

### Redaktion

M. Cierpka, Heidelberg  
P.L. Janssen, Dortmund  
B. Strauß, Jena

# Checkliste zur Bewertung der wissenschaftlichen Qualität kontrollierter psychotherapeutischer Interventionsstudien (CPI)

**R**andomisierte kontrollierte Studien gelten als das Optimum zur wissenschaftlichen Erforschung des Nutzens medizinischer Interventionen. Die Qualität solcher Studien hängt vor allem davon ab, daß bestimmte Voraussetzungen und Regeln eingehalten werden, wie beispielsweise die Kriterien der Cochrane-Collaboration bzw. Evidence-Based-Medicine (Cooke u. Sackett 1996, vgl. Tabelle 1), die das Ausmaß der erreichten wissenschaftlichen Evidenz bestimmen (z.B. Crombie 1998). Zur Überprüfung, inwieweit diese Voraussetzungen erfüllt werden, gibt es in den letzten Jahren eine zunehmende Zahl von Checklisten, mit denen systematisch Studien überprüft werden können (z. B. The Standard of Reporting Trials Group 1994; Moher et al. 1998). Das Ziel besteht meist darin, die Studienqualität abschließend zu bewerten. In jüngster Zeit gibt es auch Versuche, diese Kriterien, die zunächst für den Bereich der Organmedizin entwickelt wurden, auch auf das Gebiet der Psychotherapieforschung zu übertragen (z.B. Kriterien für EBMH-Publikationen 1998).

Grundsätzlich ist diese Entwicklung, methodenkritisch und mit nachvollziehbaren Kriterien den Nutzen bestimmter psychotherapeutischer Interventionen zu bestimmen, sehr zu begrüßen. Um nur einen positiven Aspekt herauszustellen: In der Regel ist es dem in der Patientenversorgung tätigen Psychotherapeuten v.a. aus Zeitgründen

nicht möglich, die korrekte methodische Durchführung einer Untersuchung zu überprüfen, auf die er möglicherweise gerade seine Therapieentscheidung begründet. Durch eine nachvollziehbare, zusammenfassende Bewertung der wissenschaftlichen Qualität einer Studie kann ihr praktischer Wert rasch und adäquat eingeschätzt werden. Dies kann rationale Entscheidungen bei der Indikationsstellung erleichtern und erhöht für Patienten die Chance, daß sie für ihre Störung die jeweils nach dem aktuellen Forschungsstand effektivste Behandlung erhalten.

Es liegt jedoch auf der Hand, daß Qualitätskriterien, die überwiegend für pharmakologische Studien entwickelt wurden, nur teilweise und dann oft mit anderen Gewichtungen auf den Bereich der Psychotherapie angewendet werden können.<sup>1</sup> Sie müssen den Erfordernissen dieses Forschungsbereichs angepaßt werden, damit eine adäquate Bewertung der wissenschaftlichen Qualität der Untersuchungen möglich wird. Bislang gibt es, nach einer umfassenden Literaturrecherche, keine publizierten Checklisten, die speziell für die Evaluation von Psychotherapiestudien nach den Kriterien der Evidence-Based-Medicine entwickelt wurden.

<sup>1</sup> Man stelle sich nur eine doppelt verblindete Psychotherapiestudie vor, bei der nicht nur der Patient, sondern auch der Therapeut nicht wissen, mit welcher Methode gerade behandelt wird

Tabelle 1

### Klassifikation der Levels wissenschaftlicher Evidenz

(Cooke u. Sackett 1996)

1. **Harte Evidenz** beruhend auf mindestens einem systematischen Review, das verschiedene kontrollierte randomisierte Studien mit gutem Design einschließt
2. **Harte Evidenz** beruhend auf mindestens einer kontrollierten randomisierten Studie angemessener Größe mit gutem Design
3. **Evidenz** beruhend auf nicht-randomisierten Studien mit gutem Design, einzelne Gruppen vor-nach, Kohorten-Studie, Serien in zeitlicher Folge oder Fall-Kontroll-Studie
4. **Evidenz** beruhend auf nicht-experimentellen Studien mit gutem Design, die von mehr als einem Zentrum oder mehr als einer Forschergruppe durchgeführt wurden
5. **Meinungen** respektierter Experten, beruhend auf kritischer Evidenz, deskriptive Studien oder Berichte von Expertenkomitees

Zielsetzung der vorliegenden Arbeit ist es, eine Checkliste zur Bewertung der wissenschaftlichen Qualität *kontrollierter* psychotherapeutischer In-

Dipl.-Psych. F. Petrak

Klinik für Psychosomatische Medizin und Psychotherapie der Johannes Gutenberg-Universität Mainz, Untere Zahlbacher Straße 8, D-55131 Mainz

Tabelle 2

**Checkliste zur Bewertung kontrollierter psychotherapeutischer Interventionsstudien (CPI)**

Quellenangabe der Studie:	bearbeitet von:			
<b>Studiendesign und -durchführung</b>				
	ja	nein	?	Punkte
1. Ist eine Randomisierung bei der Fragestellung der Studie <i>sinnvoll</i> ?				
a) JA: Es erfolgte eine randomisierte Zuordnung. Die Randomisierung ist beschrieben und methodisch zuverlässig. ....	10	0	0	.....
b) NEIN: Die Vergleichsgruppen sind im Hinblick auf die Behandlung angemessen zusammengesetzt und ausgewählt. (Vorsicht: Es kann nur 1a oder 1b gewertet werden!) .....	10	0	0	.....
2. Es erfolgte eine Katamnese über einen klinisch relevanten Zeitraum. ....	10	0	0	.....
3. Interventions- und Kontrollbedingungen sind nachvollziehbar beschrieben. ....	5	0	0	.....
4. Die Stichprobengröße ist der Fragestellung angemessen. ....	5	0	0	.....
5. Es ergeben sich keine Anhaltspunkte für bedeutsame Abweichungen vom geplanten Studiendesign. ....	5	0	0	.....
<b>Operationalisierung und Messung der Konstrukte</b>				
6. Reliable und valide Messungen sind wahrscheinlich. ....	5	0	0	.....
7. Die Outcomevariablen sind klinisch relevant. ....	5	0	0	.....
8. Die zentralen Outcomevariablen / Hypothesen wurden a priori definiert. ....	2	0	0	.....
9. Es erfolgte eine angemessene „Verblindung“ der Interventions- und Kontrollbedingungen für die Patienten? .....	3	0	0	.....
10. Es erfolgte eine „blinde“ Evaluation der Ergebnisse. ....	3	0	0	.....
11. Die Möglichkeit negativer Behandlungseffekte wurde angemessen berücksichtigt (z. B. Varianzerweiterungs- und Deteriorationseffekt). ....	2	0	0	.....
<b>Stichprobeneffekte</b>				
12. Die Drop-out-Rate ist bis zum letzten Meßzeitpunkt niedrig. Bei hohen Abbruchraten (> 20 %) und/oder deutlicher Unterschiede in den Vergleichsgruppen erfolgte eine Intention-to-treat-Analyse zusätzlich zur Completer-Analyse. ....	10	0	0	.....
13. Der Verbleib aller rekrutierten Patienten ist dargestellt. Gründe für Studienabbrüche werden ggf. angegeben. ....	2	0	0	.....
14. Selektionseffekte:				
a) Die Art der Institution in der die Behandlungen stattfinden wurde spezifiziert (Spezialklinik, Allgemeinkrankenhaus, Praxen, usw.). ....	1	0	0	.....
b) Die Diagnosekriterien sind nachvollziehbar (z. B. DSM-III-R, ICD-10). ....	2	0	0	.....
c) Erkrankungscharakteristika der Stichprobe sind angegeben. ....	1	0	0	.....
d) Bedeutsame Selektionseffekte (z. B. aufgrund von Ausschlußkriterien) sind unwahrscheinlich. ....	1	0	0	.....
<b>Auswertung</b>				
15. Die verwendeten statistischen Verfahren sind angegeben. ....	2	0	0	.....
16. Die statistischen Verfahren sind der Fragestellung angemessen. ....	2	0	0	.....
17. Statistische Basisdaten sind angegeben (Punktschätzer: Anzahl, Mittelwert, Median o. Prozente; Streuungsmaße, p-Werte o. Konfidenzintervalle). ....	2	0	0	.....
18. Die Vergleichsgruppen sind vor der Behandlung bezüglich bedeutsamer Charakteristika gleich. Falls nicht, wurde es angemessen berücksichtigt. ....	2	0	0	.....

Tabelle 2 Fortsetzung

**Interpretation**

19. Bedeutsame Implikationen sind aus den Ergebnissen ableitbar (z. B. Therapieempfehlungen). .....	10	0	0	.....
20. Es kann entschieden werden, auf welche Patienten sich die Ergebnisse übertragen lassen. ....	3	0	0	.....
21. Es erfolgte eine adäquate Bewertung der Ergebnisse im Vergleich zum Forschungsstand. ....	2	0	0	.....
22. Bei fehlender Wirksamkeit einer Behandlung erfolgte eine methodenkritische Bewertung der Wirksamkeitsdefizite. ....	2	0	0	.....
23. Es gibt kein Anhaltspunkt für bedeutsame Effekte, die vom Autor übersehen wurden. ....	2	0	0	.....
<b>Gesamtpunktwert</b> .....				.....

**Abschließende Gesamtbewertung**

**Qualität der Studie**  
Gesamtpunktwert

**Angestrebtes „Cochrane-Level“**

**Level erreicht?**

- > 66 = hoch .....
- > 33 = mäßig .....
- ≤ 33 = gering.....

- 2
- 3

- ja
- nein

terventionsstudien (CPI), also *Untersuchungen, die neben der Behandlungsbedingung mindestens eine Vergleichsgruppe beinhalten*, zunächst kurz vorzustellen (vgl. Tabelle 2). Eine ausführliche Erläuterung zu den einzelnen Kriterien und der Begründung für ihre Auswahl befindet sich in Vorbereitung. Es sei ausdrücklich betont, daß es sich bei der CPI um eine erste Arbeitsversion handelt, die Gegenstand einer kritischen Diskussion sein sollte. Insbesondere ist die Punktevergabe für die Erfüllung der einzelnen Kriterien als vorläufig zu betrachten.

**Struktur der Checkliste**

Die Checkliste umfaßt fünf Kriterienbereiche (vgl. die nachfolgenden Punkte 1–5), die zur Evaluation der wissenschaftlichen Qualität kontrollierter Interventionsstudien anhand detaillierter Fragen bearbeitet werden. Das „Abhaken“ einzelner Kriterien ist dabei in den meisten Fällen eher das Ergebnis einer inhaltlichen Auseinandersetzung als das formale Prüfen einzelner Punkte auf ihr Vorhandensein. Somit ist klar, daß methodische Grundkenntnisse vorausgesetzt werden.

**1. Studiendesign- und durchführung**

Im wesentlichen werden Fragen zur sinnvollen Zusammensetzung und Größe der Vergleichsgruppen, zu möglichen Selektionseffekten, zur Randomisierung und Verblindung sowie der Kamnosedauer gestellt. Als exemplarischer Punkt, der die Unterschiede zwischen den Anforderungen von Medikamenten- und Psychotherapiestudien verdeutlicht, soll hier nur das Beispiel der Randomisierung herausgestellt werden. Der Vorteil einer Randomisierung bei der Evaluation konkurrierender Therapieformen besteht darin, systematische Verzerrungen zu vermeiden, indem neben den untersuchten Einflußvariablen auch mögliche unbekannte oder nicht kontrollierte Einflußvariablen zufällig auf die Vergleichsgruppen verteilt werden. In pharmakologischen Studien wird dabei die Wirkung unterschiedlicher Medikamente überprüft. Dieses Forschungsparadigma sollte auch in der Psychotherapieforschung eingesetzt werden, wenn die untersuchten Therapieformen weitgehend vergleichbar sind (z.B. ambulante Verhaltenstherapiegruppe vs. ambulante psychodynamische Gruppe). Sind die untersuch-

ten Verfahren jedoch nicht vergleichbar, wie beispielsweise bei ambulanter vs. stationärer Therapie, wird eine Randomisierung der Fragestellung nicht gerecht. Hier können sich durch die Zufallszuweisung zu den Behandlungen unterschiedlich hohe Zustimmung- oder Dropout-Raten in Abhängigkeit von Patientenpräferenzen ergeben (Saarento et al. 1995; McKay et al. 1995).

In einer Studie von Krauthauser und Bassler (1997) „fügten“ sich nach der vorherigen Aufklärung über die Behandlungszweige signifikant häufiger die schwerer gestörten Patienten der Zufallszuweisung zu einer stationären Behandlung, auch wenn sie eine ambulante Therapie bevorzugt hätten. Die weniger beeinträchtigten Patienten stimmten einer stationären Therapie v.a. dann zu, wenn sie bereits vor der Randomisierung eine stationäre Therapie für die günstigere Methode hielten. Eine von mehreren systematischen Verzerrungen besteht darin, daß in der ambulant behandelten Gruppe überproportional mehr Patienten vertreten waren, welche die ihrer Meinung nach „richtige“ Therapie erhielten. Damit bewirkte die Randomisierung gerade das Gegenteil des erwünschten Effekts,

nämlich die Zufallsverteilung möglicher Moderatorvariablen zwischen den Behandlungsgruppen.

Hinsichtlich der *Studiendurchführung* wird überprüft, inwieweit diese analog zum geplanten Design erfolgte. Erhebt die Untersuchung den Anspruch einer „verblindeten“ Studie, wird kritisch bewertet, inwieweit die „Verblindung“ z.B. der Patienten hinsichtlich ihrer Behandlungsart, tatsächlich durchgehalten werden konnte.

## 2. Operationalisierung und Messung der Konstrukte

Besondere Beachtung erfahren in diesem Abschnitt Fragen zur psychometrischen Qualität der eingesetzten Meßverfahren und zur klinischen Relevanz der Outcomevariablen. Zudem wird überprüft, ob die Evaluation der Therapien durch einen „verblindeten“ Beurteiler erfolgte.

## 3. Stichprobeneffekte

Es wird überprüft, inwieweit die Ergebnisse durch Stichprobeneffekte verzerrt sein könnten. Von zentraler Bedeutung ist in diesem Zusammenhang das Thema der Studienabbrucher: Bei hohen oder deutlich unterschiedlichen Dropout-Raten in den Vergleichsgruppen sollte eine Intention-to-treat-Analyse erfolgen. Außerdem wird geprüft, ob es Hinweise auf Selektionseffekte gibt.

## 4. Auswertung

In diesem Abschnitt geht es um die Nachvollziehbarkeit und Angemessenheit der herangezogenen statistischen Verfahren (in Zweifelsfällen sollte der Rat eines Statistikers eingeholt werden).

## 5. Interpretation

Das zentrale Kriterium in diesem letzten Abschnitt betrifft die Frage, ob sich aus den Ergebnissen der Studie bedeutende Implikationen ableiten lassen, z.B. Therapieempfehlungen für eine gut definierbare Patientenpopulation.

## Auswertung der Checkliste

Jeder CPI-Frage ist ein Wert, entsprechend seiner Bedeutsamkeit für die Gesamtbewertung der Studienqualität, zu-

geordnet. Ist das Kriterium erfüllt, wird das Kästchen in der „Ja-Spalte“ angekreuzt. Ist das nicht gegeben oder ist die Frage aufgrund fehlender Angaben im Artikel nicht beurteilbar, wird das Kästchen in der „Nein-“ bzw. „?-Spalte“ markiert. In der „Punkte-Spalte“ wird jeweils die entsprechende Punktzahl notiert. Die Summe aller Punkte ergibt den Gesamtpunktwert der Studie. Beurteilt wird folgendermaßen: Punktwerte  $>66$  gelten qualitativ als „hoch“, solche mit Werten  $>33$  und  $\leq 66$  als „mäßig“ und Untersuchungen die nur Werte  $\leq 33$  erhalten, werden hinsichtlich ihrer Qualität als „gering“ eingestuft.

Die zweite zusammenfassende Beurteilung bezieht sich darauf, welcher Grad der wissenschaftlichen Evidenz (vgl. „Cochrane-Kriterien“) durch das Studiendesign bzw. die Darstellung der Autoren angestrebt wurde. Aufgrund der nun abgeschlossenen methodenkritischen Bewertung wird eingeschätzt (im ja/nein-Kästchen ankreuzen), ob das beanspruchte wissenschaftliche Evidenzniveau tatsächlich vorausgesetzt werden kann. Es sei ausdrücklich betont, daß es sich bei der Einteilung nach den „Cochrane-Levels“ um eine Zusatzinformation handelt, die lediglich *angibt, auf welcher Stufe der wissenschaftlichen Evidenz* die methodenkritische Bewertung der Studie erfolgt. Als Negativbeispiel seien Untersuchungen genannt, die sich als randomisierte kontrollierte Studien darstellen, jedoch bei näherer Betrachtung eine methodisch fehlerhafte Randomisierung beinhalten. Hier würde das Level 2 „angestrebt“, es würde jedoch als „nicht erreicht“ eingeschätzt. Die Kombination beider Bewertungen – wissenschaftliche Qualität und „Cochrane-Levels“ – erlaubt demnach Aussagen wie z.B. „...auf der Ebene der nicht-randomisierten kontrollierten Studien (Level 3) ist die wissenschaftliche Qualität der vorliegenden Untersuchung als hoch bzw. mäßig oder gering einzuschätzen“.

Natürlich ergeben gut durchgeführte randomisierte Studien, in denen die Zufallszuweisung auch inhaltlich sinnvoll war, eine „härtere“ Evidenz als solche der Levels 3–5. Bei alleiniger Betrachtung der „Cochrane-Levels“, ohne zu überprüfen inwieweit die Zuordnung zu diesem Evidenzlevel berechtigt ist, besteht jedoch die Gefahr, z.B. Studien mit gravierenden methodischen

Mängeln in der Randomisierung fälschlicherweise als höherwertig einzuschätzen als gut gemachte Untersuchungen ohne Randomisierung.

## Ausblick

Eine Checkliste, die geeignet sein soll, die wissenschaftliche Qualität von Studien zu überprüfen, muß sich natürlich auch selbst einer kritischen Überprüfung ihrer eigenen Qualität stellen. Zu diesem Zweck ist eine Überprüfung der Interraterreliabilität unabhängiger Beurteiler bereits in Vorbereitung. Darüber hinaus sind uns kritische Anmerkungen und Rückmeldungen sehr willkommen.

Nach unserer bisherigen Erfahrung kann, nach einer gewissen Einarbeitungsphase, die Bearbeitung von Studien für die Entwicklung von EBM-gestützten Leitlinien mit der vorgestellten Checkliste erheblich erleichtert und systematisiert werden.<sup>2</sup>

## Literatur

- Cooke IE, Sackett DL (1996) Evidence-based obstetrics and gynaecology. Baillieres Clin Obstetrics Gynaecol 10:551–567
- Crombie IK (1998) The pocket guide to critical appraisal: A handbook for health care professionals. BMJ-Publishing Group, London
- Krauthauser H, Bassler M (1997) Das Problem der Randomisierung bei Psychotherapiestudien. Psychother Psychosom Med Psychol 47:279–284
- Kriterien für EBMH-Publikationen (1998) Evidence-Based Mental Health 4:98–99
- McKay JR, Alterman AI, McLellan AT, Snider EC, O'Brian CP (1995) Effect of random versus nonrandom assignment in a comparison of inpatient and day hospital rehabilitation for male alcoholics. J Consult Clin Psychol 63:70–78
- Moher D, Pham B, Jones A, Cook D, Jadad AR, Moher M, Tugwell P, Klassen TP (1998) Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? Lancet 352:609–613
- Saarento O, Hakko H, Pirkola S, Väisänen E, Tienari P (1995) First contact with the psychiatric treatment organisation and later utilization of care services. Randomized trial of inpatient psychiatric treatment versus later outpatient care with a 5-year follow-up. Nordisk J Psychiatry 49:69–73
- The Standard of Reporting Trials Group (1994) A proposal for structured reporting of randomized controlled trials. JAMA 272:1926–1931

<sup>2</sup> Wir bereiten die Durchführung von Trainingsseminaren zum Einsatz der Checkliste vor. Information und Anmeldung beim Erstautor (e-mail=petrak@psychosomatik.klinik.uni-mainz.de)