

Stellungnahme

Zu Verfahrensregeln des Wissenschaftlichen Beirats Psychotherapie zur Beurteilung der wissenschaftlichen Anerkennung von Methoden und Verfahren der Psychotherapie (Version 2.1)

Zusammenfassung, Folgerungen und allgemeine Änderungsvorschläge zu den Verfahrensregeln des WBP

1. Verfahrensregeln schaffen Transparenz und werden begrüßt

Die mit dem Entwurf verfolgte Absicht, die Beurteilung der wissenschaftlichen Anerkennung von psychotherapeutischen Methoden und Verfahren durch definierte Vorgehensweisen und Beurteilungskriterien transparent und nachvollziehbar zu gestalten, ist sehr zu begrüßen.

Auch die – wenn auch unzureichende – Berücksichtigung von Studien mit Aspekten externer Validität sowie die Berücksichtigung von Studien mit Stichproben über mehrere Diagnosegruppen sind ein wichtiger Schritt in Richtung Berücksichtigung der Versorgungsrelevanz der untersuchten Wirksamkeiten.

2. Unterscheidung von bereits zugelassenen und neu zuzulassenden Verfahren sinnvoll

a. Bei der Beurteilung der wissenschaftlichen Anerkennung von Verfahren, die noch nicht im deutschen Gesundheitswesen zugelassen sind, auf experimentelle Nachweise der Wirksamkeit zu fokussieren, erscheint insofern zweckmäßig, als Studien

- an eher kleinen Populationen
- mit eher kurzen Behandlungsdauern und Katamnesedauern
- mit eher systematisiertem Behandlungsvorgehen

unter pragmatischen Gesichtspunkten leichter zu realisieren sind und hinreichend Anhaltspunkte für eine Bewährung im Versorgungsalltag liefern können. Die formulierten hohen Anforderungen an methodischer Qualität und interner Validität sind bei Studien an kleinen Populationen notwendig, um überhaupt verlässliche Aussagen generieren zu können.

Allerdings sind für noch nicht zugelassene Verfahren, die nicht primär auf isolierbaren und manualisierbaren Interventionen aufbauen sondern auf therapeutische Beeinflussung komplexer seelischer Regulationssysteme, auch Kompensationsstrategien der Operationalisierung zu definieren, die verfahrensspezifischen nicht manualisierten Behandlungsstrategien in Studien adäquat berücksichtigen. Bei der Beurtei-

lung der Wissenschaftlichkeit ist dabei Methodentreue gegen Praxistreue regelhaft abzuwägen.

b. Für bereits im Gesundheitssystem zugelassene Verfahren sind – soweit eine Überprüfung der wissenschaftlichen Anerkennung erforderlich ist - teilweise deutlich abweichende Überprüfungskriterien zu entwickeln, die in hohem Maß den Kriterien der Praxisbedingungen gerecht werden. Dabei sind naturalistische Untersuchungen und Versorgungsforschung an möglichst unselektierte Populationen mit Anwendungsformen der Methoden/Verfahren im Versorgungssetting stärker zu gewichten als experimentell hergestellte Methodenreinheit. Als Endpunkte sind z.B. Lebenszufriedenheit, Mortalität, Morbidität, Inanspruchnahme medizinischer und sozialer Maßnahmen etc. besonders zu gewichten.

Dabei sind auch Operationalisierungen der Beurteilung und Gestaltung von Untersuchungsdesigns zu entwerfen, die Aussagen zur Wirksamkeit von Langzeittherapien ermöglichen.

Fazit: Für bereits zugelassene Verfahren sind sehr komplexe Forschungsdesigns im Sinne der Versorgungsforschung zu fordern.

3. Verfahrensregeln können sich auch auf pragmatische Begründungen stützen

Es sollte unbedingt auf Begründungen zum Vorgehen (Verfahrensregeln) verzichtet werden, die

- eine wissenschaftlich nicht herstellbare „Objektivität“ bei der Wirksamkeits- und Unbedenklichkeitsbeurteilung suggerieren,
- andere Untersuchungsmethoden (wie z.B. Prozessforschung, Versorgungsforschung) gegenüber den Studien zur experimentellen Wirksamkeit als weniger wichtig oder weniger aussagekräftig darstellen und damit ihre Nichtverwendung legitimieren. Soweit es sich um pragmatische Gründe für das gewählte Vorgehen handelt, sollten diese lediglich als solche benannt werden.

Auf die Gefahr eines selbstreferenziellen Systems, bei dem nur das für wissenschaftlich erklärt wird, was der experimentellen Untersuchungsmethodik entspricht (der Untersuchungsgegenstand wird dem Untersuchungsinstrument angepasst), samt der Folge der Einschränkung wissenschaftlicher Erkenntnis und praktischer Anwendung wird hingewiesen.

4. Keine Beurteilungen von Verfahren für einzelne Störungen

Auf eine Beurteilung der wissenschaftlichen Anerkennung **primär** für einzelne Methoden für einzelne Störungen sollte verzichtet werden. Es gibt keine wissenschaftliche Begründung für diesen Primat. Im Gegenteil ermöglichen experimentelle Wirksamkeitsnachweise für umgrenzte Populationen mit umgrenzten Störungen (bisher) keine sicheren Belege der Übertragbarkeit auf die diagnostische Gesamtgruppe, erst recht nicht unter realen Praxisbedingungen. Die Legitimation kann sich lediglich auf eine leichtere Operationalisierbarkeit stützen.

5. Begrenzte Anwendung von Manualen in der Versorgung – Nicht nur RCT

Es sollte auf eine programmatische Definition der psychotherapeutischen Vorgehensweisen mittels Manual oder manual-ähnliche Behandlungsrichtlinien als Regelfall verzichtet werden. Im Versorgungsalltag spielen Manuale eine eher untergeordnete Rolle. Eine Einschränkung auf RCT-Studien sollte nicht vorgenommen werden.

6. Externe Validität zentraler Faktor bei der Beurteilung - Konzepte müssen noch entwickelt werden

Begrüßt werden die Ansätze der Berücksichtigung von externer Validität. Allerdings wird bei der vom WBP vorgeschlagenen Vorgehensweise nur ein Kompromiss auf Kosten der externen Validität angeboten, so dass mit den operationalisierten Kriterien die effectiveness nicht sicher genug beurteilt werden kann. Die Reziprozität von interner und externer Validität müsste problematisiert werden und daraus konsistentere Folgerungen und Umsetzungen bei der Operationalisierung der Kriterien abgeleitet werden.

7. Weitere Vorschläge zum konkreten Vorgehen

a. Wegen der erheblichen Probleme der Herstellung von brauchbaren Kontrollgruppen zumindest in Deutschland (freie Therapeutenwahl, Patientenanspruch auf bewährte Therapieformen, bias in Richtung geringerer Behandlungsdringlichkeit der Kontrollgruppenpopulation) sollte noch mehr, als mit Punkt 2.5. Nr.4. geschehen, die entsprechenden Forderungen (einer Kontrollbedingung) relativiert werden und dem Untersuchungsgegenstand angemessene Ersatzstrategien vorgeschlagen werden.

b. Für ältere Studien ist ein späterer Zeitpunkt als der 1.1.90 für die Gültigkeit der unter 2.7. gefassten Beurteilungskriterien zu wählen.

c. Überprüft werden soll, ob es überhaupt möglich ist, zugleich mit einer Studie sowohl die Kriterien für die interne als auch für die externe Validität zu erfüllen. Ansonsten sollte es bei drei Studien pro Anwendungsgebiet bleiben, wobei jeweils eine mindestens eine hinreichende interne bzw. externe Validität aufweisen sollte.

d. Es ist ein eigener Kriterien-Katalog zu operationalisieren für naturalistische Studien und Versorgungsstudien, sowohl was die Anforderungen an allgemeiner methodischer Qualität betrifft (mit Kompensationen für Nichterfüllung von nur experimentell herzustellenden Bedingungen) als auch die Anforderungen an den Nachweis von externer Validität.

Anhang 1

enthält Einzel-Kommentare und detaillierten Änderungsvorschläge zu den konkreten Textstellen des WBP-Entwurfs zu den Verfahrensregeln, Version 2.1. vom 02.01.07

Anhang 2

enthält eine grundlegende Untersuchung des Verhältnisses von experimentellen Studien zur Versorgungspraxis und zum Untersuchungsgegenstand Psychotherapie. Nur anhand der dort dargelegten Inhalte wird der Stellenwert der von uns vorgetragenen Änderungsvorschlägen und ihre wissenschaftliche Legitimation logisch stringent nachvollziehbar.

Zur Annäherung an die abzuwägenden wissenschaftlichen Fragestellungen ein Zitat von F. Caspar:

„Die Einführung des RCT-Standards für psychotherapeutische Wirksamkeitsforschung bietet verschiedene Vorteile: In der politischen Diskussion können Psychotherapien einfacher im Vergleich zu alternativen Behandlungsformen als ‚einheitliche Droge‘ dargestellt werden; dies erhöht ihre Konkurrenzfähigkeit. Nach wissenschaftlichen Gesichtspunkten erfüllen RCTs in höherem Maße Anforderungen an methodische Strenge und interne Validität. Und natürlich sind kausale Wirksamkeitsbelege nur durch experimentelle Studien zu erreichen.

Die Konzentration auf die Wirksamkeit von psychotherapeutischen Interventionen und die Dominanz des RCT-Standards sind jedoch auch in einer Reihe von Punkten kritisch zu bewerten:

Ein wesentlicher Kritikpunkt ist ein Mangel an externer Validität: so verlangen beispielsweise RCTs bei der Patientenselektion eine starke Ausrichtung an klaren Diagnosen, in der Regel ohne Komorbidität. Diese ist jedoch epidemiologisch und in der Alltagspraxis nicht repräsentativ. Gleichzeitig kann die starke Reduktion von Heterogenität durch Selektion von Patienten bewirken, dass durch Reduktion von Prä-Varianz Effektstärken unrealistisch vergrößert werden. Bei gleichem absolutem Betrag der therapeutischen Veränderung, aber natürlicher Variation der Erfolgsvariablen, wären die gefundenen Effekte ceteris paribus in der Regel bescheidener (...)

Eine Reihe weiterer Kritikpunkte am RCT-Ansatz aus der mittlerweile umfangreichen Literatur dazu können wie folgt zusammengefasst werden:

- Zu undifferenzierter, oft tendenziöser Vergleich von Verfahren in einem Teil der Studien (horse race- Forschung’),*
- unangemessene Reduktion der betroffenen Patienten auf meist eine Diagnose und Vernachlässigung von Komorbiditäten;*
- Nicht-Berücksichtigung der schwierigsten bzw. für die Alltagspraxis typischen Patienten;*
- Vernachlässigung kleinerer, in der Summe aber bedeutender Patientengruppen;*
- Vernachlässigung von (z.B. interpersonalen) Eigenschaften, die über Diagnosen im engeren Sinne hinausgehen;*

- Bericht von Ergebnissen nur für Patienten mit abgeschlossener Therapie statt für alle Patienten, die behandlungsbedürftig bzw. –willig („intention to treat“) wären;
- Unmöglichkeit, den ganzen Bereich psychischer Störungen mit spezifischen Manua-len abzudecken;
- Unmöglichkeit, Psychotherapeuten in mehr als einigen wenigen störungsspezifi-schen Verfahren gut auszubilden;
- Unzureichende Nutzung des outcome-Varianzanteils, der nicht von Techniken, son-dern z.B. der therapeutischen Beziehung abhängt;
- hohe Kosten für Randomisierte Klinische Studien (RCTs), welche diesen For-schungsansatz als flächendeckende Strategie unrealistisch machen;
- beschränkter Beitrag zu Wissen, wie Psychotherapie wirkt – was zur Optimierung der Behandlung am einzelnen Patienten ebenso wie der Techniken als solcher notwendig wären.“

(Deutsche Forschungsgemeinschaft: Franz Caspar: Entwicklung und Bewertung von anwen-dungsorientierter Grundlagenforschung in der Psychologie- Forschungsdesigns in der Psychotherapie-forschung: Die Diskussion um Randomisierte Klinische Studien – Akademie Verlag Berlin 2006)

Freiburg, den 11.02.07



Dr. Birgit Clever



Norbert Bowe

Anhang 1

Anmerkungen und Änderungsvorschläge zu einzelnen Textstellen der Entwurfsfassung

Zu Punkt 1. Wissenschaftliche Anerkennung (S. 3 – 8)

- a. Mit den Eingangsausführungen fokussiert der WBP ganz auf die experimentelle Wirksamkeitsforschung, mit folgender Begründung:

„Aus der Erklärung der Wirkungsweise eines Verfahrens oder seiner Theoretischen Begründung lässt sich nicht zwingend auf seine Wirksamkeit und Unschädlichkeit schließen.“

Diese Feststellung trifft aber in vergleichbarem Umfang auch auf die experimentelle Wirksamkeitsforschung zu, wie bereits aus der Pharmakazulassung bekannt ist und wie erst recht bei den erheblichen methodischen Problemen bei der Überprüfung von Psychotherapieverfahren (s. III., Punkt 1.-7.) unterstellt werden muss.

Gerade wegen der großen methodischen Unsicherheiten muss die Bedeutung der Erklärung von Wirkungsweisen ein höherer Stellenwert zugemessen werden. Denn diese Kenntnisse ermöglichen im Versorgungsalltag therapeutische Strategien zu entwickeln in Behandlungsabschnitten, in denen standardisierte Behandlungsstrategien nicht (mehr) weiterhelfen können. Es spricht alles dafür, diese Quelle für - den individuellen Behandlungsverläufen angepasste - Interventionen nicht in die Beurteilung einzubeziehen.

- b. Experimentelle Wirksamkeitsnachweise sind auch nicht von vorn herein geeigneter, **Unbedenklichkeit nachzuweisen**, sie können allenfalls eine begründete Anfangsvermutung der Unbedenklichkeit ergeben. Ernst zu nehmende Bedenken bezüglich der Auswirkungen bei Anwendung im Versorgungsalltag können nur mit breit angelegten Versorgungsforschungsstudien ausgeräumt werden. Die einseitige Bezugnahme auf experimentelle Studien ist in dieser Hinsicht nicht adäquat.

Bei der Beurteilung der Unbedenklichkeit bezieht sich beispielsweise die Verfahrensordnung des Gemeinsamen Bundesausschusses (GBA) auf Faktoren wie **Morbidität, Mortalität und Lebensqualität**. Diese outcome-Kriterien können nur anhand einer auch im Langzeitverlauf anhaltenden Wirksamkeit und auf Dauer ausbleibenden nachteiligen Auswirkungen („Nebenwirkungen“, Komplikationen, Wiedererkrankungen) an größeren Populationen in der Versorgung beurteilt werden. Dazu sind v.a. auch andere Studiendesigns (epidemiologische, naturalistische) unerlässlich. Mehr als für Arzneimittelstudien gilt, dass empirisch-experimentelle Wirksamkeitsstudien, auch bei Erfüllung der WBP-Kriterien für externe Validität, keine sichere Beurteilung der Wirksamkeit und Unbedenklichkeit in der breiten klinischen Anwendung erlauben: Nebenwirkungen und Risiken (ob bei Breitenanwendungen vermehrt Therapieabbrüche, Krisen, Suizidalität etc. auftreten) sind mindestens so bedeutungsvoll für eine Methodenbeurteilung, können aber naturgemäß nicht schon im Anerkennungsverfahren hinreichend valide überprüft werden.

- c. Die Ausdrucksweise: „*In der Wissenschaft gilt daher ein Verfahren nur dann als **positiv** wirksam, wenn seine Wirksamkeit und seine Unbedenklichkeit **nachgewiesen** wurden*“ ist wissenschaftlich bedenklich (es gibt nur wirksam, aber nicht positiv wirksam), wenn dadurch ein genereller wissenschaftlicher Vorrang für die empirisch-experimentelle Wirksamkeitsforschung unterstellt wird und eine Herstellbarkeit objektiver Nachweise behauptet wird. (s. auch Ausführungen zur Unbedenklichkeit).
- d. Die vom WBP in den Beurteilungs-Algorithmus nicht einbezogenen Aspekte der „Erklärung der Wirkungsweise“ und „der verfahrensbezogene Grundlagenforschung und Versorgungsforschung“ stellen u. E. eine wissenschaftlich nicht begründete Einschränkung dar. Diese Aspekte, insbesondere die Versorgungsforschung, liefern nicht nur entscheidend wichtige Beiträge zur bisherigen und zukünftigen Entwicklung der Psychotherapie. Für die wissenschaftliche Beurteilung bereits sozialrechtlich zugelassener Verfahren sind sie sogar unverzichtbar. Nur mittels der Versorgungsforschung können die erheblichen Nachteile und Schwächen der Aussagekraft der experimentellen Studiendesigns (Entfremdung von der Versorgungsrealität, Einschränkungen der externen Validität) – in Studien über lange Beobachtungszeiträume an großen Populationen - kontrolliert und korrigiert werden.

Die Kritik richtet sich nicht grundsätzlich gegen die hier gewählte Form der Operationalisierung zur Zulassung neuer Methoden/Verfahren. Aber es sollte unbedingt auf eine positivistische Einengung der Begrifflichkeit (nur behauptete Objektivität) aus Gründen der Wissenschaftlichkeit verzichtet werden. Denn ohne die Ergänzungen durch nicht-experimentelle Untersuchungen an den Versorgungsrealitäten würde einem selbstreferenziellen System Vorschub geleistet, mit der Folge einer Strangulation wissenschaftlichen Fortschritts und nachfolgend einer medizinisch nicht zu rechtfertigenden Beschneidung psychotherapeutischer Behandlungsmöglichkeiten für betroffene Patienten. Wenn der Untersuchungsgegenstand Psychotherapie gezwungen wird, sich primär den Untersuchungsmethoden der experimentellen Wirksamkeitsforschung anzupassen, dann schneidet die „wissenschaftliche Methode“ auf Dauer nach Art eines Prokrustes-Bettes die Aspekte der Psychotherapie ab, die durch diesen Untersuchungstyp nicht hinreichend erfasst werden. Da es höchst unwahrscheinlich ist, dass nur mit der experimentellen Psychotherapieforschung mit klaren Manualisierungen und festen Definitionen von therapeutischem Vorgehen die Psychotherapie versorgungsadäquat in allen wesentlichen Aspekten abgebildet und weiterentwickeln werden kann (s. Punkt III., 5.), für Langzeittherapien sogar zwingend andere Wege zu beschreiten sind (s. Punkt III., 6.), muss eine derartige Einengung des Verständnisses wissenschaftlicher Untersuchungen und wissenschaftlicher Begrifflichkeiten unbedingt vermieden werden.

„Die Wirksamkeitsforschung und ihre Ergebnisse sind nur in eingeschränktem Maße nützlich, d.h. diese Forschung stellt nicht in ausreichender Weise Wissen über die Ätiologie und Behandlung psychischer Störungen bereit, das von Therapeuten umgesetzt wird und zu einer nachgewiesenermaßen erfolgreichen Praxis führt.“ (2)

„Ohne klinische Erfahrung riskiert die ärztliche Praxis durch den bloßen Rückgriff auf die Evidenz ‚tyrannisiert zu werden, da selbst exzellente Forschungsergebnisse für den individuellen Patienten nicht anwendbar oder unpassend sein können (Sackett, Mitbegründer der EbM)“ (3)

Aus der Prozessforschung sind inzwischen Systematisierungen (OPD und Heidelberger Unstrukturierungsskala) entstanden, deren Anwendungen in der Forschung valide Ergebnisse ermöglichen, die z.T. allerdings gegenüber den WBP-Kriterien modifizierte Operationalisierungsstrategien sinnvoll machen. Beispielsweise können diese Instrumente auf anderem Weg als über ICD 10 Störungen von Krankheitswert nachweisen und statt über Manualisierung durch Prozesscharakterisierungen eine treatment integrity ohne weitreichende Verfremdung des therapeutischen Vorgehens wahrscheinlich machen. Der Nachteil von Einschränkungen der methodischen Qualität ist gegen den Nachteil der Praxisferne abzuwägen. Es sollte der Fehler vermieden werden, durch zusätzliche methodische Auflagen den Vorteil der Praxisnähe wieder zu unterlaufen.

Zu Punkt 2. Psychotherapeutische Methoden und Verfahren (S.6 – 7)

Die vorgeschlagenen grundlegenden Definitionen der Begriffe „Technik, Methode, Verfahren“ (S.6) erscheinen sinnvoll und brauchbar als generell zu übernehmende Konvention im Bereich der Psychotherapie.

Problematisch erscheint der mit der Version 2.1 vorgenommene Einschub zu den allgemeinen Therapiefaktoren insofern, als dort der spezifische Beitrag eines Psychotherapieverfahrens zum Umgang mit der Therapeut-Patienten-Beziehung, der insbesondere bei Langzeittherapien von außerordentlicher Bedeutung ist, behandelt wird, als sei es eine unspezifische, allen Verfahren zuzuordnende Größe, die quasi urwüchsig und unkontrolliert (als Placeboeffekt) zustande käme. Dem Therapeuten Mittel an die Hand zu geben, sich auch unter therapeutischen Gesichtspunkten für seine Patienten zu interessieren und ihnen Heilungschancen anzubieten, wenn sie nicht innerhalb der begrenzten Interventionszeit zu einer hinreichenden seelischen Stabilisierung kommen, erfordert spezifisch zu vermittelnde Inhalte der Behandlungsgestaltung und der Gestaltung der therapeutischen Beziehung, die mit allgemeinen Therapeutenvariablen nicht allein zu erfassen sind.

Zu Punkt 3. Anwendungsbereiche der Psychotherapie (S. 7 – 8)

Im Text heißt es: *„Bei Anwendung dieses Vorgehens in den vergangenen Jahren wurde jedoch deutlich, dass eine Generalisierung von Aussagen über die Wirksamkeit einer Methode bei einer bestimmten Störung (z.B. Soziale Phobie) auf alle Störungen dieses Anwendungsbereichs (im Beispiel Angststörungen) nicht immer wissenschaftlich gerechtfertigt ist.“* Dieser Satz geht u. E: noch weit an den Versorgungsrealitäten vorbei. Am Beispiel der depressiven Störungen sei nur kurz darauf hingewiesen, dass eine verlässliche Zuordnung z.B. bei Vorliegen einer depressiven Symptomatik mittels üblicher Screening-Methoden nur schwer möglich ist, dass z.B. der therapeutische Umgang mit einer rezidivierenden Depression mit Phasen ohne erkennbare oder nur sehr allgemeine Auslösesituationen völlig anders zu behandeln ist als eine Dysthymie mit erkennbaren lerngeschichtlich oder psychodynamisch bedeutsamen Bedingungs- und Auslösezusammenhängen. Es müssten letztlich eine Unzahl von Studien generiert werden und eine Unzahl von Anerkennungsverfahren. Es kann im Bereich depressiver Störungen überhaupt nicht von Wirksamkeitsstudien über Kurztherapien an 20 oder weniger Personen davon ausgegangen werden, dass

sich die verschiedenen ICD-Differenzierungen wieder fänden, erst recht nicht die enorme Zahl von Störungscharakteristika, die mit dieser Systematisierung gar nicht erfassbar sind.

Eine Generalisierung ist nicht nur „nicht immer“ gerechtfertigt, sondern die mangelnde Generalisierungsmöglichkeit ein durchgängiges Charakteristikum u.a. aufgrund phänomenologisch ausgerichteter Diagnostikkriterien. Das ist ein systematischer Grund, weshalb über Wirksamkeitsnachweise mit Manualisierung und strengen Diagnosedefinitionen es generell nicht möglich erscheint, das Spektrum psychischer Störungen adäquat abzubilden.

Die vom WBP gezogene Folgerung: *„Beurteilungen der wissenschaftlichen Anerkennung sollten **daher primär** für einzelne Methoden für einzelne Störungen oder Anwendungsbereiche vorgenommen werden.“* ist von daher nicht nur nicht zwingend, sondern eher nicht logisch: da die Eingrenzungen auf bestimmte Störungen immer nur kleine Abschnitte des Spektrums der einzelnen Anwendungsbereiche abdecken, müsste den Studien, die mit größeren Teilnehmerzahlen größere Anwendungsgruppen abgreifen, ohne dadurch alle Validitätskriterien des WBP unter Berücksichtigung der Realisierbarkeit einhalten zu können, eine höhere Relevanz bei der Wirksamkeitsbeurteilung zugesprochen werden.

Mit der Bemerkung: *„zusammenfassende Bewertung nicht vollständig aus den einzelnen Forschungsergebnissen ableitbar“*. wird indirekt auf die nicht herstellbare Stringenz der Entscheidungsfindung hingewiesen.

Zu Punkt 1. : Verfahren zur Beurteilung der wissenschaftlichen Anerkennung von Methoden und Verfahren der Psychotherapie (S. 10 – 12):

Methode der Formulierung der Fragestellung und Zusammenstellung der Studien ist sehr gut und transparent dargelegt.

Einzig bei der Ausmusterung der Studien anhand der abstracts sollte bei den zu beurteilenden Verfahren die Möglichkeit vorgesehen werden, ggf. durch Vorlegen der gesamten Studie eine Revision des Ausschlusses zu ermöglichen.

Zu Punkt 2. Beurteilung einzelner Studien zur Wirksamkeit von Psychotherapiemethoden (S. 13 – 18):

2.1. Überprüfte psychotherapeutische Vorgehensweise:

„Die psychotherapeutische Vorgehensweise ist in der Regel durch ein Manual oder manual-ähnliche Behandlungsrichtlinien konkretisiert.“ (S.13)

Diese Bedingung schließt von vorn herein anerkannte andere psychotherapeutische Vorgehensweisen aus. Sie greift in das Konzept und in die Art der verfahrensspezifischen Gestaltung des Therapeut-Patienten-Verhältnisses bei mehreren Psychotherapieverfahren ein i.S. einer Beeinträchtigung der verfahrensspezifischen Wirkfaktoren. Hier wird das Erfordernis deutlich, dass neben der „manual-ähnlichen Behand-

lungsrichtlinie“ und stattdessen Kriterien zu definieren sind, die den Nachteil geringerer Standardisierung der Vorgehensweise adäquat ausgleichen.

Eine Manualisierung ist nur sinnvoll anwendbar, soweit Kurztherapien untersucht werden, die entweder typisierbare Krankheitsverläufe behandeln oder abgrenzbare Störungen bzw. Störungsaspekte. Damit ist auch ihre beschränkte Anwendungsbreite in der Psychotherapie charakterisiert. Die Wirkung verfahrensspezifischer Beiträge zur Entwicklung der therapeutischen Beziehung und des therapeutischen Prozesses darf nicht durch Zwang zur Standardisierung konterkariert werden.

2.1. Zur Feststellung des Indikationsbereichs (S. 13):

*„Dies betrifft in erster Linie die Diagnosen im Sinne des DSM- oder ICD-Schlüssels. Es können jedoch **ergänzend** auch Informationen, welche die untersuchte Patientstichprobe einer Studie zusätzlich kennzeichnen, festgehalten werden.“*

Soweit diese Verfahrensregeln auch auf bereits in der GKV-Versorgung befindliche Verfahren angewandt werden sollen, ist auf jeden Fall zu ergänzen, dass – v.a. auch bei naturalistischen Studien - als Diagnosekriterium (entsprechend Kriterium A. 2. und C.1.) auch die Bewilligung im Gutachterverfahren als Nachweis einer diagnostisch relevanten Störung mit Krankheitswert adäquat bewertet wird. Bei untersuchten großen Populationen kann dann mittels Stratifizierung ggf. eine Zuordnung zu einer oder mehreren Indikationsgruppen erfolgen. Diese Möglichkeit sollte auf jeden Fall im Text beschreiben werden.

2.2. Beurteilung der allgemeinen methodischen Qualität einer Studie (S. 14)

Da Wirksamkeit und Unbedenklichkeit beurteilt werden soll, müsste auch berücksichtigt werden, dass in naturalistischen Studien aus der Versorgungspraxis zwar methodisch weniger stringent, aber praxisrelevanter vom Umfang des Störungsspektrums, der Anwendung von Interventionen und Verfahren im Praxisalltag und von der Breite zusätzlicher Einflussgrößen (psychosoziale, Komorbiditäten, beeinflussende Lebens- und Beziehungsereignisse etc.) die Wirksamkeit nachgewiesen werden kann.

Es sollte unbedingt eine gesonderte methodische Beurteilung mit einer eigenen Kriterienliste für naturalistische Studien ergänzt werden. Ansonsten entstünde eine Überbewertung experimenteller, praxisferner Studien unter Vernachlässigung der eingangs genannten Differenzierung zwischen efficacy und effectiveness.

Durch das vorgesehene Erfordernis, dass der Durchschnitt der Bewertungen aller A-Kriterien kleiner als 2,25 sein muss, haben fast nur Studien der Evidenzstufe 1 Zugang zur positiven Bewertung. Damit werden relevante Studien für den Nachweis von Wirksamkeit von vorn herein ausgeschlossen und auch entsprechende Forschungsbemühungen behindert.

2.4. Beurteilung der externen Validität (S. 15)

Zunächst einmal ist zu begrüßen, dass diese Kategorie explizit Berücksichtigung findet.

Nicht günstig ist es, zur Beurteilung der outcome - Maße auf das Kriterium B.12. zurückzugreifen. Stattdessen sollte ein eigenständiger Punkt zur outcome-Messung unter C. eingeführt werden, der den Vergleich mit einer Kontrollgruppe als nur eines der möglichen Messinstrumente vorsieht. Da Kontrollgruppenbedingungen – zumindest wenn sie eine hinreichende Vergleichbarkeit mit der Interventionsgruppe gewährleisten sollen – eine deutliche Beeinträchtigung der externen Validität hervorrufen können, sollten alternativ anzuwendende Standards hier vorgesehen werden.

Als problematisch wird auch die prominente Bedeutung der Punkte C.10 bis C.13. beurteilt werden.

C.11. fragt die Herstellbarkeit der notwendigen Behandlerqualifikation ab: Wenn die Behandlerqualifikation mit der Ausbildung in einem Verfahren nur herstellbar ist, dann ist sie nur mit großem Zeitaufwand herstellbar. Dass eine zeitaufwändigere Strukturqualität aber als Nachteil gewertet werden soll gegenüber mit geringem Zeitaufwand herzustellender Qualifikation, erscheint wenig plausibel.

Ebenso muss C.13. als den Anforderungen der externen Validität unangemessen angesehen werden. Herstellbarkeit der Treatmentmerkmale (z.B. durch Manuale) setzt wieder einen experimentellen Ansatz voraus, der den Praxisbedingungen, unter denen Interventionen nach Manualen in der Regel in andere Interventionsformen und andere Arten der therapeutischen Haltung eingebaut werden, oder verfahrensspezifisch ohne Manuale gearbeitet wird, zu sehr entfremdet ist. Auch mit dieser Vorgabe wird die Messung der externen Validität erheblich eingeschränkt.

Da aber der Wert für C.10. – C.13. t unter 2,25 bleiben muss, sind Bedingungen vorgegeben, die eine hohe externe Validität verhindern und Verfahren benachteiligen, die Manualisierungen nicht oder nur als ergänzende Interventionsformen vorsehen

Ohne Korrektur beinhalten die unter 2.4. genannten Kriterien erhebliche Beeinträchtigungen der Beurteilbarkeit der externen Validität (effectiveness). Sie stellen eine Kompromissbildung dar, bei der Evidenzstufenhöhe (experimentelle Wirksamkeit) auf Kosten von Versorgungsrelevanz (externe Validität) gesichert wird.

Das Spannungsverhältnis zwischen interner Validität und externer Validität wird zwar versucht, durch Operationalisierungen ex juvantibus ‚irgendwie‘ zu berücksichtigen, es werden aber die Grenzen des eigenen Vorgehens nicht hinreichend problematisiert. Für den bisher unaufgelösten Konflikt zwischen möglichst exakter Kontrolle von Gegenstand und Einflussgrößen und dadurch bedingter reziproker Verringerung der externen Validität erscheint eine offenere Abwägung erforderlich, und dementsprechend auch Ausformulierungen alternativer Operationierungskriterien.

2.5. Beurteilung der Ergebnisse einer Studie (S. 16)

Die Anforderungen unter III. 1. sind insofern problematisch, als unter Bedingungen des deutschen Gesundheitssystems eine unbehandelte Kontrollgruppe ethisch nicht vertretbar ist, da jeder Patient einen Anspruch auf Behandlung hat und eine reine Wartegruppe den Anforderungen einer Kontrollgruppe i.S. dieser Kriterien insofern nicht entspricht, als bei verantwortungsvoller Handhabung der Vergabe von Behandlungsplätzen schwerer Erkrankte zuerst in Behandlung genommen werden und außerdem individuelle Vorauswahlkriterien unkontrolliert vom Behandler zur Anwendung gebracht werden.

Eine placebo-Gruppe ist in der Psychotherapie weder herstellbar, noch ist der Versuch ethisch vertretbar (es sei denn, es liegen keine Störungen von Krankheitswert vor). Darum sollte dieses Kriterium auch nicht erwähnt werden.

Problematisch ist der immanente Zwang zum Vergleich mit einer als wirksam erwiesenen Behandlungsform im Sinne der Gleichwertigkeit. Damit werden neue Methoden und Verfahren dazu gezwungen, vergleichbare Messinstrumente zu verwenden. Es kann aber sein, dass neue Verfahren mit anderen Messinstrumenten andere Effekte nachweisen, die additiv zu den Effekten zu beurteilen sind, die im Vergleich mit dem Referenzverfahren beurteilt wurden. Dann kann beispielsweise eine größere Patientenzufriedenheit, die eine höhere Compliance ermöglicht, ein Verfahren mindestens genau so effektiv und versorgungsrelevant erscheinen lassen als ein Verfahren mit guten Wirksamkeitsnachweisen aber schmalerer Akzeptanz.

Bei der Beurteilung der Ergebnisse einer Studie sind weitere qualitativ aussagekräftige Ergebniskriterien adäquat in die Bewertung mit einzubeziehen.

Sehr zu begrüßen ist die in unter Nr. 4. vorgesehene Möglichkeit, auch Studien ohne Kontrollgruppe zu berücksichtigen. Allerdings ist unklar, inwieweit dieser Gesichtspunkt überhaupt zum Tragen kommt: Soweit die „hohe externe Validität“ an den Kriterien von 2.4. gemessen wird, haben Studien mit hoher Praxisnähe und damit verbundenen Einschränkungen bei der methodischen Qualität keine Chancen.

2.7. Berücksichtigung älterer Studien (S. 17):

Unter Nr. 4. kann nicht die fehlende Kontrollbedingung zum Ausschlusskriterium gemacht werden, zumal auch bei neueren Studien fehlende Kontrollbedingungen toleriert werden (s. 2.5., III., Nr.4.).

Überprüft werden sollte auch der Cut-off-Zeitpunkt des 1. Januar 1990 für die Publikation älterer Studien. Da die hier zugrunde gelegte Kriterien erst im Laufe der 90er Jahre sich allgemein im medizinischen Bereich durchgesetzt haben, erscheint eine später angesetzte Trennlinie angebracht (z.B. 1.1.1998).

Zu Punkt 3. Wissenschaftliche Anerkennung von Psychotherapiemethoden (S. 19 - 20)

- a. Zu begrüßen ist die Kontrolle der Haltbarkeit der Therapieerfolge. Eine Follow-up-Messung 6 Monate nach Therapieende ist zwar als Kriterium für die Erst-Anerkennung der Wissenschaftlichkeit akzeptabel, um unter pragmatischen Gesichtspunkten nicht mit höheren Anforderungen einer Verhinderung der Anerkennung neuer Behandlungsmethoden Vorschub zu leisten.

Zur Beurteilung der Bewährung in der Behandlungspraxis unter Versorgungsbedingungen erscheint allerdings diese Zeitspanne völlig ungeeignet, eine belastbare Stabilität der erreichten Behandlungserfolge nachzuweisen. Hier sind follow-up-Zeiträume von mindestens 1 Jahr, eher aber über 2 und mehr Jahre erforderlich, um adäquat zu beurteilen, welche Wirksamkeit und welchen Nutzen eine Behandlung erbracht hat.

- b. Das Kriterium Punkt 3. Nr. 3. und Punkt 4. Nr. 4. ist sprachlich missverständlich formuliert.

Vorgeschlagen wird:

.. „und wenn zugleich 3. (bzw.4.) höchsten 10% der Studien, in der Regel höchstens zwei Studien, für welche.....nachgewiesen haben, dass die Methode erhebliche schädliche Effekte hat.“

Zu Punkt 4. Wissenschaftliche Anerkennung für einzelne Anwendungsbereiche (S. 21 - 22)

Nr.1: Es erscheint zumindest fraglich, inwieweit eine Studie sowohl die Kriterien der internen Validität als auch die der externen Validität zugleich erfüllen kann. Wenn dieses i.d.R. nicht erfüllt werden kann, wird damit im Endeffekt die Zahl der Studien zum Wirksamkeitsnachweis auf vier 4 erhöht und damit der Zugang für neue Verfahren erschwert. Zur Vermeidung weiterer Hürden sollte es bei 3 Studien bleiben, wobei jeweils eine mindestens eine hinreichende interne bzw. externe Validität aufweisen sollte.

Zu Punkt 5. Empfehlungen für Vertiefungsgebiete (S. 23)

Diese Empfehlungen greifen dem Stand der Forschung und Wissenschaft vor, sind (auf jeden Fall noch) nicht wissenschaftlich begründbar. Da die externe Validität i.S.d. effectiveness nicht mit überwiegend kontrollierten Designs, kleinen Stichproben und manualisierten Vorgehensweisen überprüft werden kann, sondern dazu breite Anwendungsforschung unter naturalistischen Bedingungen erforderlich ist, stützen sich die hier vorgetragenen Empfehlungen nur auf die bereits problematisierte Annahmen, dass durch Addition von 4 oder 5 Methoden, deren experimentelle Wirksamkeiten an selektierten Patientengruppen unter Anwendung isolierter **Interventionsformen** nachgewiesen wurden, eine umfassende psychotherapeutische Kompetenz garantiert werden könne - auch für die gesamte Breite des Störungsspektrums und auch für Behandlung von Störungen mit hohen Komplexitätsgraden.

Es wird bei der Empfehlung möglicherweise auch außer Acht gelassen, dass Patienten mit Multimorbidität und komplexe Störungen oft nur mit Langzeittherapien geholfen werden kann, für die spezifische Interventionstechniken in Kurzzeittherapie-Anwendungen keinerlei Rüstzeug bieten. In der Ausbildung bzw. Weiterbildung zum Psychotherapeuten sind Fähigkeiten zur Behandlung solcher Patienten zu vermitteln und Konzepte für das therapeutische Vorgehen jenseits der strukturierten Interventionen und Techniken zur Bewältigung damit verbundener Belastungen. Psychotherapieverfahren haben dabei die Funktion, entsprechend Störungsmodelle, Modelle seelischer Regulationsvorgänge und Strukturen sowie Verlaufs- und Behandlungsprozessmodelle zur Verfügung zu stellen

Auf die unter Punkt 5. gemachten Empfehlungen sollte daher verzichtet werden.

5.1. Für die Ausbildung zum Psychologischen Psychotherapeuten...

5.1.1.

Der hier vorgeschlagene Algorithmus wird begrüßt und als hinreichend adäquat angesehen – im Gegensatz zu dem vom GBA angewandten. Er wird eher dem Kriterium der Versorgungsrelevanz und der wichtigen Forderung der hinreichenden Anwendungsbreite gerecht als eine einseitige und nicht zu begründende Festlegung auf 3 Indikationsgebiete, die nur noch auf den Aspekt der allgemeinen epidemiologischen Inzidenz der ausgesuchten Diagnosegruppen rekurriert.

5.1.2.

Diese neu eingeführte Kategorie ist sehr zu begrüßen. Hiermit wird den von uns vortragenen Bedenken partiell Rechnung getragen. Allerdings verengen die A.- B.- und C.-Kriterien des Anhangs wieder zu sehr auf experimentelle Bedingungen.

Hier erscheint es unverzichtbar, naturalistische Studien mit großen Fallzahlen und großer externer Validität (nicht i.S. der hier vorgeschlagenen Operationalisierung) zum Wirksamkeitsnachweis systematisch einzubeziehen.

Das Verhältnis von 3 (bzw.4) anwendungsbereichsbezogene wissenschaftliche Anerkennungen über (experimentelle) Wirksamkeitsnachweise zu einem Ersatz durch 3 Studien an diagnostisch heterogenen Patientenstichproben ist für die Anerkennung neuer Verfahren vermutlich insofern hinreichend, als experimentelle Studien für noch nicht im Versorgungssystem befindliche Verfahren leichter durchführbar und ethisch begründbar sind.

Bei der Beurteilung bereits im System befindlicher Psychotherapie-Verfahren erscheint diese Gewichtung aber nicht sinnvoll, da dann hinsichtlich der externen Validität aussagekräftigere naturalistische Studien an in der Routine-Versorgung befindlichen großen Patientenpopulationen möglich sind. Dann muss auch ein anderes Verhältnis (2 zu 2) möglich sein.

Zu Anhang 1. Kriterienkatalog:

A. allgemeine methodische Qualität (S. 26 – 27)

Unter A.2.

Sollte auch ein peer-review-Verfahren (z.B. Gutachterverfahren Richtlinien-therapie) als adäquater Nachweis des Vorliegens von Störungen mit Krankheitswert vorgesehen werden.

Zu A.5.

Die Verbindung von hohen Stichprobengrößen und praxisnahem Patientenzugang sollten methodische Schwächen bei anderen Kriterien kompensieren können, damit auch naturalistische oder versorgungsnahe Studien einbezogen werden können.

„Sobald man Stichprobengrößen nach zuvor definierten klinisch relevanten Unterschieden berechnet, sind für eine hinreichende Robustheit der Ergebnisse auch größere Stichproben erforderlich. Beispielsweise „für die viel zitierte Catie-Studie, in der vier verschiedene atypische Antipsychotika miteinander verglichen wurden, kam man hierbei auf eine Größe von 300 Patienten pro Gruppe.“ (Psychopharmako-Therapie, 13. Jahrgang, Heft 6, 2006).

Angesichts der enormen Vielzahl der Variablen, die mit randomisierten Zuweisungen ausgeglichen werden müssen, spielt die Stichprobengröße eine große Rolle.

Die a priori - Selektionen, die allein durch die Art des Studiendesigns, den Ort der Durchführung und das Angebot an die Behandlungen durchführenden Therapeuten provoziert wird und damit beeinflusst, welche Patientenpopulationen von vorn herein zu gewinnen sind oder nicht, multiplizieren dabei noch einmal die Einschränkungen hinsichtlich der Robustheit von Ergebnissen, die mit kleinen Stichproben (n unter 100) gewonnen wurden

Unter A.8.

sind explizit auch allgemeine Parameter zu benennen, die auf bessere psychische Stabilität schließen lassen: AU-Tage, Medikation, Klinikaufenthaltstage, Arztbesuche, Rentenantragsstellung etc. sowie Erfassung von Patientenzufriedenheit und Lebensqualität.

Zusammenfassend zu A.:

Die hier dargelegten Kriterien zur allgemeinen methodischen Qualität sind für die Beurteilung experimenteller Studien u.E. geeignet und machen die Beurteilung durch den WBP transparent – im Gegensatz zum Vorgehen des GBA beim sozialrechtlichen Anerkennungsverfahren der GPT.

B. Kriterien zur Bewertung der internen Validität (S. 28 – 29)

Zu B.8.

Die Bindung der internen Validität an das Kriterium der Randomisierung/mindestens Parallelisierung in dieser zwingenden Form stellt ein einseitige Fokussierung auf RCT-Standard dar.

Vorgesehen werden sollte z.B. auch ein Vergleich 2 Jahre vor der Psychotherapie zu 2 Jahre nach Beendigung der Psychotherapie hinsichtlich allgemeiner Versorgungsparameter (Medikamentenverbrauch, Arztbesuche, Klinikaufenthalte).

C. Kriterien zur Beurteilung der externen Validität (S. 29 – 31):

Kritikpunkte wurden bereits unter 2.4. aufgeführt. Die externe Validität wird nur in Hinsicht auf RCT-Studien oder Studien mit parallelisierten Untersuchungsarmen operationalisiert. An dieser Stelle fehlen Kriterien für Versorgungsstudien/ naturalistische Studien. Dabei sind Abstriche hinsichtlich der Kontrollbedingungen hinzunehmen, dafür aber Kriterien der externen Validität schärfer zu fassen. Folgender Lösungsansatz ist denkbar:

- Der Kriterienkatalog C wird für die Beurteilung der externen Validität von am experimentellen Studientyp ausgerichteten Studien – unter Berücksichtigung der vorgeschlagenen Modifikationen – angewendet.
- Für die Beurteilung naturalistischer Studien wird ein eigener Kriterienkatalog mit Anforderungen der methodischen Qualität, der internen und externen Validität entworfen. Dabei könnten Gesichtspunkte u.a. sein:
 - keine Spezifikation von Treatmentmerkmalen
 - erheblich höhere Stichproben
 - ggf. Behandlungsprozess-Kriterien (z.B. Heidelberger Umstrukturierungsskala)
 - größeres Gewicht von Katamnese-Untersuchungen (deutlich länger als ein halbes Jahr, mindestens 2 Jahre)
 - größeres Gewicht von allgemeinen outcome- Maßen wie Patientenzufriedenheit, Lebenszufriedenheit, Inanspruchnahme von medizinischen-Reha-, sozialen Versorgungsleistungen etc.

Fydrich und Schneider zu Naturalistischen Studien: „Naturalistische Studien im klinischen Feld werden hier im Sinne von Versorgungsstudien verstanden, in denen zwar eine direkte Interventions-Wirkungsaussage erschwert ist, die dafür aber – im Sinne erhöhter externer Validität – psychotherapeutische Interventionen deutlich repräsentativer abbilden, als dies im Rahmen von RCT Studien möglich ist. Naturalistische Studien sind nicht grundsätzlich von geringerer Qualität als RCT-Studien. Beide Studienformen ermöglichen jedoch unterschiedliche Aussagen (vgl. Leichsenring & Rüger, 2004). Wird in den klinischen Studien (RCTs) die Wirkung(efficacy) einer Intervention erfasst und die interne Validität hoch gehalten, so haben naturalistische Studien die Potenz einer hohen externen Validität und es kann damit die Wirksamkeit (effectiveness) von therapeutischen Behandlungen unter Alltagsbedingungen geprüft werden.

Leichsenring und Rüter (2004) haben als Alternative bzw. Ergänzung für die versorgungsnahe Psychotherapieforschung u.a. die folgenden methodischen Kriterien vorgeschlagen:

- Möglichst klinisch repräsentative Problemstellungen und Behandlungen
- klare Ein- und Ausschlusskriterien für Patienten mit klar definierten und diagnostizierten Krankheitsbildern
- Nutzung angemessener und zuverlässiger diagnostischer Methoden
- möglichst weitgehende Sicherung der internen Validität durch angemessene Design, Vorhersage komplexer Ergebnis-Muster
- Einsatz von nicht-randomisierten Vergleichsbedingungen, möglichst jedoch mit Matching- und Stratifizierungsstrategien
- möglichst blinde Beurteilung der Erfolgskriterien“ (3)

Norbert Bowe

Anhang 2

Grundsätzliche Gesichtspunkte zum Verhältnis von experimentellen Studien zur psychotherapeutischen Praxis

Einleitung:

„Im Liber Augustalis, dem ersten mittelalterlichen Versuch einer staatlichen Kodifizierung der Rechtsordnung, schrieb im Jahr 1231 Kaiser Friedrich II: „Da die Medizin niemals erfolgreich sein kann...ohne die Kenntnis der Logik, befehlen wir, dass keiner Medizin studiere, der nicht vorher mindestens drei Jahre Logik betrieben habe.“ (zitiert nach T. Sawicki: Evidenzbasierte Medizin: Keine ‚Zwangsjacke‘ für den Arzt).Dieser Auffassung folgend sollen in Auseinandersetzung mit der vom WBP vorgeschlagenen Verfahrensweise einige grundsätzlicheren Gesichtspunkte dargelegt werden, die auf ihre Stichhaltigkeit, d.h. Richtigkeit ihrer Voraussetzungen und Richtigkeit der darauf angewandten Logik, zu prüfen sind.

Bei Verfahrensregeln zur wissenschaftlichen Anerkennung sollten die dem Entscheidungsalgorithmus zugrunde liegenden Annahmen und Abwägungen, die nicht durch wissenschaftliche Erkenntnisse abzusichern sind, offen dargelegt werden. Z.B. sind Vorzüge und Schwachpunkte verschiedener Untersuchungsdesigns zur Beurteilung von Wirksamkeit, Unbedenklichkeit und Langzeiteffekten gegeneinander abzuwägen und Entscheidungen zu legitimieren, für welche zu beurteilenden Teilaspekte welche Studientypen als am besten geeignet herangezogen werden.

In den Anwendungsbereichen der Psychotherapie liegen zum einen keine eindeutig definierten Krankheiten wie im Bereich der somatischen Medizin (z.B. Herzinfarkt, Diabetes) sondern nur Systematisierungen von Symptom/Kriterien-Kombinationen. Zum anderen besteht die Behandlung nicht in der Verabreichung einheitlicher Drogen wie bei Arzneimittelstudien. Die damit einhergehenden Einschränkungen der Aussagekraft von Studien sind zu problematisieren. Daraus ergeben sich Konsequenzen hinsichtlich der Frage, welche Art der Wirksamkeitsnachweise und den Unbedenklichkeitsnachweise adäquat sind. Eins zu eins Übertragungen aus dem Arzneimittelsektor sind schon deshalb als fehlerbehaftet und dem Gegenstand nicht ganz angemessen anzusehen.

1. Allgemeine Probleme bei der Beurteilung der Wirksamkeit und Unbedenklichkeit – erläutert am Beispiel von Studien zur Wirksamkeit von Medikamenten

Bei Neueinführungen und Vergleichsstudien zu Medikamenten mit gleichen Indikationsbereichen sind vor einer Zulassungsentscheidung Nachweise hinreichender Wirksamkeit und fehlende Hinweise auf häufige oder gravierende Nebenwirkungen oder Kontraindikationen erforderlich. Gültige, effektive Überlegenheitsnachweise oder auch Gleichwertigkeitsnachweise neuer Medikamente sind bei Einführung nicht zu erbringen; wenn sie durch Studien „nachgewiesen“ werden, stützen diese sich nicht selten auf Effekte, die anhand von fragwürdigen (zu niedrigen: bei Wirkungsnachweisen; zu hohen: bei Nachweisen von günstigeren Nebenwirkungsprofilen) Dosierungen der Vergleichssubstanzen (bewährtes Medikament) geführt werden. Außerdem stellt sich die optimale Dosierung einer neuen Substanz oft erst nach weiteren Anwendungserfahrungen nach Zulassung im klinischen Alltag heraus. Erst in

einem mehr als ein Jahrzehnt dauernden Prozess nach Zulassung kann anhand vorliegender Anwendungserfahrungen und versorgungsrelevanten Untersuchungen an großen Populationen, Untersuchungen auch zu verschiedenen Untergruppen von Diagnosegruppen (z.B. Risikogruppen), mit Untersuchungen auch zu selteneren Nebenwirkungen und Sammeln von Einzelfall-Erfahrungen Fragen der Mortalität und Morbidität (z.B. (Mit-) Verursachung anderer Erkrankungen) beantwortet werden. Erst dann werden z.T. ältere Medikamente als Mittel zweiter Wahl geführt oder als Medikamente für eingeschränkte Indikationen oder aus dem Handel gezogen, bzw. neuere Medikamente als zu risikoreich aus der Versorgung herausgenommen oder als teurere me-too Präparate als verzichtbar bezeichnet.

Beispielhaft für die Schwierigkeit, Wirksamkeit, Unbedenklichkeit und Nutzen wissenschaftlich sicher zu begründen, sei hier das Catie-Projekt erwähnt („Clinical Antipsychotic Trials of Intervention Effectiveness“- Projekt des National Institute of Mental Health). Darin wurden Raten und Zeiten bis zum Abbruch der Behandlung unter doppelblinden Bedingungen von vier neueren Neuroleptica (sog. Atypische) gegen Perphenazin untersucht und keine Unterschiede der (erheblich teureren) neuen gegenüber der Vergleichssubstanz Perphenazin gefunden. Zwar werden gegenüber den Folgerungen aus diesen Befunden und zum methodischen Vorgehen Bedenken angemeldet, dennoch kann aus dem Vorgang insgesamt als Hinweis entnommen werden, welchen Schwierigkeiten eine Urteilsfindung über Nutzen und Unbedenklichkeit begegnet. Weder die hohen Kosten der neuen Neuroleptica, noch die Compliance noch die erheblichen Nebenwirkungen der alten Neuroleptica (Parkinsonoid) noch der neuen wie z.B. des Olanzapin (massive Gewichtszunahmen/ metabolisches Syndrom) haben bisher zu einer Rücknahme vom Markt geführt oder zu einer eindeutigen Festlegung von Präferenzen bei der Verordnung.

Folgerungen:

- a. Wenn schon bei Medikamenten, die im Gegensatz zur Psychotherapie eindeutig identifizierbar und in ihrer Zusammensetzung identisch sind, und deren Indikationsbereich viel eindeutiger durch Krankheits-Diagnosen umrissen werden kann, es so schwer zu überprüfen und zu entscheiden ist, welche der Präparate besser, schlechter oder nur begrenzt einsetzbar sind, so müssen sich die aus entsprechenden Studien im Bereich der Psychotherapie zu ziehenden Schlüsse folgerichtig noch zurückhaltender ausfallen, die daraus abzuleitenden Konsequenzen hinsichtlich Zulassungs- und Indikationsentscheidungen noch sorgfältiger abgewogen und durch weitere Kriterien ergänzt werden.
- b. Zulassungsentscheidungen können lediglich anhand von begründeten Annahmen erfolgen. Eine eindeutige Beweislage ist nicht herzustellen.
- c. Die Frage der Zurücknahme einer Zulassung kann nicht ausschließlich an Wirksamkeitsnachweisen nach RCT-Bedingungen entschieden werden. Sie muss sich auf Untersuchungen an großen Kollektiven stützen und auch mit naturalistischen Untersuchungsdesigns Fragen der Abbruchraten, der Spätfolgen, der Langzeitwirkungen beantworten. Es bleibt ein nicht zu beseitigender Interpretationsspielraum.

d. Erst langjährige klinische Praxisanwendungen bringen einen Shift hin zu den im Versorgungsalltag sich bewährenden Medikamenten. In diesem Prozess kann sich auch erst herausstellen, welches Medikament für welche Subgruppe, bei welchen Patientenvariablen (Präferenzen zur Therapieart, Empfindlichkeiten, Risikofaktoren, Komorbiditäten, Nebenwirkungs-(in-)toleranzen etc.) geeignet ist (Klärung einer Differentialindikation). Derartige Gesichtspunkte sind auf den Bereich der Psychotherapie zu übertragen.

2. ‚Droge Psychotherapie‘ ist nicht herstellbar – beim Versuch degeneriert immer ein Teil der essentiellen psychotherapeutischen Wirkfaktoren.

Psychotherapie kann nicht auf gleiche Interventionseinheiten, die 3 x täglich oder 1 x wöchentlich verabreicht werden, reduziert werden, d.h. im Gegensatz zum Medikament besteht eine nicht aufzulösende Differenz zwischen den in der experimentellen Studie angewandten manualisierten Technik und der im Versorgungsalltag angewandten Psychotherapie (Ungleichheit der ‚Droge‘ unter Studien- und Versorgungsbedingungen). Soweit aber keine Identität herstellbar ist, müssen die anhand von experimentellen Studien gewonnenen Aussagen hinsichtlich des Aussagewertes zur Bewährung in der Versorgungssituation noch sehr viel vorsichtiger eingeschätzt werden als bei Arzneimittelstudien.

Auch lässt sich die Intervention der experimentellen Studie nicht von den Einträgen der therapeutischen Beziehung reinigen. Im Prinzip ist schon ein Versuch, dies zu tun, ethisch fragwürdig oder gar unerlaubt. Denn Patienten haben einen Anspruch auf eine umfassende und nicht künstlich reduzierte Therapie. Dieses mag sich in einem Land wie den USA möglicherweise etwas anders darstellen, wo viele Patienten unterversichert und unterversorgt sind und daher als Pool ohne Chancen auf eine baldige adäquate Regelversorgung für experimentelle Untersuchungen zur Verfügung stehen und von daher bei Studienteilnahme auch keine umfassende Therapie vorenthalten bekommen.

3. Für Indikation und Ausrichtung psychotherapeutischer Behandlungen sind über die Achse I – Diagnosen hinaus andere Kriterien wesentlich

Die psychotherapeutisch zu behandelnde Krankheit ist – im Gegensatz zu leichter zu umreißenden somatischen Erkrankungen wie Tuberkulose, Eisenmangelanämie oder Herzinfarkt – i.d.R. nicht mit einer Diagnose auf einer Achse auch nur annähernd zu beschreiben. Studien mit kleinen Fallzahlen und strengen Ein- und Ausschlusskriterien erlauben am ehesten noch valide Effektaussagen bei Vorliegen eindeutig bestimmender ‚Krankheitsentitäten‘. Tatsächlich stellen die deskriptiven Kriterien des Kapitel F der ICD 10 eine völlig ungenügende Systematisierung für den Personenkreis dar, der psychotherapeutisch behandlungsbedürftig ist.

Grawe : „Fast alle Patienten mit schweren psychischen Störungen haben eine Vorgeschichte von Verletzungen ihres Bindungs- und Kontrollbedürfnisses in ihrer frühen Kindheit, wie es in einem unsicheren Bindungsstil zum Ausdruck kommt. Ihr Problem lässt sich deshalb nicht auf die zuletzt entwickelte Störung reduzieren. Sie ist nur ein letztes Glied in einer langen Kette verletzender Erfahrungen, die schließlich zu so hoher Inkonsistenz geführt haben, dass ein qualitativ neues Mittel zu ihrer

Reduktion erforderlich wurde. Für die Patienten selbst ist es auch klar, dass die psychopathologische Störung im engeren Sinn nur einen Teil ihrer behandlungsbedürftigen Probleme ausmacht. Nur 9 % aller Patienten geben die Besserung ihrer Störung als einziges Therapieziel an. Drei von vier Patienten geben Probleme zwischenmenschlicher Natur als eines ihrer drei wichtigsten Probleme an.“

Die folgende tabellarische Übersicht soll nur andeuten, in welcher Breite sich z.B. depressive Syndrome, entsprechend den F 3.-Diagnosekriterien, im ambulanten Versorgungsalltag darstellen, verbunden mit jeweils deutlich unterscheidbaren Behandlungskonsequenzen.

Depressive Symptome

Symptom	Pathogenese/ Ätiologie	Behandlung
Rezidivierende Depression	Erbliche Komponente Organische Komponente?	Medikamente, Therapie zur Krankheitsbewältigung
Dysthymie (neurotische Entwicklung)	Anhaltender neurotisierender Einfluss	u.U. komplexe charakterliche Verankerung – langfristige Arbeit
Dysthymie (Persönlichkeit)	Angeborene u. erworbene Verarbeitungsweisen	sthenische Arbeit am Charakterwiderstand
Dysthymie (Biographie)	Folgestände von Missbrauch, Vergewaltigung	Scham, Verschweigen von dyn. wichtigen Aspekten, Beziehungsverwicklung
Dysthymie (Biographie)	Folge fortgesetzter traumat. Kindheitsbedingungen	Sehr unterschiedliche Verläufe, Rekonstruktions-Arbeit, Entwicklungsarbeit
Dysthymie (Beziehungsst.)	Bez.-Muster der Kindheit	System. Aspekte, (Gegen-) Übertragung
Anpassungsstörung Entwicklungsschritt	Umschriebenes belastende Kindheitsbedingen, aktuelle auslösende Situation	Herstellen von Bezügen zw. Aktueller Situation und biographischen Faktoren
Anpassungsstörung Psychosoziale Belastung	Charakterliche Einflüsse, soziale Einflüsse.	Psychosoziale Komplikationen, Arbeit an aktuellen Bewältigungsstrategien
Anpassungsstörung Verlustreaktion	charakterliche Einflüsse, neurotische und situative	Trauerarbeit, Verzicht und Verinnerlichung Von der Verleugnung zur Erinnerung.
Anpassungsstörung Emigration, Milieuwechsel	Charakterliche Einflüsse, kulturelle und situative	Übersetzungsarbeit für die Transponierung Integration von Milieuaspekten
Abhängige Persönlichkeit	Angeborene und erworbene Charaktermerkmale	Nur mit langfristig angelegten Therapien zu behandeln,

In der psychotherapeutischen Praxis steht nicht selten am Anfang die Diagnose einer depressiven Störung, während die weitere Diagnostik und Behandlung erst die hier dargestellte Differenzierung ermöglicht. Das bedeutet, dass hinter der Diagnose depressive Störung sich sehr unterschiedliche seelische Störungen verbergen, die sowohl die Möglichkeit einer einfachen und hinreichend reliablen Gruppenbildung erheblich erschwert als auch die Passung zwischen manualisierbarer Intervention und Störung.

4. Ohne Doppelverblindung sind Einflussfaktoren auf die Effekte schwerer zu kontrollieren

Gegenüber Arzneimittelstudien weisen Studien zur Wirksamkeit von Psychotherapie weitere erhebliche Handicaps auf, die die Aussagekraft der RCT-Studien einschränken:

- Bei psychotherapeutischen Interventionen ist keine Doppelverblindung möglich
- Bei Wirksamkeitsstudien zur Psychotherapie kann eine randomisierte placebo-Gruppe – zumindest unter Bedingungen des deutschen Gesundheitswesens kaum hergestellt werden.

Caspar: „So ist beispielsweise Verblindung bei Psychotherapien praktisch nicht durchführbar. Auch Wartekontrollgruppen erwecken den Verdacht, es würden gar nicht richtige Patienten untersucht, weil beim Bestehen alternativer Behandlungsmöglichkeiten und gutem Informationsstand ein Patient mit Leidensdruck kaum sehr lange ohne Behandlung verharret.“ (2)

Fydrich/Schneider: „Im Kontext der Psychotherapieforschung ist deutlich, dass zumindest der Therapeut weiß und wissen muss, welche Behandlung er durchführt.

Bei randomisierter Zuordnung muss daher darauf geachtet werden, dass die Kontrollbedingungen für die Patienten glaubhaft und für den Therapeuten authentisch sind. Hierdurch kann jedoch eine Konfundierung von Therapeutenvariablen und den Behandlungsbedingungen entstehen, die im Untersuchungsdesign schwer zu kontrollieren ist.“ (3)

5. Geringe Fallzahl der Studien und hohe Zahl der Varianzen

Die Übertragbarkeit der Aussagekraft von RCT-Studien mit begrenzten Patientenzahlen (knapp über 30 oder sogar weniger als 30) auf die zu versorgende Gesamtpopulation wird um so fragwürdiger, je höher die Zahl der Variablen ist, die das Störungsgeschehen und den Behandlungsbedarf der Patienten trotz eng durch Aus- und Einschlusskriterien definierter Gruppe maßgeblich beeinflussen. Folgende Zitate geben Anhaltspunkte für die Breite der die Effekte modulierenden Variablen:

„Zu den relevanten Patientencharakteristika, für die Einflüsse auf den Therapieprozess gezeigt werden konnten, gehören (siehe z.B. Clarkin und Levy 2004, Schneider 2006):

- Die Diagnose bzw. Charakteristika der Störung (z.B. Schweregrad und das Ausmaß an funktioneller Behinderung u.a. bezogen auf die Arbeitsfähigkeit, Freizeit, Partnerschaft und Familie)
- die Komorbidität
- soziodemographische Variablen (Alter, Geschlecht, sozioökonomischer Status und ethnische Herkunft)

- Persönlichkeitsmerkmale oder – für die psychodynamische Perspektive – das Vorliegen intrapsychischer Konflikte
- Interpersonelle Variablen wie Beziehungsstile der Patienten
- Krankheitsverarbeitung und Psychotherapiemotivation.“ (3)

Iver Hand: „In diesem Kontext sind Achse-I – Hauptdiagnosen nach ICD nur *ein* Kriterium für die Therapieplanung – neben „Komorbidität auf Achse I, II und III (oder mehr), neben intrapsychischer und interaktioneller Funktionalität, neben Veränderungsmotivation und biographisch bedeutsamen Einflussfaktoren etc.! Die dabei anfallende Komplexität der Informationen ist mit realistischen Patientenzahlen in Studien, die den Cochrane Kriterien genügen, nicht mehr zu prüfen.“ (4)

Für verlässlichere Studienaussagen über die zu versorgenden Population sind also weit größere Stichproben erforderlich. Der Aussagewert der experimentellen Studien mit meist kleinen Stichproben relativiert sich damit erheblich.

5. Schwächen der Vereinheitlichung der Intervention durch Manualisierung

Während RCTs auf Ursache-Wirkungszusammenhänge abzielen - „RCTs haben die Herausarbeitung von Ursachen-Wirkungszusammenhängen zum Ziel und sollen möglichst ‚reine‘ Effekte (efficacy) erfassen.“ (3) - spielen sich geistige und psychische Regulationsmechanismen in komplex verschalteten Systemen von Regelkreisen ab, bei denen reine Kausalzusammenhänge nur in Zwischengliedern vorkommen.

Spitzer: „Zum zweiten wird in der Neuroinformatik und Psychologie seit zwei Jahrzehnten die Theorie neuronaler Netzwerke auf die unterschiedlichsten psychologischen Phänomene angewendet (...) und es gehört zum Wesen dieser Modelle, dass man die Frage nach der Bewusstheit der untersuchten geistigen Prozesse gar nicht stellt. Der systematische Grund hierfür ist, dass bewusstes Denken regelgeleitet operiert, Netzwerke hingegen Regeln zwar implementieren können, aber nicht funktionieren, indem sie Regeln folgen. Als dritte Wurzel des neuen Unbewussten lässt sich die Sozialpsychologie ausmachen, deren Experimente zu (unbewussten) Vorurteilen, Entscheidungen und Handlungen eine Fülle von Daten zu unbemerkt ablaufenden, aber zugleich für das menschliche Zusammenleben sehr bedeutsamen, Phänomenen hervorgebracht haben.“ (1)

„Bewusstes Denken ist nicht nur sehr stark begrenzt, es ist auch stark fokussiert und arbeitet eher konvergent als divergent...Die im Folgenden diskutierten Untersuchungen zeigen beispielhaft auf, dass die im Gehirn dauernd unbemerkt ablaufende Parallelverarbeitung zu nachweisbaren Ergebnissen führt. Auch das unbewusste Denken kommt also zu Ergebnissen, ohne dass wir vom Prozess etwas bemerken. ‚Es denkt‘ in uns, unbemerkt, anders (als bewusstes Denken funktioniert).“ (1)

„Die Natur solcher Erlebnisse macht es schwierig, sie experimentell zu untersuchen. Dennoch hat man es versucht und es gibt sogar eine recht lange Tradition, jedoch meist mit negativen Resultaten. Dies wundert nicht: Plötzliche Einsichten in komplexe Probleme können lange Zeit brauchen (länger als ein Experiment erfor-

dert). Ganz allgemein gilt nämlich, dass Entscheidungen durchaus vom Nachdenken profitieren können, denn sie sind durch unser bewusstes Denken gleichsam durchschaubar. Bei allzu vielen zu beachtenden Variablen jedoch (seit Georg Miller wissen wir genau: bei mehr als 7 plus minus 2) stürzt unser bewusstes Denken ab, kommt nicht mehr mit der Komplexität klar und stört die Entscheidung, weil es sich auf nur wenig konzentrieren und/oder die Dinge falsch gewichtet.“ (1)

Damit sind implizit auch Grenzen bewusster Therapieplanung und –gestaltung angesprochen. Wenn der Raum für Einfühlung und kreativen Einfällen durch Manualisierung des Therapiegeschehens eingeengt wird, wird ein ganz wesentlicher Teil der therapeutischen Arbeit ausgeblendet.

Bewusste Prozesse und Systematisierungen therapeutischen Vorgehens sind damit noch nicht a priori unwirksam oder weniger wert. Aber Wirksamkeit von Verfahren nur an solchen Elementen der Therapie festzumachen, erscheint nicht mit der Funktionsweise des Gehirns und dem Ablauf komplexer seelischer Vorgänge kompatibel.

Folgende problematische Implikationen lassen sich trotz reproduzierbarer guter Effekte an kleinen Populationen in methodisch einwandfreien Studien nicht sicher ausschließen:

a. Patientenbezogen:

- die manualisierten Interventionen führen u.U. zu einer „Pseudosicherheit“ bei Therapeut und Patient, die zu einer vorläufigen Entlastung führt, deren Effekt sich aber erst auf längere Sicht als unzureichend hinsichtlich des Störungsgeschehens erweist.
- Der Effekt der manualisierten Interventionen, der über die allgemeinen Therapieeffekte hinausgeht, ist durch Verzerrungen bedingt, die durch besondere Trainingsmaßnahmen zur Herstellung der treatment integrity im Interventionsarm erforderlich ist (Motivierung für das Spezifische).
- Patienten mit bestimmten Persönlichkeitsmerkmalen zeigen (keine) Bereitschaft, sich an einer Therapie mit manualisierten Interventionen bzw. an einer Studie zur Erprobung eines Verfahrens zu beteiligen (Selektionsmechanismus). Generelle Übertragbarkeit ist damit nicht gegeben.
- Besonders schwer Erkrankte, situativ zusätzlich Belastete, deren Behandlung von vorn herein Zusatzkompetenzen (über die manualisierte Intervention hinaus) erforderlich machen, sind in der untersuchten Population (erheblich?) unterrepräsentiert.
- die manualisierte Intervention erweist sich zwar als gut wirksam, lässt sich bei Übertragung auf die zu versorgende Population nur in den selteneren Fällen anwenden und nur in deutlich modifizierter Form.

- die bei Anwendung in der Versorgungspraxis notwendigen Modifikationen bedürfen einer zusätzlichen Kompetenz, die nicht durch manualisierte Interventionsanwendungen gelernt werden kann, sondern durch Methoden, die Einflussnahme auf komplexere seelische Funktionsstrukturen zum Gegenstand haben.
- die manualisierte Intervention ist zwar bei einem großen Teil von Patienten relativ gut anwendbar, zeigt aber bei einer zahlenmäßig kleineren, gefährdeten Gruppe hohe Abbruchraten mit erhöhter Gefahr (auch suizidaler) Krisen, Verschlechterungen des Leidens.
- die manualisierte Form ist bei einer kleineren Risikogruppe innerhalb der Diagnosegruppe, die erst mit größeren Populationen erfasst werden kann, ungünstig, da ihre Anwendung über eine Belastung der therapeutischen Beziehung bzw. Behinderung alternativer Vorgehensweisen zu einer Heilungsverzögerung oder Verschlechterung führt.

Diese Liste möglicher Gefahren- bzw. Fehlerquellen ist sicher nicht vollständig. Wie in der Anwendungserprobung von Medikamenten sind sie an breiter angelegten Untersuchungen an großen Populationen in der Versorgungsrealität zu überprüfen.

b. Bezogen auf die zu prüfenden Psychotherapieverfahren

- die manualisierte Intervention stellt eine/keine mit den therapeutischen Methoden/ Techniken des zu untersuchenden Behandlungsverfahrens kompatible Herangehensweise dar.
- die manualisierte Intervention passt gut/weniger gut zu den verfahrenstypischen Störungs- und Therapiemodellen (auf beobachtbare Verhaltensänderungen und Symptomreduktion ausgerichtete Psychotherapie vs. Veränderungen der Selbstregulation, Bearbeitung innerer Konflikte)
- Der parallel erreichte Nutzen/die Effekte von Vergleichsverfahren außerhalb der in den Studien abgebildeten Effekte wird nicht erfasst.

Soweit manualisierte Interventionen zur Grundbedingung gemacht werden dafür, dass Studien überhaupt in die Bewertung mit einbezogen werden, stellt eine solche Vorgabe eine systematische Verzerrung zugunsten experimenteller Therapieansätze dar und zu Ungunsten von Verfahren, für die Manualisierungen eine Verfremdung darstellt und deren Behandlungskonzepte primär auf Verstehen und (Selbst-) Veränderung komplexere seelische Regelungssysteme abzielen.

c. bezogen auf das in ambulanter Praxis zu behandelnden Gesamtspektrum psychischer Störungen mit Krankheitswert

- bei isolierten Störungen/Diagnosen wirksame manualisierte Interventionen sind nicht sicher auf komorbide Störungen zu übertragen

- bei an isolierten Störungen bewährten Interventionen wäre eine nicht absehbar hohe Anzahl von jeweils störungsspezifischen Vorgehensweisen für Diagnosegruppen und Subgruppen zu erlernen und auszuüben, was zu Konfusionen beim Therapeuten führen kann.
- Erst während der Behandlung zu gewinnende zusätzliche diagnostische Erkenntnisse können unerwartet Änderungen des manualisierten Vorgehens erforderlich machen, die ohne Anwendung nicht störungsspezifischen Behandlungsmodelle zu einem Bruch im Behandlungsgeschehen führen.

Die Praxistauglichkeit von Verfahren/Methoden kann trotz nachgewiesener guter Effektivität in zahlreichen methodisch guten RCT-Studien daher erheblich eingeschränkt sein, allein mit RCT-Studien sind die hier aufgezeigten Probleme nicht auszuräumen.

Fydrich/Schneider: „Für einige psychotherapeutische Interventionsmethoden wurde die Wirksamkeit für spezielle psychische Problemstellungen und psychische Störungen überzeugend belegt (...). Aus methodischen Gründen ist jedoch strittig, ob die diesen störungsspezifisch orientierten „Kataloge“ (alleinige) Grundlage für eine angemessene psychotherapeutische Versorgung sein sollen und dürfen.“ (3)

„Die amerikanische Kommission der American Psychological Assoziation (APA) gelten als in ihrer Wirksamkeit durch hinreichende Nachweise belegt:

Kognitive Therapie und die Interpersonale Therapie nach Klerman und Weissmann bei Depressionen, die Reizkonfrontation mit Reaktionsverhinderung bei Zwangsstörungen und die kognitive Verhaltenstherapie bei Panikstörungen und Agoraphobie.“ (3)

Caspar: „Es ist unrealistisch, den Anspruch zu formulieren, für alle denkbaren Störungen und ihre Kombinationen jeweils manualisierte Vorgehensweisen zu entwickeln. Therapeuten müssen auch nach allgemeinen Heuristiken und Wirkprinzipien handeln können. Die Forschung hierzu ist noch wenig entwickelt.

Neues Wissen findet seinen Eingang in die Praxis über den Therapeuten. Die Person des Therapeuten und sein ‚inneres Funktionieren‘ spielen dabei eine zentrale bisher aber vernachlässigte Rolle (Caspar 1997). Darüber ist empirisch noch wenig bekannt.“ (2)

6. (Zu) kurze Behandlungsdauern im Studiendesign

Es gibt einen inneren Zusammenhang zwischen Manualisierung und Dauer der Behandlung: Je kürzer eine Behandlung angelegt ist, desto eher kann eine Manualisierung durchgehalten werden bzw. ist eine Manualisierung vertretbar. Je länger eine Behandlung dauert (aufgrund entsprechender Störungen), desto weniger ist eine verlässliche Manualisierung möglich, desto mehr spielen Fragen der Qualität der therapeutischen Beziehung eine Rolle, deren Einfluss wiederum mit dem Gebot der Standardisierung der Intervention konfligiert. Das führt dazu, dass bei RCT-Studien Studiendesigns bevorzugt werden mit kurzen Behandlungsabschnitten und raschem

Eintritt von Wirksamkeitseffekten, während der Stabilität solcher Effekte durch relativ kurze Beobachtungs- und Katamnesezeiträume eine geringere Bedeutung zugemessen wird.

Sowohl die bei längerer Therapiedauer sich ausdifferenzierenden Behandlungsbedürfnisse des Patienten, als auch die im Therapeuten darauf bezogenen Varianten therapeutischer Antworten bewirken zunehmende Grade von Modifikationen/Abweichungen von Standards, die nicht mehr als Einflussfaktoren kontrolliert werden können.

Es hat sich in der Jahrzehntelangen Behandlungspraxis als erforderlich herausgestellt, dass für verschiedene Störungsbilder sehr unterschiedliche Behandlungsdauern, und sehr oft längere, als bei experimentellen Studien angesetzt, erforderlich sind. Wenn sich wissenschaftlich erhärten ließe, dass diese Erfahrungen auf Irrtümern beruhten (z.B. weil Therapeuten ihre Patienten lediglich aus nicht therapierelativen Gründen hielten), müsste dieses erst einmal mit einer hohen Wahrscheinlichkeit nachgewiesen werden. Bisher liegen nur Feldstudien vor, die das Gegenteil wahrscheinlich machen, nämlich dass Langzeitbehandlungen ohne Druck von außen zu den stabilsten Behandlungsergebnissen führt (consumer-Report, Studie von Zepf et al.). Bevor aber nicht nachgewiesen wurde, dass Langzeitbehandlungen nicht wirksam sind, dürfen sie den Patienten, die einen Anspruch auf adäquate Behandlung haben, nicht vorenthalten werden.

Folgerungen: Langzeitbehandlungen, bei denen das therapeutische Vorgehen sich nicht adäquat mit Manualisierungen abbilden lässt, deren Behandlungsziele sich nicht an Symptomreduktionen der Achse I – Kriterien adäquat beschreiben lassen und deren Wirkungen und Langzeiteffekte sich nur in längeren Zeiträumen realisieren lassen, als sie den meist präferierten experimentellen Studientypen zugrunde liegen, können nicht adäquat oder nur eingeschränkt mit einer Methodik wie vom WBP vorgesehen beurteilt werden.

Bekanntere größere Schwierigkeiten einer Operationalisierung von Kriterien für Wirksamkeitsnachweisen von Langzeitbehandlungen dürfen nicht zum Anlass genommen werden, diesen mangelnde Wirksamkeit zu unterstellen. Ein solch unerlaubtes Vorgehen würde rasch Einschränkungen der Patientenversorgung mit Langzeitbehandlungen provozieren. Hier muss statt Überstülpung einer inadäquaten Untersuchungsmethodik eine adäquate Methodik noch entwickelt werden.

Unter Patienten mit längeren Behandlungsdauern befinden sich (zumindest auch) vermehrt schwerere Erkrankte oder schwieriger zu Behandelnde. Zur Realisierung des entsprechenden Versorgungsangebotes sind Psychotherapieverfahren wichtig, die nicht nur eingegrenzte, unterkomplexe Veränderungsziele behandeln können, sondern ein methodengestütztes Vorgehen und in der Prozessforschung erprobte Modelle für komplexe Interventions- und Änderungsstrategien entwickelt haben, mit denen diese Prozesse beschrieben und überprüft werden können. Hochkomplexe, mit Langzeittherapien zu behandelnde Störungen sind eher nicht mit den Bedingungen der RCT-Studien zu untersuchen. Die Wirksamkeit der dabei angewandten Verfahren ist mit Hilfe von epidemiologischen Studien bzw. naturalistischen Studien nachzuweisen. Dazu müssen „Kompensationen“ definiert werden für die nicht herstellbare Erfüllung von Kriterien der internen Validität.

7. Kontrolle von Langzeiteffekten und -wirksamkeit

In der ambulanten Versorgung ist – verfahrensunabhängig – die Erfahrung bekannt, dass bei guter Entwicklung der Therapeut-Patient-Beziehung in der Therapie sich anfänglich rasch eine Symptomreduktion entwickelt. Danach besteht eine Neigung der Patienten, aufgrund der Besserung die Therapie zu beenden und sich nicht mehr auf das Bearbeiten von komplex gestörten systemischen Regelungsmechanismen, wie z.B. mit Modifikationen der Selbstregulation, der Affekte der Konfliktbereitschaften etc. einzulassen, auch wenn diese mitbedingend für die Aufrechterhaltung des Störungsgeschehen sind. Es gibt eine entsprechende Neigung auch der behandelnden Therapeuten, sich nicht auf die Bearbeitung dieser schwierigeren Störungsanteile einzulassen, zumal wenn dadurch Gefühle von Nichtbeherrschung, persönlicher Belastung oder mangelndem Erfolg vermieden werden können.

Erreichte Anfangsverbesserungen können - auch aufgrund der damit verbundenen gegenseitigen narzisstischen Aufwertung (von Patient und Therapeut) - längere Phasen psychischer Stabilität einleiten, die erst im Gefolge einer neuen auslösenden Situation/Krise/psychosozialen Belastungssituation von einer erneuten psychischen Dekompensation abgelöst werden. Allein aufgrund dieser sehr häufig beobachteten Phänomene müssen Katamneseuntersuchungen im Abstand von einem halben Jahr als ungeeignet angesehen werden, um die Dauerhaftigkeit erreichter psychotherapeutischer Effekte zu beurteilen. Studien mit Halbjahres-Katamnesen können auf keinen Fall Fragen der Linderung oder Heilung seelischer Störungen mit Krankheitswert schlüssig beantworten. Der kurzfristig erreichbare Effekt, der aus der affirmierenden Erfahrung rasch zu erreichender Symptomlinderung resultiert und der durch rationale Überschaubarkeit manualisierter Vorgehensweisen und Anbieten einer ‚machbaren‘ technischen Bewältigung eine primäre Überzeugungskraft bei Patient und Therapeut verstärken kann, ist von langfristig stabilen Effekten zu unterscheiden, die sich erst anhand von Bewährung in psychosozialen Belastungssituationen sicherer abgrenzen lassen. Nur bei sehr viel längeren Zeiträumen und Beobachtungen an größeren Populationen unter Versorgungsbedingungen können stabile Effekte unter wiederholten „Bewährungssituationen“ im Zeitverlauf nachgewiesen werden.

8. Zusammenfassung und Folgerungen aus Punkten 2. bis 7.:

Wenn schon Studien zur Wirksamkeit von Arzneimitteln so wenig sicher Auskunft geben können darüber, welches der bei gleicher Indikation eingesetztes klar definierbares Medikament im Endeffekt besser ist als ein anderes und wenn dabei Nutzen, differenzielle Einsetzbarkeit, Nebenwirkungsprofil, langfristigen (Neben-)Wirkungen etc. erst nach langen Anwendungszeiten abwägend beurteilt werden können, dann muss gegenüber Wirksamkeitsnachweisen von psychotherapeutischen Methoden aus experimentellen Studien eine noch viel größere Zurückhaltung der Bewertung an den Tag gelegt werden.

Da weder eine einheitliche ‚Droge Psychotherapie‘ herstellbar ist (s. 2.), noch auf eindeutige Krankheitseinheiten wie in der somatischen Medizin rekurriert werden kann (s. 3.), da keine Doppelverblindung möglich ist und Randomisierungen zumindest im deutschen Versorgungsbereich erheblich erschwert sind (s. 4.), da bei Versuchen der Vereinheitlichung der untersuchten Interventionen es zu erheblichen Ab-

weichungen von der Wirkung im Therapiealltag kommt sowie zu Entstellung von Therapiemethoden (s. 5.) und da mit den vereinheitlichten Interventionen nur kurze Behandlungszeiten abgegriffen werden können (s. 6.) sowie für Langzeittherapien keine geeigneten Systematisierungen zur Verfügung gestellt werden können (s.7.), sind die Wirksamkeitsnachweise aus experimentellen Studien nur als erste Hinweise auf die Brauchbarkeit im Versorgungsalltag zu interpretieren. Sie erlauben in keinem Fall, damit umfassend oder annähernd adäquat die Wirksamkeit und der therapeutische Nutzen eines Behandlungsverfahrens zu erfassen und zu beurteilen.

Nur mit Studien, die längere Zeiträume der klinischen Anwendung abdecken, auch mit naturalistischem Design, können sicherere Nachweise von beispielsweise ungünstigeren „Nebenwirkungsprofilen“, auf Dauer nicht anhaltende Effekte, oder gar erst katamnestisch zu erfassende Verschlechterungen liefern. Erst nach langjähriger Anwendungserfahrung sind – wie bei den Medikamenten – mit Hilfe von Katamneseuntersuchungen relevante Endpunkte zu beurteilen mit Überprüfung von Mortalität (z.B. Suizidraten) und Morbidität (Häufigkeit von schweren Krisen, Häufigkeit von Krankenhausaufenthalten etc.), Lebensqualität etc.

Dazu Ivo Hand: „Die EbM Anforderungen aus der ‚Somatomedizin‘ sind nicht direkt auf die ‚Psychofächer‘ übertragbar, wie das Scheitern der Depressionsstudie des National Institute of Mental Health in den USA als Beispiel zeigt: erst waren Pharmakotherapie und interpersonelle Psychotherapie die ‚Gewinner‘, in den Katamneseuntersuchungen dann die VT und die Placebogruppe. Außer der Erkenntnis von hohen Kosten waren gültige Schlussfolgerungen daraus nicht möglich.“ (4)

Norbert Bowe

Literatur:

(1) Das neue Unbewusste II –kreativ denken und richtig entscheiden, M. Spitzer Nervenheilkunde 9/2006)

(2) Deutsche Forschungsgemeinschaft: Entwicklung und Bewertung von anwendungsorientierter Grundlagenforschung in der Psychologie- Forschungsdesigns in der Psychotherapieforschung: Die Diskussion um Randomisierte Klinische Studien – Franz Caspar.(Akademie Verlag Berlin 2006)

(3) T. Fydrich, W. Schneider in: „Psychotherapeut“ 1/2007, Evidenzbasierte Psychotherapie

(4) Iver Hand, Prof., jahrelanger Klinikdirektor an der UKE Hamburg in einem Interview mit dem bvvp-magazin 1/2007, S..

(5) Psychopharmakotherapie, 13. Jahrgang, Heft 6, 2006, S. 234.