

Modellversuch „Neue psychotherapeutische Interventionsprogramme und Evaluationskonzepte im Schweizer Strafvollzug“

Zusammenfassung des Schluss- und Evaluationsberichts

Astrid Rossegger und Michael Weber, Universität Konstanz

04. Oktober 2017

Hintergrund: Der Modellversuch

Ziele und Fragestellung des Modellversuchs

Die zentrale Zielsetzung des Modellversuchs (MV) bestand in der Implementierung und Wirksamkeitsprüfung der Gruppentherapieprogramme „Reasoning and Rehabilitation“ (R&R; Ross, Fabiano, & Ross, 1986; R&R2; Ross, Hilborn, & Liddle, 2007) und „Anti-Sexuelle-Aggressivität-Training@Suisse“ (ASAT@Suisse; Falk & Steffensen, 2014). Zudem sollten geeignete Strategien zur Evaluation von Behandlungsprogrammen für Straftäter implementiert werden. Die primäre Fragestellung des MV war, inwieweit eine störungs- und deliktorientierte Behandlung von Gewalt- und Sexualstraftätern zu messbaren Veränderungen von Einstellungen, Werthaltungen, Verhalten und persönlichkeitsbezogenen Variablen sowie des Rückfallrisikos führt.

Ablauf des Modellversuchs

Abbildung 1 zeigt den Ablauf des Modellversuchs in der Übersicht.

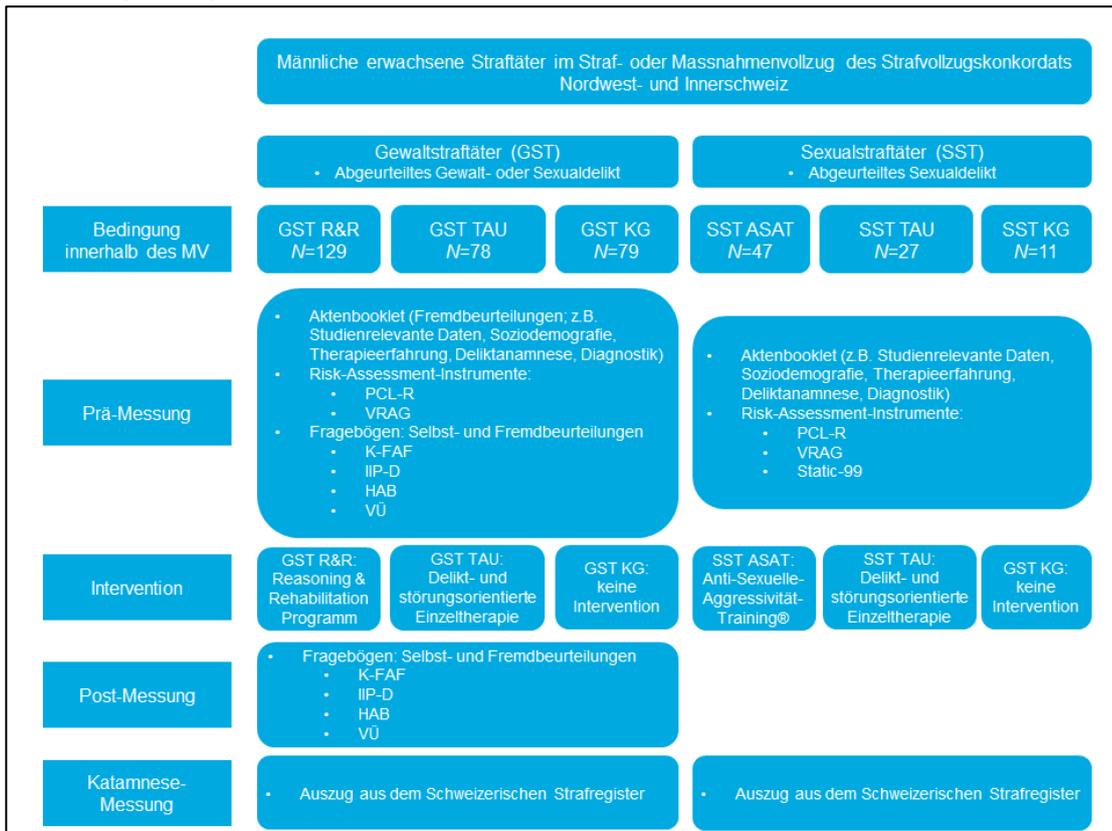


Abbildung 1. Schematischer Ablauf des Modellversuchs

Ein- und Ausschlusskriterien

Grundsätzlich handelte es sich bei den eingeschlossenen Probanden um volljährige männliche verurteilte Gewalt- oder Sexualstraftäter. Für die Teilnahme an einer der beiden Experimentalbedingungen (GST R&R für Gewalt- und/oder Sexualstraftäter

und SST ASAT für Sexualstraftäter) sowie den Vergleichsbedingungen (GST TAU bzw. SST TAU) musste eine Therapie gerichtlich oder vollzugsseitig angeordnet worden sein. Die Kontrollgruppen umfasste Gewalt- und/oder Sexualstraftäter (GST KG) bzw. Sexualstraftäter (SST KG), die sich nicht in psychotherapeutischer Behandlung befanden.

Ausschlusskriterien für eine Teilnahme waren ungenügende Deutschkenntnisse, eine Intelligenzminderung sowie eine akute Alkohol- oder Drogen-Intoxikation.

Methodisches Design

Der Versuchsplan umfasste eine longitudinale Datenerhebung zu verschiedenen Messzeitpunkten: Eine Prä-Messung vor Beginn der Intervention, eine Post-Messung nach Beendigung der Intervention (jeweils Fragebogen) sowie eine Katamnese-Messung (Erhebung der Rückfälligkeit). Die Analyse wurde über Differenzwerte vollzogen.

Methode der Evaluation

Fokus und Hypothesen

Das Behandlungsprogramm R&R(2) wurde im vorliegenden Bericht anhand der folgenden zwei Fragestellungen evaluiert:

- 1) Auswirkung einer Teilnahme am R&R(2)-Programm auf Messwerte in Fragebögen, die auf die Erfassung von risikoassoziierten Persönlichkeitseigenschaften, Einstellungen und Verhaltensweisen zielen. Dabei handelte es sich um den Kurzfragebogen zur Erfassung von Aggressivitätsfaktoren (K-FAF; Heubrock & Petermann, 2008), das Inventar zur Erfassung interpersonaler Probleme (IIP-D; Horowitz et al., 2000), den Hostile Attribution Bias (HAB; Tremblay & Belchevski, 2004) sowie den Fragebogen zur Verantwortungsübernahme (VÜ; Gabriel et al., 2005; Oswald & Bütikofer, 2002).
- 2) Auswirkung einer Teilnahme am R&R(2)-Programm auf die Rückfälligkeit gemäß Auszügen aus dem Schweizerischen Strafregister (VOSTRA).

Das Behandlungsprogramm ASAT@Suisse wurde ausschließlich über die Auswirkung einer Programmteilnahme auf die Rückfälligkeit evaluiert.

Ergebnisse

Stichprobe

Tabelle 1 zeigt die Anzahl der für die vorliegende Auswertung grundsätzlich berücksichtigten Probanden nach Ausschluss der im Evaluationsbericht genannten Fälle.

Tabelle 1. Stichprobenzusammensetzung: Für die Evaluation berücksichtigte Fälle

GST Bedingung	N	SST Bedingung	N
GST R&R	129	SST ASAT	47
GST TAU	78	SST TAU	27
GST KG	79	SST KG	11
GST Gesamt	286	SST Gesamt	85

Anmerkungen. GST: Gruppe der Gewaltstraftäter; SST: Gruppe der Sexualstraftäter; R&R: Reasoning and Rehabilitation Programm; ASAT: Anti-Sexuelle-Aggressivität-Training@Suisse; TAU: Treatment as Usual (Vergleichsgruppe); KG: Kontrollgruppe.

Ergebnisse R&R

Fragestellung 1: Veränderungen in den Messwerten der Fragebögen in Abhängigkeit von der Therapie

Für die Beantwortung dieser Frage wurden die Differenzwerte zwischen der Prä- und der Postmessung in den Messwerten verschiedener Fragebögen inferenzstatistisch zwischen den drei Bedingungen der GST-Gruppe verglichen. In der vorliegenden Zusammenfassung werden lediglich die zentralen Ergebnisse der Selbstbeurteilung durch die Probanden wiedergegeben.

Kurzfragebogen zur Erfassung von Aggressivitätsfaktoren (K-FAF)

In den Differenzwerten des Aggressivitäts-Summenwertes zeigten sich signifikante Unterschiede zwischen den drei Bedingungen der GST-Gruppe: $F(2, 206) = 3.52$, $p = .031$; $\eta^2 = .033$. Die Probanden aller drei Bedingungen gaben zum zweiten Messzeitpunkt eine geringer ausgeprägte Aggressivität an ($\mu_{\text{GST-R\&R}} = -9.2$; $\mu_{\text{GST-TAU}} = -2.2$; $\mu_{\text{GST-KG}} = -1.3$). Ein post hoc durchgeführter Tukey-Kramer-Test wurde auf einem Alpha-Niveau von .05 nicht signifikant. Der bedeutsamste Unterschied zeigte sich auf dem Niveau eines Trends ($p = .064$) zwischen der Experimental- und der Kontrollgruppe zugunsten der Experimentalgruppe.

Inventar zur Erfassung interpersoneller Probleme (IIP-D)

Die acht Skalen des IIP-D bilden Schwierigkeiten in verhaltensnahen Aspekten der Persönlichkeit ab. Weder in den Differenzwerten der einzelnen Skalen noch des Gesamtwertes zeigten sich signifikante Unterschiede zwischen den drei Bedingungen der GST-Gruppe.

Hostile Attribution Bias (HAB)

Mit dem HAB wird das Ausmaß an feindseliger Interpretation des Verhaltens eines fiktiven potenziellen Aggressors erfasst. Dazu geben die Probanden an, mit welcher

Wahrscheinlichkeit sie in verschiedenen Situationen feindseliges Verhalten zeigen würden. Bezogen auf den Summenwert der Wahrscheinlichkeit für feindseliges Verhalten in Situationen, die eine eindeutige Provokation des Gegenübers beschreiben, zeigten sich signifikante Unterschiede: $F(2, 213) = 4.03, p=.019; \eta^2 = .036$. Ein post hoc durchgeführter Tukey-Kramer-Test zeigte, dass sich Experimental- und Kontrollgruppe auf einem Alpha-Niveau von .05 signifikant zugunsten der Experimentalgruppe unterscheiden. Die mit R&R(2) behandelten Probanden berichten zum zweiten Messzeitpunkt eine reduzierte Wahrscheinlichkeit für feindseliges Verhalten, die unbehandelte Kontrollgruppe hingegen eine leicht höhere Wahrscheinlichkeit ($\mu_{\text{GST-R\&R}} = -4.2; \mu_{\text{GST-TAU}} = -1.5; \mu_{\text{GST-KG}} = 0.3$). Bezogen auf die einzelnen Aspekte feindseligen Verhaltens zeigte sich diese geringere Wahrscheinlichkeit zum zweiten Messzeitpunkt vor allem auf den Dimensionen „Wahrnehmung als absichtliche Provokation“ ($\mu_{\text{GST-R\&R}} = -0.5; \mu_{\text{GST-TAU}} = -0.3; \mu_{\text{GST-KG}} = 0.4$) und „Anschreien/Beschimpfen“ ($\mu_{\text{GST-R\&R}} = -0.8; \mu_{\text{GST-TAU}} = -0.3; \mu_{\text{GST-KG}} = 0.5$).

Keine signifikanten Gruppenunterschiede im Gesamtwert oder in einzelnen Dimensionen des HAB zeigten sich hingegen in Situationen mit eindeutig nicht provozierender sowie unklarer Intention des Gegenübers.

Fragebogen zur Verantwortungsübernahme (VÜ)

In den Differenzwerten des VÜ-Gesamtwerts als Maß für die selbst berichtete Verantwortungsübernahme für das begangene Delikt zeigten sich signifikante Unterschiede ($F(2, 186) = 3.17, p=.044; \eta^2 = .033$). Ein post hoc durchgeführter Tukey-Kramer-Test zeigte auf einem Alpha-Niveau von .05 einen signifikanten Unterschied zwischen der Experimental- und der Kontrollgruppe zugunsten der Experimentalgruppe. Die mit R&R(2) behandelten Probanden geben zum zweiten Messzeitpunkt eine erhöhte Verantwortungsübernahme an, die Probanden der unbehandelten Kontrollgruppe hingegen eine verminderte ($\mu_{\text{GST-R\&R}} = -2.4; \mu_{\text{GST-TAU}} = -0.7; \mu_{\text{GST-KG}} = 0.5$).

Fragestellung 2: Rückfälligkeit

Tabelle 2 zeigt die Ergebnisse der Analyse zur Rückfälligkeit in der GST-Gruppe unter Berücksichtigung aller für diese Fragestellung eingeschlossenen $n=278$ Probanden. Von diesen befanden sich zum Zeitpunkt des Strafregisterauszugs 63 in Freiheit (22.7%). Diese waren einer vergleichbaren Time at Risk von durchschnittlich 20 Monaten ausgesetzt. Zwölf Probanden (19.1%) wurden im zur Verfügung stehenden Beobachtungszeitraum rückfällig, davon fünf in der Experimental-, vier in der Vergleichs- und drei in der Kontrollgruppe. Exakte logistische Regressionen, trotz der geringen Fallzahl zur Überprüfung auf Unterschiede zwischen den drei Bedingungen berechnet, führten weder in Bezug auf die allgemeine Rückfälligkeit ($p=.772$) noch in Bezug auf die Rückfälligkeit wegen eines Gewalt- oder Sexualdelikts ($p=.261$) zu signifikanten Ergebnissen.

Tabelle 2. Rückfälligkeit GST (Intent-to-Treat-Analyse)

GST Bedingung	Fehlend [% (N)]	N	Time at risk [Monate] [M (SA)]	p	Rückfällig [% (N)]	p	Rückfällig (Gewalt- oder Sexualdelikt) [% (N)]	p
GST R&R	0% (N=0)	30	M=19.8 (SA=10.1)		16.7% (N=5)		3.3% (N=1)	
GST TAU	0% (N=0)	15	M=21.0 (SA=14.9)	.924 ^{a)}	26.7% (N=4)	.772 ^{b)}	13.1% (N=2)	.261 ^{b)}
GST KG	0% (N=0)	18	M=19.6 (SA=13.4)		16.7% (N=3)		0% (N=0)	
GST Gesamt	0% (N=0)	63	M=20.0 (SA=12.1)		19.1% (N=12)		4.8% (N=3)	

Anmerkungen. GST: Gewaltstraftäter; R&R: Reasoning and Rehabilitation Programm (2); TAU: Treatment as Usual (Vergleichsgruppe); KG: Kontrollgruppe; TAR: Time at Risk; ^{a)} Ergebnis eines Kruskal-Wallis-H-Tests; ^{b)} Ergebnis einer exakten logistischen Regression.

Ergebnisse ASAT®Suisse

Tabelle 3 zeigt die Ergebnisse der Analyse zur Rückfälligkeit in der SST-Gruppe unter Berücksichtigung aller in die Evaluation eingeschlossenen n=85 Probanden. Von diesen befanden sich zum Zeitpunkt des Strafregisterauszugs lediglich neun in Freiheit (10.6%). Diese waren einer vergleichbaren Time at Risk von durchschnittlich knapp zwei Jahren ausgesetzt. Ein Proband aus der Vergleichsgruppe wurde rückfällig. Eine statistische Analyse ist aufgrund der geringen Fallzahlen nicht möglich.

Tabelle 3. Rückfälligkeit Gruppe SST (Intent-to-Treat-Analyse)

SST Bedingung	Fehlend [% (N)]	N	TAR [Monate] [M (SA)]	p	Rückfällig [% (N)]	p	Rückfällig (Gewalt- oder Sexualdelikt) [% (N)]	p
SST ASAT	0% (N=0)	2	M=21.0 (SA=11.7)		0% (N=0)		0% (N=0)	
SST TAU	0% (N=0)	6	M=21.1 (SA=14.9)	.475 ^{a)}	16.7% (N=1)	N/A	0% (N=0)	N/A
SST KG	0% (N=0)	1	M=39.3 (SA=0)		0% (N=0)		0% (N=0)	
SST Gesamt	0% (N=0)	9	M=23.1 (SA=13.9)		11.1% (N=1)		0% (N=0)	

Anmerkungen. SST: Sexualstraftäter; ASAT: Anti-Sexuelle-Aggressivität-Training®Suisse; TAU: Treatment as Usual (Vergleichsgruppe); KG: Kontrollgruppe; TAR: Time at Risk; ^{a)} Ergebnis eines Kruskal-Wallis-H-Tests.

Diskussion

Gefundene Unterschiede zwischen den Bedingungen

Zum Effektkriterium Rückfälligkeit lagen weder für die Evaluation des R&R(2)- noch des ASAT®Suisse-Programms hinreichende Fallzahlen für einen aussagekräftigen Vergleich zwischen den Bedingungen vor.

Darüber hinaus zeigten sich in Ansätzen kleine positive Behandlungseffekte des R&R(2)-Programms im Vergleich zur weitestgehend unbehandelten Kontrollgruppe. Dies betrifft insbesondere die selbst berichtete Aggressivität sowie, mit Einschränkungen, die selbst berichtete Verantwortungsübernahme. Teilweise konnten diese Hinweise auch in Bezug auf die selbst berichtete Zuschreibung von Feindseligkeit in einer potenziell konflikt-auslösenden Situation beobachtet werden. Sämtliche gefundene Unterschiede sind jedoch zu wenig robust, um von einer Überlegenheit des R&R(2)-Programms gegenüber der Kontrollgruppe oder der Standardbehandlung sprechen zu können. Über diese unmittelbaren Ergebnisse hinaus können die tendenziell niedrigeren Abbrecher-Quoten sowohl der R&R(2)- als auch der ASAT®Suisse-Teilnehmer im Vergleich zur Standardbehandlung als Indiz für bessere motivationale Eigenschaften interpretiert werden.

Stärken des MV

Eine große Stärke stellt der realisierte Multicenter-Ansatz und die damit einhergehende beachtenswerte Stichprobengröße in der GST-Gruppe dar. Der MV bildet ein großes Stück Realität im Schweizer Straf- und Maßnahmenvollzug ab. Es wurden zahlreiche Bemühungen ersichtlich, eine angemessene Programmintegrität zu gewährleisten. Als sehr zielführend erwies sich die Wahl von risikorelevanten Verhaltens-, Einstellungs- und Persönlichkeitsmaßen in der GST Gruppe als Ergänzung zum schwierig zu erhebenden Effektkriterium der Rückfälligkeit. Das gewählte quasi-experimentelle Design stellt eine adäquate Untersuchungsmethode für die Wirksamkeit von Behandlungsprogrammen für Gewalt- und Sexualstraftäter dar.

Limitationen: „Lessons learned“

Wie die Mehrzahl der durchgeführten Evaluationsstudien sah sich auch der MV mit einer Vielzahl typischer Schwierigkeiten in diesem Feld konfrontiert. Die limitierte Follow-Up-Dauer begrenzte die Möglichkeit aussagekräftiger Rückfallraten. Trotz der zahlreichen Bemühungen wie der Wahl eines Multicenter-Ansatzes war am Ende eine limitierte Stichprobengröße zu verzeichnen. Dies ist auch einer der Hauptgründe für die geringe statistische Power, die zu einer reduzierten Möglichkeit führt, tatsächlich bestehende Unterschiede im Rahmen des MV statistisch nachweisen zu können. Viele beabsichtigte Auswertungen konnten aus verschiedenen Gründen letztlich nicht durchgeführt werden. Problematisch erwies sich der Umstand, dass im Laufe des MV eine vollständige Umstellung von der Langversion des R&R-Programms auf die

Kurzversion stattfand, sodass Probanden innerhalb der Experimentalbedingung zwei unterschiedliche Varianten der Intervention durchliefen.

Ausblick

Auch wenn im Rahmen der Evaluation die erwarteten Effekte nicht im erhofften Maß aufgezeigt werden konnten, sollte dies nicht als Widerlegung des Nutzens von strukturierten kognitiv-behavioralen Therapieprogrammen für Gewalt- und Sexualstraftäter verstanden werden. Aus verschiedenen Gründen konnten die Hypothesen nur bedingt überprüft werden. Die Schwierigkeiten bieten allerdings Implikationen für die zukünftige Forschung im Bereich forensischer Behandlungsprogramme, die es sowohl in der Planungs- als auch der Durchführungs- und Evaluationsphase von Studien zu berücksichtigen gilt. Eine hinreichend lange Förderdauer von Evaluationsstudien erscheint wesentlich, um robuste Aussagen zur Wirksamkeit von Therapien treffen zu können. Die Wichtigkeit von Kooperationen ist zu betonen. Zentral erscheint die Dissemination von bestehendem „Systemwissen“ in der forensischen Community. Nicht zuletzt ist bei der Wirksamkeitsüberprüfung von Behandlungsprogrammen für Straftäter Augenmerk auf Kosten-Nutzen-Aspekte zu richten.