

Thema

Erfolgseinschätzung bei der Evaluation von Interventionen

Christoph Liel

Vor allem in Bereichen, in denen aggressives selbst- oder fremdschädigendes Verhalten von Klienten bearbeitet wird, kann die Erfolgseinschätzung als zentrale Aufgabe bei der Evaluation sozialer Praxis gesehen werden. Zugleich stellen sich hier regelmäßig besondere Herausforderungen (z. B. eine geringe Häufigkeit des schädigenden Verhaltens). Ausgehend von einem Evaluationsvorhaben im Feld der Partnerschaftsgewalt, das am Deutschen Jugendinstitut e.V. entwickelt wurde, erörtert der Beitrag Probleme und Chancen bei Erfolgseinschätzungen.

Wie können Erfolg und Misserfolg von Interventionen zuverlässig gemessen werden? Diese Frage stellt sich regelmäßig bei der Evaluation Sozialer Arbeit. Gängiges Verfahren ist es, die Sichtweise von Nutzerinnen und Nutzern, der durchführenden Fachkräfte und des Gemeinwesens auf das Angebot zu erheben. Spätestens bei der Evaluation von Programmen, die selbst- und fremdschädigendes Verhalten behandeln, stößt dieses Vorgehen an Grenzen.

Schädigendes Verhalten bearbeiten und evaluieren Es liegt auf der Hand, dass Zufriedenheitskriterien beispielsweise bei einem Anti-Aggressivitäts-Training allein nicht ausreichen, da hier Verhaltensprobleme mit hohem Schädigungspotential bearbeitet werden. Niemand würde ernsthaft von einem Erfolg sprechen wollen, wenn zwar alle Beteiligte eine solche Maßnahme positiv bewerten, sich aber herausstellt, dass die Teilnehmer unverändert gewaltbereit sind. Für die Erfolgseinschätzung werden daher zusätzlich objektivierbare Kriterien benötigt. Schnell wird bei diesen Programmen nach einer Rückfallquote gefragt, deren Erhebung häufig als Goldstandard bei der

Erfolgseinschätzung angesehen wird.

Das gesellschaftliche Interesse an einem Erfolgsnachweis ist berechtigt, weil durch das Schädigungsverhalten ein hohes Rechtsgut gefährdet wird: Die Unversehrtheit von Klienten selbst, von Menschen im unmittelbaren sozialen Umfeld und von Dritten bzw. der Öffentlichkeit. Deshalb haben Gesellschaft, Geldgeber, überweisende Einrichtungen und Hilfesystem ein erhöhtes Interesse an einem Erfolgsbeleg, verglichen mit anderen Angeboten der Sozialen Arbeit. Das öffentliche Interesse an einer belegten Arbeitsweise besteht zudem auch deshalb, weil Strafverfolgung und Sanktionen sowie weiteres staatliches Eingriffshandeln (besonders bei Kindeswohlgefährdung) je nach der Verfügbarkeit Erfolg versprechender Interventionen anders eingesetzt werden können. Beispielsweise können straf- und familiengerichtliche Entscheidungen die Programmteilnahmen begünstigend berücksichtigen oder zu ihnen verpflichten. Die Programme müssen als Kooperationspartner im Hilfesystem größtmöglichen Verlass bieten. Auch die Fachkräfte, die in den Programmen arbeiten, benötigen Rückfallinformationen. Schließlich bestehen Ziel und Auftrag der Interventionen in der Unterbrechung und Beendigung des Problemverhaltens.

Als Goldstandard der Evaluation kann die Aufklärung der Häufigkeit von Rückfällen daher zum einen dabei helfen, Erfolg versprechendes Programmhandeln von weniger Erfolg versprechendem abzugrenzen. Zum anderen können Informationen darüber gesammelt werden, welche Klientinnen und Klienten in Programmen besser profitieren, um so die Passgenauigkeit der Angebote zu sichern und gegebenenfalls Problemgruppen innerhalb des Klientelspektrums zu identifizie-

ren, für die die Angebote möglicherweise nicht indiziert sind. Beide Aspekte verdienen Beachtung bei der Evaluation, weil ungeeignete Interventionen nicht nur schlechte Erfolgsquoten haben, sondern auch zu einer Problemverschärfung beitragen können.

Bemüht sich die Evaluationsforschung allerdings, diesen Goldstandard zu erreichen, beispielsweise durch methodisch anspruchsvolle Rückfallstudien, die Kontrollgruppen einbeziehen, so können teilweise nur geringe Effekte nachgewiesen werden (Wottawa, Thierau 1998). Das bereits angesprochene Anti-Aggressivitäts-Training ist ein gutes Beispiel, über das eine quasi-experimentelle Rückfallerhebung im Jugendstrafvollzug nur wenig Unterschiede im Vergleich zu anderen inhaftierten und behandelten Gewalttätern nachweisen konnte (Ohlemacher et al. 2001). Quasi-experimentell heißt in diesem Zusammenhang, dass die Versuchsgruppen nicht vorher zufällig ausgewählt, sondern erst im Nachhinein parallelisiert wurden. 77 Trainingsteilnehmer wurden mit Probanden aus der gleichen Haftanstalt verglichen, die ähnliche delikttypische Merkmale aufwiesen. Absolute Rückfälle, die Häufigkeit je Proband und der Zeitabstand des Rückfalls zur Maßnahme waren nahezu identisch.

Deuten diese und ähnliche Befunde darauf hin, dass die evaluierten Interventionen „nichts bringen“ bzw. geringe Effekte das Maximum des Erreichbaren darstellen? Oder könnte es sein, dass Rückfallquoten allein verzerrte bzw. lückenhafte Informationen liefern und daher als Alles-oder-Nichts-Kriterium für die Bewertung von Interventionen nicht ausreichen?

Schwierigkeiten beim Nachweis von Rückfällen Ein Problem bei der Erfolgsmessung allein über Rückfalldaten ergibt

sich aus der eher geringen Zuverlässigkeit diesbezüglicher Klientenangaben mittels Interviews oder Fragebögen, die bei der Evaluation Sozialer Arbeit überwiegend genutzt werden. Für die Erhebung von Rückfällen erweisen sich Klientenbefragungen (teilweise auch Angehörigenbefragungen) als nicht ausreichend zuverlässig, da es sich bei Aggressivität um sozial höchst unerwünschtes Verhalten handelt, was zu einer eingeschränkten Offenheit der Befragten führt. Falsche negative Angaben bleiben vor allem bei Gewalt im sozialen Nahraum in der Regel unentdeckt, so dass sich bezogen auf den Einzelfall wenig zuverlässige und bezogen auf alle Teilnehmer zu geringe Raten gefundener Rückfälle ergeben.

Ein Beispiel für solche Befragungsergebnisse zeigt die grafische Darstellung aus einer Evaluation eines Täterprogramms bei häuslicher Gewalt. 25 Klienten wurden schriftlich befragt, ob sie nach Abschluss der Maßnahme noch einmal gewalttätig geworden sind. Es ist wenig wahrscheinlich, dass eine Rückfallquote von 4% der Wirklichkeit entspricht.

Um Ergebnisverzerrungen durch Klientenbefragungen zu entgehen, ist versucht worden, öffentlich bekannt gewordene Rückfalldaten z. B. bei Polizei und Staatsanwaltschaft zu erheben. Zwar sind die Daten aus dem Hellfeld verlässlicher als Klientenbefragungen im Dunkelfeld. Sie bilden in sensiblen Bereichen wie häuslicher Gewalt aber nur einen sehr kleinen Teil tatsächlicher Rückfälle ab. Der Vergleich von Hellfelddaten mit repräsentativen Dunkelfeldbefragungen von Frauen lässt den Schluss zu, dass ein größerer Teil überhaupt nicht öffentlich wird. Daher besteht auch hier das Problem

niedriger Häufigkeiten bekannt gewordener Schädigungsereignisse.

Wenn es keine Möglichkeit gibt, Rückfälle zuverlässig zu erheben, mindert dies die Chance, allein auf diesem Weg ein aussagekräftiges Bild des tatsächlichen Erfolgs von Interventionen zu erhalten.

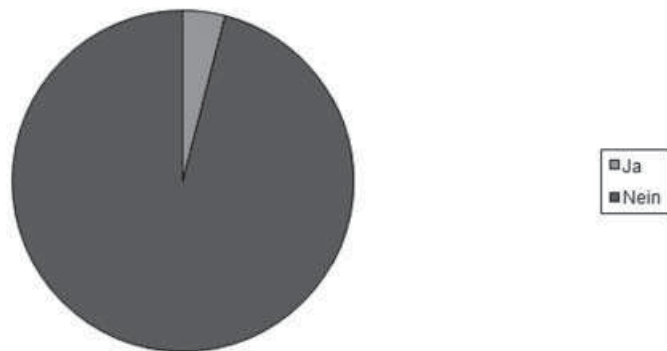


Abb. 1: Frage: „Haben Sie nach Abschluss der Beratung [...] innerhalb der Familie noch einmal gewalttätig gehandelt?“

Weiterhin ist zu berücksichtigen, dass Gewaltereignisse, selbst wenn sie zuverlässig erhoben werden könnten, wohl meist auch bei unbehandelten Gewalttätern mit einer relativ geringen Grundrate auftreten, d. h. auch diese Personen, begehen überwiegend nicht jede Woche ein Gewaltdelikt. Die aus den genannten Gründen niedrigen erhobenen Rückfallraten beeinflussen deshalb die Chance in Studien statistisch bedeutsame Unterschiede zwischen Interventions- und Kontrollgruppen zu messen, denn es werden sehr große Untersuchungsgruppen benötigt (Gondolf 2004).

Zwar zeigen zumindest Meta-Analysen, also zusammenfassende Untersuchungen mehrerer Rückfallstudien, dass mit Interventionen eine Verringerung von Rückfallraten erreicht werden kann. Über Täterarbeit bei Partnerschaftsgewalt international mitt-

lerweile vorliegende Metaanalysen (Feder, Wilson 2005; Babcock et al. 2004) können schwache bis moderate Effekte nachweisen. Es ist aber unsicher, ob diese Studien für sich genommen den tatsächlichen Erfolg nicht unterschätzen.

Ein weiterer Aspekt ist ethischer Natur: Echte Kontrollgruppen können in Rückfallevaluationen nicht

gebildet werden, weil staatliches Eingreifen in jedem Fall erforderlich ist (Babcock et al. 2007). Zwei Vorgehensweisen haben methodische Vor- und Nachteile: Experimentelle Studien verwenden randomisierte, also zufällig eingeteilte, Interventions- und Kontrollgruppen. Oft werden in der Interventionsgruppe aber Absolventen und Abbrecher zusammengefasst und die Probanden in Kontrollgruppen haben bereits andere Interventionen erfahren (z. B. Inhaftierung). Quasi-experimentelle Studien nutzen hingegen den Vergleich von Programmabsolventen mit Abbrechern, d. h. letztere werden zur Kontrollgruppe. Sie erheben daher nicht randomisiert und können so bessere, aber weniger gesicherte Erfolgsraten nachweisen.

Rückfallerhebungen im Kontrollgruppendesign bleiben trotzdem ein erstrebenswertes Ziel, um verlässliche Aussagen über die

Wirksamkeit von Programmen zu erhalten. Diese Untersuchungsmethodik ermöglicht die Beurteilung, ob ein Interventionsansatz für eine Problemstellung grundsätzlich geeignet ist. Daher ist es ein Problem, wenn Rückfallstudien gänzlich fehlen, wie es in Deutschland beispielsweise bei häuslicher Gewalt der Fall ist.

Ein Ausweg aus den methodischen Schwierigkeiten mit Rückfällen als alleinigem oder hauptsächlichem Erfolgskriterium könnte sein, Rückfälle nicht kategorial (eingetreten – nicht eingetreten) sondern dimensional zu begreifen. Neben Auftreten oder Ausbleiben eines Rückfalls spielen bei der Erhebung dann der Zeitraum der Gewaltfreiheit sowie die genaue Art und Intensität des Vorfalls eine Rolle. Eine zweite, hier bevorzugte Alternative stellt eine Erhebung von so genannten Proxy-Variablen dar, d. h. Fallaspekten, die in einem belegbaren Zusammenhang zur Rückfallhäufigkeit stehen, aber leichter erhoben werden können.

Proxy-Variablen als Alternative bei der Wirkungsevaluation Nicht jede Evaluation ist mit einem umfangreichen Auftrag zur Rückfallerhebung und entsprechenden zeitlichen und finanziellen Ressourcen ausgestattet. Auch deshalb ist es sinnvoll, darüber nachzudenken, wie vorbereitend und ergänzend zu kontrollierten Studien die Logik der Rückfallerfassung auch weniger aufwendig umgesetzt werden kann und wie diese Evaluationen aussagekräftige Ergebnisse erzielen können. Viele Evaluationsvorhaben, besonders wenn sie einzelne Praxisprojekte untersuchen, arbeiten zudem mit eher kleinen Gruppen und haben keine Möglichkeit, die für Rückfallstudien erforderlichen Stichprobengrößen zu erreichen. Trotzdem stehen sie unter Erfolgsdruck.

Ein Ausweg kann über die Wahl von Evaluationskriterien erreicht werden, die eine Messung im Vorfeld von tatsächlichen Rückfällen ermöglichen. Solche Kriterien bei der Erfolgsmessung werden Proxy-Variablen (Stellvertreter) oder Vorfeldmaße genannt, weil sie einen empirisch belegten Zusammenhang zur Rückfallwahrscheinlichkeit aufweisen. Es gibt statische, unveränderbare Vorfeldmaße (z. B. männliches Geschlecht) und dynamische, also veränderbare Proxy-Variablen (z. B. Problemsicht), mit Hilfe derer durch die Intervention erreichte Wirkungen messbar werden. Verändern sich Proxy-Variablen im Verlauf einer Intervention in positiver Richtung, wird also ein Rückfall unwahrscheinlicher, kann dies als Erfolg gewertet werden. Dies ist nicht der Fall, wenn sich keine Veränderungen ergeben oder ein Rückfall sogar wahrscheinlicher wird.

Dynamische Proxy-Variablen, die für eine Erfolgseinschätzung von Interventionen geeignet sind, können für verschiedene Bereiche aggressiven und schädigenden Verhaltens gebildet werden. Bei genereller Gewaltkriminalität haben sich beispielsweise ein antisozialer Lebensstil, geringes Schuldbewusstsein und wenig Opferempathie als Vorfeldmaße einen dynamischen Zusammenhang zum Rückfallrisiko empirisch bestätigt (Mulder et al. 2010). Hohe Opferempathie und Behandlungcompliance, also die Bereitschaft, an der Intervention mitzuwirken, gelten bei Gewalt- wie Sexualdelikten hingegen als sehr vorhersagekräftig für Rückfallfreiheit. Für Rückfälle bei sexueller Kindesmisshandlung sind Defizite in der Fähigkeit zu Intimität, kognitive Verzerrungen gegenüber sexuellen Übergriffen, Zugang zu potentiellen Opfern und eine negative

Lebenseinstellung als Vorhersagevariablen empirisch belegt (Kindler 2001). Bei Kindesmisshandlung und -vernachlässigung werden dynamische Vorfeldmaße mit hohem Rückfallzusammenhang besonders bezogen auf Fürsorgeverhalten beschrieben. Beispiele sind rigide oder grob unpassende Erziehungsvorstellungen oder geringe Feinfühligkeit (für eine Übersicht: Hagemann-White et al. 2010).

Erste Erfahrungen mit dieser Vorgehensweise bei der Evaluation liegen mittlerweile auch in Deutschland vor. Bei der Therapie von sexuellen Kindesmisshandlern konnte eine proxy-gestützte Evaluation beispielsweise positive Effekte der Intervention belegen (Kindler 2001; 2009). Gegenwärtig wird diese Evaluationsform im Rahmen eines Projekts für Täterprogramme bei Partnerschaftsgewalt erprobt (vgl. nächster Abschnitt).

Die Abschätzung von Interventionserfolgen über eine Erhebung von Proxy-Variablen für die Rückfallgefahr lässt sich natürlich in ausgefeiltere Evaluationskonzepte einbinden. So ist es etwa manchmal sinnvoll, Fallschweregrade zu bilden, d. h. zu Hilfebeginn bereits erkennbare Untergruppen mit hohem oder geringem Rückfallrisiko zu unterscheiden. Eine solche Unterscheidung kann sinnvoll sein, da nicht ohne Weiteres anzunehmen ist, dass ein Interventionskonzept bei allen Fallschweregraden gleichermaßen erfolgreich ist. Eine Bildung von Fallschweregraden, ist besonders in Interventionsfeldern mit einem breiten und heterogenen kliententypologischen Spektrum wie bei Partnerschaftsgewalt von Interesse. Ergänzend sinnvoll zu einer summativen Evaluation ist es zudem, die Prozessqualität der Programme zu beschreiben (formativer Aspekt). Denn bisher ist wenig bekannt darüber, welches

Programmhandeln förderlich ist im Vergleich zu anderem. In der internationalen Forschung ist man deshalb dazu übergegangen, günstige Bedingungen für Programmabschlüsse verglichen mit Abbrüchen näher zu untersuchen.

se ein Selbstevaluationskonzept zu einem sozialen Training für Partnerschaftsgewalttäter, durchgeführt von der Arbeiterwohlfahrt und der Diakonie Düsseldorf, zu erstellen (Liel, Kindler 2009). Das DJI interessierte sich für die



Abb. 2: Vorgehen im Evaluationsprojekt

Arbeit mit Partnerschaftsgewalttätern – ein proxy-gestütztes Selbstevaluationskonzept Im Folgenden wird diese Evaluationsproblematik fokussiert auf Täterarbeit bei häuslicher Gewalt (BMFSFJ 2008). Es wird ein Evaluationsvorgehen vorgestellt, das im Rahmen eines wissenschaftlichen Projekts am Deutschen Jugendinstitut e.V. (DJI) für die Praxis von Täterprogrammen entwickelt wurde und aktuell an drei Standorten erprobt wird. Auch im Feld der Partnerschaftsgewalt, wie häuslichen Gewalt synonym genannt wird, gilt es als gleichermaßen schwierig, Verhaltensänderungen praktisch herbeizuführen wie Erfolge methodisch zu messen. Dennoch können selbst geringe Verbesserungen für Betroffene und Gemeinwesen praktisch sehr bedeutsam sein.

Ausgangspunkt des Projekts war der Auftrag durch das Jugendamt Düsseldorf an das DJI, im Rahmen einer wissenschaftlichen Experti-

Anfrage, weil Täterprogramme in der Jugendhilfe zunehmend auch für Kinderschutzfälle angefragt werden, kontrollierte Rückfallstudien in Deutschland aber bisher fehlen. Zur Strukturierung des Auftrags wurden Beschreibungs- und Einschätzungsaufgaben definiert und empirisch geprüft, die von den Fachkräften bei der Selbstevaluation zu erfüllen sind. Diese Aufgaben betreffen die Bereiche Bewertung der Konzeptqualität, Dokumentation von Falldaten und diagnostische Einschätzungen, zu denen auch die Erfolgsbewertung zählt.

Kriterien für die für die Erfüllung der Beschreibungs- und Bewertungsaufgaben durch die Fachkräfte wurden durch eine Abfrage des internationalen Forschungsstandes gewonnen. Es wurde nach Variablen gesucht, die Rückfallfreiheit und einen Programmabschluss prognostisch vorhersagen. Über fünfzig empirische Studien und Forschungs-

berichte wurden gesichtet. Sofern der Forschungsstand für einzelne Evaluationsaufgaben keine Bewertungskriterien liefern konnte, wurde auf einen Praktikerkonsens und auf im Arbeitsfeld geltende Qualitätsstandards zurückgegriffen. Die Fachkräfte vor Ort waren durch eine Befragung über die bisherige Wahrnehmung der Beschreibungs- und Bewertungsaufgaben und die Diskussion der Ergebnisse in den Prozess eingebunden. Für die Erfolgseinschätzung lieferten vier Rückfallstudien empirische Messkriterien. Auf dieser Grundlage wurden standardisierte Erhebungsinstrumente erstellt und zum Teil unter Rückgriff auf forschungserprobte und validierte Instrumente ergänzt. Bewertungsgrundlage ist weitgehend eine Fachkräfteeinschätzung.

Auf dem Rückfallrisiko basierende Baseline-Erhebung und Erfolgsbewertung Eine Baseline-Erhebung stellt vergleichbare Informationen über die Programmteilnehmer bereit, die Grundlage nicht nur für eine diagnostische Eignungseinschätzung im Programm ist, sondern auch für die spätere Erfolgsbewertung. Die hier erhobenen Daten bilden den Nullpunkt auf der Bewertungsskala, von dem aus Veränderungen nach der Intervention als positiv oder negativ bewertet werden können.

Gängiges Procedere in den Praxisprojekten ist ein Aufnahmeverfahren mit einer Fall- und Beratungsstatistik. Belegbar prognostisch relevant ist dabei die Frage nach dem Ausmaß der Partnerschaftsgewalt. Die richtige Erhebung von Daten wird in der Forschungsliteratur jedoch viel diskutiert, z. B. indem an standardisierten Selbstbefragungsinstrumenten kritisiert wird, das Gewaltproblem verzerrt und oberflächlich abzubilden. Daher wurde

Grad der Verantwortungsabwehr		Codierung
(1)	Leugnet jegliche Gewalt und fühlt sich als Opfer <i>Ankerbeispiel:</i> Berichtet dass die Partnerin mit den Kindern heimlich ins Frauenhaus gegangen sei, und weiß nicht warum, außer dass sie ihm schaden wolle	4
(2)	Räumt nur vergleichsweise zu den Tatvorwürfen harmlose Gewalt ein, sieht zudem die Verantwortung anderswo <i>Ankerbeispiel:</i> Gesteht ein, Partnerin zweimal mit der flachen Hand geschlagen zu haben, leugnet aber massive Misshandlungen oder generelle Kontrolle der Partnerin	3
(3)	Räumt einzelne von mehreren bekannt gewordenen Gewaltakten ein, übernimmt nur partielle Verantwortung <i>Ankerbeispiel:</i> Berichtet, Partnerin geschlagen zu haben, es sei aber ganz anders gewesen als in den Unterlagen dargestellt	2
(4)	Räumt die meiste der bekannt gewordenen Gewalt ein, sieht die Verantwortung gleich verteilt <i>Ankerbeispiel:</i> Sagt, er habe seine Partnerin zwar am Hals gepackt, aber nicht zgedrückt und gewürgt, oder bekräftigt wiederholt, auch über die Gewalt reden zu wollen, die er durch seine Partnerin erfahren hat	1
(5)	Übernimmt die Verantwortung für die gesamte bekannt gewordene Gewalt <i>Ankerbeispiel:</i> Gesteht seine Gewalt ein und seine Tatschilderung ist weitgehend deckungsgleich mit den Unterlagen	0

Abb. 3: Instrumentenbeispiel (dynamisches Proxy):
Änderungssensivität des Klienten

hier eine Befragung der Klienten zu Gewaltformen, Häufigkeit und Anzahl betroffener Partnerinnen (Straus et al. 1996) gewählt, die eine Fachkräftebewertung der Information einschließt und sie um Daten zur Einbindung der Gewalt in nichtkörperliche Kontrollmuster laut Partnerin (Johnson, Leone 2005) ergänzt. Besonders schwere Formen von Partnerschaftsgewalt und behandlungsbedürftige Verletzungen der Partnerin sechs Monate vor Programmbeginn erhöhen die Rückfallwahrscheinlichkeit signifikant. Deshalb eignen sie sich auch als Proxy, um die Beschreibung von Fallschweregraden im Aufnahmeverfahren vorzunehmen. Als weitere Proxy-Variablen für die Bestimmung des Fallschweregrades wurden Ärgerstörungen und Impulsivität außerhalb von Partnerschaften, generelle Gewaltkriminalität, Alkohol- und Drogenprobleme, ein junges Alter und geringe Änderungssensivität, d. h. Problemeinsicht und Veränderungsbereitschaft, des Klienten gefunden

(Eckhardt et al. 2008; Benett et al. 2007; Murphy et al. 2003).

Bei Ärgerstörungen konnten Eckhardt et al. (2008) einen engen Zusammenhang zum Rückfallrisiko, zu in der Herkunftsfamilie miterlebter Partnerschaftsgewalt sowie zu psychopathologischen Auffälligkeiten nachweisen. Ein Rückgriff auf dieses Kriterium und die Verknüpfung mit genereller Gewaltkriminalität sind interessant, weil die Testung psychopathologischer Auffälligkeiten für eine Evaluation in der Praxis meist zu aufwendig ist. Ärgerstörungen und generelle Gewaltkriminalität lassen sich wahrscheinlich einfacher erfragen. Für beides wurden Instrumente konzipiert. In der Forschungsliteratur nicht unkritisiert bleiben Erhebungsmaße für die Proxys Alkohol- und Drogenkonsum und Änderungssensivität des Klienten. Einfache Erhebungen wie der CAGE-Test für Suchtmittelabhängigkeit oder die Stages of Change, einem Maß für Veränderungszyklen, setzen eine hohe Offenheit voraus und sind

zu wenig eindeutig. Zumindest für Suchtmittelkonsum wurde ein aussagekräftiges Instrument in einer deutschen Evaluation (Helfferich et al. 2006) gefunden und für Änderungssensivität wurde ein Forschungsmaß aus der Studie von Benett et al. (2007) weiterentwickelt, das die Veränderungsbereitschaft indirekt über die Problemsicht der Klienten erhebt: Zum Teil werden auch nach Programmaufnahme vertiefte Risikoabwägungen für bereits identifizierte Klienten benötigt. Für Gefährdungseinschätzungen im Einzelfall und für eine Überprüfung der Erhebung von Fallschweregraden wurde daher ein sehr validiertes Instrument zur Gefährdungsabschätzung, der Domestic Violence Risk Appraisal Guide (Hilton et al. 2010), übersetzt.

Da Änderungssensivität eine dynamische Proxy-Variable ist, eignet sie sich auch zur Erfolgseinschätzung nach dem Täterprogramm. Dabei hat es sich praktisch nützlich erwiesen, nicht nur die Verantwortungsübernahme für

die Gewalt selbst, sondern auch die Vorstellungen von Rückfallvermeidung für Gefahrensituationen zu bewerten, z. B. wieweit Notfallpläne ausgearbeitet sind. In empirischen Studien wurde auch die Sichtweise der gewaltbetroffenen Partnerin als vorhersagekräftig für die Rückfallgefährdung identifiziert. Ein Befund von (Heckert, Gondolf 2004) belegt angegebenes Sicherheitsgefühl der Partnerin und von der Partnerin eingeschätzte Rückfallgefährdungen als prospektiv ähnlich vorhersagekräftig wie einfache validierte Risikoeinschätzungsverfahren. Die Erhebungsmaße dieser Studie wurden deshalb für die Erfolgseinschätzung im Prä-Post-Vergleich übernommen.

Ergänzend zur Fachkräfte- und Partnerinneneinschätzung kann auch der Selbstbericht von Klienten Informationen für die Erfolgseinschätzung liefern. In vielen Schädigungsfeldern (z. B. bei Sexualstraftätern) hat sich die Empathiefähigkeit als Zusammenhangsmaß zum Rückfallrisiko bewiesen. Zwar wurde keine Überprüfung dieser Proxy-Variable auf ihre Vorhersagekraft bei Partnerschaftsgewalttätern gefunden. Jedoch bestätigte eine Studie einen Selbstberichtsfragebogen zur Empathiefähigkeit bei Partnerschaftsgewalttätern, den Interpersonal Reactivity Index, der deshalb auch für die Erfolgseinschätzung genutzt werden kann (Sartin 2004). Eine weitere Proxy-Variable betrifft das Potential für Kindeswohlgefährdungen. Weil Partnerschaftsgewalt selbst auch als Indikator für Kindeswohlgefährdungen gilt, können diesbezügliche Veränderungen auch als Wirkungsaussagen von Interventionen verstanden werden. Das Child Abuse Potential Inventory, das am besten validierte Selbstbeurteilungsmaß für Eltern zu Kindeswohlgefährdungen, wurde

bereits an Partnerschaftsgewalttätern erprobt (Deegener et al. 2009). Beide Fragebögen wurden für einen Prä-Post-Vergleich in das Evaluationskonzept aufgenommen.

Perspektiven des Evaluationsprojektes Das Konzept der Selbstevaluation mittels Proxy-Variablen eignet sich für eine externe Datenauswertung von mehreren Standorten. Nach Abschluss der DJI-Expertise wurden die entwickelten Instrumente vom Jugendamt Düsseldorf zur weiteren Datengewinnung dankenswerter Weise freigegeben und durch den Autor weiter ausgearbeitet. Zwei weitere, charakteristisch unterschiedliche Einrichtungen in München und Rosenheim signalisierten Interesse an der Evaluation. Die Datenerhebung hat im Jahr 2010 an drei Standorten begonnen. Beweisen sich die Erhebungsinstrumente in einem Pre-Test in Düsseldorf, München und Rosenheim, wird eine Untersuchung mit weiteren Projektstandorten angestrebt. Ergänzend zu der Erhebung des Rückfallrisikos mit Vorfeldmaßen ist eine Rückfallerhebung in einem Follow-up-Zeitraum mittels polizeilicher Datenabfrage und Klienten- bzw. Partnerinnenbefragung vorgesehen.

Als Beleg für ein proxy-gestütztes Vorgehen bei häuslicher Gewalt sei zum Schluss auf eine große amerikanische quasi-experimentelle Rückfallstudie zu Täterprogrammen verwiesen (Gondolf 2002). Mit Hilfe zusätzlicher Beschreibungs- und Vorhersagekriterien für das Rückfallrisiko und dem Vergleich von vier Programmstandorten konnte die Studie signifikante Effekte nachweisen, beispielsweise ein um 44-64% geringere Rückfallwahrscheinlichkeit bei Programmabschlüssen gegenüber Abbrüchen.

Literatur

- Babcock, Julia C.; Brittany, E. Canady; Graham, Katherine; Schart, Leslie 2007: The Evolution of Battering Interventions: From the Dark Ages Into the Scientific Age. In: Hame, John; Nicholls, Tonia (Hg.): Family Interventions in Domestic Violence: A Handbook of Gender-Inclusive Theory and Treatment. New York: Springer Publishing Company
- Babcock, Julia C.; Green, Charles E.; Robie, Chet 2004: Does Batterers' Treatment Work? A Meta-Analytic Review of Domestic Violence Treatment. In: Clinical Psychology Review, Jg. 23, S. 1023-1053
- Benett, Larry W.; Stoops, Charles; Call, Christine; Flett, Heather 2007: Program Completion and Re-Arrest in a Batterer Intervention System. In: Research on Social Work Practice, Jg. 17, H. 1, S. 42-54
- Bundesministerium für Familie, Senioren Frauen und Jugend 2008: Standards und Empfehlungen für die Arbeit mit männlichen Tätern in interinstitutionellen Kooperationsbündnissen gegen Häusliche Gewalt. Materialien zur Gleichstellung Nr. 109/2008. Berlin
- Deegener, G.; Spangler, G.; Körner, W.; Becker, N. 2009: Eltern-Belastungs-Screening zur Kindeswohlgefährdung. Deutsche Form des Child Abuse Potential Inventory (CAPI) von Joel S. Milner. Göttingen: Hogrefe Verlag
- Eckhardt, Christopher I.; Samper, Rita E.; Murphy, Christopher M. 2008: Anger Disturbances among Perpetrators of Intimate Partner Violence. Clinical Characteristics and Outcomes of Court-Mandated Treatment. In: Journal of Interpersonal Violence, Jg. 23, H. 11, S. 1600-1617
- Feder, Lynette; Wilson, David B. 2005: A Meta-Analytic Review of Court Mandated Batterer Intervention Programs: Can Courts Affect Abusers' Behavior? In: Journal of Experimental Criminology, Jg. 1, S. 239-262
- Gondolf, Edward W. 2002: Batterer Intervention Systems. Issues, Outcomes and Recommendations. Thousand Oaks: Sage Publications
- Gondolf, Edward W. (2004): Evaluating Batterer Counseling Programs: A Difficult Task Showing Some Effects and Implications. In: Aggression and Violent Behavior, H. 9, S. 605-631
- Hagemann-White, Carol; Kavemann, Barbara; Kindler, Heinz; Meysen, Thomas; Puchert, Ralf 2010: Review of Research

on Factors at Play in Perpetration. Report for the European Commission

Heckert, D. Alex; Gondolf, Edward W. 2004: Battered Women's Perceptions of Risk Versus Risk Factors and Instruments in Predicting Repeat Reassault. In: Journal of Interpersonal Violence, Jg. 19, H. 7, S. 778-800

Helfferich, Cornelia; Barz; Monika 2006: Häusliche Gewalt beenden: Verhaltensänderung von Tätern als Ansatzpunkt. Eine Evaluationsstudie zum Vorgehen und Wirkung von Täterprogrammen im Kontext von Interventionsprojekten gegen häusliche Gewalt in Baden-Württemberg. Stuttgart: Landesstiftung Baden-Württemberg

Hilton, N. Zoe; Harris, Grant T.; Rice, Marnie E. (2010): Risk Assessment for Domestically Violent Men. Tools for Criminal Justice, Offender Intervention, and Victim Services. Washington DC: American Psychological Association

Johnson, Michael P.; Leone, Janel M. (2005): The Differential Effects of Intimate Terrorism and Situational Couple Violence. Findings from the National Violence Against Women Survey. In: Journal of Family Issues, Jg. 26, H. 3, S. 322-349

Kindler, Heinz 2001; 2009: Evaluation der ambulanten Arbeit mit sexuellen Kindesmisshandlern. Expertise zum Projekt und Zusammenfassung der Ergebnisse. München

Liel, Christoph; Kindler, Heinz 2009: Selbstevaluation des sozialen Trainings. Wissenschaftliches Konzept. München: Deutsches Jugendinstitut e.V.

Mulder, Eva; Brand, Eddy; Bullens, Ruud; van Marle, Hjalmar (2010): A Classification of Risk Factors in Serious Juvenile Offenders and the Relation between Patterns of Risk Factors and Recidivism. In: Criminal Behaviour and Mental Health, Jg. 20, S. 23-38

Murphy, Christopher M.; Morrel, Tanya M.; Elliot, Jeffery D.; Neavis, Tara M. 2003: A Prognostic Indicator Scale for the Treatment of Partner Abuse Perpetrators. In: Journal of Interpersonal Violence. In: Journal of Interpersonal Violence, Jg. 18, H. 9, S. 1087-1105

Ohlemacher, Thomas; Sögdig, Dennis; Höynck, Theresia; Ethé, Nicole; Welte, Götz 2001: Anti-Aggressivitätstraining und Legalbewährung: Versuch einer Evaluation. Hannover: Kriminologisches Forschungsinstitut Niedersachsen (Forschungsbericht Nr. 83)

Sartin, Robert M. 2004: Characteristics Associated with Domestic Violence Perpetration: An Examination of Factors Related to Treatment Response and the Utility of a Batterer Typology. Lincoln. University of Nebraska

Straus, Murray A.; Hamby, Sherry L.; Boney-McCoy, Sue; Sugarman, David B. 1996: The Revised Conflict Tactics Scale (CTS2). Development and Preliminary Psychometric Data. In: Journal of Family Issues, Jg. 17, H. 3, S. 283-316

Wottawa, Heinrich; Thierau, Heike 1998: Lehrbuch Evaluation. 2. Aufl. Bern: Huber



Christoph Liel, Jg. 1973, Dipl.-Sozialarbeiter, arbeitet im Münchner Informationszentrum für Männer e.V. und war wissenschaftlicher Referent in 2 Forschungsprojekten am Deutschen Jugendinstitut e.V., Interessen- und Forschungsschwerpunkte: Interventionen und Risikoeinschätzung bei Gewalt und Kindeswohlgefährdung. liel@gmx.de