

«Evidenzbasierung» als Kriterium der Psychotherapie-Selektion?

Über eine gutes Konzept – und seine missbräuchliche Verwendung

Jürgen Kriz

Psychotherapie-Wissenschaft 9 (2) 42–50 2019

www.psychotherapie-wissenschaft.info

CC BY-NC-ND

<https://doi.org/10.30820/1664-9583-2019-2-42>

Zusammenfassung: In diesem Beitrag wird zunächst hervorgehoben, dass «Evidenzbasierung» ein gutes und nützliches Konzept ist, aber durch eine Reduktion auf RCTs inzwischen oft missbräuchlich verwendet wird. Zwar ist das RCT-Design durchaus ein hervorragendes Modell zur Untersuchung experimenteller Forschungsfragen. Dies gilt aber nur dann, wenn klar definierbare Ursachen und ebenso klar definierbare Wirkungen hinreichend isolierbar sind und diese dabei die Untersuchungsrealität brauchbar abbilden. Je grösser der Handlungsspielraum zur Gestaltung der Ursachen (Interventionen) ist und je komplexer die relevanten Wirkungen sind, desto weniger tauglich ist dieser Ansatz. Wie sodann aufgezeigt wird, eignen sich RCT-Designs in der Psychotherapieforschung besonders dann, wenn manualisierte *Programme angewendet* werden. Dies ist primär für die Verhaltenstherapie der Fall. Für psychotherapeutisches Vorgehen, das wesentlich in der prozess-spezifischen *Entfaltung von Prinzipien* begründet ist, sprengt die grosse Variabilität der Detail-Ursachen (Interventionen) die erforderliche massenstatistische Homogenität für ein experimentelles Design. Die Forderung, Wirksamkeit vor allem mit RCT-Designs zu belegen, läuft dort auf methodische Artefakte hinaus, beziehungsweise stülpt eine bestimmte psychotherapeutische Präferenz, nämlich für die Verhaltenstherapie, auf dem Wege der Wissenschaftsmethodik anderen Psychotherapieansätzen über. Abschliessend werden noch einige weitere Reduktionen und Missinterpretationen im Zusammenhang mit RCT- und Metastudien-Designs problematisiert.

Schlüsselwörter: Effektstärke, Evidenzbasierung, Passung, psychotherapeutische Verfahren, RCT, Wirkfaktoren

Das «eigentlich» gute Anliegen der Evidenzbasierung

Mit dem Attribut «evidenzbasiert» ist gewöhnlich das Konzept «Evidenzbasierte Medizin» (EbM) einer Forschergruppe um David Sackett (Sackett et al., 1997) gemeint. Es war das Anliegen dieser Gruppe, professionalisierte Hilfestellung durch wissenschaftlich gesicherte Informationen in Form von Übersichten und Bewertungshilfen zur Verfügung zu stellen. Weder PraktikerInnen noch einzelne ForscherInnen können heutzutage die ungeheure Fülle an Literatur mit klinischen Studien noch überblicken und in ihrer Aussagekraft und Güte kritisch bewerten. Mit dem Programm der Evidenzbasierung sollen diese besser in die Lage versetzt werden, den aktuellen Stand der Forschung in einem bestimmten Fachgebiet mit ihrer ärztlichen bzw. psychotherapeutischen Erfahrung und der individuellen PatientInnenperspektive so zusammenzuführen, dass der/m jeweiligen Patientin/en die beste Behandlung angeboten werden kann. So weit der fraglos gute Ansatz der EbM, der sich rasch verbreitet hat: Besonders von der internationalen Cochrane Collaboration (CC) werden Bewertungen und Zusammenfassungen (sogenannte *Reviews*) von wissenschaftlichen Studien im Bereich der Medizin und teilweise auch der Psychotherapie vorgenommen.

Ob dieser nützliche *Überblick* über real durchgeführte *Forschungen* nach bestimmten Kriterien (s. nachfolgend) nun aber auch ein nützliches *Kriterium* dafür sein kann, welche Methoden *in der Praxis angewendet bzw. unterlassen* werden sollten, hängt von der Erfüllung bestimmter Voraussetzungen ab. Dazu zählt zum einen insbesondere die Frage, wie weit das Spektrum realer Behandlungen tatsächlich und hinreichend biasfrei im Spektrum dieser Forschung wiedergespiegelt wird. Zum anderen geht es um die Frage, wie brauchbar sich das methodische Modell dieser Forschung nicht nur für die Beurteilung von Pillen und Apparaten sondern auch für Psychotherapie eignet.

Schon hinsichtlich der ersten Frage werden oft erhebliche Zweifel geäussert, die mit Interessenkonflikten zusammenhängen (vgl. z.B. Kienle et al., 2003). Denn Forschung orientiert sich nicht einfach wertfrei an inhaltlich relevanten Fragen der KlinikerInnen bzw. an zu klärenden Fragen für die PatientInnen. Für Forschung zu Medikamenten, Apparatemedizin etc., die weitgehend von Firmen finanziert werden, zeigt sich vielmehr, dass überwiegend Therapien erforscht werden, die patentierbar und gewinnversprechend sind. Zudem werden pharmazeutische Unternehmen begünstigt, die mit wenigen Medikamenten grossen Umsatz machen. Diese sind die Gewinner im EbM-Biotop. Verlierer sind dagegen Therapien ohne Aussicht auf breite Vermarktung oder ohne

finanzstarke industrielle Rückendeckung, wie zum Beispiel nicht-pharmakologische Therapien.

Aber auch bei Forschung im Bereich von Psycho- und Sozialtherapie, Reha-Massnahmen etc., die primär über universitäre und öffentliche Mittel finanziert wird, steht die Kritik oft mangelnder Neutralität im Raum, denn es werden bevorzugt Fragestellungen erforscht, die im Rahmen von Diplom-, Doktor- und Habilitationsarbeiten angegangen werden und vergleichsweise schnell und einfach publiziert werden können – das heisst Arbeiten, die gut in universitäre Karrierestrukturen passen. So lässt sich beispielsweise eine spezielle Vorgehensweise (die dann dem Spektrum Verhaltenstherapie zugerechnet werden kann) an einer speziellen PatientInnengruppe sehr gut im Rahmen einer Dissertation experimentell untersuchen. Die Wirkung einer langfristigen, kaum manualisierbaren Vorgehensweise – wie etwa der Humanistischen Psychotherapie oder der Langzeit-Psychoanalyse – lässt sich in diesem Rahmen so praktisch nicht untersuchen bzw. muss in ein für diese Ansätze unpassendes Prokrustesbett zurechtgeschnitten werden, bei dem Wichtiges, wenn nicht gar Wesentliches verloren geht. Dies wird besonders deutlich, wenn man einen aktuellen Beitrag von Strauß (2019) betrachtet. Strauß – immerhin alternierender Vorsitzender des deutschen Wissenschaftlichen Beirates Psychotherapie (WBP), verweist dabei auf die Bedeutsamkeit der sogenannten «kontextuellen Faktoren» für das Verständnis der Wirksamkeit von Psychotherapie, die sich aus der internationalen Forschung der letzten Jahrzehnte herauskristallisiert haben. Drei der sechs wichtigsten kontextuellen Faktoren (vgl. Tabelle in Strauß, 2019, S. 6) entsprechen exakt den drei zentralen Bedingungen, die Rogers (1957) für die Personzentrierte Psychotherapie als ein Ergebnis seiner Forschung herausgestellt hat und die inzwischen für die Humanistische Psychotherapie essenziell sind. Derselbe WBP hat aber gerade der Humanistischen Psychotherapie die wissenschaftliche Anerkennung verweigert und die Personzentrierte Psychotherapie mit dem Hinweis, sie sei nach derzeitigen Kriterien nicht mehr wissenschaftlich anerkannt, massiv diskreditiert (auch wenn dieses Werturteil selbst beim WBP auf dem vermeintlichen «Fehlen» einer einzigen Studie beruht).

Der durchaus informative Überblick über die Forschungsliteratur durch EbM oder CC spiegelt somit keineswegs nur die *Nützlichkeit von Interventionsmethoden* wider, sondern in erheblichem Ausmass *die Nützlichkeit der Forschung unter kommerziellen oder Wissenschaftskarriere-Aspekten*. Noch bedenklicher ist allerdings, dass, trotz eindringlicher Warnungen von Sackett und seiner Gruppe selbst, EbM inzwischen missbräuchlich in ein Instrument zur Selektion und Massregelung umfunktioniert wird. So werden unter dem Begriff «evidenzbasierter Gesundheitsversorgung» die Prinzipien der EbM auf immer mehr Entscheidungen zur Steuerung des Gesundheitssystems angewandt.

Aktuelles Beispiel ist die gemeinsame Stellungnahme der Deutschen Gesellschaft für Psychologie (Prof. Spinath), des Fakultätentages Psychologie (Prof. Antoni) und der Kommission Psychologie und Psychotherapieausbildung (Prof. Rief) vom Mai 2019 zum deutschen Gesetzentwurf

für das neue Studium zur Approbation in Psychotherapie: Darin wird die vollständige Auflösung der psychotherapeutischen Verfahren zugunsten eines Baukasten-Prinzips aus störungsspezifischen, evidenzbasierten Interventionen gefordert. Dabei wird übersehen, dass diese Interventionen im Rahmen der Untersuchung ihrer Evidenz fast alle von TherapeutInnen durchgeführt wurden, die keineswegs theorie- oder schulenfrei agierten. Die Evidenz dieser Interventionen im Rahmen einer reinen Baukastentherapie ist somit gar nicht untersucht worden. Dazu muss bedacht werden, dass ohnehin der Anteil spezifischer Techniken am Therapieergebnis überaus gering ist – Norcross und Lambert (2011) schätzen ihn auf unter 10 %. Über 90 % dessen, was in der Psychotherapieforschung als wirksam gilt, soll also nach solchen Vorstellungen in Zukunft den PatientInnen vorenthalten werden – oder durch intuitiven Dilettantismus ersetzt werden. Das aber ist ganz sicher nie das Anliegen von Evidenzbasierung gewesen.

Für die Praxis ähnlich schädlich ist, dass die von Sackett konzipierten sogenannten «Evidenzklassen», die immerhin noch Informationen aus unterschiedlichen Studiendesigns berücksichtigen, durch eine fast ausschliessliche Fokussierung auf Randomisierte kontrollierte Studien (RCT) ersetzt wurden. Bei diesen RCTs geht es allein um die Wirksamkeit von abgegrenzten Einzelfaktoren bezogen auf abgegrenzte Symptome, unter streng kontrollierten Laborbedingungen. Der inzwischen übliche Etikettenschwindel, «evidenzbasiert» irrtümlich oder missbräuchlich allein auf RCTs zu reduzieren, hat die methodische Monokultur ganz erheblich vorangetrieben. Denn es gelten faktisch nur noch Belege, die mit experimentellen Designs erbracht wurden.

Aus diesem Grund ist es angezeigt, die Vor- und Nachteile der experimentellen Konzeption hier näher zu untersuchen (wobei die Argumentation Aspekte aus Kriz, 2000, 2007, 2008 mit aufgreift – vgl. auch Schulthess, 2019 und besonders in Kriz, 2014a, b, 2015 teilweise ausführlicher dargestellt wird).

Grundlagen des Experiments

Das zentrale Fundament, auf dem in den empirischen Wissenschaften und den von ihnen durchdrungenen gesellschaftlichen Diskursen der Geltungsanspruch von Aussagen beruht, ist das Experiment. Dieses beruht auf dem Grundschemata, dass ein Untersuchungsgegenstand bestimmten Manipulationen ausgesetzt wird, die als «unabhängige Variable» (UV) bezeichnet werden. Die Reaktionen des Untersuchungsgegenstandes werden an den Veränderungen von «abhängigen Variablen» (AV) registriert. Sofern man alle weiteren Einflüsse ausschalten kann, lässt sich diese Veränderung der AV als Funktion der UV verstehen, also $AV = f(UV)$.



Abb. 1: Experimentelles Grundschemata

In experimentellen Forschungsdesigns der Psychologie, Psychotherapie, Medizin und Pharmakologie (am Menschen) sind Untersuchungsgegenstände – je nach Nomenklatur – «Versuchspersonen» (Vpn), «PatientInnen», «Untersuchungsteilnehmer» (UT) usw. Gerade im Rahmen moderner Psychologie beziehen sich zahlreiche Fragen, die gestellt werden, auf einen sehr spezifischen, klar umrissenen Forschungsgegenstand, bei dem einzelne Wirkvariablen im psychologischen Labor gut isolierbar sind. Ein nicht unbeträchtlicher Teil der Reputation der Psychologie als Wissenschaft beruht auf einer überwältigenden Fülle an Befunden zu Fragen über menschliche(s) Wahrnehmung, Lernen, Gedächtnis, Denken, Sprache usw., die ganz überwiegend durch eine experimentelle Herangehensweise untersucht wurden.

Das ist allerdings keineswegs so unproblematisch, wie es zu sein scheint. Denn die Vorgehensweise setzt voraus, dass die unabhängige Variable in Abbildung 1 eben auch (hinreichend) unabhängig ist – und dies wiederum setzt voraus, dass die Untersuchungsgegenstände (hinreichend) adäquat als Objekte behandelt werden können. Sobald aber Untersuchungsgegenstände eher als Subjekte gesehen werden müssen, die die Beschaffenheit der unabhängigen Variable erheblich beeinflussen, indem sie diesen unterschiedliche Bedeutungen geben, sind die UV eben nicht als «unabhängig» sondern als «abhängig» anzusehen.

Nun hat schon vor über 100 Jahren die klassische Gestaltpsychologie der Berliner Schule (Wertheimer, Koffka, Köhler, Lewin, Goldstein u. a.) auf vielfältige Weise zeigen können, dass in vielen Situationen nicht so sehr die «objektiven» (von Physikern beschreibbaren) UV einer Reizwelt für Menschen beim Wahrnehmen, Erinnern, produktivem Denken etc. entscheidend sind, sondern die *Bedeutungen* (in Form von «Gestalten») die diesen Reizkonfigurationen jeweils durch den Menschen als Subjekt gegeben werden. Komplementär zu der Beschreibungsperspektive durch WissenschaftlerInnen, BeobachterInnen oder Manuale müsste somit auch immer die des Subjekts in seiner Lebenswelt (vgl. Kriz, 2017) mitberücksichtigt werden – ein grosses Manko in den gegenwärtigen Psychotherapie-Diskursen.

Eigentlich sollten sich gerade PsychologInnen über diese Einwände gegenüber dem Konzept unabhängiger Variablen nicht wundern, beruht doch eines ihrer Lieblingsparadigmen auf der klassischen Konditionierung nach Pawlow – bei der eben physikalisch gesehen ein und dieselbe Glocke in Abhängigkeit von den Erfahrungen des Hundes die Bedeutung von «Glocke» als «Futter» erhält (diese Wörter sind natürlich unsere anthropomorphen Bezeichnungen). «Lernen» meint eben, dass die angeblich unabhängigen Variablen von der Lerngeschichte abhängig sind. Prinzipiell lassen sich zwar auch diese Lerngeschichten bisweilen kontrollieren oder gar selbst zum Gegenstand einer experimentellen Frage machen. Aber es gibt stets auch eine Fülle nicht erfasster Lerngeschichten der Untersuchten – und damit deren unterschiedlichen subjektiven Bedeutungszuweisungen –, die das Konzept unabhängiger Variablen obsolet machen.

Das wird vielleicht noch eindrucksvoller am Beispiel der Frühphase der kognitiven Verhaltenstherapie (VT): Meichenbaum (1979, S. 109ff.) wollte die Effizienz des VT-Standardverfahrens (Wolpe & Lazarus, 1966) zur Beseitigung von Schlangenphobien durch eine Erweiterung der Intervention um kognitive Anteile steigern. Neben der Versuchsgruppe, die die «richtige» Reihenfolge erhielt, und wo angesichts der Schlange negative Verstärkung erfolgte, wenn die PatientInnen sich entspannte, gab es noch eine Kontrollgruppe. Diese erhielt die negative Verstärkung genau dann, wenn die PatientInnen von der Schlange angewidert war. Eigentlich müsste sich nach den klassischen Lerntheorien bei der Kontrollgruppe die Schlangenphobie erhöhen oder zumindest stabilisieren. Überraschenderweise waren aber beide Programme gleich wirksam. Durch Befragung fand Meichenbaum heraus, dass die PatientInnen der Kontrollgruppe die Situation und die Aussagen bezüglich der Angst erzeugenden Selbstanweisung einfach für sich umgedeutet hatten. Die Lerntheorien beschrieben somit die «objektive Reizstruktur» der Intervention, nach der die Phobie hätte verstärkt werden müssen, die PatientInnen aber organisierten sich diese Struktur eigenständig in eine «subjektive Rezeptionsstruktur» um, die völlig anders wirkte. Dieses Beispiel sollte vielleicht manchen, der zu sehr auf Manualisierung setzt und zu sehr an «unabhängige Variablen» im zuvor angeführten experimentellen Design glaubt, nachdenklich machen.

RCT-Forschung als Repräsentant des experimentellen Ansatzes

Wie bereits zuvor betont wurde, sind zahlreiche Fragen, die im Rahmen moderner Psychologie gestellt werden, optimal durch eine experimentelle Herangehensweise nach dem Schema in Abbildung 1 untersuchbar. Allerdings passt dieses Modell keineswegs für alle Fragen. So werden in Experimenten und der entsprechenden Statistik Beobachtungsdaten als empirische Realisation theoretischer, «wahrer» Parameter (oder Naturgesetze) verstanden: Der Fall einer Kugel ist zum Beispiel eine Realisation des Fallgesetzes. Empirisch ergeben sich zwar immer kleine Abweichungen der Messwerte von den theoretisch erwarteten Werten. Diese Abweichungen werden jedoch Störeinflüssen zugeschrieben, die möglichst zu vermeiden sind.

Für die Behandlung von bakteriellen Infektionen ist dieses Modell noch hinreichend brauchbar, weil beim Befall einer/s Patientin/en mit zum Beispiel Pneumokokken diese konkreten Pneumokokken eben als Realisation aus der «wahren» Population dieser Bakterien angesehen werden können. Ein Antibiotikum gegen diesen Bakterientyp wird somit weitgehend unabhängig davon wirken, ob die Beziehung der/s Patientin/en zur Ärztin bzw. zum Arzt gut oder schlecht ist – jedenfalls wäre der Einfluss der Beziehung eine leicht zu eliminierende Störgrösse.

Schwieriger wird der Einsatz dieser Forschungslogik allerdings selbst bei bakteriellen Infekten schon dann, wenn

das Medikament nicht die Bakterien selbst bekämpfen, sondern die Abwehrkräfte steigern soll: Von was wären gemessene «Abwehr-Daten» (z. B. Immun-Parameter) denn als Realisationen zu verstehen? Und was sind hier Störgrößen?

Noch schwieriger wird es, wenn es um psychische Störungen geht: Ist das, was wir bei als «depressiv» diagnostizierten PatientInnen Störung nennen, die konkrete Realisation einer theoretischen Grösse (Krankheit) namens «Depression»? Und – noch wichtiger – kann die Behandlungsweise «b» gegen diese Störung als eine konkrete Realisation der Behandlungsmethode «B» verstanden werden (die dann, mit anderen Methoden für andere Störungen zusammen Verfahren «X» ergibt)?

Doch trotz solcher Zweifel, führt eine (klassisch) experimentelle Perspektive in der obigen Form (UV à M à AV) geradezu *notwendig* in die Logik der RCTs.



Abb. 2

Denn wenn Abbildung 2 Sinn machen soll, so gilt *notwendig*:

- X *muss* zwangsläufig weitgehend manualisiert werden, denn man will ja nicht irgendeine UV testen, sondern X.
- Es *muss* eine Stunden-Anzahl mit X vorgegeben sein, sonst wäre nicht X die Wirkung, sondern die Behandlungsdauer.
- Es *müssen* Kontroll- bzw. Vergleichsgruppen gebildet werden, denn in der massenstatistischen Analyse (für E) können sonst viele Einflüsse (z. B. Spontanremission) und nicht X die Wirkung hervorrufen.
- Ebenso *muss* randomisiert (bzw. bei kleinem N besser: parallelisiert) werden, denn sonst wirken statt X ggf. die Unterschiede in den Gruppen.
- Um die Wirkung von X im Sinne von Effektstärken (E) sinnvoll erfassen zu können, ist eine weitere notwendige Voraussetzung, dass die Symptom- und auch die Heilungsdynamik hinreichend linear – zumindest ohne grosse Instabilitäten und Nichtlinearitäten – verläuft: Ansonsten würde bei der formal vorgenommenen Aggregation der Daten im Rahmen der statistischen Auswertung über unterschiedliche Verläufe aggregiert, wo ggf. dieselbe Dosis von X einmal fast nichts, ein anderes Mal extrem viel verändert. Mittelwerte würden dann weitgehend Artefakte der zufälligen Mischung unterschiedlicher Phasen abbilden und weniger über X (im Vergleich z. B. zur Kontrollgruppe oder zu einem anderen Ansatz Xa) aussagen.
- Eine letzte der hier keineswegs erschöpfend (vgl. Kriz, 1996, 2006) aufzuführenden impliziten Voraussetzungen ist, dass ausser X keine weiteren bedeutsamen Wirkvariablen im Experiment wirken. Wie ist es aber beispielsweise mit der therapeutischen Beziehung? Inzwischen wird deren Bedeutsamkeit auch für die VT nicht bestritten. Doch wo findet sich in der VT-

Theorie eine genaue Herleitung dessen, was mit «therapeutische Beziehung» in diesem Ansatz gemeint ist? Und da wohl kein/e VerhaltenstherapeutIn die Sicht für akzeptabel hielt, dass VT lediglich ein *Add-on* zur Wirkvariable «therapeutische Beziehung» ist, wie sie beispielsweise in der Gesprächspsychotherapie fundiert theoretisch hergeleitet und empirisch erforscht ist: Wo sind die Dismantling-Studien, die den Einfluss dieser Wirkvariablen von der «eigentlichen» VT bestimmen können?

Um Missverständnisse zu vermeiden sei an dieser Stelle explizit betont, dass die Diskussion mit Verweisen auf die VT nicht deshalb erfolgt, weil diese als besonders kritikwürdig angesehen wird, sondern weil die VT von allen Ansätzen noch am ehesten mit dem klassisch-experimentellen Paradigma übereinstimmt und sich in der Betonung von RCT-Designs auch primär auf dieses beruft. Daher ist es wichtig zu zeigen, dass auch die VT jenseits idealisierter Grundlagenforschung mit diesem Paradigma in Schwierigkeiten gerät, wenn es um Aussagen über die Wirkung realer Psychotherapie geht. Diese Vorbehalte gelten daher mindestens ebenso bei anderen Richtungen der Psychotherapie – allerdings haben diese auch gar nicht den Anspruch erhoben, in ihren wesentlichen Aspekten durch das klassisch-experimentelle Paradigma abgebildet zu werden.

Es gibt somit viele implizite aber *notwendige* Voraussetzungen für die Angemessenheit der klassisch-experimentellen Perspektive auf Fragen der Psychotherapie. Dass diese auch nur hinreichend erfüllt sind, darf nach der obigen Diskussion besonders dann bezweifelt werden, wenn man nicht über die Bedingungen im Labor, sondern über ein Verfahren in der realen Anwendung urteilen will. Für die Psychotherapie ist daher zwar Grundlagen- und Laborforschung wichtig – aber eben nicht nur: Denn nur wenige PsychotherapeutInnen werden ForscherInnen im engeren Sinne. Die meisten gehen in die reale Praxis. Daher gilt es, sich in einem Behandlungs- und Anforderungsalltag zu orientieren, in dem die PatientInnen eben nicht als Realisation von reinen Störungen betrachtet werden können (und das auch nicht modellhaft angemessen), sondern wo die Komplexität ihrer in Störungen manifestierten Lebens- und Leidensgeschichten mit dem Begriff «Komorbidität» belegt wird.

Dabei ist selbst die Bezeichnung «Komorbidität» irreführend. Denn PatientInnen «haben» nicht zwei oder mehrere reine Störungen – so wie Fusspilz und Grippe – sondern ihre komplexen Leiden sind diagnostisch nur in mehreren Kategorien erfassbar. Das ist ein wesentlicher Unterschied. Deshalb ist die für klassisch-experimentelle Forschungsdesigns so notwendige Manualisierung selbst für die VT in der Alltagspraxis auch recht fragwürdig, weil die weiteren Verläufe und Dynamiken der Behandlung nur sehr bedingt Standardverläufen entsprechen – ja, oft nicht einmal über längere Verläufe geplant werden, sondern ständig individuell neu adjustiert werden müssen. Nicht umsonst betont auch der Verhaltenstherapeut Zarbock (2010, S. 126): «Verhaltenstherapie stellt eben nicht die

Behandlung von Diagnosen mit medikationsähnlichen Standardinterventionen dar.» Die – ohne Zweifel wichtigen und interessanten – Ergebnisse aus der Psychotherapieforschung im klassisch-experimentellen Paradigma haben somit nur sehr artifizielle Aussagekraft für die reale Psychotherapiepraxis.

Konsens besteht ohnehin unter WissenschaftstheoretikerInnen dahingehend, dass Psychotherapie-Praxis nicht als Anwendung von Psychotherapie-Theorie verstanden werden kann. Selbst für die klassische VT – bei der der Gedanke der Anwendung von Theorie vielleicht noch am ehesten naheliegen könnte – hat schon Westmeyer (1980) gezeigt, dass sie (bzw. Teile davon) nicht als «angewandte Lerntheorien» verstanden werden kann. Stattdessen plädiert er, wie auch Perrez (1982) dafür, Theorie im Rahmen von Psychotherapie bestenfalls als Bereitstellung von «technologischen Regeln» zu verstehen, nämlich als Aufforderungen, in bestimmten Situationen bestimmte Massnahmen zu ergreifen, um bestimmte Ziele zu erreichen.

Das Bias des RCT-Ansatzes

Gleichwohl lässt sich die Frage auch andersherum stellen: Unter welchen Bedingungen passt das Modell der experimentellen Forschung und das RCT-Design zumindest hinreichend brauchbar zur Realität? Geht man dieser Frageperspektive nach, so ist jenseits des Labors ein solches Modell für die Praxis dann hinreichend brauchbar, wenn diese Praxis eben entsprechend den Bedingungen gestaltet werden kann bzw. typischerweise so gestaltet wird. Dies nun wiederum ist besonders bei der Entwicklung und Evaluation von psychotherapeutischen *Programmen* der Fall, in denen die zuvor angeführten Aspekte eben auch als standardisierte Vorgehensweisen in der Praxis umgesetzt sind. Sind zum Beispiel in einer Klinik für einen sechswöchigen Aufenthalt zehn Sitzungen für eine bestimmte Störung vorgesehen, in denen jeweils ein sehr klar vorgegebenes Programm durchgeführt wird, so sind die zuvor angeführten Bedingungen brauchbar erfüllt. *Um die Wirksamkeit der Anwendung von Programmen zu prüfen, sind RCT-Studien gut brauchbar.*

Wie brauchbar ist dieser Ansatz aber zur Beurteilung der Wirksamkeit von Psychotherapie, bei der die Behandlung typischerweise nicht nach einem vorgegebenen Programm erfolgt; die also auf einer dynamischen Passung von eher allgemeinen Prinzipien an die gegenwärtige Problem-Situation beruht? Wobei in die «Problem-Situation» recht viele Aspekte einfließen können: von der genetischen Disposition über biografische Strukturierungen, Lerngeschichten, aktuelle Belastungen, Symptomstabilisierungen im sozialen Umfeld etc. etc. Wird also entsprechend der spezifischen aktuellen Situation das konkrete therapeutische Vorgehen aus bestimmten Prinzipien erst entfaltet und in passendes Handeln umgesetzt, so sind die zentralen Bedingungen typischerweise nicht anwendbar.

Fasst man diese Argumente zusammen, so lässt sich sagen: *Um die Wirksamkeit von Therapieformen zu prü-*

fen, die in spezifischer Passung Kriterien entfalten, sind RCT-Studien unangemessen.

Natürlich kann man auch diese Therapieformen auf programmartige Vorgehensweisen, unter den zuvor angeführten Standardbedingungen reduzieren und damit die Forderung nach RCTs für diese «Varianten» erfüllen. Gezwungenermassen wurden solche Studien auch für Vorgehensweisen durchgeführt, die nicht zu dieser Forschungslogik passen. Denn auch in den USA ist eine Gruppe sehr einflussreich, die unter dem Begriff «Empirically Supported Treatment» (EST; empirisch gestützte Behandlungen) nur RCTs als einzige Form der empirischen Forschung ansieht – obwohl die Lehrbücher über empirische Forschungsmethoden auch in den USA zahlreiche andere Ansätze enthalten. Es sollte aber klar sein, dass ein solcher Forschungsansatz eben unangemessen ist und das Wesentliche verfehlt. Es ist ein wenig so, als wenn die Autolobby lediglich den Nachweis gelten lässt, wie schnell man auf Autobahnen vorankommt – und dabei das Essenzielle ignoriert, dass dies für die Beurteilung von PKW und LKW sinnvoll sein mag, aber nicht für Pferde – auch wenn man diese zwingen kann, sich am Wettrennen auf Autobahnen zu beteiligen. Doch dies ist nicht nur unangemessen, sondern entwertet auch die Leistung der Pferde, zum Beispiel über Hecken, Zäune und in unwegsamem Gelände voranzukommen und somit auch solche Menschen zu erreichen, die nicht an den Hauptstrassen angesiedelt sind. (ebenso erreicht man mit der *Mainstream-Psychotherapie* nicht alle Menschen).

Als ein sehr wesentliches Zwischenergebnis muss somit festgehalten werden, dass die Forderung, Wirksamkeit vor allem mit RCT-Designs zu belegen, Ansätzen, die vor allem auf einer situationsspezifischen, am realen psychotherapeutischen Prozess orientierten Entfaltung von Prinzipien ausgerichtet sind, damit eine bestimmte psychotherapeutische Präferenz, nämlich für die VT, überstülpt. Im Kleide von Forschungsmethodik wird hier also ein behaviorales Weltbild durchgesetzt.

Ein weiterer Aspekt des Bias des RCT-Ansatzes liegt in der Nicht-Berücksichtigung der Passungsproblematik nach dem allgemeinen Modell von Psychotherapie (AMP) nach Orlinsky & Howard (1987): Eigentlich gehört es heute zum Minimalstandard einer angemessen komplexen Diskussion, zumindest auf diese Passung einzugehen und diese in den Designs zu beachten. Es geht dabei um

- die Passung Behandlungsmodell und Störungsmodell,
- die Passung PatientIn und Behandlungsmodell,
- die Passung TherapeutIn und PatientIn und
- die Passung TherapeutIn und Störung der/s Patientin/en.

Die faktische, an RCTs und EbM orientierte Forschung ist aber ausschliesslich auf den ersten Aspekt ausgerichtet (Relation zwischen Behandlungsmodell und Störungsmodell). Und selbst dieser Aspekt ist auf die Interventionstechniken reduziert, die auf PatientInnensymptome wirken. Wiederrum ist die Fokussierung auf störungsspezifische Interventionstechniken als Untersuchungsmethodik für Programme noch vertretbar, für andere Konzepte von Psychotherapie hingegen ist diese Sicht unzulässig beschränkt.

«Effektstärke» – ein irreführender Begriff

In den Diskursen zur Bewertung von Psychotherapie spielt – spätestens bei der Zusammenführung der Studien zu Metaanalysen – eine Masszahl eine zentrale Rolle, die «Effektstärke» (ES) genannt wird. Diese ES ist üblicherweise wie folgt definiert und operationalisiert (vgl. Reinecker, 1996): $ES = (Me - Mk) / Sk$, wobei:

- Me = Mittelwert von T der *Experimental-*(= *Behandlungs-*)*gruppe* in der Post-Messung
- Mk = Mittelwert von T der *Kontrollgruppe* in der Post-Messung
- Sk = Standardabweichung von T der *Kontrollgruppe* in der Post-Messung
- T = Test-/Messvariable (z. B. ein klinischer Test)

Dieser ES wird deshalb eine so bedeutsame Stellung eingeräumt, weil die Ergebnisse unterschiedlicher RCTs auf sehr unterschiedlichen Vergleichsgruppen, Ausgangsbedingungen, Zielgrössen etc. beruhen. Der Begriff «Effektstärke» suggeriert, dass hier auch tatsächlich die Stärke des Behandlungseffektes gemessen bzw. vergleichbar gemacht wird – und so wird er in den Diskursen auch meist verwendet.

In Wirklichkeit kann ES aber nur recht bedingt etwas über Effektivität aussagen. Der Grund liegt darin, dass aus der obigen Formel hervorgeht, dass ES dann gross ist, 1. wenn $Me - Mk$ gross ist – das ist weitgehend wohl auch das, was man sich unter «Effektivität» vorstellt; 2. wenn Sk klein ist – das aber hat fast nichts mit Effektivität zu tun, sondern hängt von Besonderheiten der Gruppe und vor allem auch der gewählten Tests bzw. Operationalisierungen ab.

Insgesamt «misst» ES somit eher *statistische Diskriminierung* zwischen zwei Gruppen und nicht das, was man sich inhaltlich *klinisch* unter «Effektivität» vorstellt. Es kann die therapeutisch relevanten Beziehungen im Vergleich ggf. sogar ins Gegenteil verkehren. In Kriz (2014) findet sich ein durchgerechnetes Demonstrationsbeispiel, bei dem eine Behandlung (A) rechnerisch eine sehr hohe ES (2.0) hat, obwohl diese klinisch kaum effektiv ist, weil 97,5 % der Behandelten nach immer noch deutlich gestörter als die Normalbevölkerung sind. Bei einer alternativen Behandlung (B) ist die berechnete ES sehr viel geringer (1.3), obwohl sie viel effektiver ist, weil diese Gruppe nach der Behandlung völlig im Bereich der Normalbevölkerung liegt.

Die Art und Weise, in der in der Literatur über ES gesprochen wird, lassen zweifeln, dass diese formalen Zusammenhänge bzw. Unterschiede hinreichend bei KlinikerInnen bekannt sind. Vielmehr wird meist so getan, als messe die ES tatsächlich und unbesehen einen therapeutischen Effekt. Es zeigt sich darüber hinaus, dass es falsch wäre, Effekte in heterogenen, störungsunspezifischen Gruppen (d. h. mit grossem Sk) als weniger bedeutsam oder aussagekräftig abzutun: Zumindes in formal-quantitativer Hinsicht ist eher das Gegenteil richtig (da störungsunspezifische Gruppen typischerweise eher eine grössere als eine kleinere Varianz gegenüber homogenen Gruppen aufweisen).

Evidenzbasierter Unsinn

Computer, mathematische Modelle und Rechenvorschriften sagen per se nichts über den Sinn oder Unsinn ihrer Verwendung aus. Daher lassen sich Daten aus berechneten Effektstärken auch dann in eines der üblichen Computerprogramme für Metaanalysen füttern, wenn die inhaltliche Fragestellung als unsinnig zu erkennen wäre – sofern man nicht einer Mystifikation von Zahlenwerten aus Computerprogrammen unterliegt sondern den eigenen fachlichen Sachverstand bemüht. Ein Beispiel für einen fragwürdigen Missbrauch von Effektstärken zum Zwecke von Metaanalysen und scheinbar evidenzbasierten Aussagen, stellt das von Norcross herausgegebene Werk mit dem Titel «Evidence-Based Therapy Relationships» (2010) dar.¹ Dieses enthält 13 Beiträge, die jeweils Metaanalysen zu «Elementen» bzw. «Faktoren» der therapeutischen Beziehung durchführen. Solche Elemente sind unter anderem therapeutische Allianz, Zielübereinstimmung, Management von Gegenübertragung, positive Wertschätzung, Kongruenz oder auch Empathie. Jedem Element ist ein Kapitel mit einer Metaanalyse gewidmet; wohlgemerkt: jedem einzelnen. Dies führt dann unter anderem zu Aussagen, dass *Empathie* (57 Studien, 3.599 Klienten) «nachweislich», positive Wertschätzung (18 Studien) aber nur «wahrscheinlich» effektiv wirkt. Und für *Kongruenz* (16 Studien, 863 Klienten) gebe es «noch keine hinreichende Evidenz» für Wirksamkeit. Das sind beachtliche Unterschiede in der Bewertung – die aber mehr für statistische Akribie denn für substanzielle Erkenntnis sprechen: Denn die mittlere Korrelation zwischen «Element» und «Behandlungserfolg» beträgt als Effektstärke für *Empathie* .31, für *positive Wertschätzung* .27 und für *Kongruenz* .24 – das sind, freundlich formuliert, inhaltlich recht irrelevante Differenzen.

Zu kritisieren ist aber nicht nur, wie artefaktanfällig solche Unterschiede sind. Wichtiger ist die Frage, ob zum Beispiel in allen 57 Studien mit Daten zur «Empathie» überhaupt hinreichend das Gleiche darunter verstanden wurde. Oder wie weit dies für «Behandlungserfolg» gilt. Und hatte nicht Rogers, auf den diese Konzepte zurückgehen, einst betont, dass es sich um drei untrennbare, wechselwirkende Aspekte *einer* Haltung handelt? Das heisst, dass Therapien mit *Empathie* ohne *Wertschätzung* oder *Kongruenz* ohne *Empathie* ganz andere Erfolgsraten haben dürften. Genauso könnte man aus den Verkaufsdaten von 100 Rembrandt-Bildern analysieren, mit welchem Ausmass die Farbe Rot und mit welchem die Farbe Braun zum erzielten Verkaufspreis beigetragen haben.

Den Unsinn, in komplexen Prozessräumen mit nicht-linearen Wechselwirkungen, wie dies für Psychotherapie typisch ist, nach isolierbaren Wirkfaktoren zu suchen, hat Plaum (1999, S. 201) in einer kurzen – fiktiven aber nicht untypischen – Fallgeschichte dargestellt:

¹ Zu finden unter <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.463.7645&rep=rep1&type=pdf> (23.06.2019) sowie, etwas erweitert, als Buch (Norcross, 2011) und, mit Wampold zusammen, auch Heft 1/2011 von *Psychotherapy*.

«Aufgrund der Distanz von zuhause war Frau X. in der Lage, ein anregendes Klima, eine reizvolle Landschaft, positiv auf sich wirken zu lassen, was durch die entspannende Wirkung der Bäder verstärkt worden ist; dass es Mitpatienten gibt, denen es offenbar schlechter ging als ihr, hat Frau X. neuen Lebensmut gegeben und sie offener für andere Menschen gemacht; die daraufhin möglichen Gespräche konnten ihr wiederum veränderte Lebensperspektiven nahebringen; schließlich hat der Masseur bei Frau X. erotische Gefühle ausgelöst, die wiederum in der geschilderten aktuellen sozialen Situation zur Realisierung eines ›Kurschattens‹ führten; dieser hat aufgrund seiner anziehenden Wirkung Frau X. zu der Überzeugung bringen können, dass es doch besser sei, die Diätvorschriften einzuhalten und nicht etwa sich am Nachmittag ins Café zu setzen, um dort die Sahnetorte zu verspeisen.»

Zu Recht fragt Plaum: «Was hat zur Verbesserung des gesundheitlichen Zustandes von Frau X. während einer Kur geführt? Bäder? Massagen? Das Klima? Die Distanz von zuhause? Der Kontakt zu Mitpatienten? Der ›Kurschatten‹?» (ebd., S. 200). Im gewählten Beispiel muss man sagen: wohl alles irgendwie. Doch nichts wirkt als isolierbarer Faktor als solcher – und auch nicht im Sinne varianzanalytischer Wechselwirkung zusammen. Denn man kann sich gut vorstellen, dass in Gruppen mit Personen unter fast genau denselben Bedingungen die Gesamtwirkung völlig unterschiedlich ausfällt. Ein Forschungsansatz, bei dem der Einfluss des Kurschattens oder des Masseurs etc. als unabhängige Variable in einem RCT-Design untersucht werden soll, wäre daher entsprechend lächerlich. Und bei anderen PatientInnen wirken vielleicht in derselben Kurklinik andere Aspekte zusammen. Man kann sogar davon ausgehen, dass die PatientInnen selbst auch etwas Kompetenz besitzen, aus den reichhaltigen Programmen und komplexen Möglichkeiten einer solchen Klinik das auszuwählen, was für sie gut ist, und dass es vielleicht ausreicht, sie bei dieser Wahl und bei der Reflexion darüber zu begleiten und zu beraten.

Freilich ist die exzessive Zerlegung therapeutischer Prozesse in metaanalytisch isolierbare Elemente keineswegs neu: In zahlreichen Lehrbüchern (übrigens auch bei Norcross) geistert ein Tortendiagramm herum, das auf Asay und Lambert (2001) zurückgeht, wonach 30 % der Verbesserungen in Therapien auf die therapeutische Beziehung zurückgehen sollen (40 % auf extratherapeutische Variablen, 15 % auf Erwartungseffekte und nur 15 % auf die gewählte Technik). Diese Zahlen sind für viele, die die Bedeutsamkeit der therapeutischen Beziehung nicht vernachlässigt wissen möchten, natürlich erfreulich. Allerdings zählt die «Erfreulichkeit» eines Ergebnisses nicht zu den Kriterien in der Wissenschaft. Diese Zahlen sind schlicht numerischer Unsinn und die Aussagen nichts weiter als pseudo-quantitative Datenbelletristik. Was würden wir zum Beispiel mit einer Aussage anfangen: «Im Mittel sind die Gewässer in Deutschland 162,7 cm tief»? Zu Recht würden wir doch fragen: Welche Gewässer?

Sind es die grossen Flüsse wie Rhein, Elbe etc., oder auch die kleinen wie Hase, Düte usw.? Und wie steht es mit Gartentümpeln oder Baggerseen? Und falls nur Flüsse gemeint sein sollen: Wo wird denn gemessen: näher an der Quelle oder an der Mündung?

Genauso klar ist, dass die Wirkanteile im Therapieprozess nicht konstant und unabhängig sind und zum Beispiel von der Art der Störung oder dem jeweiligen Ansatz abhängen: Sie sind (hoffentlich) im humanistischen Ansatz eher höher als bei verhaltenstherapeutischen Vorgehensweisen, wo andersherum die dort zentralen Techniken vermutlich mehr als 15 % des Erfolgs ausmachen. Jede quantitative Aussage hängt somit von der Mischung an unterschiedlichen therapeutischen Vorgehensweisen ab.

Die Aussage von Asay und Lambert ist daher nicht falsch – aber sie bezieht sich auf das Ergebnis einer Reanalyse mit *einer* ganz bestimmten Mischung an unterschiedlichen therapeutischen Vorgehensweisen. Aber für *eine* ganz bestimmte Mischung an Gewässern wäre eben die obige Aussage von 167,2 cm für die mittlere Tiefe deutscher Gewässer ebenfalls nicht falsch. Allerdings würden wir für einen anderen Mix auch andere Anteile der Faktoren erwarten bzw. eine andere mittlere Gewässertiefe. Und genau wie bei den Gewässern gibt es auch für Psychotherapie per se keinen «richtigen» oder zumindest «angemessenen» Mix hinsichtlich globaler Gesamtaussagen. Denn woran sollte man auch die Richtigkeit eines solchen Mixes messen? Die Quantifizierung ist daher nicht falsch oder richtig, sie ist schlicht sinnlos.

Dieser alte – aber übermässig oft zitierte – Unsinn von Asay und Lambert ist also in Norcross' Sammlung nur nochmals potenziert. Dass bei dieser missbräuchlichen Anwendung massanalytischer Vorgehensweisen renommierte Forscher wie Wampold, Elliott, Greenberg oder Lambert mitgewirkt haben, macht die Sache keineswegs besser, sondern eher noch bedenklicher. Offenbar ist der Druck, in den Diskursen über evidenzbasierte Ergebnisse nicht unterzugehen, derart gross, dass selbst solche sonst seriösen Forscher artefaktanfällig werden.

In der Untersuchung der (fiktiven) Kurklinik von Plaum wäre es daher weit angemessener, statt nach dem evidenzbasierten Wirkanteil eines Kurschattens zu suchen, sich der zahlreichen Möglichkeiten zu bedienen, den Erfolg einer Kurklinik bei den PatientInnen zu evaluieren um sicherzustellen, dass mit Steuer- und Krankenkassengeldern nicht einfach Unwirksames bezahlt wird. Es ist ja ohnedies mehr als merkwürdig, dass in deutschen Kliniken sehr wohl vieles von dem eingesetzt und im Kassensystem abgerechnet wird, was angeblich «nicht wirksam» sein soll – obwohl (oder gerade weil) Kliniken mit sehr spitzem Bleistift rechnen müssen und sich keine überflüssigen Ausgaben leisten können. Aber sie wissen, wie gute HausärztInnen, um einen eher ganzheitlichen Heilerfolg, der nicht beliebig in Einzelfaktoren zerlegt und mit RCT-Designs untersucht werden kann – so wie eine Körperzelle in einzelne materiellen Bestandteile zerlegt werden kann, die aber jeweils kein «Leben» ausmachen.

Resümee

Resümierend lässt sich sagen, dass EbM im Ansatz und in der Intention zwar prinzipiell positiv zu sehen ist, in der Praxis aber entstellt und zweckentfremdet eingesetzt wird. Aus dem guten Anliegen, für PraktikerInnen einen informativen Überblick über die von ihnen schwer zu bewältigende Flut an Forschungsergebnissen bereitzustellen, wurde missbräuchlich ein Instrument zur Selektion und Massregelung entwickelt. Der Begriff «Evidenzbasierung» ist so (wortwörtlich) zu einem Schlagwort verkommen, um inhaltliche Argumente und Diskurse zu unterbinden: Gut zwei Jahrzehnte nach Einführung von EbM durch die Gruppe um David Sackett, werden die eindringlichen Warnungen des Gründers (z. B. Sackett et al., 1996) ignoriert und oft eine entstellte Form, die lediglich RCTs berücksichtigt, irreführend als «EbM» ausgegeben.

Diese starke Betonung von RCTs macht dort Sinn, wo experimentelle Designs dem Untersuchungsgegenstand angemessen sind – dies gilt neben der Apparate- und Pharmaforschung in der Medizin vor allem dann, wenn klar definierbare und operationalisierbare Programme angewendet werden können. Für Teile der VT und recht bedingt auch für andere Psychotherapieansätze ist das noch hinreichend gewährleistet. Für viele andere psychotherapeutische Ansätze hingegen, die wesentlich auf der situationsspezifischen und passungsgerechten Entfaltung von Prinzipien beruhen, ist diese Methodik weitgehend inadäquat. Die Reduktion von EbM auf RCT-Designs schliesst somit nicht nur bedeutsame Teile wissenschaftlicher Methodik von vornherein aus (wie z. B. das gesamte Spektrum qualitativer Forschung), sondern enthält zusätzlich ein starkes Bias für nur eine der vier zentralen psychotherapeutischen Grundorientierungen – nämlich für die verhaltenstherapeutische auf Kosten der psychodynamischen, humanistischen und systemischen Grundorientierung.

Darüber hinaus werden durch die überstarke Fokussierung auf RCTs wesentliche Erkenntnisse internationaler Psychotherapieforschung weitgehend ignoriert – so beispielsweise die Bedeutsamkeit der Passung für die therapeutische Beziehung und die Erforschung der Wirksamkeit. Ignoriert werden in RCT-Designs weiterhin die essenziellen Auswirkungen nicht-linearer Entwicklungsverläufe in Variablennetzwerken, die aber für psychotherapeutische Veränderungen typisch sind.

Zu diesen Mängeln und Einseitigkeiten kommt neben der missbräuchlichen Reduzierung der «Evidenzbasierung» auf RCTs eine Fehlinterpretation von Konzepten wie «Effektstärke» – ein Begriff, der vorgaukelt, ein Mass des klinischen Effektes von Interventionen zu sein, in Wirklichkeit aber nur die statistische Diskriminierung zweier Populationen misst. Durch eine leider nicht untypische Mystifikation von numerischen Ergebnissen, die Computerprogramme geliefert haben, bleibt auch hier bisweilen der inhaltliche Sachverstand von KlinikerInnen und TherapeutInnen auf der Strecke.

Insgesamt gesehen ist das Ausmass an methodischem Unverständnis und faktischem Missbrauch der guten

Idee von «Evidenzbasierung» gerade auch von einflussreichen Gremien derart gravierend, dass EbM daher für die Bewertung von Psychotherapie sehr kritisch gesehen werden muss. Sie dient inzwischen wohl auch eher einer interessengeleiteten Selektion bestimmter psychotherapeutischer Ansätze als der Information über den Stand wissenschaftlicher Forschung oder gar dem Wohl von PatientInnen. Bedauerlich ist, dass viele Forschungsfragen, die für psychotherapeutisches Handeln wichtig sind – und die daher PraktikerInnen weit mehr interessieren als die «Rechtfertigungsforschung» gegenüber der Konkurrenz aus Pharmaindustrie und Medizin –, aufgrund dieser Fokussierung auf RCT-Designs stark vernachlässigt werden. Es ist sowohl aus wissenschaftlicher Sicht wie auch im Hinblick auf die Interessen der TherapeutInnen und ihrer PatientInnen überaus bedenklich, welche ungeheuren Ressourcen in den letzten Jahrzehnten für den Nachweis aufgewendet wurden, dass ein bestimmter Ansatz mindestens so wirksam ist wie die Konkurrenz (d. h. andere Ansätze und/oder Psychopharmaka). Und wie wenig vergleichsweise den vielfältigen Fragen zur Wirkungsweise von zahlreichen Aspekten im psychotherapeutischen Prozess nachgegangen wurde.

Literatur

- Asay, T.P. & Lambert, M.J. (2001). Empirische Argumente für die allen Therapien gemeinsamen Faktoren: quantitative Ergebnisse. In M. Hubble, B.L. Duncan & D. Miller (Hrsg.), *So wirkt Psychotherapie* (S. 41–81). Dortmund: Verlag modernes lernen.
- Kienle, G.S. et al. (2003). Evidenzbasierte Medizin: Konkurs der ärztlichen Urteilskraft? *Deutsches Ärzteblatt*, 100, C1688–C1692.
- Kriz, J. (1996). Grundfragen der Forschungs- und Wissenschaftsmethodik. In R. Hutterer-Krisch et al. (Hrsg.), *Psychotherapie als Wissenschaft – Fragen der Ethik* (S. 15–160). Wien: Facultas.
- Kriz, J. (2000). Perspektiven zur Wissenschaftlichkeit von Psychotherapie. In M. Hermer (Hrsg.), *Psychotherapeutische Perspektiven am Beginn des 21. Jahrhunderts* (S. 43–66). Tübingen: DGVT-Verlag.
- Kriz, J. (2006). Wie evident ist klinische Evidenz? In A. Bührmann et al. (Hrsg.), *Gesellschaftstheorie und die Heterogenität empirischer Sozialforschung* (S. 32–43). Münster: WD-Verlag.
- Kriz, J. (2007). Wie lässt sich die Wirksamkeit von Verfahren X wissenschaftlich begründen? Versuch der Präzisierung einer methodologischen Hintergrunds-Kontroverse. *Psychotherapeutenjournal*, 6, 258–261.
- Kriz, J. (2008). Vermessene Wissenschaftlichkeit. Kritische Aspekte und bedenkliche Tendenzen des Methodenpapiers. *Psychotherapeutenjournal*, 7, 117–119.
- Kriz, J. (2014a). Aspekte «wissenschaftlicher» Faktengläubigkeit – und ihr Nutzen. *geographische revue*, 16(1), 7–19.
- Kriz, J. (2014b). Wie evident ist Evidenzbasierung? Über ein gutes Konzept – und seine missbräuchliche Verwendung. In S. Sulz (Hrsg.), *Psychotherapie ist mehr als Wissenschaft. Ist hervorragendes Expertentum durch die Reform gefährdet?* (S. 154–185). München: CIP-Medien.
- Kriz, J. (2015). Psychotherapieforschung – und ihre Beschränkung durch einen schulenspezifischen Bias. *Resonanzen. E-Journal für biopsychosoziale Dialoge in Psychotherapie, Supervision und Beratung*, 3(2), 101–113.
- Kriz, J. (2017). *Subjekt und Lebenswelt. Personzentrierte Systemtheorie für Psychotherapie, Beratung und Coaching*. Göttingen: Vandenhoeck & Ruprecht.
- Meichenbaum, D.W. (1979). *Kognitive Verhaltensmodifikation*. München: Urban & Schwarzenberg.
- Norcross, J.C. (2011). *Psychotherapy Relationships that Work: Evidence-based Responsiveness*. New York: OUP.

- Norcross, J.C. & Lambert, M.J. (2011). Evidence-based therapy relationships. In J.C. Norcross (Hrsg.), *Psychotherapy Relationships that Work: Evidence-based Responsiveness* (S. 279–300). New York: OUP.
- Orlinsky, D.E. & Howard, K. (1987). A generic model of psychotherapy. *J. Integrative Eclectic Psychotherapy*, 6, 6–27 [dt.: (1988). Ein allgemeines Psychotherapiemodell. *Integrative Therapie*, 4, 281–308].
- Perrez, M. (1982). Was nützt Psychotherapie? *Psychol. Rundschau*, 33, 121–126.
- Plaum, E. (1999). Weshalb fährt der IC 781 am 26. Geburtstag von Sabine M. um 13.49 Uhr mit einer Geschwindigkeit von 82,5 km/h durch den Bahnhof Eichstätt? Oder: Das Elend mit der Suche nach reinen Wirkfaktoren in einer hochkomplexen Realität. *Gestalt Theory*, 21, 191–207.
- Reinecker, H. (1996). Therapieforschung. In J. Margraf (Hrsg.), *Lehrbuch der Verhaltenstherapie* (S. 31–47). Heidelberg: Springer.
- Rogers, C.R. (1957). The Necessary and Sufficient Conditions of Therapeutic Personality Change. *Journal of Consulting Psychology*, 21, 95–103.
- Sackett, D.L. et al. (1996). Evidence based Medicine: What It Is and What It Isn't. *British medical journal.*, 312, 71–72.
- Schulthess, P. (2019) Kriterien der Wissenschaftlichkeit. *à jour! Psychotherapie-Berufsentwicklung*, 5(1), 23–27.
- Strauß, B. (2019) Innovative Psychotherapieforschung – Wo stehen wir und wo wollen wir hin? *Psychotherapeutenjournal*, 1, 4–10.
- Westmeyer, H. (1980). Zur Paradigmadiskussion in der Psychologie. In W. Michaelis (Hrsg.), *Kongreßbericht DGP* (S. 115–126).
- Wolpe, J. & Lazarus, A. (1966). *Behavior therapy techniques*. London: Pergamon.
- Zarbock, G. (2010). Der Master in Psychologie als optimale Basisqualifikation für Psychotherapeuten. *Psychotherapeutenjournal*, 2, 125–129.

«Evidence-based» as a criterion for psychotherapy selection? Concerning a good concept – and its misuse

This paper first points out that «evidence-based» is a good and useful concept but is frequently misused today due to reduction to RCTs. While the RCT design is absolutely an outstanding model for studying experimental research questions, this only applies if it is possible to isolate clearly definable causes and effects to a sufficient degree and if these provide a usable representation of the reality of the investigation. The greater the scope for designing the causes (interventions) and the more complex the relevant effects are, the less this approach is useful. We will show that RCT designs are suitable for psychotherapy research in particular if manualized *programs are applied*. This is primarily the case for behavioral therapy. For psychotherapeutic approaches that are mainly founded in the process-specific *development of principles*, the great variability of the detail causes (interventions) bursts the required bulk-statistical homogeneity for an experimental design. The requirement of proving effectiveness mainly with RCT designs amounts to methodical artifacts in this case or imposes a certain psychotherapeutic preference, i.e. for behavioral therapy, onto other approaches by means of scientific methodology. Finally, we discuss a few additional

reductions and misinterpretations in connection with RCT and metastudy designs.

Key words: strength of effect, evidence-based, fit, psychotherapeutic procedures, RCT, effect factors

«Evidence-basing» come criterio della selezione della psicoterapia? Un concetto buono e il suo uso improprio

Prima di tutto, questo documento sottolinea che l'approccio «evidence-basing» è un concetto buono e utile, ma spesso viene utilizzato in modo improprio a causa di una riduzione a RCT. Il progetto RCT è certamente un modello eccellente per lo studio di questioni sperimentali di ricerca. Tuttavia, si applica solo se cause chiaramente definibili ed effetti altrettanto chiaramente definibili possono essere sufficientemente isolati e se rappresentano in tal modo la realtà dello studio in modo utile. Maggiore è lo spazio di manovra per la creazione delle cause (interventi) e più sono complessi i relativi effetti, meno adatto è questo approccio. Come verrà mostrato in seguito, i progetti RCT sono particolarmente utili nella ricerca in psicoterapia quando vengono *utilizzati programmi* manualizzati. Questo è principalmente il caso della terapia comportamentale. Per le procedure psicoterapeutiche, che si basano essenzialmente sullo *sviluppo di principi* specifico, la grande variabilità delle cause di dettaglio (interventi) supera l'omogeneità statistica di massa richiesta per un progetto sperimentale. Il requisito di dimostrare l'efficacia, in particolare con i progetti RCT, si traduce in artefatti metodici oppure una certa preferenza psicoterapeutica, vale a dire per la terapia comportamentale, sulla strada verso la metodologia scientifica, impone altri approcci psicoterapeutici. Infine, vengono problematizzate alcune ulteriori riduzioni e interpretazioni errate nel contesto degli RCT e dei metastudi.

Parole chiave: dimensione dell'effetto, evidence-basing, adattamento, procedura psicoterapeutica, RCT, fattori

Der Autor

Jürgen Kriz, Prof. Dr., ist Emeritus für Psychotherapie und Klinische Psychologie an der Universität Osnabrück, hatte aber über 25 Jahre auch Professuren in Statistik und Forschungsmethoden inne. Er ist Ehrenmitglied etlicher psychotherapeutischer Fachverbände, war Gastprofessor u. a. in Wien, Zürich, Berlin, Moskau, Riga und den USA, und ist Autor von über 20 Büchern und 300 Fachbeiträgen. Zu seinen Ehrungen zählt neben dem Viktor-Frankl-Preis (2004) aktuell der Schweizer Dr. Margrit Egnér-Preis (2019).

Kontakt

E-Mail: kriz@uos.de