



Landespsychotherapeutenkammer
Baden-Württemberg

Stellungnahme
der Landespsychotherapeutenkammer
Baden-Württemberg
zum
Methodenpapier des
Wissenschaftlichen Beirat Psychotherapie

„Verfahrensregeln des Wissenschaftlichen Beirats Psychotherapie zur
Beurteilung der wissenschaftlichen Anerkennung von Methoden und Ver-
fahren in der Psychotherapie“

Version 2.0 vom 8.12.2006

R. Nübling, D. Munz & M. Klett

Stuttgart, 28. Februar 2007

Zusammenfassung

Die Landespsychotherapeutenkammer Baden-Württemberg begrüßt den Vorschlag des Wissenschaftlichen Beirats Psychotherapie (WBP), den Prozess der wissenschaftlichen Anerkennung psychotherapeutischer Verfahren und Methoden durch Verfahrensregeln festzulegen und transparent zu machen. Der im Entwurf vom 8.12.2006, Version 2.0 enthaltene Versuch, die externe Validität von Studien gegenüber der internen Validität stärker zu gewichten, wird ausdrücklich unterstützt. Die darin indirekt zum Ausdruck gebrachte Abweichung von der in §18, Abs. 3 der Verfahrensordnung des Gemeinsamen Bundesausschusses (G-BA, 2005) festgelegten Evidenzstufen, in denen Randomized Controlled Trials (RCT-Studien) als beste Bewertungsgrundlage für die Zulassung von Verfahren und Methoden definiert werden, findet volle Zustimmung.

Aus Sicht der LPK Baden-Württemberg geht der Vorschlag des WBP allerdings nicht weit genug. In der hier vorgelegten Version der Verfahrensregeln besitzen RCT-Studien nach wie vor eine höhere Gewichtung hinsichtlich Wissenschaftlichkeit als naturalistische Studien. Sie sind aus Sicht des WBP unverzichtbare Voraussetzung. Diese Haltung wird von der LPK Baden-Württemberg grundsätzlich in Frage gestellt. Vielmehr ist eine absolute Gleichstellung naturalistischer Studien mit RCT-Studien zu fordern und parallel Evidenzstufen für naturalistische Studien zu entwickeln bzw. zu definieren. Der experimentelle Nachweis der Wirksamkeit wird nicht als unverzichtbares Kriterium gesehen. Die vom G-BA in der Verfahrensordnung zugrundegelegten „levels of evidence“ werden als zu einseitig für eine Bewertung der Übertragbarkeit eines Verfahrens oder einer Methode auf die klinische Versorgung abgelehnt. Die Argumente hierzu werden im folgenden ausgeführt.

1 Grundfragen der Wissenschaftlichkeit von Psychotherapiestudien – Experimentelle vs. naturalistische Designs

Randomized Clinical Trails (RCTs) werden als „gold-standard“ der Therapieforschung und evidenz-basierten Vorgehens bezeichnet. Sie führen die „levels of evidence“ an, die beiden höchsten Stufen umfassen RCTs (level II) oder Metaanalysen von RCTs (level I). Diese aus der pharmakologischen Forschung bzw. der evidence based medicine vom G-BA übernommene Klassifikation der „Wertigkeit“ wissenschaftlicher Studien ist Grundlage für die Anerkennung und Zulassung therapeutischer Verfahren und Methoden. Dabei haben Verfahren und Methoden, die Studien der diesem Sinne definierten höheren Evidenz vorlegen können, eine größere Wahrscheinlichkeit, für die Routineversorgung des deutschen Gesundheitswesens zugelassen zu werden (vgl. auch IQWiG, 2006).

Ziel von RCTs ist die Herausarbeitung von Ursachen-Wirkungs-Zusammenhängen sowie die Erfassung „reiner“ Effekte (vgl. z. B. Rossi et al., 1999). Dies soll u.a. erreicht werden durch Zufallszuweisung zu Behandlungs- und Kontrollgruppen, manualisierte Behandlung, Verblindung sowie die Kontrolle so genannter Störgrößen. All diese Maßnahmen sollen dazu führen, die interne Validität von Studien zu steigern, also zur Beantwortung der Frage, ob gemessene Effekte tatsächlich der Behandlung zuzuschreiben oder ob sie eher auf unspezifische oder zufällige Variablen zurückzuführen sind. Eine ausreichende interne Validität ist ohne Zweifel eine wesentliche Voraussetzung für die Feststellung der Wissenschaftlichkeit von Studien.

Nicht weniger wichtig für die Bewertung der Wissenschaftlichkeit eines Therapieverfahrens oder einer Therapiemethode ist – und das wird (zwar nicht grundsätzlich im WBP-Methodenpapier, aber) in den von GB-A und IQWiG zugrundegelegten levels of evidence vernachlässigt – die externe Validität, also die Generalisierbarkeit der Befunde auf den klinischen Alltag. Die sich in den Evidenzstufen ausdrückende, in der Tendenz einseitige

Schwerpunktlegung auf die interne Validität von Studien wird zunehmend (wieder) als Überbewertung experimenteller Designs kritisiert (vgl. z.B. Porzsolt et al., 2006; Zurhorst, 2003; Leichsenring, 2004a, b; Tschuschke, 2005). Diese Kritik ist umso mehr beachtenswert, als sie zur Zeit auch aus der – dem Paradigma kontrollierter Studien stärker verpflichteten – Pharmaforschung kommt (vgl. Porzsolt et al. 2006). Porzsolt et al. (2006) halten dazu fest: „Es wäre äußerst ungewöhnlich, wenn diese Methode (RCTs, d. Verf.) nur Vorteile, aber nicht auch ebenso Nachteile hätte. Sorgfältige Überlegungen, ob eine Randomisation tatsächlich anderen Methoden überlegen ist, werden kaum angestellt. ... Beobachtungen legen nahe, dass auch randomisierte Studien eine „Kehrseite der Medaille“ haben, die wir aber aus Überzeugungsgründen selten analysieren“ (S. 359). Die Kritik an RCTs ist nicht neu. So wurde bereits in den 1970er und 1980er Jahren die Kontroverse um den „experimentellen und korrelativ-naturalistischen“ Ansatz geführt (Kiesler 1977; vgl. auch das Modell der Datenboxen von Wittmann, 1985, 1988). Smith (1980) z.B. diskutierte die Durchführbarkeit (feasibility) und Erwünschtheit (desirability) von experimentellen Designs in der Evaluationsforschung, Gelso (1979) thematisierte im Rahmen seiner „bubble hypothesis“, dass die Vorteile der einen Richtung (interne Validität) auf Kosten der anderen (externe Validität) geht und umgekehrt. Verschiedentlich wurde auch vom „Mythos Experiment“ gesprochen (Nübling, 1992).

Hauptkritik ist, dass die Ergebnisse von RCTs nur sehr begrenzt repräsentativ sind für die klinische Praxis, also nicht ohne weiteres generalisiert werden können (Beutler, 1998; Fonagy, 1999; Howard et al., 1996; Seligman, 1995; Heekeren, 2005; Bayer, 2007). Dies gilt nicht nur für die Psychotherapie, sondern auch für andere Bereiche der Gesundheitsversorgung. Fydrich & Schneider (2007) nennen eine Reihe inhaltlicher und methodischer Probleme mit RCTs im Kontext der Psychotherapieforschung, v.a. die nicht vorhandene aktive Therapeutenwahl, die möglichen Kontroll- und Placebobedingungen, die fragliche Nähe zur Versorgungspraxis, die Art der Patientenselektion bzw. deren Repräsentativität hinsichtlich der Versorgungspraxis, den Umgang mit Behandlungsabbrüchen oder den Publikationsbias. Revenstorf (2005) spricht von „notorischer Irrepräsentativität“ der RCTs und vertritt die Auffassung, dass mit ihnen ein Kuckucksei aus der Medikamenten- in die Psychotherapieforschung übernommen wurde. Wie Wittmann (1985) ausführt, wird in RCTs unkritisch davon ausgegangen, dass die Reliabilität des Treatments = 1 ist, d.h. in der Behandlungsgruppe eine hinsichtlich Treatmentstärke und Treatmentintegrität homogene, einheitliche Behandlung „verabreicht“ wird. Es mag in pharmakologischen Studien möglich sein, die „Dosis“ für alle Personen der Behandlungsgruppe konstant zu halten, in der Psychotherapie ist dies – auch bei manualisiertem Vorgehen – eher die Ausnahme. In RCTs wird mit Grunde alles behindert bzw. ausgeschlossen, was gute klinische Praxis bzw. Anwendungsrealität ist: Patienten mit komplexen Störungsbildern, wie sie in der klinischen Praxis die Regel sind, die sich ihren Behandler aussuchen; Therapeuten, die mit den Therapieverfahren behandeln, die sie üblicherweise auch anwenden; eine Therapiedauer, die sich nach klinischen Erfordernissen richtet und nicht konstant für alle Patienten gehalten wird u.v.m. (Leichsenring, 2004a, b).

All diese Kritik bezieht sich nicht darauf, RCTs grundsätzlich als ungeeignet für die Erfassung der Outcomes psychologischer bzw. psychotherapeutischer Interventionen zu halten, sie wird hier v.a. deshalb ausgeführt, weil hierdurch die Bedeutung der Ergebnisse naturalistischer Studien für die Versorgung besser eingeschätzt werden können. Gerade im Hinblick auf die Psychotherapie werden die Vorteile naturalistischer Studien massiv unterschätzt. Die unvergleichlich höhere Praxisnähe und Generalisierbarkeit der Befunde führen zu einer verbesserten Prognose dessen, was ein Verfahren in der konkreten klinischen Alltagspraxis zu leisten vermag und welche gesundheitspolitische und gesundheitsökonomische Bedeutung ihm zugeschrieben werden kann. RCTs sind wegen ihrer Künstlichkeit keine adäquate Grundlage für eine solche Prognose. Wie Fydrich & Schneider (2007) hervorheben, sind naturalistische Studien nicht grundsätzlich von geringerer Qualität, sie verfolgen unterschiedliche Zielrichtungen und ermöglichen unterschiedliche Aussagen.

Ein gutes Beispiel für die potentielle Bedeutung naturalistischer Studien für die Bewertung eines Versorgungsbereichs stellt die vor kurzem abgeschlossene und im Frühjahr 2007 erscheinende Metaanalyse der Effekte psychosomatischer Rehabilitation (Steffanowski et al., in press) dar. In sie wurden 65 Studien, die im Versorgungsbereich der stationären Rehabilitation psychischer Erkrankungen in den vergangenen 20 Jahren durchgeführt und publiziert wurden, einbezogen. Fast alle Studien (n=63) sind naturalistisch angelegt, d.h. es wurden konkrete Inanspruchnahmepopulationen aus der alltäglichen Klinikroutine gewonnen. Die in der Meta-Analyse untersuchten Studien umfassten insgesamt über 25.000 Patienten, deren mittlere Behandlungseffekte (Effektstärken) über alle Studien und Ergebnismaße zwischen $d=0,51$ (Entlassungszeitpunkt) und $d=0,41$ (1-Jahres-Katamnese), bei Betrachtung von Patienten spezifischer Störungsbereiche (z.B. depressive Patienten) unter Verwendung störungsspezifischer Messinstrumente zwischen $d=1,07$ und $d=0,76$ liegen. Gesundheitsökonomisch kann von einer Kosten-Nutzen-Relation von bis zu 1:4 ausgegangen werden, der gesamtgesellschaftliche Nutzen der jährlich etwa 100.000 Behandlungen liegt bei über 3 Milliarden € pro therapiertem Patientenjahrgang. In der Sichtweise der RCT-focussierten Evidenzstufen würden diese Studien wegen „mangelnder Evidenz“ gar nicht oder nur sehr bedingt zu einer Bewertung im Sinne der Verfahrensordnung des G-BA ausgewählt werden. Es kann allerdings nicht ernsthaft behauptet werden, diese in der alltäglichen Versorgung durchgeführten Studien dieser Größenordnung (Stichprobengröße) hätten keine Evidenz. Ganz im Gegenteil belegen sie die Wirksamkeit sowie auch die Effizienz der durchgeführten Behandlungen für diesen Versorgungsbereich eindrücklich. Vor diesem Hintergrund ist die Frage der Evidenz bzw. der Evidenzstufen neu oder zumindest anders zustellen, das Absolutheitskriterium der „Unverzichtbarkeit“ des experimentellen Nachweises wird ad absurdum geführt.

Ein Ausweg aus dem Dilemma zwischen den Ansätzen kann, wie letztlich auch in den vorgeschlagenen Verfahrensregeln implizit konzipiert (S.5), darin liegen, dass ein Stufenmodell, wie es im pharmakologischen Prüfmodell praktiziert wird, auch für die Bewertung psychotherapeutischer Interventionen zugrundegelegt wird (Kächele, 2006; Shadish, Matt, Navarro & Siegle, 1997; Heckerens, 2005, Campell et al., 2000). Dieses umfasst aufeinander aufbauend:

1. Pilotstudien zur Klärung von Effekten, Risiken, Anwendbarkeit u.a.m.
2. Kontrollierte klinische Studien
3. Erprobung der Interventionen an speziellen Populationen
4. Evaluation im Rahmen des öffentlichen Gesundheitswesens und
5. Erprobung unter realen Praxisbedingungen.

Eine konsequentere Möglichkeit wird von Leichsenring (2004a, b) unter Bezug auf die wissenschaftstheoretischen Grundlagen thematisiert. Er fordert, levels of evidence **getrennt** für RCTs und für naturalistische Studien zu formulieren. Leichsenring führt die „Verabsolutierung“ von RCTs und der implizit damit verbundenen Theorie einer universellen Gültigkeit auf ein überholtes wissenschaftstheoretisches Verständnis der Struktur von Hypothesen und Theorien zurück. Zentral ist der Gedanke Westmeyers (1982), dass es keine kontextunabhängige Hypothesen geben kann (vgl. auch Slife, 2004: „there is no escape from assumption“ ter Bezug auf Jaspers' „there is no escape from philosophy“). Daraus folgt, dass es von den intendierten Anwendungen und Fragestellungen abhängt, ob kontrollierte (wenn sich die Hypothese auf Laborkontexte bezieht) oder naturalistische Studien (wenn sich die Hypothese auf natürliche Bedingungen bezieht) durchgeführt werden. Und hieraus wiederum, dass RCTs grundsätzlich keine Evidenz höheren Niveaus liefern als naturalistische Designs (und umgekehrt), weshalb naturalistische Studien nicht länger als zweitklassig in Bezug auf die gelieferte Evidenz angesehen werden könnten.

Die LPK Baden-Württemberg schließt sich dieser Auffassung uneingeschränkt an. Die derzeitige Praxis des IQWiG, zur Überprüfung bzgl. sozialrechtlicher Zulassung, mit wenigen Ausnahmen nur RCT-Studien zur Bewertung von Verfahren und Methoden zuzulassen sowie die Verfahrensordnung des G-BA halten wir für dringend modifizierungsbedürftig, dies nicht

nur für die psychotherapeutische Versorgung. Dies sollte sich auch in der Neufassung der Methodenkriterien für die Anerkennung von Verfahren und Methoden durch den WBP niederschlagen. Die LPK Baden-Württemberg erwünscht sich vom Wissenschaftlichen Beirat eine noch stärkere Gewichtung naturalistischer Studien und ggf. die Vertretung dieser gegenüber G-BA und IQWiG.

2 Konkrete Vorschläge zu den Verfahrensregeln des WBP

Wie ausgeführt, fordert die LPK Baden-Württemberg den WBP dazu auf, über eine Neubewertung der wissenschaftlichen Bedeutung von naturalistischen Studien, insbesondere im Vergleich zu RCTs, nachzudenken und Kriterien für eine **gleichwertige** Evidenz zu entwickeln. Eine gute Grundlage hierfür stellt nach Ansicht der LPK Baden-Württemberg der Ansatz von Leichsenring (2004a, b) dar, der Evidenzstufen auch für naturalistische Studien analog bzw. parallel zu RCTs vorgeschlagen hat.

Im folgenden werden darüber hinausgehend einige Vorschläge bzgl. der Einzelkriterien für die allgemeine methodische Qualität von Studien sowie für die interne und externe Validität gemacht.

2.1 Allgemeine methodische Qualität

Die vom WBP vorgeschlagenen allgemeinen methodischen Kriterien für die Studienbewertung werden von der LPK Baden-Württemberg geteilt, die Ausschlusskriterien grundsätzlich unterstützt. Ein weiteres Ausschlusskriterium könnte aus unserer Sicht bzgl. der Stichprobengröße definiert werden, in dem Studien mit Stichprobengrößen unter $n=10$ (auch bei Gruppendesigns) nicht zugelassen werden.

Der Vorschlag, neben den Ausschlusskriterien einen Mittelwert der verbleibenden Kriterien zu bilden, wird ebenfalls unterstützt.

Vorschläge bzgl. einzelner Kriterien:

- Kriterium 3 und 4: die Höhe des maximalen Drop-Outs sollte konkret festgelegt/definiert werden (was bedeutet Drop-Out-Quote „vergleichbar“ mit Studien entsprechender Patientengruppen), ggf. für einzelne bzw. die wichtigsten Patientengruppen. Dabei ist zu berücksichtigen, dass Studien unter Praxisbedingungen in der Regel höhere Drop-Out-Raten aufweisen als experimentelle Studien.
- Kriterium 5: die Stichprobengrenzen sollten höher gesetzt werden, z.B. $n < 10$ Ausschluss, $n = 10-30 \rightarrow 3$, $n = 30-50 \rightarrow 2$ und $n > 50 \rightarrow 1$
- Kriterium 6 ist formuliert bezogen auf Kontrollgruppendesigns, eine entsprechende Formulierung für naturalistische Studien fehlt

2.2 Kriterien zu Bewertung der internen Validität

Die Kriterien zur Bewertung der internen Validität beziehen sich vorwiegend auf Kontrollgruppendesigns. Es wird von der LPK Baden-Württemberg vorgeschlagen, die Kriterien für die interne Validität getrennt für RCTs und für naturalistische Studien zu formulieren. Alle auf Kontrollgruppendesigns bezogenen Kriterien (4, 5, 8, 9, 12, insbesondere Kriterium 8 als Ausschlusskriterium) können nur auf Kontrollgruppendesigns angewandt werden, diese sind per definitionem für naturalistische unerfüllbar. Dies bedeutet, dass naturalistischen Studien von vornherein eine geringere interne Validität zugesprochen wird, als RCTs.

Vorschläge zu einzelnen Kriterien:

- als neues Kriterium zur Erfassung von Treatmentstärke und Treatmentintegrität wird für die Bewertung naturalistischer Studien die umfassende quantitative Dokumentation der durchgeführten Behandlungen (Treatment-Box, Wittmann, 1988) im Sinne einer Basisdokumentation vorgeschlagen (vgl. Schmidt et al., 1992, Nübling & Schmidt, 1998)
- Kriterium 11: die Follow-up-Messung(en) sollten wenigstens einen Messzeitpunkt \geq ein Jahr nach Behandlungsende umfassen
- Kriterium 12: die Outcome-Messung sollte zur gesundheitsökonomischen Abschätzung der Behandlungseffekte auch kostenrelevante Kriterien umfassen (Arbeitsfehlzeiten, Krankenhaustage, Arztbesuche, Medikamentenkosten)

2.3 Kriterien zur Bewertung externer Validität

Wie eingangs ausgeführt, wird das Einbeziehen von Kriterien der externen Validität zur Bewertung psychotherapeutischer Verfahren und Methoden von der LPK Baden-Württemberg ausdrücklich begrüßt. Auch die vorgeschlagenen Operationalisierungen der Kriterien können ausnahmslos mitgetragen werden. Lediglich für Kriterium 9 (Outcome-Messung) gilt die gleiche Anmerkung wie für das entsprechende Kriterium der Liste zur internen Validität (Kriterium 12).

2.4 Beurteilung der Ergebnisse einer Studie

Die an dieser Stelle des Bewertungsprozesses – und nur an dieser – vorgeschlagene bzw. formulierte annähernde Gleichrangigkeit von RCTs und naturalistischen Designs („und/oder“-Verknüpfung in 2.5, II.) entspricht der oben skizzierten Grundauffassung der LPK Baden-Württemberg und findet Zustimmung. Wünschenswert für die in III. aufgeführten Kriterien bei kontrollierten Studien (Punkte III., 1.-3.) ist darüber hinaus die Forderung des Nachweises der „klinischen Bedeutsamkeit“ der gemessenen Veränderungen analog Kriterium III./4. für naturalistische Studien.

2.5 Berücksichtigung älterer Studien

Grundsätzlich teilt die LPK Baden-Württemberg die Auffassung des WBP, aktuellere Studien stärker zu gewichten. Das Schwellenkriterium 1990 kann gut akzeptiert werden. Auch das Kriterium, dass mehr als die Hälfte der Studien neueren Datums sein sollen, wird unterstützt. Allerdings ist die Begründung, dass ältere Studien nicht den heute gültigen methodischen Standards genügen nicht per se plausibel. Auch ältere Studien können – wie die aktuellen – unter Anwendung der Kriterienlisten hinsichtlich ihrer allgemeinen methodischen Qualität sowie hinsichtlich ihrer internen und externen Validität beurteilt und dann ein- oder ausgeschlossen werden. Darüber hinaus ist nicht nachvollziehbar, mit welcher Begründung für ältere Studien eine Kontrollbedingung gefordert wird (2.7, Kriterium 4). Naturalistische Studien vor 1990 werden damit ausgeschlossen. Dies wird von der LPK Baden-Württemberg abgelehnt; gefordert wird eine Formulierung, die es erlaubt auch naturalistische Studien, die vor 1990 durchgeführt bzw. publiziert wurden, einzuschließen.

3 Literatur:

- Bayer, K. (2007). Methodische und begriffliche Überlegungen zur evidenzbasierten Medizin (EbM) im Hinblick auf deren Eignung als Werkzeug, um die bisherige Medizin zu reformieren. *bvvp-Magazin*, 1, 2007, 22-24.
- Beutler, L. E. (1998). Identifying empirically supported treatments: What if we didn't? *Journal of Consulting and Clinical Psychology*, 66, 113-120.

- Campbell, M., Fitzpatrick, R., Haines, A., Kinmonth, A. L., Sandercock, P., Spiegelhalter, D. & Tyrer, P. (2000). Framework for design and evaluation of complex interventions to improve health. *British Medical Journal*, 321: 694–696
- Fonagy, P. (1999). Process and outcome in mental healthcare delivery: a model approach to treatment evaluation. *Bulletin of Menniger Clinic*, 63, 288-304.
- Fydrich, T. & Schneider, W. (2007). Evidenzbasierte Psychotherapie. *Psychotherapeut*, 52, 55-68.
- Gelso, C. J. (1979). Research in counseling: methodological and professional issues. *The Counseling Psychologist*, 8, 7-35.
- Gemeinsamer Bundesausschuss G-BA (2006). Verfahrensordnung. *Bundesanzeiger* 2006, S. 4876; download: <http://www.g-ba.de/cms/upload/pdf/richtlinien/2006-07-07-VerfO.pdf>.
- Heekerens, H.P. (2005). Vom Labor ins Feld. Die Psychotherapieevaluation geht neue Wege. *Psychotherapeut*, 50, 357-366.
- Howard, K. I., Moras, K., Brill, P. L., Martinovich, Z. & Lutz, W. (1996). Evaluation of psychotherapy. Efficacy, effectiveness and patient progress. *Journal of Consulting and Clinical Psychology*, 51, 1059-1064.
- Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen IQWiG (2006). Methoden. Version 2.0 vom 19.12.2006. Köln, IQWiG, 2006, Internetveröffentlichung. http://www.iqwig.de/download/2006_12_19_IQWiG_Methoden_V-2-0.pdf.
- Kächele, H. (2006). Wirksamkeitsnachweise – Das Bessere ist der Feind des Guten. *Psychotherapeutenjournal*, 5, 2006, 140.
- Leichsenring, F. (2004a). Randomized controlled versus naturalistic studies: A new research agenda. *Bulletin of Menniger Clinic*, 68, 137-151.
- Leichsenring, F. (2004b). „Empirically supported Treatments“: Wissenschaftstheoretische und methodische Aspekte kontrollierter vs. naturalistischer Studien. *Zeitschrift für Klinische Psychologie, Psychiatrie und Psychotherapie*, 52, 209-222.
- Nübling, R. (1992). Psychotherapiemotivation und Krankheitskonzept. Frankfurt a. M.: VAS.
- Nübling, R. & Schmidt, J. (1998). Interne Qualitätssicherung in der stationären psychosomatischen Rehabilitation. Erfahrungen mit einem „zweigleisigen Modell“. In A.-R. Laireiter & H. Vogel (Hrsg.), *Qualitätssicherung in der Psychotherapie und psychosozialen Versorgung – Ein Werkstattbuch* (S. 335-353). Tübingen: dgvt-Verlag.
- Orlinsky, D. E., Grawe, K., & Parks, B. K. (1994). Process and Outcome in Psychotherapy - Noch einmal. In S. L. Garfield & A. E. Bergin (Eds.), *Handbook of Psychotherapy and Behavior Change* (pp. 270-376). New York: Wiley.
- Porzsolt, F., Ackermann, M. & Amelung, V (2006). Konzept zur Bewertung des intangiblen Nutzens von Gesundheitsleistungen. *Gesundheitsökonomie und Qualitätsmanagement*, 11, 353-364.
- Revenstorf, D. (2005). Das Kuckucksei - Über das pharmakologische Modell in der Psychotherapie-Forschung. Internetveröffentlichung. http://www.revenstorf.de/meg/downloads/Das_Kuckucksei.pdf
- Rossi, P. H., Freeman, H. E. & Lipsey, M. W. (1999). *Evaluation. A systematic approach*. 6th Edition, Thousand Oaks, Sage.
- Schmidt, J; Nübling, R; Lamprecht, F (1992): Möglichkeiten klinikinterner Qualitätssicherung (QS) auf der Grundlage eines Basis-Dokumentations-Systems sowie erweiterter Evaluationsstudien. *Gesundh-Wes* 54, 70-80.
- Seligman, M. E. P. (1995). The effectiveness of psychotherapy. The Consumer Reports Study. *American Psychologist*, 50, 965-974
- Shadish, W. R., Matt, G., Navarro, A. M. & Phillips, G. (2000). The effects of psychological therapies under clinically representative conditions: a meta-analysis. *Psychological Bulletin*, 126, 512-529.
- Slife, B. D. (2004). Theoretical challenges to therapy practice and research: the constraint of naturalism. In Lambert, M. L. (Ed.), *Bergin and Garfields Handbook of Psychotherapy and Behaviour Change* (pp. 44-83). 5th Edition. New York: Wiley & Sons.
- Smith, N. L. (1980). The feasibility and desirability of experimental methods in evaluation. *Evaluation and Program Planning*, 3, 252-256.
- Steffanowski, A., Löschmann, C., Schmidt, J., Nübling, R. & Wittmann, W. W. (in press). Metaanalyse der Effekte psychosomatischer Rehabilitation. Bern, Huber.
- Tschuschke, V. (2005). Psychotherapie in Zeiten evidenzbasierter Medizin. Fehlentwicklungen und Korrekturvorschläge. *Psychotherapeutenjournal*, 4, 106-115.
- Westmeyer, H. (1989)(Ed.). *Psychological theories from a structuralist point of view*. New York, Springer.
- Wittmann, W. W. (1985). *Evaluationsforschung*. Berlin: Springer.
- Wittmann, W. W. (1988). Multivariate reliability theory. Principles of symmetry and successful validation strategy. In J. R. Nesselroade & R. B. Cattell (Hrsg.), *Handbook of multivariate experimental psychology* (pp. 505-560). New York: Plenum.
- Zurhorst, G. (2003). Eminenz-basierte, Evidenz-basierte oder Ökologisch-basierte Psychotherapie. *Psychotherapeutenjournal*, 2, 97-104.