



Schriftenreihe des Instituts für Prävention und psychosoziale Gesundheitsforschung  
(Nr. 03/P11)

## **Kodierregeln 1.4**

**Kurzkodierbogen und Kodierbogen zum Methodenpapier des  
Wissenschaftlichen Beirats Psychotherapie nach § 11 PsychThG (Version 2.7 / 2.8)**

Zur Bewertung von Wirksamkeitsstudien eines psychotherapeutischen Verfahrens bzw.  
einer psychotherapeutischen Methode

Melanie Ratzek  
Luisa von Hauenschild

Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

Diese Kodierregeln sind als Nr. 03/P11 in der Schriftenreihe des Instituts für Prävention und psychosoziale Gesundheitsforschung (ISSN 1436-2090) erschienen.

**Zitiervorschlag:**

Ratzek, M. & v. Hauenschild, L. (2011). Kodierregeln 1.4: Kurzkodierbogen und Kodierbogen zum Methodenpapier des Wissenschaftlichen Beirats Psychotherapie nach § 11 PsychThG (Version 2.7 / 2.8) - Zur Bewertung von Wirksamkeitsstudien eines psychotherapeutischen Verfahrens bzw. einer psychotherapeutischen Methode. (Schriftenreihe des Instituts für Prävention und psychosoziale Gesundheitsforschung: Nr. 03/P11) Berlin: Freie Universität Berlin.

Dieser Bericht steht als kostenloser Download zur Verfügung unter:

<http://www.ewi-psy.fu-berlin.de/einrichtungen/arbeitsbereiche/ppg/publikationen/forschung/Schriftenreihe/index.html>

Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

**Beurteilerin:**

**Datum der Beurteilung:**

**Autoren:**

**Titel:**

**Publikationsjahr:**

**Datenzugriff:**

Originalstudie	<input type="checkbox"/>
Reanalyse	<input type="checkbox"/>
Replikation	<input type="checkbox"/>

**Weitere Charakteristiken und methodologische Eigenschaften der Studie:**

Anwendungsbereich	
<b>Affektive Störungen:</b>	Benennung der Diagnosen (Major Depression, Bipolare affektive Störung, Dysthymia, postnatale Depression etc. inkl. %-Angaben)
<b>Gemischte Störungsgruppen:</b>	Auflistung aller Störungen, die in die gemischte Störungsgruppe(n) eingehen (Hauptdiagnosen) inkl. %-Angaben (falls vorhanden)
Zentrale Patientenmerkmale:	
	z.B. <ul style="list-style-type: none"> <li>- Geschlechterverteilung</li> <li>- Altersverteilung</li> <li>- Herkunft</li> <li>- Bildungsgrad</li> </ul> etc.

Anwendungsform der psychoanalytisch begründeten Verfahren		
Benennung in Studie:		Benennung der Behandlung entsprechend der Benennung in der Studie (Zitat)
Einordnung nach Psychotherapierichtlinien (plus Psychoanalyse)	Sitzungsanzahl/Dauer und Frequenz	
tiefenpsychologisch fundiert		<b>Mehrfachrating möglich!</b> Rating in Anlehnung an einschlägige Literatur. <b>Orientierungswerte:</b> <ul style="list-style-type: none"> <li>• 100 Sitzungen, 1-2/Woche, sitzend → TP</li> <li>• 300 Sitzungen, 2-3/Woche, liegend → AP</li> </ul> In Abhängigkeit von den Treatmentarmen kann eine Studie durchaus bspw. tiefenpsychologisch fundierte und analytische Therapien untersuchen.
analytische Psychotherapie		
plus Psychoanalyse		
Gehen unterschiedliche Anwendungsformen der psychoanalytisch begründeten Verfahren in <u>einen</u> Treatmentarm ein oder werden die Anwendungsformen <u>separat</u> voneinander untersucht?		Wird bspw. innerhalb <u>eines</u> Treatmentarms sowohl tiefenpsychologisch fundierte als auch analytische Psychotherapie durchgeführt?
Setting:		
Einzeltherapie		
Gruppentherapie		
a priori festgelegte Sitzungsanzahl		
ja		Ist die Sitzungsanzahl bzw. Dauer der Behandlung im Rahmen der Studie limitiert?
nein		
störungsunspezifisches vs. störungsspezifisches Vorgehen		
störungsunspezifische Anwendung der analytisch begründeten Therapie		Anwendung einer störungsspezifischen Methode des psychoanalytisch begründeten Verfahrens? <sup>1</sup>
störungsspezifische Anwendung der analytisch begründeten Therapie		

<sup>1</sup> Vgl. Beutel, Doering, Leichsenring & Reich (2010); DGPT (2009); Reimer & Rüger (2006).

Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

Kontroll-/Vergleichsgruppendesign		Sitzungszahl/Dauer und Frequenz	
unbehandelte Kontrollgruppe (z.B. Warteliste)			<b>Mehrfachrating möglich!</b>
Placebo-Kontrollgruppe (stützende Gespräche, nonspecific treatment)			
TAU-Kontrollgruppe			
Aktive Kontrollgruppe (z.B. supportive Interventionen)			
verfahrensexterne und etablierte Vergleichsbehandlung(en) <sup>2</sup>			„Verfahrensextern“ bedeutet ein/e Verfahren/Methode, das <u>nicht</u> zu den psychoanalytisch begründeten Verfahren gehört
verfahrensexterne Vergleichsbehandlungen(en) <sup>2</sup> - <u>kein</u> bereits etabliertes Treatment			
verfahrensinterne Vergleichsbehandlungen(n) (z.B. Vergleich zweier Methoden der analytisch begründeten Verfahren)			„Verfahrensintern“ bedeutet ein/e Verfahren/Methode, das zu den psychoanalytisch begründeten Verfahren gehört
Vergleichsgruppe mit ausschließlich psychopharmakologischer Behandlung			
keine Vergleichsgruppe			
Wie viele Behandlungsarme insgesamt?- Welche?			Aufzählung <u>aller</u> Treatmentarme (d.h. der Intervention und der Kontroll- und/oder Vergleichsgruppen). Hier werden auch die Behandlungsarme innerhalb eines verfahrensinternen Vergleichs mitberücksichtigt und aufgezählt (so sie separat untersucht wurden).

<sup>2</sup> Als „etabliertes“ Treatment werden Verfahren betrachtet, die zum aktuellen Zeitpunkt (01/2011) als wissenschaftlich anerkannte Verfahren gelten (Verhaltenstherapie, Gesprächspsychotherapie, Systemische Therapie).

Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

Messzeitpunkte		prospektives vs. retrospektives Design		
Prä-Post		Wenn kein expliziter Post-Zeitpunkt („kollektiver Messzeitpunkt“) → Prä-Katamnese	<b>prospektiv</b> (vor Beginn des zu evaluierenden Treatments wird das Design festgelegt: Treatmentarme bzw. zu vergleichende Gruppen; Messzeitpunkte; Outcomekriterien)	-
Prä-Katamnese				
Prä-Post-Katamnese				
Post-Katamnese			<b>retrospektiv</b> (das/die Treatment/s sind bereits erfolgt und die Auswahl der Treatmentarme, die Bestimmung der Messzeitpunkte und die Festlegung der Outcomemaße erfolgt erst nach Beendigung [ggf. während] des/r Treatments)	
nur Post				
nur Katamnese(n)				
<b>Katamnesezeitraum (Post-Katamnese):</b>				Falls mehr als eine Katamnese-messung, <u>alle</u> aufzählen
<b>Gruppenzuweisung</b>				
randomisierte Zuteilung	verfahrensinterne Vergleichsgruppe/n		Hier werden auch die Behandlungsarme <u>innerhalb eines verfahrensisernen Vergleichs</u> mitberücksichtigt, d.h. die Zuweisung zu verfahrensisernen Vergleichsgruppen geratet (so sie separat untersucht wurden).	
	verfahrensexterne Vergleichsgruppe/n <i>oder</i> Kontrollgruppe/n (Warteliste, Placebo, TAU, aktive Kontrolle)			„Verfahrensiserne Vergleichsgruppen“ sind Gruppen, die mit einem Verfahren / einer Methode, das/die zu den psychoanalytisch begründeten Verfahren gehört, behandelt wurden.
teilweise randomisiert und teilweise Selbstzuteilung	verfahrensinterne Vergleichsgruppe/n		„Verfahrensexterne Vergleichs-/Kontrollgruppen“ sind Gruppen, die mit <u>keinem</u> Verfahren/ <u>keiner</u> Methode, das/die zu den psychoanalytisch begründeten Verfahren gehört, behandelt wurden.	
	verfahrensexterne Vergleichsgruppe/n <i>oder</i> Kontrollgruppe/n (Warteliste, Placebo, TAU, aktive Kontrolle)			

Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

patientenseitige Selbstzu- teilung unter Anwendung von Strategien wie z.B. Parallelisierung, Stratifizie- rung, Matching	verfahrensinterne Vergleichsgruppe/n		Parallelisierung: Balancierung der Gruppen hinsichtlich bekannter Störvariable, so dass Mittelwert und Streuung dieser Variable annähernd gleich in den Grup- pen Stratifizierung: Anzahlmäßige Gleichverteilung von Störvariab- len (z.B. Geschlechterverteilung in allen Gruppen gleich) Matching: Paarbildung hinsicht- lich Matchingvariable, so dass in Gruppen gleichverteilt
	verfahrensexterne Vergleichsgruppe/n <i>oder</i> Kontrollgruppe/n (Warteliste, Placebo, TAU, aktive Kontrolle)		
patientenseitige Selbstzu- teilung <u>ohne</u> Anwendung von Strategien wie z.B. Parallelisierung, Stratifizie- rung, Matching	verfahrensinterne Vergleichsgruppe/n		-
	verfahrensexterne Vergleichsgruppe/n <i>oder</i> Kontrollgruppe/n (Warteliste, Placebo, TAU, aktive Kontrolle)		
<b>Standardisierungsgrad des Treatments</b>			
Verwendung von Manual bzw. manualähnlichen Behandlungsrichtlinien (behandlungsprinzipienbasiert)			Primär bezogen auf Gruppe mit psychoanalytisch begründetem Verfahren:
nicht manualisiert bzw. keine Verwendung manualähnlicher Behandlungs- richtlinien (behandlungsprinzipienbasiert)			„Manualähnliche Handlungs- richtlinien“ <sup>3</sup>
explizites Therapeutentraining zwecks Studiendurchführung			-
kein explizites Therapeutentraining			-
Implementationskontrolle			Hiermit ist die Kontrolle der Behandlungsimplementierung gemeint (treatment integrity)
keine Implementationskontrolle			
<b>Selektivität der Stichprobe</b>			
Ausschluss subklinischer Symptomausprägungen	ja		-
	nein		
Ausschluss komorbider Störungen	ja		Ausschluss epidemiologisch rele- vanter Störungen
	nein		

<sup>3</sup> Vgl. Beutel, Doering, Leichsenring & Reich (2010); DGPT (2009); Reimer & Rüger (2006).

**Von den Autor/innen formulierte Fragestellung (Untersuchungsziel):**

	<p>v.a. im Abstract und unter dem Abschnitt „aim of the study...“ zu finden</p>
--	---

**Studiendesign:**

B.8.	Gruppenzuweisung	angemessene Randomisierung (inkl. Cluster-Randomisierung) bei ausreichender Stichprobengröße ( $n > 30$ /Gruppe <sup>4</sup> ), die das Gelingen der Randomisierung hinsichtlich bekannter und unbekannter (nicht erfasster) prognostisch relevanter Merkmale sicherstellt	(1)	<p><b>Eingruppendesigns und verfahrensinterne Vergleichsgruppen werden hier mit „3“ geratet.</b></p> <p>Mit „<math>n &gt; 30</math>/Gruppe“ sind die Größen pro Gruppe im Rahmen der randomisierten Zuteilung gemeint.</p>
		Parallelisierung oder teilweise randomisiert oder quasi-randomisiert oder Stichprobengröße $n < 30$ /Gruppe	(2)	<p><u>Quasi-randomisiert</u>; z.B. jeder 2. aufgenommene Patient gelangt in Gruppe A.</p>
		keine randomisierte oder parallelierte Zuweisung <b>(obligatorisches Kriterium für interne Validität (&lt;3))</b>	(3)	-

<sup>4</sup> es handelt sich um Anhaltzahlen [WBP, 2009, 2010].



Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

C.8.	Repräsentativität der patienten-seitigen Freiheit hinsichtlich der Wahl der Intervention  Echte Warteliste: „9“	Patienten entscheiden sich selbst für eine der angebotenen Therapieformen	(1)	<b>Auch bei Eingruppendesigns (inkl. Designs mit verfahrensinternen Vergleichsgruppen) zu raten!</b>
		ein Teil der Patienten (z. B. Randomisierungswillige) wird der Therapie zufällig zugewiesen	(2)	
		alle Patienten werden der Therapie (zufällig) zugewiesen	(3)	
A.5.	Stichproben-größe pro Gruppe	n pro Gruppe > 30	(1)	Gemeint sind die Stichprobengrößen zu Beginn (ab Randomisierung oder sonstiger Gruppenzuteilung)
		n pro Gruppe 10-30	(2)	
		n pro Gruppe < 10	(3)	
A.6.	Vergleiche der (sofern vorhanden) Behandlungsgruppen und der Messzeitpunkte a priori definiert	a priori Definition der Vergleiche der (sofern vorhanden) Behandlungsgruppen und Messzeitpunkte erfüllt	(1)	<b>Bei Eingruppendesigns nur auf Messzeitpunkte bezogen!</b>  A priori Definition der Vergleichsgruppen und Messzeitpunkte
		teilweise post-hoc Definition der Vergleiche	(2)	Teilweise post-hoc Definition der Vergleichsgruppen und Messzeitpunkte
		ausschließlich post-hoc definierte Vergleiche	(3)	Ausschließlich post-hoc Definition der Vergleichsgruppen und Messzeitpunkte
A.7.	a priori Definition der primären und sekundären Zielkriterien	a priori definierte primäre und gegebenenfalls sekundäre Zielkriterien	(1)	<u>A priori</u> Definition <u>und</u> <u>Differenzierung</u> primärer und sekundärer Zielkriterien
		a priori Nennung der Zielkriterien ohne Differenzierung in primäre und sekundäre Zielkriterien	(2)	<u>A priori</u> Definition aber <u>keine</u> Differenzierung primärer und sekundärer Zielkriterien
		a posteriori Definition der Zielkriterien; keine Angaben zu Zielkriterien	(3)	Alle Zielkriterien <u>a posteriori</u> definiert <u>oder</u> keine Angabe zu Zielkriterien

Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

B.10.	Definition der Messzeitpunkte (Prospektive Messung; Follow-up-Messung)	mehrere vorab festgelegte Messzeitpunkte über den Therapieverlauf incl. Prä-Post-Messungen	(1)	Mehrere <u>prospektive</u> Messungen inkl. Prä-Post-Messung z.B. SCL-90-R, GBB, BL etc.
	<ul style="list-style-type: none"> <li>• Prä-Katamnesemessung inkl. mehrerer Messzeitpunkte: „4“</li> <li>• Prä-Katamnesemessung: „5“</li> <li>• Nur Katamnesemessung: „6“</li> </ul>	ausschließlich Prä-Post-Messung	(2)	Nur <u>Prä-Post-Messung</u> , überwiegend <u>prospektive</u> Messungen z.B. SCL-90-R, GBB, BL etc.
		ausschließlich Post-Messung	(3)	Nur <u>Post-Messung</u> <i>oder</i> überwiegend <u>retrospektive</u> Messungen (z.B. „wie stark waren Ihre Symptome zu Beginn der Therapie? Inwieweit hat sich das Problem seither verändert?“ , bzw. GAS oder VEV)
B.11.	Follow-up-Messung	zeitlich störungsangemessene Katamnese mit hoher Ausschöpfung <sup>5</sup> der Stichprobe	(1)	Es müssen <u>beide</u> Bedingungen erfüllt sein (Ausschöpfungsquote und Katamnesezeitraum) Die Ausschöpfungsquote bezieht sich auf alle Untersuchungsgruppen
		Katamnese mit fraglich angemessenem Zeitraum bzw. niedriger Ausschöpfung der Stichprobe	(2)	Sobald eine der beiden Bedingungen nicht erfüllt ist → (2)
		keine Katamnese	(3)	Keine Katamnese
	Katamnesezeitraum (Zeitraum zwischen Beendigung der Therapie und letzter Katamneseerhebung):			-
	Katamnesemessungen auf <u>allen</u> Treatmentarmen?	ja nein		-
	Weitere/r katamnestische/r Messzeitpunkt/e (vor letzter Katamnese-messung liegend)? Wenn ja, wann?	ja nein		-

<sup>5</sup> Gemeint ist die Ausschöpfung der Stichprobe zwischen Prä- und Katamnese-messung.

Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

(weiter) B.11.	Einbeziehung longitudinaler Informationen: Wurde die Inanspruchnahme weiterer psychotherapeutischer Interventionen über das gesamte Katamneseintervall erhoben (Inanspruchnahmedaten, Arztbriefe, Krankenkassendaten)?	ja		Es sind in erster Linie psychotherapeutische oder psychiatrische Inanspruchnahmen gemeint
		nein		
	Direkte persönliche Untersuchung (vs. Telefon- oder postalische Fragebogenuntersuchung)?	ja		-
		nein		
	Erhebung explizit psychodynamischer Outcomekriterien (Konfliktdynamiken, strukturelle Aspekte)?	ja		-
		nein		
	Stabilitätsmessung? (d.h. differenzierte Betrachtung der Responder und Non-Responder)	ja		Sind die Responder zum Postzeitpunkt bis zur Katamnese stabil geblieben?
		nein		
	Wird der in der Studie gewählte Katamnesezeitraum explizit begründet? Wenn ja, wie?	ja		-
		nein		
C.10.	Spezifikation und Herstellbarkeit notwendiger Settingbedingungen (Praxistransfer)	Notwendige Settingbedingungen herstellbar (z. B. Infrastruktur, Kooperation, Team)	(1)	Hier werden besonders aufwändige Behandlungen, in die etwa mehrere Teammitglieder integriert werden und die einer besonderen Infrastruktur (bestimmte Räumlichkeiten, videografische Ausstattung etc.) bedürfen, mit (3) bewertet
		Notwendige Settingbedingungen nur begrenzt herstellbar (z. B. Infrastruktur, Kooperation, Team)	(2)	
		Notwendige Settingbedingungen nicht herstellbar	(3)	
C.11.	Spezifikation und Herstellbarkeit der notwendigen Behandlerqualifikation (Praxistransfer)	Notwendige Behandlungsqualifikation eindeutig beschrieben und herstellbar	(1)	z.B. die reguläre Ausbildungspflicht
		Notwendige Behandlungsqualifikation eindeutig beschrieben, aber nur mit sehr großem Zeitaufwand herstellbar	(2)	Hier sind nicht die regulären Ausbildungszeiten zum Psychologischen Psychotherapeuten gemeint, sondern die über die Fortbildungspflicht eines PP deutlich hinausgehenden
		Notwendige Behandlungsqualifikation nicht beschrieben oder praktisch nicht herstellbar	(3)	-
C.12.	Spezifikation und Erfassbarkeit relevanter Patientenmerkmale (Praxistransfer)	Relevante Patientenmerkmale (z.B. Alter, genetische Marker) praktisch erfassbar	(1)	-
		Relevante Patientenmerkmale nur mit erheblichem Aufwand erfassbar	(2)	bspw. strukturelle Beeinträchtigungen mittels aufwändiger psychodiagnostischer Erhebungsinstrumente
		Relevante Patientenmerkmale praktisch nicht erfassbar	(3)	bspw. bildgebende Verfahren

Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

C.13.	Spezifikation und Herstellbarkeit relevanter Treatmentaspekte (Praxistransfer)	Relevante Treatmentmerkmale (Art der Interventionen, Reihenfolge, Dauer) praktisch herstellbar (z.B. durch Manual)	(1)	Behandlung in Versorgungspraxis/-alltag integrierbar
		Relevante Treatmentmerkmale nur schwer herstellbar (z.B. tägliche Behandlung, Parallelbehandlung)	(2)	In Versorgungspraxis/-alltag schwierig integrierbar (z.B. tägliche Behandlung)
		Relevante Treatmentmerkmale praktisch nicht herstellbar	(3)	z.B. 2x tägliche Behandlung oder unnachvollziehbare eklektische Behandlung

**Population/Patienten:**

C.1.	Stichprobe von Patienten mit Störungen mit Krankheitswert	ausschließlich Patienten mit Störung mit Krankheitswert (z. B. ICD-, DSM-Diagnosen)	(1)	-
		Stichprobe von Patienten mit wahrscheinlicher klinischer Störung (z. B. Menschen nach Trauma) oder bis zu (maximal) 20% der Patienten mit lediglich erhöhter Symptomausprägung, z.T. subklinisch	(2)	-
		Patienten ohne festgestellte Störung mit Krankheitswert <b>(Stufe 3 = Ausschlusskriterium)</b>	(3)	-
C.2.	Art der Rekrutierung der Stichprobe	Patientenzugang überwiegend durch gängige klinische Routinen (Überweisung, Primärzugang, etc.); keine Selektionseffekte aufgrund der Zugangswege	(1)	<b>Es steht der "Zugang zur Behandlung" im Vordergrund:</b> Neben „gängige klinische Routine“: <ul style="list-style-type: none"> <li>• Wenn angeschriebene Studientherapeuten eines/r bestimmten Verbandes/Fachgesellschaft (DGPT, DPG, DGIP etc.) ihre Pat. zur Studien vermitteln sollen → 1 (keine Selektionseffekte i.S.v. "Zugang zur Behandlung")</li> </ul>
		Patientenzugang überwiegend durch gängige klinische Routinen (Überweisung, Primärzugang, etc.); Selektionseffekte aufgrund der Zugangswege	(2)	<ul style="list-style-type: none"> <li>• Wenn Zugang über Hochschulambulanz, bestimmte Krankenkassen etc., → 2 (Selektionseffekte i.S.v. "Zugang zur Behandlung")</li> </ul>
		Patientenzugang überwiegend über Aufforderungen der Forschergruppe (z. B. Anzeigenwerbung)	(3)	-

Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

B.1.	Spezifizierung der Wenn keine Ein-/Ausschlusskriterien existent (z.B. bei Rekrutierung über Therapeuten): „4“	eindeutige Spezifizierung der Ein- und Ausschlusskriterien	(1)	
		Ein- oder Ausschlusskriterien teilweise uneindeutig beschrieben	(2)	
		Ein- und Ausschlusskriterien sind nicht eindeutig definiert	(3)	
	Ein-/Ausschlusskriterien:			
C.3.	Selektivität der Stichprobe aufgrund der Ausschlusskriterien	keine Selektionseffekte aufgrund der Ausschlusskriterien: Einschluss aller Patienten	(1)	Einschluss <u>aller</u> Patienten
		mittlere Selektionseffekte aufgrund der Ausschlusskriterien (z. B. Ausschluss einiger epidemiologisch relevanter komorbider Störungen)	(2)	z.B. Ausschluss einiger <u>epidemiologisch relevanter</u> komorbider Störungen (z.B. bei Affektiven Störungen: Angststörungen etc.)
		deutliche Selektionseffekte aufgrund der Ausschlusskriterien	(3)	z.B. Ausschluss zahlreicher / aller komorbider Störungen

**Eingangsdiagnostik:**

A.2.	Objektive und reliable Diagnosestellung (mittels (teil-) standardisierter Interviews)	Diagnosestellung mittels strukturiertem klinischen bzw. voll standardisiertem Interview (z.B. SKID, DIPS)	(1)	z.B. SKID, DIPS, DIA-X, CIDI etc.
		Diagnosestellung mittels Diagnosechecklisten oder nachvollziehbarem klinischen Urteil	(2)	z.B. Diagnosechecklisten (ICD- oder DSM-Diagnosechecklisten) oder globales klinisches Urteil nach erfolgter Anamnese (meist in Arztbrief festgehalten)
		keine adäquate Diagnosestellung ( <b>Stufe 3 = Ausschlusskriterium</b> )	(3)	Keine Angaben zur Diagnosestellung
	Diagnostische Instrumente:			Hier sind <u>keine</u> zusätzlichen Erhebungsinstrumente gemeint, die im Rahmen des Ein-Ausschlusskatalogs eingesetzt werden (z.B. BDI)

Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

B.2.	Erhebung der spezifizierten Einschluss- und Ausschlusskriterien mittels valider Methoden	Die Ein- und Ausschlusskriterien sind sämtlich klar operationalisiert und werden mittels valider Methoden erfasst (z. B. komorbide Störungen als Ausschluss werden mittels strukturiertem klinischen Interview erfasst; Ausschlusskriterien beziehen sich auf eindeutig objektivierbare Merkmale wie Alter, Geschlecht etc.)	(1)	<p>Komorbide Störungen mittels klinischer Interviews, wie z.B. SKID, DIPS, DIA-X, CIDI etc.</p> <p>z.B. Schweregradeinstufung diagnostizierter Störungen mittels dafür entwickelter mindestens „zufriedenstellend“ <u>valider</u> Inventare (z.B. BDI)</p> <p><i>oder</i></p> <p>nur objektivierbare Merkmale</p>
		Die Validität der Erhebungen von Teilen der Ein- oder Ausschlusskriterien ist teilweise eingeschränkt (z. B. komorbide Störungen als Ausschluss werden mittels globalem klinischen Urteil eingeschätzt) und wirkt sich jedoch nur in geringem Umfang auf die Zusammensetzung der Stichprobe aus	(2)	<p>z.B. globales klinisches Urteil oder Diagnosechecklisten nach anamnestischem Interview (ICD- oder DSM-Diagnosechecklisten) (meist in Arztbrief festgehalten)</p> <p>z.B. Schweregradeinstufung diagnostizierter Störungen mittels dafür entwickelter lediglich „ausreichend“ <u>valider</u> Inventare</p>
		Die Validität der Erhebungen von Teilen der Ein- oder Ausschlusskriterien ist deutlich eingeschränkt und wirkt sich differenziell auf die Zusammensetzung der Behandlungsgruppen aus	(3)	<p>keine adäquate Diagnosestellung (Diagnosestellung nicht nachvollziehbar)</p> <p>z.B. Schweregradeinstufung diagnostizierter Störungen mittels dafür entwickelter „unzureichend“ <u>valider</u> Inventare</p> <p><i>oder</i></p> <p>Wenn B.1.=3 → 3</p>

Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

B.9.	Vergleichbarkeit der Gruppen zur Baseline im Hinblick auf prognostisch relevante Merkmale	Weder statistisch noch klinisch relevante Unterschiede zwischen den Gruppen hinsichtlich prognostisch relevanter oder potentiell konfundierender Variablen	(1)	-
	Eingruppendesigns (inkl. Designs mit verfahrensinternen Vergleichsgruppen): „9“	Vergleichbarkeit hinsichtlich der meisten prognostisch relevanten Merkmale ist weitgehend gegeben; signifikante Unterschiede hinsichtlich relevanter prognostischer Merkmale zwischen den Gruppen werden statistisch angemessen kontrolliert	(2)	-
		Keine angemessene Überprüfung der Vergleichbarkeit oder Vergleichsgruppen unterscheiden sich erheblich hinsichtlich mehrerer prognostisch relevanter Merkmale und eine angemessene statistische Kontrolle des Einflusses dieser Merkmale fehlt	(3)	-
	Hinsichtlich Prognostizität untersuchte Merkmale:			-

Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

**Intervention:**

B.3.	Operationale Definition der Interventionen (Experimental- und ggf. Kontrollgruppe)	Therapiemanual, bei dem die Interventionen so beschrieben sind, dass das therapeutische Vorgehen vergleichbar und replizierbar ist <sup>6</sup>	(1)	Verweis auf Manuale/ manualähnliche Behandlungsrichtlinien, das/die in den jeweiligen Treatments auch offensichtlich eingesetzt wurden <sup>6</sup>
		Therapiebeschreibung, ohne nähere Spezifikation der einzelnen Interventionen (z.B. Fehlen gegebenenfalls für das psychotherapeutische Verfahren oder die psychotherapeutische Methode notwendiger Entscheidungskriterien)	(2)	<u>Kompakte Beschreibung</u> der Treatments im Rahmen der Studienpublikation <i>oder</i> Wenn Manual/ manualähnliche Behandlungsrichtlinie nur bsp.haft angeführt wird
		Die Intervention ist nicht klar beschrieben, beschränkt sich auf die Benennung des Psychotherapieverfahrens bzw. der Psychotherapiemethode	(3)	Lediglich Benennung der Behandlung(en), etwa „psychodynamische Therapie“ plus ggf. verfahrenstypischer Aspekte wie „Arbeit mit unbewussten Konflikten“ o.ä. <i>oder</i> Wenn kein bekannter Manual-/Behandlungsrichtlinientitel, dann danach recherchieren. Ggf. nachsehen ob Intervention in Artikel beschrieben → 2 (sonst 3)
Welches Manual bzw. welche manualähnlichen Behandlungsrichtlinien wurden verwendet?		-		
B.6.	Manualtreue, Treatment Integrity	Manualtreue/ Treatmentintegrität durch externe Beobachter (z.B. videogestützt) belegt	(1)	Hier gilt die Kontrolle der Manualtreue/Treatmentintegrität auch für die ggf. realisierte Vergleichsbehandlung (komparative Studien) und ebenfalls für Studien <u>ohne</u> Manual
		Manualtreue / Treatmentintegrität durch Fragebögen belegt	(2)	
		keine Maßnahmen zum Monitoring der Manualtreue oder Hinweise auf substantielle Abweichungen	(3)	

<sup>6</sup> Bzgl. psychoanalytisch begründeter Behandlungen vgl. Beutel, Doering, Leichsenring & Reich (2010); DGPT (2009); Reimer & Rüger (2006).



Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

C.4.	Klinische Repräsentativität der Intervention hinsichtlich Vorgehen und Dauer	Intervention wie in klinischer Alltagspraxis	(1)	<p><b>Bewertung entsprechend Psychotherapierichtlinien primär die <u>Therapiedauer</u> betreffend. Vorsicht bei Bewertung des <u>Vorgehens</u>, da selbst innerhalb Richtlinienverfahren uneinheitlich vorgegangen wird.</b></p> <p>Hinsichtlich Dauer wie in Alltagspraxis des untersuchten Verfahrens (z.B. TP: mindestens 25 Sitzungen) Wenn nur <u>Range</u> der Dauer abweicht (Hinweis auf Ausreißer) → 1</p>
		Intervention gegenüber klinischer Alltagspraxis teilweise verändert	(2)	Leichte Abweichungen hinsichtlich Dauer des untersuchten Verfahrens, z.B. wenn <u>SD</u> der Dauer abweicht
		Intervention gegenüber klinischer Alltagspraxis stark verändert	(3)	Starke Abweichungen hinsichtlich Dauer des untersuchten Verfahrens (z.B. TP: 8 Sitzungen), bzw. wenn <u>mittlere</u> Dauer von max. Richtlinien-dauer (100 TP, 300 AP) nach oben hin abweicht
C.5.	Art des Therapie-Monitorings (Einfluss auf Therapeutenverhalten)	keine Veränderung des Therapeutenverhaltens durch Therapie-Monitoring (z.B. durch Therapie-Supervision; Ausnahmen: Audio- oder Video-Aufzeichnungen ohne Feedback an Therapeuten, nur zur späteren Auswertung sind erlaubt)	(1)	<p>Therapiesupervision, die explizit zum Zwecke der Studie durchgeführt wird, wird als das Therapeutenverhalten beeinflussend betrachtet → (2) oder (3)</p> <p>Keinen Einfluss: - reguläre Supervision/Intervision - Audio-/Videoaufzeichnungen ohne Feedback an Therapeuten</p>
		mittlere Veränderung des Therapeutenverhaltens durch Therapie-Monitoring (gelegentliche Rückmeldungen an Therapeuten)	(2)	gelegentliche, unregelmäßige Rückmeldungen bei Bedarf
		starke Veränderung des Therapeutenverhaltens durch Therapie-Monitoring (durch kontinuierliche Rückmeldungen)	(3)	kontinuierliche, regelmäßige Rückmeldung
B.4.	<p>Operationale Definition der Kontrollbedingungen</p> <ul style="list-style-type: none"> <li>• Eingruppendesigns (inkl. Designs mit verfahren-internen Vergleichsgruppen): „9“</li> <li>• Echte VG: „8“</li> </ul>	Prospektive Festlegung und umfassende Beschreibung der Kontrollbedingung	(1)	<p><b>Hier sind explizit Kontrollbedingungen gemeint!</b></p> <p>Prospektive Festlegung <u>und</u> umfassende Beschreibung</p>
		Ex post facto Beschreibung der Kontrollbedingungen	(2)	Ex post facto Beschreibung (Beschreibung im Rahmen eines retrospektiven Designs)
		keine Beschreibung der Kontrollbedingung	(3)	keine Beschreibung

Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

B.5.	Strukturelle Äquivalenz bei Kontrollbedingungen • Eingruppendesigns (inkl. Designs mit verfahren-internen Vergleichsgruppen): „9“ • Echte VG: „8“	hinsichtlich des Umfangs an therapeutischer Zuwendung und der Settingbedingungen in der KG besteht Äquivalenz	(1)	Hier sind explizit Kontrollbehandlungen gemeint!
		der Umfang der therapeutischen Zuwendung in der KG ist reduziert, die Settingbedingungen weichen von der IG ab	(2)	
		der Umfang der therapeutischen Zuwendung in der KG ist deutlich reduziert, die Settingbedingungen weichen wesentlich von der IG ab	(3)	
B.7.	Zulässigkeit, Dokumentation und Analyse des Einflusses begleitender nicht-randomisierter Interventionen (insbesondere Pharmakotherapie) Eingruppendesigns (inkl. Designs mit verfahren-internen Vergleichsgruppen): „9“	Ausschluss begleitender nicht-randomisierter Interventionen	(1)	Eliminierung der Störvariablen
		begleitende nicht-randomisierte Interventionen sind zulässig, werden jedoch detailliert dokumentiert und die Analysen weisen auf keinen substanziellen, differenziellen Einfluss der begleitenden Interventionen hin	(2)	Dokumentation und Analyse und ggf. statistische Kontrollen der begleitenden nicht-randomisierten Interventionen
		begleitende nicht-randomisierte Interventionen sind zulässig, werden jedoch nicht dokumentiert oder die Analysen weisen auf eine differenzielle Inanspruchnahme von begleitenden Interventionen und deren Einfluss auf das Behandlungsergebnis hin	(3)	Keine Dokumentation <i>oder</i> Dokumentation und Analyse weisen auf differenziellen Effekt der nicht-randomisierten Begleitintervention hin und es wird keine statistische Kontrolle durchgeführt
C.6.	Zulässigkeit begleitender Interventionen (z. B. Pharmakotherapie)	keine Einschränkungen	(1)	-
		begleitende in der Routine-Praxis übliche Interventionen teilweise ausgeschlossen	(2)	-
		alle begleitenden in der Routine-Praxis üblichen Interventionen ausgeschlossen	(3)	-
C.7a.	Qualifikation der Behandler: Klinische Tätigkeit der Therapeuten	Therapeuten sind praktizierende Kliniker	(1)	Behandler: mit abgeschlossener Psychotherapieausbildung und (hauptberuflich) praktizierend.
		klinische Forscher, die überwiegend Forschung betreiben und seltener auch Patienten behandeln; (Ausbildungskandidaten)	(2)	Behandler: in Psychotherapieausbildung <i>oder</i> mit abgeschlossener Psychotherapieausbildung, jedoch hauptsächlich in Forschung statt Klinik tätig
		keine Kliniker oder Kliniker, die keine Psychotherapeuten sind	(3)	Behandler: ohne Psychotherapieausbildung

Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

C.7b.	Qualifikation der Behandler: Breite der Klinischen Tätigkeit der Therapeuten (Problemheterogenität)	Therapeut behandelt Patienten mit verschiedenen Problemen innerhalb und außerhalb der Studie	(1)	Tätigkeit in breitem Anwendungsbereich, keine Tätigkeit in nur einem abgegrenzten Bereich (Sucht, Schmerz, Essstörungen)
		Therapeut behandelt überwiegend Patienten mit ähnlichen Problemen innerhalb und außerhalb der Studie	(2)	Tätigkeit hauptsächlich in einem begrenzten Anwendungsbereich, jedoch durchaus mehrere Störungsgruppen
		Therapeut behandelt nur Patienten mit ähnlichen Problemen innerhalb und außerhalb der Studie (z.B. Therapeut, der nur Schmerzpatienten in einer Schmerzambulanz behandelt; Psychotherapeuten, die außerhalb der Studie keine Pat. behandeln)	(3)	Tätigkeit in abgegrenztem Anwendungsbereich (Sucht, Schmerz, Essstörungen) oder keine therapeutische Tätigkeit außerhalb der Studie
C.7c.	Qualifikation der Behandler: Spezifisches Training der Psychotherapeuten in einer Behandlungsmethode für die Studie	kein spezifisches Training für die Studie (z.B. Therapeuten wenden die von ihnen üblicherweise angewendete Therapie an)	(1)	-
		kurzes Training für die Studie / intensives Training nur einiger Therapeuten	(2)	-
		intensives Training vor der Studie	(3)	-

**Outcomes/Outcomediagnostik:**

A.8.	Reliable und valide Messung zumindest der <b>primären</b> Zielkriterien	reliable und valide Outcome-Verfahren	(1)	(in Primärstudie werden ausschließlich Outcome-Verfahren verwendet, die mit „ <b>exzellent</b> “, „ <b>gut</b> “ oder „ <b>zufriedenstellend</b> “ bewertet wurden)
		nur eingeschränkte Reliabilität und/oder Validität der Messverfahren	(2)	In Primärstudie werden - u.a. Outcome-Verfahren verwendet, die als „ <b>unzureichend</b> “ bewertet wurden, die jedoch einen Anteil von 25% der Gesamtanzahl verwendeter Outcomemaße nicht übersteigen <i>oder</i> - Outcome-Verfahren, von denen mindestens eines als „ <b>ausreichend</b> “ bewertet wurde
		Reliabilität und Validität der Messverfahren nicht überprüft oder Gütekriterien der Messverfahren sind unzureichend <b>(Stufe 3 = Ausschlusskriterium)</b>	(3)	(in Primärstudie werden mehr als 25% Outcome-Verfahren verwendet, die als „ <b>unzureichend</b> “ bewertet wurden)

Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

(weiter) A.8.	Wurde die Reliabilität (i.d.R. die Interrater-Reliabilität) in der Studie selbst untersucht und dargestellt?		ja	-
			nein	-
	Verändern die in der Studie berichtete(n) Reliabilität(en) die Gesamtbewertung eines/mehrerer Inventars/Inventare? Wenn ja, verbessernd oder verschlechternd?		ja	-
			nein	-
A.10.	Multiple Informationsquellen (z.B. Patient, Therapeut, Laborwerte)	multidimensionale Erfassung der Zielkriterien – drei oder mehr Informationsquellen	(1)	<b>Auf primäre und sekundäre Zielkriterien bezogen</b> z.B. Selbstbeurteilung Patient, Fremdbeurteilung Therapeut, Fremdbeurteilung durch externe Beurteiler (Kliniker oder Angehörige)
		zwei Informationsquellen	(2)	-
		eine Informationsquelle	(3)	-
A.11.	Sofern Fremdeinschätzungsverfahren: externe Beurteiler (blind für die Gruppenzugehörigkeit) <ul style="list-style-type: none"> <li>• Wenn keine Fremdeinschätzungsverfahren durch externe Beurteiler eingesetzt: „9“</li> <li>• Eingruppendesigns (inkl. Designs mit verfahrensinternen Vergleichsgruppen): mit externen Beurteilern: „4“</li> </ul>	validiertes Fremdeinschätzungsverfahren angewendet von trainierten, für die Gruppenbedingungen blinden externen Beurteilern	(1)	<b>Auf primäre und sekundäre Zielkriterien bezogen</b> Validiert, trainiert und blind
		validiertes Fremdeinschätzungsverfahren angewendet von trainierten, nicht-blinden externen Beurteilern	(2)	Validiert, trainiert aber <u>nicht-blind</u>
		validiertes Fremdeinschätzungsverfahren angewendet – Rater sind weder trainiert noch blind für die Gruppenzugehörigkeit der Patienten	(3)	Validiert aber <u>nicht-trainiert</u> und <u>nicht-blind</u> <i>oder</i> <u>nicht-valide</u> entsprechend A.8.

Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

C.9.	Primäre Zielkriterien beziehen sich auf patientenrelevante Parameter (insbesondere Schwere der Symptomatik, Leiden, Beeinträchtigung/Lebensqualität, Inanspruchnahme von Diensten des Gesundheitswesens)	Zielkriterien beziehen sich auf mehrere Dimensionen patienten- bzw. störungsrelevanter Parameter unter Einbezug von Beeinträchtigung/Lebensqualität und Inanspruchnahme von Diensten des Gesundheitswesens)	(1)	z.B. symptom-, persönlichkeits-, strukturbezogene Outcomemaße, Outcomemaße zur Lebensqualität, Inanspruchnahmedaten
		Zielkriterien beziehen sich nur auf eine Dimension	(2)	z.B. symptom-, persönlichkeits-, strukturbezogene Outcomemaße, aber weder Outcomemaße zur Lebensqualität noch der Inanspruchnahme
		Zielkriterien beziehen sich ausschließlich auf Surrogatparameter (z.B. Kontrollüberzeugung) <b>(Stufe 3 = Ausschlusskriterium)</b>	(3)	-
B.12.	Erzielte Veränderungen auf den primären und sekundären Zielkriterien ggf. im Vergleich zur Kontrollgruppe (Signifikanz, Größe und Relevanz der Effekte <sup>7</sup> )	vollständige Darstellung der erzielten Veränderungen auf den Zielkriterien inklusive der Signifikanz, Größe der Effektmaße und Ausmaß der klinisch relevanten Zielerreichung (ggf. im Vergleich zur Kontrollgruppe)	(1)	Sowohl Indikatoren der <i>Veränderungen</i> als auch der <i>Zielerreichungen</i> sind in Bezug auf die verwendeten Outcomemaße dargestellt  <b>Achtung:</b> <i>Zielerreichung</i> weniger streng bewerten, wenn Cut-off angewandt
		Darstellung des Behandlungsergebnisses nur durch Veränderungs- oder Zielerreichungsmaße oder beides ist (ggf. im Vergleich zur Kontrollgruppe) bei einigen Kriterien unvollständig	(2)	Es sind nur Indikatoren der <i>Veränderungen</i> oder nur der <i>Zielerreichungen</i> in Bezug auf die verwendeten Outcomemaße dargestellt
		weitgehend unvollständige oder inadäquate Darstellung der Outcome-Kriterien (ggf. im Vergleich zur KG) <b>(Stufe 3 = Ausschlusskriterium)</b>	(3)	-

<sup>7</sup> *Veränderungen*: Neben Prä-Post-Differenzen:

- Prozentuale Angabe der Probanden, die entsprechend dem „Reliable change index“ eine signifikante Veränderung erreicht haben
- Effektstärkeberechnungen (als Indikator des Ausmaßes der Veränderung)

*Zielerreichung*:

- Klinische Relevanz: Reliable change index plus Cut-Off-Wert (Vergleich mit Normpopulation oder dysfunktionaler Population)

(vgl. Schulte, 1993; Möller, Laux & Kapfhammer, 2005).

Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

A.9.	Klinische Bedeutsamkeit der Outcome-Messung (z.B. das Konzept der klinischen Signifikanz <sup>8</sup> )	klinische Bedeutsamkeit des Therapieeffekts (z. B. im Sinne des Konzepts der klinischen Signifikanz) ist feststellbar	(1)	<ul style="list-style-type: none"> <li>- Normativer Vergleich</li> <li>- RCI + Cutoff-Wert</li> <li>- Soziale Validierung/Subjektive Evaluation</li> <li>- Diagnosefreiheit</li> </ul> <p><b>Achtung:</b> Hier klinische Bedeutsamkeit strenger bewerten: Wenn Cutoff <u>und</u> RCI → 1, wenn nur Cutoff <u>oder</u> RCI → 2</p>
		klinische Bedeutsamkeit des Therapieeffekts (z. B. im Sinne des Konzepts der klinischen Signifikanz) ist nur eingeschränkt feststellbar	(2)	<ul style="list-style-type: none"> <li>- Normativer Vergleich nur mittels klassischer Signifikanztests</li> <li>- nur RCI oder nur Cutoff-Wert</li> </ul>
		klinische Bedeutsamkeit des (z. B. im Sinne des Konzepts der klinischen Signifikanz) Therapieeffekts ist nicht feststellbar	(3)	Die klinische Bedeutsamkeit wurde über keine der Outcomemaße berichtet
A.12.	Vollständige Darstellung der Ergebnisse zu allen Outcomemaßen und zu allen relevanten Messzeitpunkten	für alle Outcome-Variablen berichtet	(1)	Diejenigen Outcomemaße, die laut Publikation zu spezifizierten Messzeitpunkten (i.d.R. Prä-Post-Katamnese) erhoben werden, sollten dementsprechend auch in der Ergebnisdarstellung auftauchen
		ausschließlich für die primären Zielkriterien berichtet	(2)	Es werden nur die Ergebnisse der <u>primären</u> Outcomemaße zu den spezifizierten Messzeitpunkten (i.d.R. Prä-Post-Katamnese) berichtet
		nicht für alle primären Zielkriterien berichtet	(3)	Die Ergebnisdarstellung bezieht sich <u>nicht mal auf alle primären</u> Outcomemaße

<sup>8</sup> Als Erhebung der klinischen Relevanz werden i.d.R. folgende Strategien betrachtet:

1. Normativer Vergleich mittels klassischer und Äquivalenzhypthesentests (vgl. Kendall, Marrs-Garcia, Nath & Sheldrick, 1999)
2. Individuelle Veränderungen (z.B. „Reliable change index“ als „statistically reliable improvement“ [Verbesserung]; „Cut-Off-Wert“ als „clinically significance change“ [Genesung]) (vgl. Jacobson & Truax, 1991; Stieglitz, 2008)
3. Soziale Validierung/Subjektive Evaluation durch Patienten selbst, Verwandte/Freunde, Fachkräfte/externe Beobachter (vgl. Kazdin, 1994; Lambert & Ogles, 2004)
4. Kein weiteres Erfüllen der Kriterien einer psychiatrischen Diagnose (vgl. Kazdin, 2008).

Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

A.13.	Erfassung unerwünschter Wirkungen	Systematische Erfassung und Bericht von unerwünschten Ereignissen, Nebenwirkungen, Verschlechterungen	(1)	Wird bspw. der %-Anteil der Verschlechterungen, Nebenwirkungen etc. berechnet, dann sollte diese Patientengruppe näher spezifiziert werden
		unsystematische Erfassung und/oder unvollständiger Bericht von unerwünschten Ereignissen, Nebenwirkungen, Verschlechterungen	(2)	Es wird bspw. nur der %-Anteil der Verschlechterungen, Nebenwirkungen etc. angegeben, ohne nähere Spezifizierung der Patientengruppe
		Unerwünschte Ereignisse, Nebenwirkungen, Verschlechterungen wurden nicht erfasst oder nicht berichtet	(3)	-

**Drop-out-Analysen:**

A.3.	Höhe der Drop-out-Quote zu Behandlungsende (sofern nicht Erfolgskriterium)	i.d.R. Drop-out-Quote kleiner 20 %	(1)	Drop-outs werden hier als <u>Studien- und/oder Therapie-Drop-outs</u> definiert: Patienten, die die Studie abbrechen (und ggf. die Therapie weitermachen) sowie Patienten, die die Therapie abbrechen (und ggf. für Studienzwecke weiter zur Verfügung stehen) werden als Drop-outs gezählt  <b>Achtung:</b> Dropouts über <u>alle</u> Treatmentarme (EG, KG) berechnen  Berechnung nach Flow Diagramm: $1 - \frac{\text{Interventions/Studiencompleter}}{\text{Zugewiesen zur Intervention}}$  Wenn keine Dropouts berichtet, ohne, dass aus Publikation hervorgeht, dass es tatsächlich keine gab → 3
		i.d.R. Drop-out-Quote zwischen 20 % und 40 %	(2)	
		i.d.R. Drop-out-Quote größer 40 %	(3)	
		Vor allem in Bezug auf Studien <u>ohne</u> Randomisierung: Zeitpunkt, ab dem Dropouts gezählt werden (Erstgespräch, probatorische Sitzungen, Therapiebeginn o.ä.):		

Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

Wenn bei Prä-Katamnese Studien oder reinen Katamnese Studien Post-Dropouts und Katamnese-Dropouts nicht separat zu eruieren sind, dann bei **A.3** alle Dropouts zwischen Prä und Katamnese kodieren mit folgenden Richtwerten:

<b>Affektiv</b> < 2 Jahre ≥ 2 Jahre	≤ 28% ≤ 32 %	<b>1</b>
<b>Affektiv</b> < 2 Jahre ≥ 2 Jahre	29% - 52% 33% - 58%	<b>2</b>
<b>Affektiv</b> < 2 Jahre ≥ 2 Jahre	> 52% > 58%	<b>3</b>

<b>Gemischt</b> < 2 Jahre ≥ 2 Jahre	≤ 36% ≤ 40%	<b>1</b>
<b>Gemischt</b> < 2 Jahre ≥ 2 Jahre	37% - 64% 41% - 70%	<b>2</b>
<b>Gemischt</b> < 2 Jahre ≥ 2 Jahre	> 64% > 70%	<b>3</b>

A.4.	Höhe der Studien-Drop-outs zur Katamnese (falls Katamneseerhebung durchgeführt) (zwischen Post- und Katamnesezeitpunkt)	<ul style="list-style-type: none"> <li>• Wenn keine Katamnese: „9“</li> <li>• Prä-Katamnese: „4“</li> <li>• Nur Katamnese: „5“</li> </ul>	deutlich besser als in Studien mit vergleichbaren Patientengruppen und vergleichbarem Katamnesezeitraum		<b>Achtung:</b> Drop-outs über <u>alle</u> Treatmentarme berechnen  Berechnung nach Flow Diagramm:  $1 - \frac{\text{Katamnesedaten vorliegend von Interventions/Studiencompletern zur Postmessung}}{\text{Interventions/Studiencompleter zur Postmessung}}$	
			Affektive Störungen	Gemischte Störungsgruppen		(1)
			weniger als 2 Jahre: ≤ 10%	weniger als 2 Jahre: ≤ 20%		
			2 Jahre oder mehr: ≤ 15%	2 Jahre oder mehr: ≤ 25%		(2)
			Drop-out-Quote vergleichbar mit Studien mit entsprechenden Patientengruppen und entsprechendem Katamnesezeitraum			
			Affektive Störungen	Gemischte Störungsgruppen		(3)
			weniger als 2 Jahre: 11-20%	weniger als 2 Jahre: 21-40%		
			2 Jahre oder mehr: 16-30%	2 Jahre oder mehr: 26-50%		
			deutlich schlechter als in Studien mit vergleichbaren Patientengruppen und vergleichbarem Katamnesezeitraum			(3)
			Affektive Störungen	Gemischte Störungsgruppen		
weniger als 2 Jahre: > 20%	weniger als 2 Jahre: > 40%					
2 Jahre oder mehr: > 30%	2 Jahre oder mehr: > 50%		Wenn keine Drop-outs berichtet, ohne, dass aus Publikation hervorgeht, dass es tatsächlich keine gab → 3			



Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

A.16.	Intention to treat-Analysen durchgeführt  Falls keine Dropouts: „9“	ITT-Analysen durchgeführt	(1)	-
		Keine ITT-Analysen bei geringem Risiko für einen attrition bias	(2)	Wenn nahezu ausschließlich Drop-outs vorliegen, von denen mit hoher Wahrscheinlichkeit <u>keine</u> Verzerrungen des Behandlungsergebnisses zu erwarten sind [z.B. Wohnortwechsel oder Versterben des Patienten; Ausfall des behandelnden Therapeuten <sup>9</sup> ]
		Keine ITT-Analysen bei deutlichem Risiko für einen attrition bias	(3)	Wenn nahezu ausschließlich Drop-outs vorliegen, von denen mit hoher Wahrscheinlichkeit Verzerrungen des Behandlungsergebnisses zu erwarten sind [z.B. Therapieabbruch seitens Patienten aus therapiebezogenen Gründen; Abbruch seitens Therapeuten wg. patientenseitigen Motivationsmangels; Therapiewechsel; zu geringer Therapieerfolg; <u>ohne Angabe von Gründen</u> ]  Wenn keine Dropouts berichtet, ohne, dass aus Publikation hervorgeht, dass es tatsächlich keine gab → 3
A.18.	Vollständige Beschreibung der Drop-Outs  Falls keine Dropouts: „9“	vollständige Beschreibung aller Drop-outs, inkl. der Gründe und des Zeitpunkts des Drop-outs	(1)	Beschreibung von zentralen Patientenmerkmalen und derjenigen Kriterien unter B.9. <u>und Gründe und Zeitpunkt</u>
		unvollständige Angaben zu Gründen oder Zeitpunkten des Drop-outs	(2)	Unvollständige <u>Beschreibung und/oder</u> unvollständige Angabe von <u>Gründen und/oder</u> unvollständige Angabe der <u>Zeitpunkte</u>
		keine Beschreibung der Drop-outs oder definitive Angabe der Anzahl der Drop-outs in einer Gruppe	(3)	Keine Beschreibung <u>und</u> keine Angabe der Gründe <u>und</u> Zeitpunkte <u>oder</u> Keine Angabe der Anzahl der Drop-outs pro Gruppe  Wenn keine Drop-outs berichtet, ohne, dass aus Publikation hervorgeht, dass es tatsächlich keine gab → 3

<sup>9</sup> Vgl. Hiller, Bleichhardt & Schindler (2009).

Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

A.19.	Drop-out-Analysen  Falls keine Dropouts: „9“	Drop-out-Analysen unter Berücksichtigung der wichtigsten prognostischen Faktoren durchgeführt; keine signifikanten Unterschiede zwischen den Vergleichsgruppen	(1)	<b>Wenn Drop-out-Quote sehr gering, dann Rating weniger streng durchführen!</b> Keine signifikanten Unterschiede zwischen den Drop-outs der Untersuchungsgruppen hinsichtlich prognostisch relevanter Faktoren oder ggf. Korrektur
		Drop-out-Analysen unter Berücksichtigung wichtigster prognostischer Faktoren durchgeführt; trotz sign. Unterschiede ist die Validität der Ergebnisse nicht wesentlich eingeschränkt	(2)	Signifikante Unterschiede zwischen den Drop-outs der Treatmentgruppen hinsichtlich prognostisch relevanter Faktoren und Einschätzung in Studie, dass <u>nicht</u> validitätsgefährdend bzgl. Ergebnisse
		Drop-out-Analysen nicht oder unter Vernachlässigung relevanter prognostischer Merkmale durchgeführt; Drop-out-Analysen stellen die Validität der Ergebnisse deutlich in Frage	(3)	Drop-out-Analysen hinsichtlich prognostisch <u>irrelevanter</u> Faktoren durchgeführt  <i>oder</i> Keine Drop-out-Analysen durchgeführt  Wenn keine Drop-outs berichtet, ohne, dass aus Publikation hervorgeht, dass es tatsächlich keine gab → 3

**Statistische Methodik:**

A.14.	Anwendungsvoraussetzungen für statistische Modelle geprüft und erfüllt	Anwendungsvoraussetzungen geprüft und erfüllt	(1)	<ul style="list-style-type: none"> <li>• Wenn z.B. bei Anova mit Messwiederholung: Sphärizität mittels Mauchly-Test und ggf. Korrektur (Greenhouse-Geisser-Korrektur)</li> <li>• Keine Überprüfung von z.B. Normalverteilungsvoraussetzung bei hinreichend großem N pro Bedingung (<b>N&gt;30 bei t-Tests, N&gt;25 bei ANOVA</b>) oder keine Überprüfung der Homoskedastizität bei gleichgroßen Stichproben</li> <li>• Voraussetzungen (v.a. bei kleinen Stichproben) überprüft und keine Verletzungen</li> </ul>
		Anwendungsvoraussetzungen geprüft und lediglich leichte Verletzungen der Voraussetzungen oder keine Prüfung der Anwendungsvoraussetzungen bei gleichzeitiger Robustheit der angewendeten statistischen Verfahren	(2)	Voraussetzungen (v.a. bei kleinen Stichproben) überprüft und leichte Verletzungen
		deutliche Verletzungen der Anwendungsvoraussetzungen oder keine Prüfung der Anwendungsvoraussetzungen bei substantiellem Risiko für deren Verletzung	(3)	<ul style="list-style-type: none"> <li>• Eindeutige Verletzung der Sphärizität ohne Korrektur sowie sonstiger Voraussetzungen (NV, Homoskedastizität)</li> <li>• Keine Überprüfung der Sphärizitätsannahme</li> <li>• Keine Überprüfung der Voraussetzungen bei „problematischen“ (kleinen, ungleich großen) Stichproben</li> </ul>
A.15.	Angemessenheit der statistischen Analysen (inklusive der Korrektur für multiple Tests)	adäquate und umfassende statistische Analysen	(1)	ggf. $\alpha$ -Fehlerkorrektur
		weitgehend adäquate statistische Analysen	(2)	-
		unangemessene statistische Analysen (fehlende Korrektur für multiple Tests, inadäquate statistische Methoden)	(3)	<ul style="list-style-type: none"> <li>- keine <math>\alpha</math>-Fehlerkorrektur, obwohl notwendig</li> <li>- Anwendung von parametrischen Verfahren, wenn nonparametrische indiziert</li> </ul>

Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

A.17.	Statistische Power der Vergleiche bei Vergleich mit bewährter Treatment-Gruppe	adäquate statistische Power der Vergleiche	(1)	<b>Achtung: Dieses Kriterium bezieht sich allein auf verfahrensexterne Vergleiche mit bereits bewährten Behandlungen</b>  Als Richtwert für einen t-Test für unabhängige Stichproben: N (pro Gruppe) >310 bzw. N (gesamt) >620 <sup>10</sup>
	Bei Eingruppendesigns (inkl. Designs mit verfahrensisernen Vergleichsgruppen), Kontrollgruppendesigns (Warteliste, Placebo, TAU, aktive Kontrolle): „9“	eingeschränkte Power der statistischen Vergleiche (.50-.80)	(2)	Als Richtwert für einen t-Test für unabhängige Stichproben: N (pro Gruppe): 136 bis 310 bzw. N (gesamt) 272 bis 620
		unzureichende statistische Power der Vergleiche (<.50)	(3)	Als Richtwert für einen t-Test für unabhängige Stichproben: N (pro Gruppe) <136 bzw. N (gesamt) <272
	Wurde die Signifikanzüberprüfung mittels eines Intervall-Nullhypothesentests (einseitiger Äquivalenztest/Non-Inferiority) oder mittels eines nil-Nullhypothesentests durchgeführt?			nil-Nullhypothesentest: $H_0: \mu_A = \mu_B$ vs. $H_1: \mu_A \neq \mu_B$  Non-Inferiority: $H_0: \mu_A - \mu_B \geq \delta$ vs. $H_1: \mu_A - \mu_B < \delta$

**Manipulation der Daten:**

A.1.	Manipulation der Daten	keine Hinweise auf Ergebnismanipulation	(1)	-
		-	(2)	-
		Hinweise auf Ergebnismanipulation <b>(Stufe 3 = Ausschlusskriterium)</b>	(3)	z.B. wenn in unterschiedlichen Publikationen zur selben Studie (selber Datensatz!) voneinander abweichende Ergebnisse berichtet werden

<sup>10</sup> Die kalkulierten Stichprobengrößen gelten für gleichgroße Stichproben. Für ungleichgroße Stichproben werden gesonderte Kalkulationen angestellt.

Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

### Subgruppenanalysen:

Wurden Subgruppenanalysen durchgeführt?

ja	
nein	

Welche?

Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

## Literatur:

Beutel, M. E., Doering, S., Leichsenring, F. & Reich, G. (2010). *Psychodynamische Psychotherapie. Störungsorientierung und Manualisierung in der therapeutischen Praxis*. Göttingen: Hogrefe.

Deutsche Gesellschaft für Psychoanalyse, Psychotherapie, Psychosomatik und Tiefenpsychologie (DGPT) e.V. (2009). *Stellungnahme zur Prüfung der Richtlinienverfahren gemäß §§ 13 – 15 der Psychotherapie-Richtlinie für die psychoanalytisch begründeten Verfahren*.

Verfügbar unter:

<http://www.dgpt.de/dokumente/DGPT%20Stellungnahme%20zur%20Pruefung%20oder%20Richtlinienverfahren%202009.pdf>. [30.08.2010].

Lambert, M. J. & Ogles, B. M. (2004). The Efficacy and Effectiveness of Psychotherapy. In A. E. Bergin & S. L. Garfield (Hrsg.), *Handbook of Psychotherapy and Behavior Change* (5. Aufl., S. 139–193). Oxford England: John Wiley & Sons.

Hiller, W., Bleichhardt, G. & Schindler, A. (2009). Evaluation von Psychotherapien aus der Perspektive von Qualitätssicherung und Qualitätsmanagement. *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie*, 57(1), 7–22.

Jacobson, N. S. & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59(1), 12–19.

Kazdin, A. E. (1994). Methodology, design, and evaluation in psychotherapy research. In: A. E. Bergin, S. L. Garfield (Hrsg.), *Handbook of psychotherapy and behavior change* (4. Aufl., S. 19–71). Oxford England: John Wiley & Sons.

Kazdin, A. E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist*, 63(3), 146–159.

Kendall, P. C., Marrs-Garcia, A., Nath, S. R. & Sheldrick, R. C. (1999). Normative comparisons for the evaluation of clinical significance. *Journal of Consulting and Clinical Psychology*, 67(3), 285–299.

Möller, H. J., Laux, G. & Kapfhammer, H. P. (Hrsg.). (2005). *Psychiatrie und Psychotherapie* (2. Aufl.). Heidelberg: Springer Medizin Verlag.

Reimer, C. & Rüger, U. (Hrsg.). (2006). *Psychodynamische Psychotherapien: Lehrbuch der tiefenpsychologisch fundierten Psychotherapieverfahren* (3. Aufl.). Heidelberg: Springer.

Schulte, D. (1993). Wie soll Therapieerfolg gemessen werden? *Zeitschrift für Klinische Psychologie*, 22(4), 374–393.

Stieglitz, R.-D. (2008). *Diagnostik und Klassifikation in der Psychiatrie* (1. Aufl.). Stuttgart: Kohlhammer.

Wissenschaftlicher Beirat Psychotherapie nach § 11 PsychThG. (2009). *Methodenpapier des Wissenschaftlichen Beirats Psychotherapie: Verfahrensregeln zur Beurteilung der wissenschaftlichen Anerkennung von Methoden und Verfahren der Psychotherapie - Version 2.7*.

Verfügbar unter: <http://www.wbpsychotherapie.de/downloads/Methodenpapier2720090709.pdf> [15.5.2010].

Wissenschaftlicher Beirat Psychotherapie nach § 11 PsychThG. (2010). *Methodenpapier des Wissenschaftlichen Beirats Psychotherapie: Verfahrensregeln zur Beurteilung der wissenschaftlichen Anerkennung von Methoden und Verfahren der Psychotherapie - Version 2.8*.

Verfügbar unter: <http://www.wbpsychotherapie.de/downloads/Methodenpapier28.pdf> [26.11.2010].

Kriteriengruppen: A: Allgemeine methodische Qualität – B: Interne Validität – C: Externe Validität

Freie Universität Berlin  
FB Erziehungswissenschaft und  
Psychologie  
AB Public Health: Prävention und psychosoziale  
Gesundheitsforschung  
Habelschwerdter Allee 45  
14195 Berlin

Projekt: Zur Wissenschaftlichkeit von  
Psychotherapie  
Ansprechpartnerin: Melanie Ratzek  
Telefon: 030-300 117-767  
E-Mail: [melanie.ratzek@fu-berlin.de](mailto:melanie.ratzek@fu-berlin.de)