche Ausrichtung des Betreuungskonzeptes wurde vom Großteil der Befragten begrüßt; der Anspruch, gemäß wissenschaftlich fundierter Grundlagen zu arbeiten, die Orientierung am individuellen Risiko der Probanden/-innen sowie ein standardisiertes und daher vergleichbares Arbeiten wurden überwiegend positiv beurteilt. Auch die intensivere Zusammenarbeit mit den Probanden/-innen durch die mitunter hohe Kontaktfrequenz wurde trotz des damit einhergehenden zeitlichen und inhaltlichen Aufwandes als wertvoll empfunden.

Nichtsdestotrotz wurden einige Schwachstellen angesprochen, die aus Sicht der interviewten Bewährungshelfer/-innen stellenweise zu Ressourcenengpässen und Belastungserleben führten. Hinsichtlich mancher Vorgaben – z.B. der Häufigkeit von Hausbesuchen und der 6-monatigen Latenzzeit für Abstufungen – wurde der Wunsch nach mehr Eigenverantwortung und Flexibilität kommuniziert. Ein möglicher Rahmen für solche eigenverantwortlichen, vom Erlass abweichenden Entscheidungen wären die wöchentlichen Fallkonferenzen, in denen zusammen mit den Kollegen/-innen in Ausnahmefällen über vorzeitige Abstufungen entschieden werden könnte.

In den Gesprächen deutete sich zudem an, dass die Mitarbeiter/-innen einem wichtigen ursprünglichen Ziel der Neuerungen, nämlich der Spezialisierung auf eine spezifische Klientel mit ähnlichen Bedarfslagen, angesichts der momentan vergleichsweise heterogenen Klientel nicht immer gerecht werden könnten. Der Frage, inwiefern die Zusammensetzung der Klientel auch hinsichtlich der Wirksamkeit der Betreuung in Zweifel gezogen werden sollte, konnte sich mithilfe quantitativer Auswertungen bereits an anderer Stelle des Berichts angenähert werden (siehe Abschnitt 4.1.2 und Abschnitt 4.3.1).

Im Hinblick auf die Frage, ob das Betreuungskonzept des Fachbereichs möglicherweise spezielle Gruppen nur unzureichend berücksichtige, wurde von vielen Befragten die weibliche Klientel genannt. Obgleich diese einen vergleichsweise kleinen Teil der betreuten Probanden/-innen ausmache, würde es hier bislang an einheitlichen diagnostischen und betreuerischen Regelungen fehlen. Der erste Schritt, um eine ausreichende und sachgemäße Berücksichtigung von Frauen zu erreichen, würde die Risikoeinschätzung betreffen: Instrumente, die unabhängig vom Geschlecht anwendbar sind, wären wünschenswert.

Während die Befragten die prognostische Arbeit mit dem SVG-5 überwiegend positiv empfanden, wurde am Erhebungsbogen deutliche Kritik geäußert. Wäh-

rend die Ökonomie und Objektivität des SVG-5 geschätzt wurde, wurden eben diese Vorzüge beim ressourcenintensiveren Erhebungsbogen vermisst. Diese Diskrepanz in der Beurteilung der beiden Instrumente drückte sich auch in den Benotungen aus, die im Zuge der Befragungen erhoben wurden. Da der Erhebungsbogen nach Ansicht der Befragten unangemessen viele Probanden/-innen der höchsten Betreuungsstufe zuordnete, führte er dadurch zur Bindung vieler Ressourcen, die zu Lasten der übrigen Klientel gehen würde. Ob angesichts dieses Befundes eine Überarbeitung oder sogar vollständige Ersetzung des Erhebungsbogens angezeigt ist, sollte unter Einbezug der quantitativen Analyseergebnisse erörtert werden (siehe Abschnitt 4.2.2 bzw. Abschnitt 5).

In den Gesprächen wurde darüber hinaus der Wunsch nach mehr Möglichkeiten der Fort- und Weiterbildung geäußert. Dies könnte dazu beitragen, die ebenfalls geäußerten Schwierigkeiten und Unsicherheiten im Umgang mit der vergleichsweise heterogenen Führungsaufsichtsklientel zu reduzieren.

Die Gespräche verdeutlichten die Herausforderung, in der betreuenden Praxis einen angemessenen und zweckdienlichen Kompromiss zwischen Formalisierung und Flexibilität zu finden, der auch für alle Mitarbeitenden tragbar ist. Vor dem Hintergrund der grundsätzlich positiven Resonanz der Mitarbeitenden in Bezug auf die Zielsetzungen des SIMA II – Spezialisierung und Risikoorientierung – erscheint ein solcher Kompromiss durchaus realisierbar.

5 Fazit

Im Rahmen des hier vorgestellten Projektes wurden Risikoeinschätzung und Betreuungspraxis des SIMA II evaluiert, um wissenschaftlich fundierte Erkenntnisse zur Effektivität und Effizienz der durch den Fachbereich angebotenen Bewährungshilfe zu erhalten. Dabei wurden unterschiedliche Perspektiven und methodische Zugänge berücksichtigt, um ein möglichst differenziertes Gesamtbild zu erhalten. Einerseits wurde die Qualität der Risikoeinschätzung untersucht, die die Basis für die Festlegung unterschiedlicher Betreuungsstufen und die daran anknüpfende Betreuungsintensität bildet. Außerdem wurde die kriminalpräventive Wirkung der im Fachbereich durchgeführten Interventionen überprüft. Ergänzt wurden diese statistisch-quantitativen Zugänge durch qualitativ ausgewertete Interviews mit im Fachbereich tätigen Bewährungshelfer/-innen, um deren Sicht auf die Arbeitspraxis mit dem neuen Betreuungskonzept abzubilden. Dadurch sollte sich auch die Möglichkeit ergeben, die quantitativen Ergebnisse in ein praxisnahes Gesamtbild einzuordnen und die in den Befragungen angesprochenen Probleme mit den quantitativ gewonnenen Erkenntnissen abzugleichen. Im Folgenden werden die Ergebnisse der Evaluation kurz zusammengefasst, um im Anschluss die Erkenntnisse aus den unterschiedlichen Projektteilen zu integrieren.

Der SVG-5 und der Erhebungsbogen erwiesen sich im Zuge der Überprüfung als reliabel (zuverlässig anwendbare, messgenaue) einsetzbare Instrumente. Insbesondere beim Erhebungsbogen wurden jedoch gleichzeitig auf Item-Ebene Probleme identifiziert. Diese Probleme deckten sich mit den Aussagen der im Kontext von Projektteil 3 befragten Bewährungshelfer/-innen des SIMA II. Insbesondere die mangelnde Ökonomie des Erhebungsbogens wurde dabei kritisch beurteilt sowie auch dessen wissenschaftliche Fundierung in Zweifel gezogen. Demgegenüber waren die Befragten mit dem SVG-5 aufgrund seiner stärker ausgeprägten Anwenderfreundlichkeit weitgehend zufrieden. Diese Befunde sprachen dafür, eine inhaltliche Überarbeitung oder eine Ersetzung des Erhebungsbogens durch ein anderes Instrument in Betracht zu ziehen. Angesichts des Wunsches der befragten Mitarbeiter/-innen nach einem ökonomischen Instrument kämen aktuarische Instrumente wie die OGRS 3 infrage, deren Validität im Rahmen dieses Projektes ebenfalls untersucht und belegt werden konnte.

Die OGRS 3 überbot den Erhebungsbogen nicht nur hinsichtlich der meisten vorhergesagten Rückfallparameter, sie war darüber hinaus ebenso ökonomisch

anwendbar wie der SVG-5. Den in den Interviews angesprochenen Verzögerungen bei den Einstufungen und dem daraus resultierenden zusätzlichen Betreuungsaufwand könnte durch die Verwendung eines Instrumentes, für dessen Bearbeitung es lediglich aktenbasierter Informationen bedarf, entgegengewirkt werden. Dies würde auch eine stärkere Vergleichbarkeit mit den restlichen Probanden/-innen sicherstellen, bei denen die Einstufung bereits rein aktuarisch erfolgte. Überdies könnte dabei das in den Befragungen angesprochene Problem aufgegriffen werden, nach dem der Erhebungsbogen zu viele Personen der höchsten Betreuungsstufe zuteilte.

Die OGRS 3 ist für Probanden/-innen mit sehr unterschiedlichen Indexdelikten und darüber hinaus sowohl für Frauen als auch Männer anwendbar. Dies würde die Möglichkeit eröffnen, sämtliche im Fachbereich einzuschätzenden Probanden/-innen auf Basis der OGRS 3 einzustufen – auch diejenigen Probanden/-innen, die bis dato mit dem SVG-5 eingestuft wurden. Beim SVG-5, der von Seiten der Anwender/-innen aufgrund seiner wissenschaftlichen Fundierung und seiner Ökonomie sehr positiv angenommen wurde und der überdies weniger problematische Items beinhaltet, erschien dies jedoch grundsätzlich weniger angezeigt als beim Erhebungsbogen. Es bedarf zur Klärung dieser Fragen nicht zuletzt auch der Berücksichtigung der zeitlichen, personellen und finanziellen Ressourcen, die für die Implementierung eines neuen Instrumentes aufgebracht werden müssten.

Bei den Befragungen wurde deutlich, dass Frauen bislang nur unzureichend berücksichtigt wurden. Die Prognoseinstrumente – primär der SVG 5 – sind nicht ohne weiteres bei Frauen anwendbar und daher auch von fraglicher Aussagekraft. Aufgrund der geringen Fallzahl von Frauen im SIMA II konnten jedoch keine aussagekräftigen differentiellen Analysen durchgeführt werden, auf deren Basis eine Einschätzung der Eignung der Risikobeurteilung und der Betreuung der weiblichen Klientel möglich gewesen wäre. Für eine solche Überprüfung bedürfte es größerer Gruppen im Fachbereich eingestufte und betreute Frauen. Auf Basis der im Rahmen des vorliegenden Projektes untersuchten Stichprobe konnte deshalb auch die Eignung der OGRS 3 für die weibliche SIMA II-Klientel nicht überprüft werden. Die OGRS 3 ist jedoch grundsätzlich bei Frauen anwendbar, da die Normierung auch an weiblichen (Sub-)Stichproben erfolgte; für Frauen liegen separate Regressionskoeffizienten vor, welche geschlechtsspezifische Aspekte der Legalbewährung (vgl. Fazel u.a. 2016; Jehle u.a. 2016) berücksichtigen.

Projektteil 2 verfolgte das Ziel, die Wirksamkeit der Interventionen im SIMA II zu überprüfen. Hierfür wurden im SIMA II betreute Probanden/-innen mit risikoäquivalenten Probanden/-innen der hessischen Bewährungshilfe verglichen, die vor der Implementierung in Betreuung waren. Als Kriterium dienten dabei verschiedene Formen neuerlicher rechtskräftiger Verurteilungen (= Rückfälle). Sowohl für die Gesamtstichprobe als auch bei diversen Subgruppen zeigten sich signifikante Unterschiede zugunsten der SIMA II-Klientel. Die gefundenen Behandlungseffekte waren dabei in etwa auf dem Niveau vergleichbarer forensischer Wirksamkeitsstudien oder lagen darüber (vgl. Koehler u.a. 2013; Lösel 2016).

Die Befragungen der Mitarbeiter/-innen des Fachbereichs zeigten gleichzeitig eine gewisse Sorge hinsichtlich der Heterogenität der betreuten Klientel. Zwar wurde das Konzept von den Befragten für geeignet befunden, den Bedarfslagen der Hauptklientel – nämlich wegen Gewaltdelikten verurteilte Personen – gerecht zu werden, jedoch wurden Zweifel hinsichtlich der Eignung für Personen mit anderen Indexdelikten kommuniziert. Die Untersuchung der rückfallpräventiven Wirkung der Betreuung zeigte allerdings, dass auch Personen, die aufgrund anderer als gewalttätiger Delikte in Betreuung kamen, von den Interventionen profitierten. Im Vergleich zur äquivalenten Subgruppe der Kontrollgruppe wurden diese Probanden/-innen deutlich seltener rückfällig. Das spricht dafür, dass die Heterogenität der Klientel zumindest hinsichtlich der Rückfallprävention keine Einschränkungen der Wirksamkeit der Betreuungstätigkeit bedeutete.

In den Interviews wurde angedeutet, dass bereits vor der Implementierung des SIMA II in Teilen – wenn auch weniger strukturiert, sondern eher implizit – risikoorientiert und an individuellen kriminogenen Faktoren orientiert gearbeitet wurde. Die gefundenen Behandlungseffekte wiesen jedoch darauf hin, dass der Strukturierungsgrad und das Ausmaß der tatsächlichen Berücksichtigung der Risikoorientierung im Sinne des RNR-Modells (Andrews/Bonta 1994; 2010), die sich mit Einführung der neuen Richtlinien auch nach Ansicht der befragten Mitarbeiter/-innen deutlich erhöhten, entscheidend zur Wirksamkeitssteigerung beitragen dürften. Auf derartige Effekte wiesen auch Studien hin, die eine sukzessive Steigerung der Behandlungseffektivität in Abhängigkeit vom Ausmaß der Berücksichtigung der RNR-Prinzipien belegten (Koehler u.a. 2013; Lösel 2016). Derartige Befunde veranschaulichen, dass eine formale bzw. konzeptionelle Berücksichtigung der RNR-Prinzipien noch wenig über die tatsächliche praktische und angemessene Ausgestaltung oder Adhärenz bzgl. dieser Prinzipien aussagt (vgl. z.B. Drawbridge u.a. 2019; Viglione 2019; Vincent

u.a. 2016). Inwieweit in Interventionsprogrammen die konzeptionellen Prinzipien tatsächlich eingehalten werden und wie sich dies auf die Effektivität der Betreuung auswirkt, wird vom sog. *Fidelity*-Prinzip beschrieben (vgl. McGrew u.a. 1994). Entsprechend dieses Prinzips könnte ein höherer Strukturierungsgrad bzw. eine stärkere Adhärenz bei der Berücksichtigung weiterer RNR-Prinzipien (Bedürfnis- und Ansprechbarkeitsprinzip) zu einer nochmaligen Steigerung der Behandlungseffektivität führen.

Die Evaluation des SIMA II fand vor dem Hintergrund statt, den neuen Entwicklungstrend hin zu einer stärkeren Orientierung am Risiko-Prinzip im Sinne des RNR-Modells (Andrews/Bonta 1994; 2010) und hin zu einer Spezialisierung der Bewährungshilfe wissenschaftlich auf seine praktischen Auswirkungen hin zu überprüfen. Dabei zeigten sich einerseits Entwicklungspotentiale, die Möglichkeiten zur Steigerung der Effektivität, der Effizienz und auch der Zufriedenheit der im Fachbereich tätigen Bewährungshelfer-/innen eröffnen. Hinsichtlich der grundsätzlichen Ausrichtung des SIMA II-Konzepts sind die Ergebnisse jedoch sehr ermutigend und deuten auf eine wirksame Prognose- und Interventionspraxis hin, die im Vergleich zu früheren Konzepten zu einer signifikanten Steigerung der rückfallpräventiven Wirkung der Bewährungshilfe in Hessen führten.

Literaturverzeichnis

- Andrews, Donald A. & Bonta, James (1994). *The psychology of criminal conduct*. Cincinnati: Anderson Publishing.
- Andrews, Donald A. & Bonta, James (2010). *The psychology of criminal conduct* (5th ed.). Albany, N.Y.: Lexis Nexis/Anderson Publ.
- Austin, Peter C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46, 3, S. 399–424. doi: 10.1080/00273171.2011.568786
- Bacher, Johann (2002). Statistisches Matching: Anwendungsmöglichkeiten, Verfahren und ihre praktische Umsetzung in SPSS. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 51, S. 38–66. Zugriff unter: <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-199039>
- Baur, Alexander & Kinzig, Jörg (Hrsg.) (2015). *Die reformierte Führungsaufsicht. Ergebnisse einer bundesweiten Evaluation*. Tübingen: Mohr Siebeck.
- Bockshammer, Tamara & Rettenberger, Martin (2015). Eine empirische Untersuchung der dezentralen therapeutischen Versorgung entlassender Sexualstraftäter. Befragung der Therapeuten von Probanden des Sicherheitsmanagements (SIMA) in Hessen. *Recht & Psychiatrie*, 33, S. 183–192.
- Bourgon, Guy & Armstrong, Barbara (2005). Transferring the principles of effective treatment into a “real world” prison setting. *Criminal Justice and Behavior*, 32, 1, S. 3–25. doi: 10.1177/0093854804270618
- Cohen, Joel E. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, N. J.: L. Erlbaum Associates.
- Coid, Jeremy; Yang, Min; Ullrich, Simone; Zhang, Tianqiang; Roberts, Amanda; Roberts, Colin; Rogers, Robert & Farrington, David (2007). Predicting and understanding risk of re-offending: The Prisoner Cohort Study. *Research Summary 6*. London: Ministry of Justice.
- Copas, John & Marshall, Peter (1998). The offender group reconviction scale: A statistical reconviction score for use by probation officers. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47, 1, S. 159–171. doi: 10.1111/1467-9876.00104
- Cottle, Cindy C.; Lee, Ria J. & Heilbrun, Kirk (2001). The prediction of criminal recidivism in juveniles. *Criminal Justice and Behavior*, 28, 3, S. 367–394. doi: 10.1177/0093854801028003005

- Craig, Leam; Beech, Anthony R. & Browne, Kevin Dominic (2008). *Assessing risk in sex offenders. A practitioner's guide*. Chichester, England, Hoboken, NJ: Wiley. Zugriff unter: <http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10297520>
- Dessecker, Axel (2015). Die Führungsaufsicht: Entwicklung, Funktionen und empirische Daten. *Neue Kriminalpolitik*, 27, 3, S. 251–265.
- Döring, Nicola & Bortz, Jürgen (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Aufl.). Berlin, Heidelberg: Springer (Springer-Lehrbuch).
- Drawbridge, Dara C.; Todorovic, Kristina; Winters, Georgia M. & Vincent, Gina M. (2019). Implementation of risk-need-responsivity principles into probation case planning. *Law and Human Behavior*, 43, 5, S. 455–467. doi: 10.1037/lhb0000351
- Duggan, Conor (2014). The empirical basis of sex offender treatment effectiveness. *Sexual Offender Treatment*, 9, 2, S. 1–13. Wien: International Association for the Treatment of Sexual Offenders.
- Eher, Reinhard; Rettenberger, Martin; Schilling, Frank & Pfäfflin, Friedemann (2008). Validität oder praktischer Nutzen? Rückfallvorhersagen mittels Static-99 und SO-RAG. Eine prospektive Rückfallstudie an 275 Sexualstraftätern. *Recht & Psychiatrie*, 26, S. 79–88.
- Eher, Reinhard; Schilling, Frank; Mönichweger, Michael; Haubner-MacLean, Tanja & Rettenberger, Martin (2012). Die revidierte Version des „Screeninginstrumentes zur Vorhersage des Gewalttrisikos“ (SVG-5): Darstellung relativer und absoluter Rückfalldaten. *Monatsschrift für Kriminologie und Strafrechtsreform*, 95, 1, S. 18–31. doi: 10.13109/9783666252884.17
- Farrington, David. P.; Gottfredson, Denise. C.; Sherman, Lawrence W. & Welsh, Brandon C. (2002). The Maryland Scientific Method Scale. In Sherman, Lawrence W.; Farrington, David. P.; Welsh, Brandon. C. & MacKenzie, Doris. L. (Hrsg.), *Evidence-based crime prevention* (S. 13–21). London: Routledge.
- Fazel, Seena; Chang, Zheng; Fanshawe, Thomas; Långström, Niklas; Lichtenstein, Paul; Larsson, Henrik & Mallett, Susan (2016). Prediction of violent reoffending on release from prison: Derivation and external validation of a scalable tool. *The Lancet Psychiatry*, 3, 6, S. 535–543. doi: 10.1016/S2215-0366(16)00103-6
- Francis, Brian; Soothill, Keith & Humphreys, Les (2007). *Development of a reoffending measure using the Police National Computer database*. Lancaster: Centre for Applied Statistics, Lancaster University.

- Gangl, Markus (2010). Nichtparametrische Schätzung kausaler Effekte mittels Matchingverfahren. In Wolf, Christof & Best, Henning (Hrsg.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (1. Aufl.) (S. 931–961). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gläser, Jochen & Laudel, Grit (2010). *Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen* (4. Aufl.). Wiesbaden: VS Verlag.
- Hanczar, Blaise; Hua, Jianping; Sima, Chao; Weinstein, John; Bittner, Michael & Dougherty, Edward R. (2010). Small-sample precision of ROC-related estimates. *Bioinformatics*, 26, 6, S. 822–830. doi: 10.1093/bioinformatics/btq037
- Hanson, R. Karl; Bourgon, Guy; Helmus, Leslie & Hodgson, Shannon (2009). The principles of effective correctional treatment also apply to sexual offenders. *Criminal Justice and Behavior*, 36, 9, S. 865–891. doi: 10.1177/0093854809338545
- Harris, Grant T. & Rice, Marnie E. (2007). Adjusting actuarial violence risk assessments based on aging or the passage of time. *Criminal Justice and Behavior*, 34, 3, S. 297–313. doi: 10.1177/0093854806293486
- Hessisches Ministerium der Justiz (Hrsg.) (2017). *Justiz-Ministerial-Blatt für Hessen*. 69. Jahrgang. Wiesbaden: Hessisches Ministerium der Justiz. Zugriff unter: https://justizministerium.hessen.de/sites/default/files/media/hmdjie/jmbl_022017_s_49_bis_88.pdf
- Howard, Philip; Francis, Brian; Soothill, Keith & Humphreys, Les (2009). OGRS 3: The revised Offender Group Reconviction Scale. *Research Summary*, 2009, 7. London: Ministry of Justice. Zugriff unter: <https://eprints.lancs.ac.uk/id/eprint/49988/1/ogrs3.pdf>
- Jehle, Jörg-Martin; Albrecht, Hans-Jörg; Hohmann-Fricke, Sabine & Tetel, Carina (2016). *Legalbewahrung nach strafrechtlichen Sanktionen. Eine bundesweite Rückfalluntersuchung 2010 bis 2013 und 2004 bis 2013* (1. Aufl.). Mönchengladbach: Forum Verlag Godesberg GmbH. Zugriff unter: http://www.bmjv.de/SharedDocs/Downloads/DE/Service/StudienUntersuchungen/Fachbuecher/Legalbewaehrung_nach_strafrechtlichen_Sanktionen_2010_2013.pdf?__blob=publicationFile&v=1
- Keßler, Achim & Rettenberger, Martin (2017). Die Wirksamkeit psychotherapeutischer Behandlung von Sexualstraftätern nach Entlassung aus dem Strafvollzug. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 46, 1, S. 42–52. doi: 10.1026/1616-3443/a000401
- Kinzig, Jörg (2015). Die Maßregel der Führungsaufsicht: Vom Stiefkind zur Avantgarde? *Neue Kriminalpolitik*, 27, 3, S. 230–250.

- Koehler, Johann A.; Lösel, Friedrich; Akoensi, Thomas D. & Humphreys, David K. (2013). A systematic review and meta-analysis on the effects of young offender treatment programs in Europe. *Journal of Experimental Criminology*, 9, 1, S. 19–43. doi: 10.1007/s11292-012-9159-7
- Landers, Richard (2015). Computing intraclass correlations (ICC) as estimates of inter-rater reliability in SPSS. *The Winnower*.
- Långström, Niklas & Grann, Martin (2002). Psychopathy and violent recidivism among young criminal offenders. *Acta Psychiatrica Scandinavica*, 106, S. 86–92. doi: 10.1034/j.1600-0447.106.s412.19.x
- Lobo, Jorge M.; Jiménez-Valverde, Alberto & Real, Raimundo (2008). AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17, 2, S. 145–151. doi: 10.1111/j.1466-8238.2007.00358.x
- Lösel, Friedrich (2012). Offender treatment and rehabilitation: What works? In Maguire, Mike; Morgan, Rodney & Reiner, Robert (Hrsg.), *The Oxford handbook of criminology* (5th ed.) (S. 986–1016). Oxford: Oxford Univ. Press.
- Lösel, Friedrich (2016). Wie wirksam ist die Straftäterbehandlung im Justizvollzug? In Rettenberger, Martin & Dessecker, Axel (Hrsg.), *Behandlung im Justizvollzug. Kriminologie und Praxis*, 71, S. 17–52. Wiesbaden: Kriminologische Zentralstelle.
- Lowenkamp, Christopher T.; Latessa, Edward J. & Holsinger, Alexander M. (2006). The risk principle in action: What have we learned from 13,676 offenders and 97 correctional programs? *Crime & Delinquency*, 52, 1, S. 77–93. doi: 10.1177/0011128705281747
- Mayer, Klaus (2018). 10 Jahre Risikoorientierte Bewährungshilfe - Wo stehen wir heute? *Bewährungshilfe*, 65, 3, S. 242–257.
- Mayring, Philipp (2010). Qualitative Inhaltsanalyse. In Mey, Günter & Mruck, Katja (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (1. Aufl.) (S. 601–613). Wiesbaden: VS Verlag für Sozialwissenschaften (GWV).
- McGrew, John H.; Bond, Gary R.; Dietzen, Laura & Salyers, Michelle (1994). Measuring the fidelity of implementation of a mental health program model. *Journal of Consulting and Clinical Psychology*, 62, 4, S. 670–678. doi: 10.1037//0022-006x.62.4.670
- Mossman, Douglas (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology*, 62, 4, S. 783–792. doi: 10.1037/0022-006X.62.4.783
- Müller, Jochen; Nixdorff, Andreas & Pirner, Jasmin (2018). Das Sicherheitsmanagement II in Hessen - Ein weiterer Schritt in Richtung verstärkter rückfallpräventiver

- Sozialarbeit in der Bewährungshilfe. *Bewährungshilfe*, 65, 1, S. 5–19. Mönchengladbach: Forum Verlag Godesberg GmbH.
- National Offender Management Service (2008). *Guidance: Offender Group Reconviction Scale. Version 3*. London: Ministry of Justice. Zugriff unter: <http://nomsintranet.org.uk/roh/official-documents/11%20-%20OGRS%203%20guidance.pdf>
- Piquero, Alex R.; Jennings, Wesley G.; Diamond, Brie & Reingle, Jennifer M. (2015). A systematic review of age, sex, ethnicity, and race as predictors of violent recidivism. *International Journal of Offender Therapy and Comparative Criminology*, 59, 1, S. 5–26. doi: 10.1177/0306624X13514733
- Quinsey, Vernon L.; Harris, Grant T.; Rice, Marnie E. & Cormier, Catherine A. (2006). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.
- Prendergast, Michael L.; Pearson, Frank S.; Podus, Deborah; Hamilton, Zachary K. & Greenwell, Lisa (2013). The Andrews' principles of risk, need, and responsivity as applied in drug abuse treatment programs: Meta-analysis of crime and drug use outcomes. *Journal of Experimental Criminology*, 9, 3, S. 275–300. doi: 10.1007/s11292-013-9178-z
- Rehn, Gerhard; Wischka, Bernd; Lösel, Friedrich & Walter, Michael (Hrsg.) (2001). *Behandlung „gefährlicher Straftäter“*. Grundlagen, Konzepte, Ergebnisse. (2. Aufl). *Studien und Materialien zum Straf- und Maßregelvollzug*, 11. Herbolzheim: Centaurus Verlag.
- Rettenberger, Martin; Briken, Peer; Turner, Daniel & Eher, Reinhard (2015). Sexual offender recidivism among a population-based prison sample. *International Journal of Offender Therapy and Comparative Criminology*, 59, 4, S. 424–444. doi: 10.1177/0306624X13516732
- Rettenberger, Martin; Mönichweger, Michael; Buchelle, Elvira; Schilling, Frank & Eher, Reinhard (2010). Entwicklung eines Screeninginstruments zur Vorhersage der einschlägigen Rückfälligkeit von Gewaltstraftätern. *Monatsschrift für Kriminologie und Strafrechtsreform*, 93, 5, S. 346–360.
- Rice, Marnie E. & Harris, Grant T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology*, 63, 5, S. 737–748. doi: 10.1037/0022-006X.63.5.737
- Rice, Marnie E. & Harris, Grant T. (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. *Law and Human Behavior*, 29, 5, S. 615–620. doi: 10.1007/s10979-005-6832-7

- Sauter, Julia; Voß, Tatjana & Dahle, Klaus-Peter (2015). Wirksamkeit ambulanter Nachsorge bei Strafvollzugsentlassenen: Erste Evaluation der Forensisch Therapeutischen Ambulanz für Gewalt- und Sexualstraftäter in Berlin. *Der Nervenarzt*, 86, 5, S. 571–578. doi: 10.1007/s00115-014-4185-5
- Schmucker, Martin (2004). Kann Therapie Rückfälle verhindern? Metaanalytische Befunde zur Wirksamkeit der Sexualstraftäterbehandlung. *Studien und Materialien zum Straf- und Maßregelvollzug*, 22. Herbolzheim: Centaurus Verlag & Media.
- Spicer, Keith & Glicksman, Alison (2004). Adult reconviction: Results from the 2001 cohort. *Home Office Online Report 59/04*.
- Thoemmes, Felix (2012). *Propensity Score Matching in SPSS*. Zugriff unter: <https://arxiv.org/ftp/arxiv/papers/1201/1201.6385.pdf>
- Viglione, Jill (2019). The Risk-Need-Responsivity model: How do probation officers implement the principles of effective intervention? *Criminal Justice and Behavior* 46, 5, S. 655–673. doi: 10.1177/0093854818807505
- Vincent, Gina M.; Guy, Laura S.; Perrault, Rachael T. & Gershenson, Bernice (2016). Risk assessment matters, but only when implemented well: A multisite study in juvenile probation. *Law and Human Behavior*, 40, 6, S. 683–696. doi: 10.1037/lhb0000214
- Wirtz, Markus Antonius & Kutschmann, M. (2007). Analyse der Beurteilerübereinstimmung für kategoriale Daten mittels Cohens Kappa und alternativer Masse. *Die Rehabilitation*, 46, 6, S. 370–377. doi: 10.1055/s-2007-976535
- Wirtz, Markus Antonius & Caspar, Franz (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität. Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe Verl. für Psychologie.

Anhang

Anhang A: Bewertungsbogen des SVG-5

Bewertungsbogen SVG – 5

Hinweis:

Nur mit Tabulator- oder Pfeiltaste sich im Formular bewegen, sonst erfolgt keine Gesamtsummenberechnung

Name des Probanden:

Datum:

Anzahl	Bezeichnung	Bewertung	Punktzahl
1	Anzahl früherer Gewaltdelikte	0 = -5 1 oder 2 = 0 3 bis 5 = +2 6 und mehr = +4	
2	Deliktfrequenz	ja = +4 nein = -3	
3	Jemals Tötung eines Opfers	ja = -5 nein = +2	
4	Psychische Auffälligkeiten	ja = -4 nein = +2	
5	Alter des Probanden zum Zeitpunkt des ersten Gewaltdelikt	älter als 40 Jahre = -5 20 bis 40 Jahre = -2 jünger als 20 Jahre = +4	
Gesamtsumme SVG 5			0

Bewertung:

Errechneter Gesamtscore	Zuordnung der Risikokategorie	Ergebnis
weniger als -14	niedrig	<input type="checkbox"/>
-14 bis -7	niedrig-moderat	<input type="checkbox"/>
-6 bis +1	moderat-hoch	<input type="checkbox"/>
+2 bis +9	hoch	<input type="checkbox"/>
+10 und mehr	sehr hoch	<input type="checkbox"/>

Anhang B: Erhebungsbogen

Erhebungsbogen zur Bestimmung der Betreuungsstufe bei Probandinnen und Probanden unter Führungsaufsicht (außer Gewalt- und Sexualstraftaten)	
Name des Probanden: <input type="text" value="Geben Sie hier den Namen ein..."/>	Anmerkungen
1. Junger Mensch bis 30 Jahre zum Zeitpunkt der Erhebung <input type="checkbox"/> Ja (3) <input type="checkbox"/> Nein (0)	Geben Sie hier Anmerkungen ein...
2. In Herkunftsfamilie oder in eigener Familie sind straffällige Familienmitglieder. Vorherrschen von Gewalt, Suchtmittelmißbrauch, Überforderung <input type="checkbox"/> Ja (1) <input type="checkbox"/> Nein (0)	Geben Sie hier Anmerkungen ein...
3. Schulbildung <input type="checkbox"/> Hauptschulabschluss oder höher (0) <input type="checkbox"/> kein Hauptschulabschluss (1)	Geben Sie hier Anmerkungen ein...
4. Regelmäßige tagesstrukturierte Beschäftigung <input type="checkbox"/> Ja (0) <input type="checkbox"/> Nein (1)	Geben Sie hier Anmerkungen ein...
5. Protektives Freizeitverhalten <input type="checkbox"/> Ja (0) <input type="checkbox"/> Nein (1)	Geben Sie hier Anmerkungen ein...
6. Rechtskräftig verurteilte Vorstrafen <input type="checkbox"/> keine oder eine Vorstrafe (0) <input type="checkbox"/> 3 Vorstrafen und mehr (3) <input type="checkbox"/> 2 Vorstrafen (1)	Geben Sie hier Anmerkungen ein...
7. Vorstrafe wegen Gewalt/Raub <input type="checkbox"/> Ja (3) <input type="checkbox"/> Nein (0)	Geben Sie hier Anmerkungen ein...
8. In der Vergangenheit erfolgreich durchlaufene Bewährungs <input type="checkbox"/> Ja (0) <input type="checkbox"/> Nein (1)	Geben Sie hier Anmerkungen ein...
9. Hinweise auf Dissozialität/kriminogene psychische Auffälligkeit <input type="checkbox"/> Ja (1) <input type="checkbox"/> Nein (0) wenn ja, <input type="checkbox"/> Dissozialität oder kriminogene psychische Auffälligkeiten laut Gutachten (2)	Geben Sie hier Anmerkungen ein...
10. Substanzmittelgebrauch <input type="checkbox"/> Substanzmittelmissbrauch ohne bekannten Straftateneinfluss <input type="checkbox"/> Nein (0) <input type="checkbox"/> Substanzmittelgebrauch bei Straftaten (3) <input type="checkbox"/> Nein (0) wenn ja, <input type="checkbox"/> Nach abgeschlossener Therapie mind. 6 Monate abstinent (-1) <input type="checkbox"/> In regelmäßiger Suchtberatung und abstinent oder in qualifizierten Methadonprogramm (-1)	Geben Sie hier Anmerkungen ein...
Summe	0
Bewertung: 9 Punkte und mehr Betreuungsstufe 1 6 bis 8 Punkte Betreuungsstufe 2 weniger als 6 Betreuungsstufe 3	
Der Proband unterliegt der Betreuungsstufe 3	

Anhang C: Informationsschreiben im Rahmen der Abfrage der Risikoeinschätzungen

KrimZ · Viktoriastraße 35 · 65189 Wiesbaden

An die
Mitarbeiterinnen und Mitarbeiter
der Bewährungshilfe Hessen

Wiesbaden, 13. Oktober 2021

Abfrage von Einschätzungen mittels SVG-5 und Erhebungsbogen

Sehr geehrte Damen und Herren,

die Kriminologische Zentralstelle Wiesbaden (KrimZ) führt eine Evaluation des Fachbereichs SiMa II durch und untersucht in diesem Zusammenhang die Zuverlässigkeit und Aussagekraft der Risikoeinschätzungen mittels SVG-5 bzw. Erhebungsbogen (Projektbeschreibung). Wir möchten Sie bitten, uns bei dieser Untersuchung auf folgende Weise zu unterstützen.

- 1) An die Mitarbeiterinnen und Mitarbeiter der Allgemeinen Bewährungshilfe, des SiMa II, des SiMa I und der Sonderdienste

Um die **Aussagekraft** der Einschätzungen überprüfen zu können, **benötigen wir die vom Fachbereich SiMa II bereits vorgenommenen Risikoeinschätzungen mittels SVG-5 bzw. Erhebungsbogen für eine Reihe von Probanden/-innen**. Diese Probanden/-innen sind in Excel-Dateien gelistet, die den (zuletzt) zuständigen Dienststellen bzw. Fachbereichen zugänglich gemacht werden.

- Für jeden Landgerichtsbezirk wurde ein Download-Ordner erstellt
- Innerhalb dieses Ordners gibt es pro Dienststelle bzw. Fachbereich eine Probandenliste im Excel-Format.¹⁴
- Die Sachgebietsleitung jedes Landgerichtbezirks erhält einen Zugriffs-Link, über den die Listen für die verschiedenen Fachbereiche und ggf. verschiedenen Dienststellen heruntergeladen werden können.

¹⁴ In jedem Landgerichtsbezirk gibt es eine Liste für das SIMA II, mindestens eine Liste für die Allgemeine Bewährungshilfe (mehrere im Falle mehrerer Dienststellen) und in manchen Fällen eine Liste für das SIMA I sowie für die Sonderdienste.

- Wir bitten die Sachgebietsleitungen darum, diesen Zugriffs-Link an alle Dienststellen bzw. Fachbereiche weiterzuleiten, die eine Liste innerhalb des Ordners haben.¹
- Die in den betreffenden Dienststellen/Fachbereichen zuständigen Mitarbeiter/innen werden gebeten, die Risikoeinschätzungen (SVG-5/Erhebungsbogen) für die gelisteten Probanden/-innen aus SoPart zu extrahieren.¹⁵
- Für jede der gelisteten Personen benötigen wir bestenfalls alle vorliegenden Einschätzungen mittels SVG-5 und/oder Erhebungsbogen.
- Um aussagekräftige Ergebnisse zu erhalten, benötigen wir die vollständigen Einschätzungen, d.h. den Gesamt-Score sowie die Werte der einzelnen Items; insofern benötigen wir die **kompletten SVG-5-/Erhebungsbögen** (vorzugsweise im PDF-Format).
- Sollten die Einschätzungen zum Teil nur in Papierform vorliegen, bitten wir Sie darum, diese einzuscannen.
- Bitte benennen Sie die extrahierten/gescannten PDF-Dokumente nach folgendem Schema: **Name¹⁶, Vorname_Geburtsjahr_Instrument.pdf** (bspw.: Meier, Hans Martin_1975_SVG 5.pdf); dies ermöglicht die richtige Zuordnung zu den restlichen Probandendaten.
- Sollten mehrere Einschätzungen für eine/n Probanden/-in vorliegen, nummerieren Sie die Dateien bitte, beginnend mit der am weitesten zurückliegenden Einschätzung (bspw. Meier, Hans Martin_1975_SVG 5_1.pdf).
- Es ist wichtig, dass die Einschätzungen mit einem Datum versehen sind. Sollte das Datum der Einschätzung im (extrahierten/gescannten) PDF-Dokument selbst fehlen, ergänzen Sie es bitte im Dateinamen (bspw. Müller, Maria_1981_Erhebungsbogen_01.04.2018.pdf).
- Für jede Dienststelle, für die eine Excel-Datei bereitliegt, wurde ein Upload-Ordner (Bsp.: „LGB_Darmstadt_Bensheim_SVG5 u. Erhebungsbögen“) eingerichtet.
- Damit Sie nicht jede Datei einzeln hochladen müssen, sollten Sie die jeweiligen Dateien zunächst getrennt nach Zugehörigkeit in einem eigenen Ordner sammeln. Anschließend erstellen Sie daraus eine zip-Datei und laden diese über den entsprechenden Upload-Link hoch.

¹⁵ Bereits abgeschlossene und archivierte Fälle wurden für diesen Zweck vorübergehend wieder für die zuletzt zuständigen Dienststellen freigeschaltet

¹⁶ des Probanden/der Probandin

- Den Zugriffs-Link zu den jeweiligen Upload-Ordnern erhält die Sachgebietsleitung, die wir bitten, diesen an die involvierten Dienststellen/Fachbereiche weiterzuleiten
- Der Upload ist nicht Zugriffsgeschützt. Wer über den Link verfügt, kann Daten hochladen.

2) An die Mitarbeiterinnen und Mitarbeiter des SiMa II

Um die **Zuverlässigkeit** der Risikoeinschätzungen zu überprüfen, soll untersucht werden, ob unterschiedliche Anwender/innen der Instrumente (SVG-5 und Erhebungsbogen) zu gleichen Ergebnissen kommen. **Um diese Überprüfung vorzunehmen, muss eine Reihe von Probanden/-innen von jeweils zwei Bewährungshelfer/-innen unabhängig voneinander eingestuft werden.** Die Übereinstimmung dieser Einschätzungen kann anschließend von der KrimZ statistisch untersucht werden.

- Je mehr solcher doppelten Einschätzungen vorliegen, desto aussagekräftiger sind die Ergebnisse. In jedem Landgerichtsbezirk sollten vom Fachbereich SiMa II jedoch mindestens 6 Gewaltstraftäter/innen doppelt mittels SVG-5 und ebenfalls mindestens 6 Nicht-Gewaltstraftäter/innen doppelt mittels Erhebungsbogen eingestuft werden. Es ist wichtig, dass diese Einschätzungen komplett unabhängig erfolgen und nicht abgeglichen werden.
- **Wir bitten daher jeden Fachbereich SiMa II, (mindestens) zwei Bewährungshelfer/innen zu bestimmen, die diese doppelten Einschätzungen unabhängig voneinander vornehmen.**
- Die Auswahl der in diesem Kontext bewerteten Probanden/-innen ist dem Fachbereich überlassen. Es sollte jedoch beachtet werden, dass bestenfalls das gesamte Spektrum des jeweiligen Instruments abgedeckt wird. Sofern möglich, sollten also weder ausschließlich Personen mit mutmaßlich besonders hohem, noch ausschließlich Personen mit mutmaßlich besonders niedrigem Rückfallrisiko berücksichtigt werden.
- Das Schema für die Benennung der Dateien entspricht dem obigen; die unterschiedlichen Einschätzungen pro Proband/in sind mit Kennzahlen zu indizieren (bspw. Meier, Hans Martin_1975_SVG 5_1.pdf und Meier, Hans Martin_1975_SVG 5_2.pdf).
- Diese doppelten Einschätzungen laden Sie bitte in den für das SiMa II Ihres LGB eingerichteten Ordner mit der Bezeichnung „*Beurteilerübereinstim-*

mung" (Bsp.: „LGB_Darmstadt_SiMa
II_Beurteilerübereinstimmung“) hoch.

- Damit Sie nicht jede Datei einzeln hochladen müssen, sollten Sie die jeweiligen Dateien zunächst getrennt nach Zugehörigkeit in einem eigenen Ordner sammeln. Anschließend erstellen Sie daraus eine zip-Datei und laden diese über den entsprechenden Upload-Link hoch.
- Den Zugriffs-Link zu diesem Upload-Ordner erhält die Fachbereichsleitung des SiMa II des betreffenden Landgerichtbezirks von der jeweils zuständigen Sachgebietsleitung.
- Der Upload ist nicht Zugriffsgeschützt. Wer über den Link verfügt, kann Daten hochladen.

Es ist folglich nicht ersichtlich, von welchem/-r Bewährungshelfer/in die Einschätzung vorgenommen wurde. Eine solche Zuordnung ist für die Studie weder erforderlich noch erwünscht. Die Auswertungen werden auch nicht für einzelne Dienststellen, sondern ausschließlich zusammenfassend vorgenommen. Spezifische Aussagen zu einzelnen Landgerichtsbezirken oder Dienststellen sind nicht Ziel der Studie.

Bitte laden Sie die gesammelten Dokumente bis spätestens **16. Juni 2019** in die vorgesehenen Ordner! Bei Fragen steht Ihnen unsere wissenschaftliche Mitarbeiterin Lisanne Breiling (Mail: l.breiling@krimz.de; Tel.: 0611-15758-16) selbstverständlich gerne zur Verfügung.

Wir möchten Ihnen bereits jetzt vielmals für Ihre Mitarbeit und Unterstützung danken!

Mit freundlichen Grüßen

Prof. Dr. Martin Rettenberger (Direktor der KrimZ)



Lisanne Breiling (wissenschaftliche Mitarbeiterin)



Anhang D: Interviewleitfaden

Evaluation der Prognose- und Interventionspraxis im Sicherheitsmanagement (SiMa) II in Hessen

Interviewleitfaden

VORSTELLUNG

- Begrüßung und Dank für die Zeit
- Vorstellung der Interviewerin und kurze Beschreibung der Zielsetzung des Interviews:
- Informationen zur Rekrutierung (Auswahl der Teilnehmer); Freiwilligkeit der Teilnahme; Verschwiegenheit / Anonymität:
- Kurze Beschreibung des Aufbaus des Interviews (Dauer, Struktur, Tonaufzeichnung):
- offene Fragen
- Einverständniserklärung, Beginn der Aufzeichnung

START DER AUFNAHME**EINSTIEGSFRAGEN:** Angaben zur Person und zum beruflichen Hintergrund

- Dauer Bewährungshilfe:
- Dauer, Stellenanteil, Position SiMa II

THEMENBEREICH 1: Auswirkungen des SiMa II-Erlasses

- Kurze Einleitung des Themenblocks
- Nennung und Beschreibung positiver praktischer Auswirkungen der Umstrukturierung
- Nennung und Beschreibungen von Nachteilen bzw. Diskrepanzen zwischen Vorgaben und Arbeitsrealität
 - Rückfrage: Was müsste sich konkret ändern, um genannte Probleme zu lösen?
- Tatsächliche Auswirkungen der Umstrukturierung auf Betreuungsarbeit

OPTIONAL 1:

- Wie wurde der Übergang im Zuge der Implementierung von SiMa II empfunden? Hinreichende Vorbereitung?
- Wie wurde das Konzept von den Bewährungshelfer/-innen angenommen?
- Gab es bestimmte Befürchtungen? Haben sich diese bewahrheitet?
- Was sind die spürbarsten Auswirkungen der neuen Richtlinie auf den Arbeitsalltag?

- Wie gut wird das Konzept den individuellen Bedürfnissen und Voraussetzungen der unterschiedlichen Probanden gerecht?
- Wie gut gelingt die Einhaltung der im Erlass festgeschriebenen zeitlichen Fristen?

THEMENBEREICH 2: Risikoeinschätzung

- Kurze Einleitung des Themenblocks
- Vorteile/Nachteile der Instrumente; Schulnoten
 - SVG-5
 - Erhebungsbogen
- Wurde bereits zuvor risikoorientiert gearbeitet? Wie wurde das ggf. realisiert?

OPTIONAL 2:

- In der Regel auf alle erforderlichen Informationen Zugriff, die für die Risikoeinschätzung notwendig sind?
- Gibt es Personengruppen, für die die Instrumente weniger gut geeignet sind? Welche und warum?
- Wie aussagekräftig sind die Einschätzungen? Bilden sie den tatsächlichen individuellen Betreuungsbedarf gut ab?
- Vernachlässigen die Instrumente relevante Aspekte (im Sinne der Festsetzung des individuellen Betreuungsbedarfs)? Welche?
- Hat der Einsatz der Instrumente Auswirkungen auf die Beziehungsgestaltung mit den Probanden/-innen? In welcher Hinsicht?

ERGÄNZUNGEN/ABSCHLUSS

- Fazit: Erleichterung oder Erschwerung der Arbeit durch Umstrukturierung?
- Anmerkungen
- Abschluss, Danksagung

Anhang E: Einverständniserklärung zur Teilnahme am Interview

Evaluation der Prognose- und Interventionspraxis des
Fachbereichs SiMa II in Hessen:
Interviews mit SiMa II-Mitarbeiterinnen und -Mitarbeitern

Kriminologische Zentralstelle
Viktoriastraße 35
65189 Wiesbaden
Lisanne Breiling
Wissenschaftliche Mitarbeiterin
Ansprechpartnerin für die
Durchführung der Evaluation
des Fachbereichs SiMa II
Tel.: 0611-15758-16
Fax: 0611-15758-10
l.breiling@krimz.de
www.krimz.de

Einverständniserklärung

Liebe Teilnehmerin / lieber Teilnehmer,

für Ihre Teilnahme am Interview benötigen wir Ihr Einverständnis, das mit einer Unterschrift dokumentiert wird. Mit dem Einverständnis bestätigen Sie zudem die Kenntnisnahme folgender Hinweise:

- Mir ist bekannt, dass die Teilnahme an der oben benannten Befragung freiwillig ist und mir bei Verweigerung der Teilnahme keine Nachteile entstehen.
- Ich kann meine Teilnahmebereitschaft jederzeit ohne Angabe von Gründen widerrufen. Erfolgt ein Widerruf, werden meine personenbezogenen Daten und Aufzeichnungen umgehend gelöscht.
- Die verantwortlichen Mitarbeiter/-innen der KrimZ, Herr Prof. Dr. Martin Rettenberger und Frau Lisanne Breiling, versichern, dass die erhobenen Daten streng vertraulich behandelt und gegen den Zugriff von Unbefugten angemessen gesichert werden.
- Ich bin damit einverstanden, dass das Gespräch aufgezeichnet wird. Diese Aufzeichnungen werden nach der Verschriftlichung und Pseudonymisierung bzw. Auswertung vernichtet.
- Mir ist bekannt, dass alle an der Studie beteiligten Mitarbeiter der KrimZ nach § 1 Absatz 1 bis 3 des Verpflichtungsgesetzes förmlich verpflichtet bzw. auf das Datengeheimnis nach § 9 HDSG verpflichtet wurden.

- Ich habe zur Kenntnis genommen und bin damit einverstanden, dass ...
 - meine Kontaktdaten (Name, Adresse, Telefonnummern, Mailadresse) zu Forschungszwecken und für eventuelle Nachfragen bis zum Abschluss der Studie passwortgeschützt gespeichert und anschließend gelöscht werden.
 - die Inhalte des Interviews – darunter meine personenbezogenen Daten – im Rahmen der oben benannten Studie pseudonymisiert und für wissenschaftliche Zwecke verwendet werden.
 - meine erhobenen Daten zu Forschungszwecken gespeichert sowie pseudonymisiert veröffentlicht werden dürfen.

Ich (Name der Teilnehmerin / des Teilnehmers in Blockschrift)

bin über den Zweck und die Rahmenbedingungen der Befragung aufgeklärt worden. Ich erkläre mich durch meine Unterschrift damit einverstanden, an der oben benannten Befragung teilzunehmen und ein Interview zu geben. Ich bin einverstanden, dass die im Rahmen des Gesprächs erhobenen Daten und Informationen zu Forschungszwecken genutzt werden. Ein Exemplar dieser Einverständniserklärung habe ich erhalten.

Ort, Datum

Unterschrift des Teilnehmers/der Teilnehmerin

Ort, Datum

Unterschrift der KrimZ-Mitarbeiterin