

# **Die Wirksamkeit des Autogenen Trainings: eine Metaanalyse**

Dr. Sirko Kupper

Schlüsselwörter:

**Autogenes Training, Entspannung, Metaanalyse, kontrollierte klinische Studien, Effektivität**

Manuskript für die Publikation in  
*Psychotherapie, Psychosomatik und medizinische Psychologie*

## Zusammenfassung

Das Autogene Training ist ein Entspannungsverfahren, das auf der passiven Konzentration auf Körperwahrnehmungen im Sinne einer gezielten Aufmerksamkeit ohne subjektive Anspannung basiert, was durch Selbstsuggestionen noch unterstützt wird. Es wurde eine Metaanalyse durchgeführt, um die klinische Wirksamkeit des AT beurteilen zu können. 73 kontrollierte Outcome- und Follow-up-Studien aus den Jahren 1932 bis 1999 wurden über „Medline“, über Angaben einschlägiger früherer Untersuchungen sowie Monographien rekrutiert. Anhand von 4 Einschlusskriterien (z. B. mindestens eine Kontrollgruppe oder –phase) und vier Ausschlusskriterien (z. B. weniger als fünf Patienten in einer der untersuchten Gruppen) wurde eine Auswahl von 60 Studien (35 randomisiert) für die quantitative Analyse getroffen. Hier wurde gefunden, dass das Autogene Training mittlere bis große klinische Effekte auf die Hauptzielsymptomatik hin bewirkt, die zum Follow-up Zeitpunkt stabil sind und die Placebo-Effekte (wahre Kontrollgruppenbedingung) deutlich überschreiten. Gegenüber anderen Behandlungsbedingungen erbringt das Autogene Training gleiche Effekte, insbesondere zum Follow-up-Zeitpunkt hin, zum Zeitpunkt des Behandlungsabschlusses leicht geringere Effekte. Auch im Bereich der unspezifischen Symptomatik (z. B. depressive Stimmung bei Kopfschmerzpatienten) wurden mit mittleren bis großen klinischen Effekten Hinweise für die Wirksamkeit des Autogenen Trainings festgestellt. In mehr als einer Studie bei Patienten mit milder bis moderater essentieller Hypertonie, Spannungskopfschmerz und Migräne, Koronarer Herzerkrankung, Asthma bronchiale, Raynaud-Syndrom und Neurasthenie/unspezifische somatoforme Schmerzstörung zeigt das Autogene Training seine Wirksamkeit im Prä-Post-Vergleich und gegenüber der wahren Kontrollbedingung. Bei acht weiteren psychosomatischen Störungsbildern (Darmerkrankung, Epilepsie, Neurodermitis, Geburtsvorbereitung u.a.) lag jeweils nur 1 Studie für die metaanalytische Bewertung vor, so dass hier künftige Untersuchungen die Wirksamkeitshypothese weiter zu falsifizieren haben werden. Bei Angststörungen, Depressionen und Schlafstörungen kann ebenfalls von einer gesicherten Wirksamkeit des Autogenen Trainings gesprochen werden. Nicht aussagekräftig sind die positiven klinischen Effekte bei Alkoholismus und Stottern, da auch bei diesen Störungsbildern Replikationen wünschenswert sind.

**Schlüsselwörter:** Autogenes Training, Entspannung, Metaanalyse, kontrollierte klinische Studien, Effektivität.

# 1 Einführung

Das Autogene Training (AT) ermöglicht durch selbst (auto) hervorgerufene (gen) Vorstellungen, objektiv nachweisbare körperliche und psychische Veränderungen herbeizuführen. Die angestoßene „allgemeine Umschaltung“ (Schultz, 1932/1991) schlägt sich neben dem Erleben von körperlicher Ruhe auch in einer Veränderung des Bewusstseinszustandes nieder. Voigt, später auch Luthé (1970a, b, c), prägten den Begriff der passiven Konzentration, durch den sich diese Bewusstseinsänderung zutreffend beschreiben lässt. Es handelt sich hierbei um den Zustand einer gezielten Aufmerksamkeit ohne subjektive Anspannung, so wie beim versonnenen Hören eines Musikstückes, wo einen Augenblick lang „alles ringsum zu versinken scheint“, was ein angenehmes und intensives Erleben provoziert. Ein nachvollziehbares Beispiel gibt Bleulers folgende Selbstbeobachtung (Münchener Wochenschrift, 1889, zit. nach Schultz, 1932/1991), wo er fühlte, „dass die Bewegung gemacht wurde, ohne irgendwelche aktive Beteiligung meines Ich“. Schultz selbst verglich diesen Prozess der „sammelnden Einengung“, der zu reflektorischen Überwältigungen führe auch mit dem Paradoxon der selbsttätigen Passivierung kurz vor dem Einschlafen, was sich mit Hoffmann, Hoffmann, Derra & Stetter (2000) auch auf das Autogene Training „zwanglos“ übertragen lässt.

Die Konzeption des autogenen Trainings habe Schultz (1932/1991) seinen eigenen Berichten zufolge in den Jahren 1908 bis 1912 entwickelt, angestoßen durch die Arbeiten von Voigt (1893-1900). Der Berliner Nervenarzt Johannes Heinrich Schultz begründete das Autogene Training im Rahmen seiner Jenaer Untersuchungen 1920 an hypnotisierten Patienten, die über Schwere- und Wärmeempfindungen insbesondere in den Extremitäten berichteten und erprobte es in den Dresdener klinischen Versuchen bis 1924 und den darauffolgenden Anwendungen auch bei gesunden Probanden in Berlin. 1932 wurde das Konzept des Autogenen Trainings in Form eines Lehrbuches publiziert.

## 2 Psychophysiologische Hintergründe des Autogenen Trainings

Nach aktuellen Forschungsergebnissen beruht die Wirkung des AT in seinen systematisch aufgebauten Übungsinhalten auf Prozessen der selektiven Aufmerksamkeit ("konzentrierte Zuwendung", Schultz, 1932/1991), wodurch die beim ruhenden Menschen spontan einsetzende körperliche Entspannungsreaktion (Vaitl, 1993) verstärkt und automatisiert wird (Stetter, 1994). Im Sinne der Kopplung kommt es, angestoßen durch die Aufmerksamkeitseinengung „auf ein sehr begrenztes Gebiet“ (Schultz, 1932/1991) zum Aufbau einer bedingt-reflektorischen Verbindung der kognitiven Inhalte mit körperlichen Reaktionen (König, di Pol & Schaeffer, 1979). In elektromyographischen Untersuchungen konnte als physiologisches Bestimmungsstück der Schwereempfindung die Hypotonie der Extremitätenmuskulatur nachgewiesen werden. In eigenen kontrollierten Studien konnte gezeigt werden, dass sich die Wärmeempfindung auf die durch Umverteilung des Blutflusses hervorgerufene Erhöhung der Hauttemperatur bei geübten im Vergleich zu ungeübten Probanden und Patienten zurückführen lässt (Mann & Stetter, 1982; Stetter, 1985). Neben der durch die zentrale „Down-Regulation“ verstärkten Hypotonie der Muskulatur und dem Anstieg der Hautdurchblutung mit Hauttemperaturerhöhung hervorgerufen durch die zentral gesteuerte Absenkung des Sympathotonus kommt es zur Herabsetzung der zerebralen Aktivierung im Sinne

einer Dämpfung, was in dem Ruherlebnis Niederschlag findet und in Elektroencephalogramm (EEG)-Untersuchungen belegt werden konnte (Stetter, 1991). Da es gleichzeitig zu einer nachweislichen Verbesserung von Reaktionsleistungen kommt, bietet sich die Konzeptualisierung dieses Zustandes als „entspannte, fokussierte Wachheit“ (Ikemi, 1988) an. Auf dem Weg der schrittweisen Fokussierung der Aufmerksamkeit auf eng umgrenzte Gebiete kommt es durch einen Rückkopplungsmechanismus zu einer spontanen Ausbreitung auf andere Körpergebiete im Sinne der „Generalisierung“ (Schultz, 1932/1991). Dies entspricht auch der Beschreibung des „ideomotorischen Prinzips“ von Carpenter (1876/1993) oder Forels (1902) „Ideoplasie“, d. h. die den Vorstellungen innewohnende Tendenz zur Wahrnehmung zu werden, so dass alle Suggestionen bereits die Möglichkeit zu einer Realisierung in sich tragen (König et al., 1979). Vergleichbar zu den physiologisch nachweisbaren Korrelaten der Ruhetönung, der Schwere- und Wärmeempfindung lassen sich in Bezug auf die Organübungen (Atmung, Herz, Bauchraum, Stirn) keine spezifischen Wirkungen belegen. Dies stimmt auch mit den Berichten von AT-Teilnehmern in Präventions- und Katamneseuntersuchungen überein, die in weitestgehender Übereinstimmung regelmäßige positive Entspannungsreaktionen in den Bereichen Ruhetönung, Schwere- und Wärmeempfindung sowie Atmung erfahren konnten.

Vor der Durchführung des AT ist es wichtig, dass die Patienten durch einen ärztlichen oder psychologischen Psychotherapeuten sorgfältig untersucht werden, damit Patienten mit einer erheblichen Einschränkung der Selbstverfügbarkeit selektiert werden können. Das Erreichen der „organismischen“ Gesamtumschaltung, die Vertiefung der Körperwahrnehmung und die Verwendung von kurzen, wandspruchartigen mitunter sich reimenden Formeln gehören zur Unterstufe des AT. Nach sechsmonatiger Erfahrung damit ist die Erweiterung auf die Übungen der Oberstufe möglich, die auf Imaginationen basieren und zu einer wachsenden „Selbstkontrolle der Innenstruktur“ (Schultz, 1932/1991) führen.

### 3 Hypothesen

In früheren Untersuchungen zur Wirksamkeit des Autogenen Trainings konnten teils zustimmende (Linden, 1994), teils ablehnende (Grawe, Donati und Bernauer, 1994) Aussagen zur Wirksamkeit des AT's formuliert werden. In Verbindung mit dem ausführlichen Studium der vorliegenden Literatur zu diesem Thema mit zahlreichen gut dokumentierten Fallbeispielen (Schulz, 1932/91; Krampen, 1992; Mensen 1994; Hoffmann et al., 2000) wurde eine Voruntersuchung durchgeführt (Stetter & Kupper, 1998), aus der sich die Basis für die Untersuchungshypothesen aufbaute. Für die vorliegende metaanalytische Untersuchung zur Wirksamkeit des Autogenen Trainings werden folgende vier Hypothesen vorangestellt.

#### **Hypothese 1:**

Das Autogene Training führt zu signifikanten Effekten im Vorher–Nachher–Vergleich (Prä-Post) bezüglich der Hauptsymptomatik.

Mit Blick auf die Teilergebnisse der Voruntersuchung (Stetter & Kupper, 1998) ist zu erwarten, dass sich die Wirkung des Autogenen Trainings im Prä-Post-Vergleich in statistisch bedeutsamen Effekten ausdrückt mit Schwerpunkt auf der Hauptsymptomatik.

### **Hypothese 2:**

Das Autogene Training ist den Kontrollgruppen-Behandlungsbedingungen signifikant überlegen.

Es soll anhand dieser Hypothese die tatsächliche Wirkung des Autogenen Trainings geprüft werden.

### **Hypothese 3:**

Die Wirksamkeit des Autogenen Trainings ist ähnlich der Wirksamkeit anderer psychologischer Behandlungsmethoden bei den hier untersuchten Störungsbildern.

Es wird hiermit postuliert, dass sich die Entspannungseffekte des Autogenen Trainings bei einer großen Reihe psychosomatischer und psychischer Störungsbilder vergleichen lassen mit den positiven Auswirkungen anderer psychologischer Behandlungsverfahren (z.B. progressive Relaxation, Biofeedback, Entspannungshypnose).

### **Hypothese 4:**

Das Autogene Training bewirkt signifikante Effekte auf die Stimmung, kognitive Leistungsfähigkeit, Lebenszufriedenheit und physiologische Variablen.

Durch diese Hypothese wird die Annahme vertreten, die auch in der Voruntersuchung bereits Unterstützung erhielt, dass sich die Wirksamkeit des Autogenen Trainings auch auf einen breiten Bereich der für die Lebenszufriedenheit und den adäquaten Umgang mit Belastungssituationen notwendiger Grundlagen bezieht.

## **4 Stichprobe und Methodik**

### **4.1 Untersuchungsstichprobe**

#### **4.1.1 Zusammensetzung und Auswahl**

Die Gesamtstichprobe setzt sich aus 73 kontrollierten klinischen Studien zur Wirksamkeit des Autogenen Trainings zusammen. 41 Studien weisen eine Randomisierung auf. Die Auswahl der Studien war breit angelegt, so dass über sprach- und kulturelle Grenzen hinweg eine realistische Einschätzung der Wirksamkeit des Autogenen Trainings möglich ist im Sinne größtmöglicher Repräsentativität. Studien sowohl aus Russland, Japan, Spanien und Italien als auch aus Tschechien, Frankreich, England, den USA und Deutschland wurden einbezogen. Andere Studien, zum Beispiel aus Polen wurden in einer Vorauswahl selektiert, da hier lediglich der Outcome des AT im Rahmen des Vorher-Nachher-Vergleiches fokussiert wurde. Ein anderes Auswahlkriterium, das bereits im Vorfeld angesetzt wurde, war der Ausschluss von Studien mit gleichen Stichprobensamensetzungen und Stichprobengrößen bei zum Teil anderer Schwerpunktsetzung und anderer Gesamtautorenschaft (McCoy, Fein, Blanchard, Witrock, McCaffrey & Pangburn, 1988; Van Dyck, Zitman, Linsen & Spinhoven, 1991). Im Anschluss an diese erste Selektionsstufe wurden diejenigen Studien ausgewählt, die sich für eine quantitative Auswertung als geeignet erwiesen (z.B. ausreichende Darstellungsobjektivität der einzelnen Parameter). Auf diese Weise entstand der abschließende Datenpool von 60 kontrollierten Studien, von welchen 35

Untersuchungen dem Modus der Randomisierung entsprachen. Die Selektion der 13 Studien im Rahmen der Feinauswahl für die quantitative Analyse begründet sich folgendermaßen: In vier Studien wurde die AT-Methode kombiniert mit einem anderen Entspannungsverfahren (z.B. Biofeedback bzw. progressive Muskelrelaxation mit Atmungsentspannung) dargeboten (ter Kuile, Spinhoven, Linssen, Zitman & Van Dyck, 1994; Zimmermann-Tansella, Dolcetta, Azzini, Zacche, Bertagni, Siani & Tansella, 1979; Bianchi, De Punzio, Ferdeghini, Fioretti, Metelli, Neri & Venticinque, 1994; Fukunishi, Hosaka, Matsumoto & Hayashi, 1997), in zwei Studien waren keine verwertbaren Daten für die AT-Gruppe angegeben (Yurenev, Patrusheva, Aivazyan, Zaitsev, Krol, Smorodina & Mokrievich, 1988; Schulte, 1983), in zwei weiteren Studien war die Gruppenstärke der Kontrollbedingung kleiner als fünf (Simma & Benzer, 1985; Deter & Allert, 1983), in einer Studie wurden keine Daten angegeben für die AT versus Kontrollgruppenbedingung (Akimenko & Gromov, 1995), in einer Untersuchung erhielten beide Gruppen AT (Keel, Bodoky, Gerhard & Muller, 1998), in einer anderen wurde ebenfalls keine „wahre“ Kontrollgruppe rekrutiert (Reed & Meyer, 1974) und in einer Studie wurden keine Schlussfolgerungen für die Wirksamkeit des AT gezogen, und lassen sich auch nicht ableiten (Stetter, Günthner, Mann & Bartels, 1994).

In fünf Studien wurde ausschließlich eine Nachbefragung durchgeführt. 55 Studien lassen die Prä-Post-Effektstärken (ES)-Berechnung zu, wobei 40 dieser Studien den Vergleich AT versus wahre Kontrollgruppe, 26 den Vergleich AT versus andere psychologische Behandlungsbedingung zulassen. In 31 Untersuchungen wurden Katamnesen durchgeführt. 29 davon lassen den Vergleich von Vorher-Katamnesezeitpunkt für die AT-Gruppe zu, 18 für die Bedingung AT versus wahre Kontrollgruppe, 17 für den Vergleich AT versus andere psychologische Behandlungsverfahren. In 7 Untersuchungen wurde AT mit nichtpsychologischen Behandlungsformen (z.B. „Elektrosleep“) verglichen. Bei allen sieben ist der Vergleich Prä versus Post für AT möglich, drei Studien lassen die ES – Bestimmung für AT versus wahre Kontrollbedingung zu, eine AT versus andere psychologische Behandlung und alle sieben den Vergleich AT versus nichtpsychologische Behandlungen zum Outcomezeitpunkt, eine auch zum Follow-up-Zeitpunkt.

Die Wirksamkeit des AT's wurde in 60 Fällen bei psychosomatischen Störungen (z.B. Hypertonie, Asthma bronchiale) überprüft, 46 davon gingen in die quantitative Analyse ein, davon waren 31 Studien randomisiert. Zur Wirksamkeit des Autogenen Trainings bei psychischen Störungen (z.B. Angst, Depression) liegen 15 Studien vor, 14 wurden in die quantitative Metaanalyse einbezogen, vier waren randomisiert. In die Rekrutierung gingen nur veröffentlichte Studien aus Zeitschriften oder Büchern, Buchartikeln ein, sie erfolgte über die Zeitschriftendatenbank „Medline“ und über die Literaturverzeichnisse einschlägiger Reviews (Grawe et al., 1994; Linden, 1994) und Monographien (Binder & Binder, 1998; Hoffmann et al., 2000).

#### *4.1.2 Selektionsbedingungen*

Um die tatsächliche Wirksamkeit des AT's zu Grunde zu legen, war es notwendig den Studienpool einzuengen unter zwei Gesichtspunkten.

#### 4.1.2.1 Einschlusskriterien

Eingeschlossen wurden alle verfügbaren Studien zur Wirksamkeit des AT's, die

- zwischen 1932 und 1999 veröffentlicht wurden (Schulz' Originallehrbuch zum AT erschien 1932),
- auf klinisch definierte Patientengruppen ausgerichtet waren,
- mindestens eine Kontrollgruppenbedingung oder Kontrollphase mit psychologisch unbehandelten Patientengruppen beinhalteten („wahre“ Kontrollgruppe bedeutet entweder Wartegruppe, oder nur medizinische Basistherapien [nicht psychiatrische Medikation], oder in drei Fällen nur „wahre“ Placebothherapie [Jessup & Neufeld, 1977, Depression: Tone only-Gruppe: Instruktion: „Sie werden einen Ton hören, der Ihnen bei der Entspannung helfen soll, lassen sie sich von dem Klang entspannen.“; Katzenstein, Kriegel & Gaefke, 1974, essentielle Hypertonie: Einfache Entspannungsgruppe Instruktion: „Ruhen sie sich im Trainingraum bequem aus und entspannen sie sich.“; Blanchard et al., 1988, essentielle Hypertonie: Selbstentspannungs-Kontrollgruppe Instruktion: „Könnten Sie versuchen sich für sich selbst zu entspannen, zirka 20 Minuten pro Tag.“]) oder mindestens eine weitere Behandlungsgruppe (z.B. auch mit anderer psychologischer Therapie),
- das AT mindestens in einer Gruppe bezüglich seines Einsatzes zu therapeutischen Zwecken untersuchte.

#### 4.1.2.2 Ausschlusskriterien

Alle Studien, die folgende Kriterien erfüllten, wurden ausgeschlossen:

- AT spielte eine untergeordnete Rolle oder nicht klar definierbare Rolle in einem umfassenden Therapieplan (Experteneinschätzung),
- kein Outcome-Kriterium steht direkt mit der untersuchten Störung oder dem Syndrom in Verbindung,
- weniger als fünf Patienten waren in einer der untersuchten Gruppen,
- andere Publikationen existieren, die sich auf die gleiche Stichprobe, jedoch detaillierter beziehen.

Es wurden alle Studien einbezogen, die im mindesten die Ruhetönung, Konzentration auf Schwere und Wärme vermittelten. Die positiven AT-Effekte sind bereits mit diesen Grundübungen verbunden (Langen, 1959; Sellers, 1974).

#### 4.1.3 Stichprobenexploration: Verteilungsparameter

Im Vorfeld der Hauptanalysen wurde eine ausführliche Datenexploration durchgeführt, weil die Normalverteilungsannahme die Basis für die Anwendung der metaanalytischen Methoden darstellt (Fricke & Treinies, 1985). Die univariaten

Statistiken, die nach der präziseren Formel von Bentler (1995) berechnet wurden, wiesen ausreichende Werte für die Schiefe und die Kurtosis für alle 24 Datensätze aus. Häufig wird beschrieben, dass bei vielen Methoden, die die Normalverteilung voraussetzen, eine ausreichende Robustheit angenommen wird (Bortz, 1995), nicht zuletzt vor dem Hintergrund der Wirksamkeit des zentralen Grenzwerttheorems der Statistik, nach welchem ab einem  $n \geq 30$  von einer empirischen Normalverteilung der beobachteten Mittelwerte auszugehen ist. Von anderen Autoren (Bentler, 1995; Harlow, 1985) wiederum wird genau davor jedoch gewarnt, die Robustheit automatisch anzunehmen, weil die Bedingungen für die Robustheit nicht leicht nachweisbar sind. Zur weiteren Prüfung der Normalverteilungsannahme wurden Mardias Koeffizienten und dazu die normalisierte Schätzung berechnet, die akzeptable Werte der Kurtosis in angemessenen Grenzen ausweisen (SPSS, 1999; Mardia, 1970, 1974; Bortz, 1995), die Fälle der Datensätze, die einen Beitrag zur multivariaten Kurtosis leisten würden, liegen in ausgesprochen geringer und damit annehmbarer Höhe vor (negative oder positive große Werte von Mardias Koeffizient bzw. der normalisierten Schätzung entsprechen signifikanten Unterschieden von Null unter Hinzuziehung des Standardfehlers, Bentler, 1995). Auch nach den Ergebnissen des Kolmogorov-Smirnov-Goodness-of-Fit-Test kann die Nullhypothese, die beobachteten Werte sind in der Grundgesamtheit normalverteilt, beibehalten werden. Die durchschnittliche Anpassung mit  $p=0.699$  und der zugehörigen Schwankung bei nur einem Datensatz  $0.1 < p < 0.2$  ( $p=0.128$ ) und einem weiteren Datensatz  $0.2 < p < 0.3$  ( $p=0.294$ ), 7 Datensätzen  $0.48 < p < 0.60$  sowie allen anderen 15 Datensätzen mit  $p > 0.60$  ist als akzeptabel einzuordnen. Andere Autoren empfehlen die Transformation in Fishers-Z-Werte (z.B. Kraemer, 1983), da sie normalverteilt bzw. approximativ normalverteilt seien, was wir jedoch in eigenen Verteilungsuntersuchungen nicht nachvollziehen konnten. Die Fisher-Z-Verteilung entspräche einer  $N(68,0.31)$ -Verteilung mit  $p=0.012$  im Kolmogorov-Smirnov-Goodness-of-Fit-Test und einer Schiefe von 1.28 und einer Kurtosis von 1.73 nach der präziseren Mardia-Berechnung (im Vergleich dazu Standardberechnung: Schiefe=1.291, Kurtosis=1.811), die damit deutlich von Null abweichen. Mardias normalisierte Schätzung beträgt zwar nur  $-0.75$ , Mardias Kappa nur  $-0.05$ , die Fälle mit dem größten Beitrag zur normalisierten multivariaten Kurtosis erreichen jedoch deutlich zu hohe Werte (z.B. 889, 42). Dennoch führt allein die von Fisher vorgeschlagene Transformationstechnik des natürlichen Logarithmus (entspricht=2.718) zu einer im entferntesten leichten Annäherung an die Normalverteilungsgerade im Rahmen eines Regressionsmodells (vgl. auch Bortz, 1995).

Zusätzlich wurden als graphische Methoden zur Prüfung der Anpassung der beobachteten empirischen Effektstärke (ES)-Verteilung an die erwartete empirische Normalverteilung der ES in der Grundgesamtheit der für die Exploration metaanalytischer Datensätze modifizierte Normal-Quantile-Plot (Wang & Bushman, 1988), der P-P-Plot nach Blom (siehe SPSS, 1999) sowie der Normal-Quantile-Plot nach Hamilton (1992) zur Einschätzung herangezogen.

## **4.2 Operationalisierung von Haupteffekten und unspezifischen Effekten**

Entsprechend dem Ziel der vorliegenden Untersuchung wurden die Effektstärken der Outcome-Variablen (Ergebnisvariablen) berechnet, die sich direkt auf die



Hauptzielsymptomatik des jeweiligen Krankheitsbildes beziehen. Die Hauptziel-Variablen wurden in zwei Gruppen unterteilt, physiologische (z.B. Blutdruck, Hauttemperatur) sowie behaviorale (z.B. protokollierte Frequenz von Panik- oder Asthmaattacken, Schlafdauer u.ä.) und psychologische (selbst durchgeführte Fragebögen). Ferner wurden unspezifische Effekte ermittelt, über die Effektstärkenberechnung der Ergebnisvariablen, die nicht direkt auf die Zielsymptome bezogen sind, unterteilt in physiologische (z.B. Herzfrequenz bei Raynaudpatienten, Surwit et al., 1982; Keefe et al., 1980) und behaviorale/psychologische (z.B. Depressivität bei Kopfschmerzpatienten, van Dyck et al., 1991).

### 4.3 Metaanalytische Prozeduren

Unser Vorgehen und die verwendeten Formeln entlehnen wir den einschlägigen Standardwerken zur Metaanalyse (Bortz & Döring, 1995; Cook, T. D., Cooper, H., Cordray, D. D., Hartmann, H., Hedges, L. V., Light, R. J., Louis, T. & Mosteller, F., 1992; Cooper, 1984; Cooper, 1998; Cooper & Hedges, 1994; Draper, Graver, Goel, Greenhouse, Hedges, Morris, Tucker & Watenaux, 1992; Durlak & Lipsey, 1991; Fricke & Treinies, 1985; Hedges & Olkin, 1985; Hunter & Schmidt, 1990; Hunter, Schmidt & Jackson, 1982; Kraemer, 1983; Wolf, 1986)

#### 4.3.1 Umgang mit unvollständigen Ergebnisangaben

Bedauerlicherweise existieren keine einheitlichen methodischen Standards bezüglich der Weiterverwendung von Ergebnissen aus Primäranalysen bei Veröffentlichungen, so dass ein Umgang gefunden wurde, um wertvolle, repräsentative Resultate einbeziehen zu können. Folgende Formeln wurden verwendet: bei der Angabe von Werten des t-Tests

$$d = \frac{2t}{\sqrt{df}} \quad , \text{ bei F-Werten} \quad d = \frac{2\sqrt{F}}{\sqrt{df(error)}} \quad \text{und bei Angabe von } r \quad d = \frac{2r}{\sqrt{1-r^2}}$$

Wenn ausschließlich das Signifikanzniveau angegeben war, so wurde der wahre t-Wert über die t-Verteilung approximiert und in die oben angegebene Formel eingesetzt, bei nicht Signifikanz wurde gemäß des konservativen Vorgehens  $p = 1.00$  gesetzt (Kenny, 1999). Graphische Abbildungen wurden im Einzelfall mit Hilfe eines Lineals zur Informationsgewinnung herangezogen. Waren ausschließlich Mittelwerte angegeben so wurden entsprechend der Empfehlung von Linden (1994) statistische Kenngrößen (insbesondere Standardabweichungen) anderer Untersuchungen mit vergleichbarer Stichprobenzahl bei gleichem Krankheitsbild und Untersuchungsjahr herangezogen.

#### 4.3.2 Vorgehen bei der Berechnung der Effektstärken

Ausgehend von der in der Originalstudie angegebenen Ergebnissen wurde die Effektstärke (ES) als standardisierte Mittelwertdifferenz in der von Hedges (1981) modifizierten Metrik von Glass (1976) folgendermaßen berechnet:

Subtraktion des Mittelwertes der Kontrollgruppe von dem Mittelwert der Experimentalgruppe dividiert durch die gepoolte Standardabweichung (g).

Obgleich Hedges'  $g$  durch unterschiedliche Stichprobenvarianzen in der Experimental- und Kontrollgruppe unbeeinflusst bleibt, tendiert es zu einer leichten Verzerrung bei kleinen Stichprobengrößen (Hedges & Olkin, 1985). Aufgrund dieser Überschätzung der ES-Parameter insbesondere bei geringen Stichprobenumfängen wurde die errechnete Effektstärke der Hedges-Korrektur (funktioniert sehr gut bei  $n > 10$ , aber sogar noch bei  $n = 5!$ ) unterzogen nach der Formel:  $g \cdot J$ , wobei  $J$  dem Ausdruck  $1 - \{3/[4 \cdot (n_i^k + n_i^e - 2) - 1]\}$  entspricht (Hedges & Olkin, 1985; Schmid, Koch & LaVange, 1991). Die ES der einzelnen Outcomevariablen wurde für jede Studie nach der Empfehlung von Bortz & Döring (1995) und Fricke & Treinies (1985) als arithmetisches Mittel ausgedrückt, was schließlich Eingang fand in die Metaanalyse sowohl getrennt für jede Fragestellung (physiologisch / verhaltensbezogen-psychologisch/ AT Vorher-Nachher-Vergleich/ AT versus wahre Kontrollgruppe/ AT versus andere psychologische Behandlungsbedingungen/ Outcome/ Follow-up) als auch als Gesamt (Haupteffekte, unspezifische Effekte). Um die Verzerrung resultierend aus unterschiedlich großen Stichprobenumfängen und Varianzen der einzelnen Studien zu minimieren, wurde zusätzlich eine Gewichtung dieser beiden Parameter eingeführt. Welcher dieser beiden Parameter gewichtet wurde und welche Formel zur gewichteten Gesamteffektstärkenbestimmung eingesetzt wurde, richtete sich nach der Homogenität der Stichprobe.

#### 4.3.3 Test auf Homogenität

Voraussetzung für die Wahl der abschließenden Berechnung der Effektstärken (ES) ist das Wissen um die Homogenität der Effektstärkenverteilung.

Besondere Aufmerksamkeit schenken wir der Fragestellung der Homogenität der Effektstärken, weil ohne vorliegende Homogenität keine sinnvolle Interpretation der Ergebnisse möglich erscheint (Fricke & Treinies, 1985; Schwarzer, 1995; Bortz & Döring, 1995). Im Falle einer ermittelten Heterogenität liegt allen Studien nicht derselbe Populationseffekt zugrunde, da es dann angebracht erscheint, für jede Studie von einem eigenen Populationseffekt auszugehen, was in Praxi heißt, dass wir so viele Populationseffekte wie Studien haben. Wenn man sich diesen Sachverhalt vergegenwärtigt ist es tatsächlich sehr einsichtig, warum es keinen Aussagewert hat über diese unterschiedlichen Populationen einen Mittelwert zu bilden. Die Schätzung der Populationseffektstärke auf der Basis des beobachteten Effekts erscheint also nur im Falle vorliegender Homogenität sinnvoll. Homogenität wurde dann angenommen, wenn die folgenden drei Konventionen erfüllt waren (Seipp, 1989; Stoffelmayr, Dillavou & Hunter, 1983; Schwarzer 1995):

1. die verbleibende Varianz der Populationseffekte sollte kleiner als ein Viertel der Populationseffektstärke (erwartungstreu geschätzt durch  $D^{W+}$ ),
2. die verbleibende Varianz der Populationseffekte sollte nicht mehr als 25% der Varianz der beobachteten Effekte ausmachen,
3. der  $\chi^2$ -Test sollte nicht signifikant werden.

Wir beschreiben nachfolgend kurz unser Vorgehen. Unter der Nullhypothese, alle Populationseffektstärken sind gleich (Gurevitch & Hedges, 1993), wurde das asymptotisch einer  $\chi^2$ -Verteilung folgende  $Q$  nach Hedges und Olkin (1985) errechnet und auf statistische Signifikanz geprüft. Ehe wir die Nullhypothese gegen die Alternativhypothese testeten legten wir das Risiko 1 fest, das wir bei der Annahme von der Nullhypothese eingehen werden. Da bei dem Risiko, im Mittel in 5 von 100 Fällen ein Fehlurteil zu fällen, bereits kleine Differenzen genügen um die

Nullhypothese zurückzuweisen ( $\alpha=0.05$ ), erhöhten wir die Sicherheit, um mit unserer Entscheidung nur in einem von 100 Fällen einen Fehler zu begehen. Zur Zurückweisung der Nullhypothese bedarf es nun schon größerer empirischer Kennwertunterschiede (vgl. Clauß & Ebner, 1989). Wir erschwerten die Annahme der Alternativhypothese auch deshalb, weil die irrtümliche Annahme der Alternativhypothese zu sehr nachteiligen Folgen führt bezüglich letztlich auch der Übersetzung der statistischen Ergebnisse auf die klinische Anwendung am Patienten. Daher haben wir konservativ das 1%-Niveau gewählt (Clauß & Ebner, 1989; Bortz & Döring, 1995). Heterogenität wurde angenommen, wenn die Prüfgröße  $Q$  größer als der zugehörige Wert der  $\chi^2$ -Verteilung war, mit anderen Worten wenn die Wahrscheinlichkeit, bei der Wahl von Zufallsstichproben aus der Grundgesamtheit eine ebenso große oder größere Differenz als die beobachtete zu finden, sehr gering ( $p \leq 0.01$ ) war, wodurch gleichzeitig dem Historizitätsproblem (Kächele, 1995; Hillecke, Krämer, Kupper, Kupper-Horster & Kächele, 1996) Rechnung getragen wurde, wonach die aus unterschiedlichen Zeiträumen stammenden Studien das Signifikanzniveau verzerren. Unsere Überlegungen haben folgenden Hintergrund:

Da der Homogenitätstest nach Hedges (1982) sehr sensibel ist und bei großen Stichprobenzahlen eher signifikant wird, kam es der Empfehlung von Fricke und Treinies (1985) folgend zur Anwendung des oben angegebenen, um die exakte Varianzberechnung und auch um die Varianzgewichtung ergänzten robusteren Homogenitätstest nach Hedges & Olkin (1985), der als Weiterentwicklung der Prüfgröße  $H$  von Hedges (1983) zu verstehen ist. Bei Signifikanz der Prüfgröße  $Q$  ist davon auszugehen, dass die Varianz unter den Effektstärken größer ist als durch den Stichprobenfehler zu erwarten ist (Cooper, 1998), weshalb die Untersuchung anderer erklärender Variablen (systematische Faktoren, Moderatorvariablen) angezeigt scheint. Im Falle einer ermittelten Heterogenität, heißt dies auch, dass die Varianz der Populationseffekte ungleich von Null ist, weshalb diese unbekannte Varianz der Populationseffekte aus den Effektstärken der einzelnen Untersuchungen geschätzt werden muss. Bei heterogenen Effekten wurde die Varianz der beobachteten Hedges korrigierten Effektstärken in eine Komponente zu Lasten des Stichprobenfehlers und eine zu Lasten der Variation in den Populationseffekten zerlegt. Da wir die in unserem ersten Auswertungsschritt (Stetter & Kupper, 2002) angewandte Methode der Varianzzerlegung nach Hedges (1982) als nicht präzise genug einschätzen, haben wir in einem zweiten Auswertungsschritt die exaktere Methode „Hedges 83 (gew.)“ nach Fricke und Treinies (1985) eingesetzt. Das Verfahren von Hedges (1983) berücksichtigt zwar die Gefahr der Unterschätzung der Stichprobenvarianz und der beobachteten Varianz, jedoch nicht das Problem, das durch unterschiedliche Stichprobengrößen auftritt. Dieses kann immer nur durch eine zusätzliche Gewichtung in den einzelnen Studien ausreichend berücksichtigt werden und die beobachtete Varianz und damit auch die Restvarianz wird so geringer ausfallen, weshalb wir uns der Meinung von Fricke und Treinies (1985) anschließen und die Methode „Hedges 83 (gew.)“ anwenden. Die Vorschläge anderer Autoren enthalten diese Überlegungen nicht oder nur zum Teil (Bortz & Döring, 1995; Hedges, 1983; Hunter & Schmidt, 1990; Hunter, Schmidt & Jackson, 1982; Kraemer, 1983).

#### 4.3.4 Abschließende Effektstärken-Berechnung

Bei Nichtsignifikanz im Homogenitätstestvergleich mit der  $\chi^2$  - Verteilung kann die Aussage formuliert werden, dass alle Studien der gleichen Effektstärke (ES) – Population entstammen, Abweichungen auf den Stichprobenfehler rückführbar sind. In diesem Falle wurde die Weighted-Integration-Method (Fixed-Effects-Model, Hedges, 1994; Hedges & Vevea, 1998; Gurevitch & Hedges, 1999) eingesetzt, wo jeder Effekt mit seiner Varianz gewichtet wird. Die Gewichtung der Effektstärken führt vor dem Hintergrund der Theorie großer Zahlen zu präziseren Ergebnissen, denn es bewirkt eine Varianzminimierung (Cooper, 1998; Fricke & Treinies, 1985; Rosenberg, Adams & Gurevitch, 2000). In einem ersten Auswertungsschritt wurde die gewichtete Effektstärke  $D^W$  nach Hedges (1982) berechnet (Stetter & Kupper, 2002) und in einem weiteren Schritt wurde die erwartungstreue Effektstärkeschätzung  $D^{W+}$  nach Hedges (1982) berechnet, die eine leichte Korrektur zu  $D^W$  enthält, da in die Berechnung nicht die Hedges korrigierte Effektstärke  $d^u$  ( $u$ =unbiased, Hedges, 1982), sondern der Wert  $D^W$  selbst eingeht (vgl. Fricke & Treinies, 1985). Diese Methode ist unserer Meinung nach anderen vorzuziehen (z.B. Hunter, Schmidt & Jackson, 1982), da sie die Korrektur im Falle kleiner Stichproben berücksichtigt und damit einer Überschätzung des ES-Parameters vorbeugt. Im Falle der Signifikanz des Homogenitätstests ist die Aussage zugelassen, dass für jede einzelne Studie ein eigener Populationseffekt angenommen und von einer Verteilung der Populationseffekte ausgegangen werden kann. Hier wurde entsprechend des varianzanalytischen Modells das Random-Effects-Model (Hedges, 1983; Hedges & Olkin, 1985) angewendet, wobei die Varianz der beobachteten Werte nach der Methode von Fricke & Treinies (1985) jeweils mit ihrer Stichprobengröße gewichtet wurde („Hedges 83 (gew.)“).

#### 4.3.5 Signifikanz der Populationseffektstärken

Um die Fehler 1 und 2 gering zu halten schlagen Seipp (1989), Fricke und Treinies (1985) wie auch Bortz und Döring (1995) vor, das 95%-Konfidenzintervall zur Einschätzung heranzuziehen. Wenn es die Null umfasst, so ist von nicht signifikanten Unterschieden auszugehen, ist die Null nicht enthalten besteht Signifikanz. Eine zweite Methode zur Einschätzung bezieht sich darauf, dass die Populationseffektstärke mindestens zweimal so groß wie die Varianz der Populationseffekte sein sollte (Seipp, 1989; Schwarzer, 1995).

#### 4.3.6 Weitere Parameter

Zur Abschätzung der klinischen Bedeutsamkeit der Effektstärkedaten wurden zwei Methoden angewandt (vgl. Lambert & Bergin, 1994): die Berechnung des Binomial Effect Size Display (BESD; Rosenthal & Rubin, 1979, 1982) und die Ermittlung der Rate derjenigen Studien, die nach der Methode von Lambert & Bergin (1994) als „verbesserte“ ( $ES > 0.5$ ), „verschlechterte“ ( $ES < 0.5$ ) oder als „unveränderte“ ( $ES = 0.5$ ) Fälle zu kennzeichnen sind (Cut-Off:  $ES = 0.5$ ). Zur ausreichenden Beachtung des Publication Bias Form 1 wurde nach Begg (1994) ein Rangkorrelationstest eingesetzt. Wir gaben hier Kendalls Tau den Vorzug vor dem Spearmanschen Rho, da nur durch Anwendung dieses Prinzips eine möglichst exakte Prüfung der Signifikanz des Koeffizienten möglich ist (Kendall, 1948; Röhr, Lohse & Ludwig, 1983; Sokal & Rohlf, 1995). Die signifikante Korrelation zwischen  $E$  (standardisierte ES) und  $n$  markiert einen Publication Bias in der Art, dass große ES in einer Richtung

(z.B. positiv) häufiger publiziert werden als kleine ES (Rosenberg et al., 2000). Zusätzlich wurde der Publication Bias Form 2 berücksichtigt (Sterling, 1959; Greenwald, 1975) indem mit dem Fail-Safe-N nach Rosenthal (1979) die Anzahl zusätzlicher Studien mit einem Nulleffekt berechnet wurde, die erforderlich sind, um die kombinierte Signifikanz auf ein Level von 5% zu reduzieren. Rosenthal (1979) empfiehlt einen Wert von  $(5 \cdot n) + 10$  als konservativen kritischen Wert, um die Fail-Safe-N-Berechnung zu bewerten. Ergänzend wurde das nach Orwin (1983) für Effektstärken modifizierte Faile-Safe-N angegeben, das aussagt, wie viel Studien mit einem Nulleffekt erforderlich wären, um die Effektstärke auf ein Level von 0.2 Standardabweichungen zu drücken. Ein großes Fail-Safe-N zeigt also, dass viele unpublizierte Studien erforderlich wären, um die statistischen Resultate zu verändern, wodurch mehr Sicherheit diesbezüglich entsteht. Eine alternative Form wurde von Rosenberg et al. (2000) vorgestellt, wo die „modernere“ Metaanalyse stärker berücksichtigt wird, indem es unabhängig von der verwendeten Effektstärke ist, sondern lediglich errechnet, wie groß das zusätzliche Gewicht sein müsste, um die Signifikanz eines beobachteten kumulativen Effekts auf ein gegebenes  $\alpha$ -Level zu reduzieren. Ein Problem dieses Ansatzes besteht mit Sicherheit darin, dass die Freiheitsgrade des t-Tests auf der Anzahl der Studien basieren, die zur Errechnung des Durchschnittwertes ES benutzt wurden, der Vorteil liegt jedoch darin, dass hiermit eine noch exaktere Bestimmung der Anzahl der nichtsignifikanten, unpublizierten Studien, die hinzugenommen werden müssten, um das Resultat von signifikant zu nicht signifikant zu verändern, vorgeschlagen wird.

#### 4.3.7 Inferenzstatistik

Als Rechenhilfe wurden das „Statistical Package for Social Sciences“, Version 10.0.5 (SPSS for Windows, SPSS, 1999), das „EQS: A Structural Equation Program“, Version 5.6 (Bentler, 1997) und das „Meta-Analyse“- Programm von Fricke und Kreft (1986) hinzugezogen. Für begleitende Analysen kamen auch die Computerprogramme „MetaWin“ (Rosenberg et al., 2000), „Easy MA2000“ (Cuchera, 2000), „META“ (Kenny, 1999) und das „Meta-Analyse“-Programm von Schwarzer (1995) zur Anwendung.

## 5 Ergebnisse und Auswertung

### 5.1 Deskriptive Auswertung

Eine Übersichtstabelle zu allen untersuchten Studien findet sich in Stetter und Kupper (1998, 2002). Dort wurden die einzelnen Studien detailliert dargestellt geordnet nach einzelnen Krankheitsbildern zuzüglich der Angabe der Anzahl der Behandlungs- und Kontrollgruppen, der Behandlungsbedingungen, der methodischen Kommentare, der Ergebnisvariablen, der Stichprobengröße, der Länge der Behandlung und der Katamnese sowie der Bewertung der klinischen Hauptsymptomatik.

### 5.2 Metaanalytische und Inferenzstatistische Auswertung

Die Ergebnisse der Metaanalyse sind dargestellt in den Tabellen 1 bis 6, Detailergebnisse in Stetter und Kupper (2002). Die Tabellen sind folgendermaßen aufgebaut: kumulierte Effektstärke (homogener Datensatz:  $D^{W+}$ , heterogener

Datensatz: „Hedges 83 (gew.)“, zugehöriges  $r$ , Binomial Effect Size Display, Rate der Verbesserungen (alle Studien  $d > 0.5$ ), Verschlechterungen (alle Studien  $d < 0.5$ ), Unveränderten (alle Studien  $d = 0.5$ ), Homogenitätsbeurteilung anhand  $Q$  und seiner Wahrscheinlichkeit, Prozentsatz der Varianz, die auf Variation in den Populationseffekten zurückzuführen ist,  $\frac{1}{4}$  der ES, Varianzzerlegung (wurde auch für homogene Datensätze angegeben, um den Prozentsatz der Varianz, die auf den Stichprobenfehler zurückzuführen ist, abzuschätzen), 95 %-Konfidenzintervall zur Signifikanzbeurteilung der Effektstärke, Kendalls Tau und seine Wahrscheinlichkeit zu Abschätzung des Publication bias, Form1 (nur große ES werden publiziert), Rosenthals Faile-Safe-N und Orwins-Faile-Safe-N zur Abschätzung des Publication bias, Form 2 (nur signifikante Studien werden veröffentlicht), Anzahl der Patienten (N), Anzahl der Studien (k), Anzahl der Einzel-ES, die pro Studie zu einer ES aggregiert wurden. Die folgende Ergebnisdarstellung ist gleichzeitig als Orientierung zum Lesen der Tabellen zu verstehen.

Das zentrale Ergebnis gibt die erste Spalte von Tabelle 3 wieder. Für den Vergleich von AT versus wahre Kontrollgruppe zum Behandlungsende (Outcome) für randomisierte Studien wird eine homogene Effektstärke von  $ES = 0.624$ , transformiert zu einem  $r = 0.298$ , ermittelt. Es erfolgt eine Erfolgssteigerung dadurch, dass sich die Patienten in der AT-Gruppe befinden von 35 auf 65 % (BESD). Wenn ein Cut-Off von  $ES = 0.5$  gesetzt wird, sind 68 % der Fälle als gebessert, 4 % als unverändert und 28 % als verschlechtert einzuschätzen. Die Homogenitätsbeurteilung erfolgt folgendermaßen: 1. das  $Q < \chi^2_{0.01, f=24}$ ,  $p_Q > 0.01$ , 2. ein Viertel der ES  $>$  Restvarianz ( $V(\Delta)$ ), 3. Die verbleibende Varianz der Populationseffekte macht nur noch 3.7 % der Varianz der beobachteten Effekte aus; alle drei Kriterien sprechen also für Homogenität. Als nächstes wird die Signifikanz der Populationseffekte beurteilt ( $H_0 = \delta = 0$ ): 1. die Grenzen des 95% Konfidenzintervalls befinden sich auf der positiven Seite, 2. die Effektstärke ist mindestens zweimal so groß wie die Restvarianz ( $V(\Delta)$ ); beide Kriterien sprechen also für vorliegende Signifikanz, d. h. die Differenz von Null kann nicht auf zufällige Stichprobenfehler zurückgeführt werden, sondern ist Ausdruck einer Verschiedenheit der Grundgesamtheit von Null. Der Betrag von Kendalls Tau ist kleiner als die Zufallshöchstwerte von Tau ( $|\tau| \leq \tau_{\max; 0.05; 25}$ ) und seine Wahrscheinlichkeit  $p_{\tau} > 0.05$ , d.h.  $D^{W+}$  und N sind unabhängig voneinander; es besteht keine Rangkorrelation, was anzeigt, dass der Publication Bias Form 1 zu vernachlässigen ist, das erzielte Ergebnis nicht nur auf die Tatsache zurückzuführen ist, das in den untersuchten Publikationen zum großen Teil nur große ES veröffentlicht wurden (fast das Gegenteil ist der Fall, denn  $p_{\tau} = 0.71!$ ). Rosenthals Fail-Safe-N beträgt 674 und ist damit größer als der kritische Wert von 135, was bezüglich des Publication Bias Form 2 aussagt, dass die Signifikanz des Ergebnisses von  $D^{W+} = 0.624$  nicht nur deshalb zustande kam, weil vielleicht eine Reihe von anderen existierenden, jedoch unveröffentlichten Studien mit nicht signifikanten Effekten nicht berücksichtigt wurden. Orwins Fail-Safe-N beträgt Null, was zur Erhöhung der Sicherheit der bereits getroffenen Schlussfolgerungen leider keinen Betrag leistet. Die Anzahl der Patienten war 1039, die Studienanzahl 25 und die zugrundegelegten Erfolgsmaße, wenn die Summe über alle Studien gebildet wird, 88, was nur zur Orientierung angegeben wird, um mit den historisch zurückliegenden Metaanalysen (z.B. Smith, Glass & Miller, 1980; Shapiro & Shapiro, 1982) grob zu vergleichen.

Die folgende Ergebnisbeurteilung stützt sich auf die randomisierten Studien ( $r$ ). Zur Orientierung wurden in Klammern zusätzlich auch die Ergebnisse der nicht

randomisierten Studien (n-r) angegeben. Der Beurteilung wurde die Konvention von Cohen (1977/1969; 0.2= kleiner Effekt, 0.5= mittlerer Effekt, 0.8= großer Effekt) zugrundegelegt.

### 5.2.1 Psychosomatische Störungen

Die Metaanalyse der einbezogenen Studien bei psychosomatischen Störungen (N= 46, davon randomisiert = 31) lässt nachfolgendes Bild entstehen.

#### 5.2.1.1 Prä-Post-/Prä-Follow-up

- Mittlerer Effekt auf die Hauptzielsymptomatik (randomisiert (r),  $D^{W+}$ : 0.74; nicht randomisiert (n-r),  $D^{W+}$ : 0.38) und mittlerer Effekt auf die unspezifische Symptomatik (r,  $D^{W+}$ : 0.52; n-r,  $D^{W+}$ : 0.45) für Outcome.
- Großer Effekt auf die Hauptzielsymptomatik (r, Hedges 83 (gew.): 0.89; n-r,  $D^{W+}$ : 0.36) und mittlerer Effekt auf die unspezifische Symptomatik (r,  $D^{W+}$ : 0.53; n-r,  $D^{W+}$ : 0.52) für Follow-up.

#### 5.2.1.2 AT versus Kontrollgruppe

- Mittlerer Effekt auf die Hauptzielsymptomatik (r,  $D^{W+}$ : 0.62; n-r,  $D^{W+}$ : 0.56) und kleiner Effekt auf die unspezifische Symptomatik (r,  $D^{W+}$ : 0.27; n-r,  $D^{W+}$ : 0.35) für Outcome.
- Kleiner Effekt auf die Hauptzielsymptomatik (r, Hedges 83 (gew.): 0.49; n-r,  $D^{W+}$ : 0.25) und kleiner Effekt auf die unspezifische Symptomatik (r,  $D^{W+}$ : 0.46; n-r: keine) für Follow-up.

#### 5.2.1.3 AT versus andere psychologische Verfahren

- Kleiner negativer Effekt auf die Hauptzielsymptomatik (r,  $D^{W+}$ : -0.32; n-r,  $D^{W+}$ : -0.03) und kein Effekt auf die unspezifische Symptomatik (r,  $D^{W+}$ : 0.008; n-r,  $D^{W+}$ : 0.30) für Outcome.
- Kleiner negativer Effekt auf die Hauptzielsymptomatik (r,  $D^{W+}$ : -0.22; n-r,  $D^{W+}$ : -0.22) und kein Effekt auf die unspezifische Symptomatik (r,  $D^{W+}$ : 0.05; n-r,  $D^{W+}$ : 0.21) für Follow-up.

### 5.2.2 Psychische Störungen

Die Metaanalyse der einbezogenen Studien bei psychischen Störungen (N= 14, davon randomisiert = 4) lässt nachfolgendes Bild entstehen.

#### 5.2.2.1 AT-Prä-Post/Prä-Follow-up

- Großer Effekt auf die Hauptzielsymptomatik (r,  $D^{W+}$ : 0.87; n-r,  $D^{W+}$ : 0.67) und großer Effekt auf die unspezifische Symptomatik (r, Hedges 83 (gew.): 1.73; n-r,  $D^{W+}$ : 1.08) für Outcome.
- Großer Effekt auf die Hauptzielsymptomatik (r,  $D^{W+}$ : 1.12; n-r,  $D^{W+}$ : 0.83) und zur Beurteilung der Wirkung auf die unspezifische Symptomatik liegt keine Effektstärke vor (r: keine ES; n-r,  $D^{W+}$ : 0.64) für Follow-up.

### 5.2.2.2 AT versus Kontrollgruppe

- Mittlerer Effekt auf die Hauptzielsymptomatik ( $r, D^{W+}$ : 0.58; n-r,  $D^{W+}$ : 1.03) und zur Beurteilung der Wirkung auf die unspezifische Symptomatik liegt keine Effektstärke vor ( $r$ : keine ES; n-r,  $D^{W+}$ : 1.01) für Outcome.
- Kleiner Effekt auf die Hauptzielsymptomatik ( $r, D^{W+}$ : 0.33; n-r,  $D^{W+}$ : 0.50) und zur Beurteilung der Wirkung auf die unspezifische Symptomatik liegt keine Effektstärke vor ( $r$ : keine; n-r: keine) für Follow-up.

### 5.2.2.3 AT versus andere psychologische Behandlung

- Kleiner negativer Effekt auf die Hauptzielsymptomatik ( $r, D^{W+}$ : -0.24; n-r,  $D^{W+}$ : 0.15) und großer Effekt auf die unspezifische Symptomatik ( $r, D^{W+}$ : 1.81; n-r,  $D^{W+}$ : 0.37) für Outcome.
- Kein Effekt auf die Hauptzielsymptomatik ( $r, D^{W+}$ : 0; n-r,  $D^{W+}$ : -0.15) und zur Beurteilung der Wirkung auf die unspezifische Symptomatik liegt keine Effektstärke vor ( $r, D^{W+}$ : keine ES; n-r,  $D^{W+}$ : 0.05) für Follow-up.

## 5.2.3 Alle Störungen

Die Metaanalyse der einbezogenen Studien bei allen Störungen ( $N= 60$ , davon randomisiert = 35) lässt nachfolgendes Bild entstehen.

### 5.2.3.1 AT-Prä-Post/Prä-Follow-up

- Mittlerer Effekt auf die Hauptzielsymptomatik ( $r, D^{W+}$ : 0.78; n-r,  $D^{W+}$ : 0.44) und mittlerer Effekt auf die unspezifische Symptomatik ( $r$ , Hedges 83 (gew.): 0.73; n-r, Hedges 83 (gew.): 0.84) für Outcome.
- Großer Effekt auf die Hauptzielsymptomatik ( $r$ , Hedges 83 (gew.) : 0.80; n-r,  $D^{W+}$ : 0.53) und mittlerer Effekt auf die unspezifische Symptomatik ( $r$ , Hedges 83 (gew.) : 0,55; n-r,  $D^{W+}$ : 0.62) für Follow-up.

### 5.2.3.2 AT versus Kontrollgruppe

- Mittlerer Effekt auf die Hauptzielsymptomatik ( $r, D^{W+}$ : 0.62; n-r, Hedges 83 (gew.): 0.71) und mittlerer Effekt auf die unspezifische Symptomatik ( $r$ , Hedges 83 (gew.) : 0.57; n-r, Hedges 83 (gew.): 1.41) für Outcome.
- Mittlerer Effekt auf die Hauptzielsymptomatik ( $r$ , Hedges 83 (gew.) = 0.53, n-r,  $D^{W+}$ : 0.31) und großer Effekt auf die unspezifische Symptomatik ( $r$ , Hedges 83 (gew.) : 1.05; n-r,  $D^{W+}$ : keine) für Follow-up.

### 5.2.3.3 AT versus andere psychologische Behandlung

- Kleiner negativer Effekt auf die Hauptzielsymptomatik ( $r, D^{W+}$ :-0.29; n-r,  $D^{W+}$ : -0.09) und kleiner Effekt auf die unspezifische Symptomatik ( $r$ , Hedges 83 (gew.) : 0.26; n-r,  $D^{W+}$ : 0.34) für Outcome.
- Kleiner negativer Effekt auf die Hauptzielsymptomatik ( $r = -0.21$ ; n-r,  $D^{W+}$ : -0.18) und kein Effekt auf die unspezifische Symptomatik ( $r, D^{W+}$ : 0.04; n-r,  $D^{W+}$ : 0.34) für Follow-up.



## 6 Diskussion

Die Untersuchung ist grundsätzlich folgender Themenstellung zuzuordnen: Metaanalyse klinischer Studien zur Wirksamkeit des Autogenen Trainings zum Outcome- und Follow-up-Zeitpunkt.

Mit dieser Fragestellung sollen die Ergebnisse nachfolgend kurz bewertet werden.

Das zentrale Ergebnis spiegelt sich in mittelgroßen Effekten der Behandlungsergebnisse wider, die Follow-up-Werte weisen bezüglich der Hauptzielsymptomatik sogar auf große Effekte des AT hin. Im Vergleich zur Kontrollgruppe sind die Effekte des AT bezüglich der Hauptzielsymptomatik sowohl zum Outcome als auch Follow-up-Zeitpunkt im mittleren Bereich, die Auswirkung auf die Stimmung, Verbesserung der Lebenszufriedenheit und kognitive Leistungsfähigkeit repräsentiert sich im großen Effekt des Autogenen Trainings. Im Vergleich zu den Auswirkungen anderer Behandlungsverfahren sind die Effekte als klinisch negativ bis klinisch positiv einzustufen, woraus sich keine deutlichen Vor- oder Nachteile zum Autogenen Training gegenüber anderen Entspannungsverfahren ableiten lassen. Wie stellt sich vor dem Hintergrund dieser Ergebnisse die Verifizierung bzw. die Falsifizierung der Hypothesen dar?

In Hypothese 1 wurde postuliert, dass das Autogene Training zu signifikanten Prä-Post- und Prä-Follow-up-Effekten führt. Diese Alternativhypothese kann angenommen werden, da das AT durchweg zu signifikanten Effekten im Vorher-Nacher-Vergleich zum Outcome- und Follow-up-Zeitpunkt führte. Bei psychischen Störungen wurden große Effekte, bei psychosomatischen Störungen mittlere Effekte gefunden. Das Ergebnis legt einen noch stärkeren AT-Effekt als in der Studie von Linden (1994; biologische Erfolgsmaße  $ES=0.43$ , psychologisch-behaviorale Erfolgsmaße  $ES=0.58$ , arithmetisches Mittel=  $0.505$ ) nahe.

Mit Hypothese 2 wurde die Annahme aufgestellt, dass das Autogene Training der Kontrollgruppenbedingung überlegen ist. Hierfür wurden eindeutige Hinweise festgestellt. Der klinische Outcome für die Patienten der AT-Gruppen ist als günstiger einzustufen, als für die Patienten der Kontrollgruppenbedingungen mit in der Regel mittleren Effekten. Dies entspricht bzw. überschreitet leicht die Ergebnisse aus der Untersuchung von Linden (1994, biologische Erfolgsmaße  $ES=0.24$ , psychologisch-behaviorale Erfolgsmaße  $ES=0.67$ , arithmetisches Mittel=  $0.455$ ).

Mit Hypothese 3 wurde unterstellt, dass das Autogene Training eine ähnliche Wirksamkeit wie andere Entspannungsverfahren der aktuellen Untersuchung aufweist. Hierfür wurde keine vollständige Bestätigung erbracht. Kleine negative Effekte zugunsten des anderen psychologischen Entspannungsverfahrens wurden festgestellt. Bei psychischen Störungen wurde zum Follow-up-Zeitpunkt kein Effekt festgestellt. Das entspricht den Angaben von Linden (1994, biologische Erfolgsmaße  $ES=0.00$ , psychologisch-behaviorale Erfolgsmaße  $ES=-0.06$ , arithmetisches Mittel=  $-0.030$ ).

In Hypothese 4 wurde angenommen, dass sich das Autogene Training signifikant auf unspezifische Bereiche des Erlebens und Verhaltens z. B. der Stimmung auswirkt. Diese Hypothese wird unterstützt in den Vergleichen AT-Prä-Post-/Prä-Follow-up

und AT versus Kontrollgruppe, wo mittlere bis große AT-Effekte festgestellt wurden. Somit darf die Hypothese als verifiziert gelten.

Die hier vorgelegten Ergebnisse zur Wirksamkeit des Autogenen Trainings stimmen sehr gut mit den Ergebnissen anderer Metaanalysen von Entspannungsverfahren überein. Für den Vergleich mit einer unbehandelten Kontrollgruppe finden Hyman, Feldman, Harris, Levin und Malloy (1989) in ihrer Metaanalyse über 55 Studien (das Kriterium Randomisierung wurde als mittel bis hoch geratet) zur Wirksamkeit von Entspannungsverfahren bei Angsterkrankungen (16.4%), Hypertonie (16.4%), Insomnie (20%) und akutem Schmerz (18.2%) für die Progressive Relaxation eine  $ES=0.63$  (21 Studien), für die Hypnose eine  $ES=0.54$  (2 Studien) und für das Autogene Training eine  $ES=0.46$  (11 Studien). Linden (1994) berichtet für den Vergleich AT Vorher-Nachher zum Outcomezeitpunkt für biologische Messungen eine  $ES=0.43$ , für psychologisch-verhaltensmäßige Messungen eine  $ES=0.58$  (arithmetisches Mittel  $ES=0.50$ ) und für den Vergleich AT versus unbehandelte Kontrollgruppe für biologische Messungen eine  $ES=0.36$ , für psychologisch-verhaltensmäßige Messungen eine  $ES=0.67$  (arithmetisches Mittel  $ES=0.51$ ) sowie für den Vergleich AT versus Placebo für biologische Messungen eine  $ES=0.51$ , für psychologisch-verhaltensmäßige Messungen eine  $ES=0.24$  (arithmetisches Mittel  $ES=0.37$ ). Das arithmetische Mittel der  $ES$  über die Vergleiche AT versus unbehandelte Kontrollgruppe und AT versus Placebo beträgt 0.44. Dieses ist der orientierende Vergleichswert für unser Ergebnis aus dem Vergleich AT versus wahre Kontrollgruppe (Erläuterung siehe oben) von  $ES=0.62$  ( $r=0.298$ , 25 Studien,  $N=1039$ , randomisiert, homogen). Im Vergleich AT Vorher-Nachher zum Outcomezeitpunkt ermittelten wir eine  $ES=0.78$  ( $r=0.362$ , 33 Studien,  $N=1530$ , randomisiert, homogen). Auch mit den Ergebnissen der kürzlich erschienenen Metaanalyse von Bongartz, Flammer & Schwonke (2002) zur Hypnose im Sinne eines Entspannungsverfahrens stimmt dieses Ergebnis gut überein. Bongartz et al. (2002) teilten für den Vergleich Hypnose versus Kontrollgruppe (Wartegruppen oder nur medizinische Standardversorgung) eine  $ES=0.51$  ( $r=0.24$ , 43 Studien,  $N=1935$ , randomisiert, heterogen) und für den Vergleich Hypnose Vorher-Nachher zum Outcomezeitpunkt einen Wert von 0.90 ( $r=0.39$ , 11 Studien,  $N=348$ , randomisiert, heterogen) mit. Für nicht randomisierte Studien ermittelten wir im Vergleich AT versus wahre Kontrollgruppe (Erläuterung siehe oben) Heterogenität, wobei die verbleibende Varianz der Populationseffekte 69 % der Varianz der beobachteten Effekte ausmacht und nur 31 % dieser Varianz auf den Stichprobenfehler zurückzuführen sind ( $ES=0.71$ ,  $r=0.335$ , 16 Studien,  $N=1423$ , nicht randomisiert, heterogen) und im Vergleich AT Vorher-Nachher zum Outcomezeitpunkt eine  $ES=0.436$  ( $r=0.213$ , 22 Studien,  $N=1720$ , nicht randomisiert, homogen). Als systematischer Faktor wurde die Störungsspezifität isoliert und die Berechnung der Effektstärken für die Krankheitsbilder führte bis auf einen Analysefall (Vergleich AT versus wahre Kontrollgruppe bei Angststörungen, randomisiert, Outcome) zu Homogenität und lässt damit entsprechende Interpretationen zu ( $ES$ : 0.33 [somatoforme Schmerzstörung] bis 1.11 [Depression]). Für Detailergebnisse verweisen wir auf Stetter und Kupper (2002). Bongartz et al (2002) errechneten für nicht randomisierte Studien in dem Vergleich Hypnose versus Kontrollgruppe (Wartegruppen oder nur medizinische Standardversorgung) eine  $ES=0.91$  ( $r=0.39$ , 17 Studien,  $N=1255$ , nicht randomisiert, heterogen) und für den Vergleich Hypnose Vorher-Nachher zum Outcomezeitpunkt einen Wert von 2.00 ( $r=0.61$ , 16 Studien,  $N=498$ , nicht

randomisiert, heterogen). Leider ist die Vergleichbarkeit unserer Daten zum Beispiel mit den Daten von Bongartz et al. (2002) der Einschränkung unterworfen, dass bis auf die Ausnahme einer Untergruppe (Hypnose als Unterstützung medizinischer Maßnahmen,  $ES=0.38$ ,  $r=0.19$ ) sämtliche ES-Daten von Bongartz et al. (2002) heterogen sind. Im Extremfalle heißt dies für die Daten zur Hauptinterpretation (Hypnose versus Kontrollgruppe) von Bongartz et al. (2002), dass die 43 Effektstärken aus 43 unterschiedlichen Populationen von Effektstärken stammen, und sich dann die Frage aufwerfen würde, was eine kummulierte ES über diese möglicherweise 43 Populationen aussagen soll. Es ist hier üblich, in der oben beschriebenen Art und Weise eine Varianzzerlegung vorzunehmen, um festzustellen, welcher Prozentsatz der Varianz auf Stichprobenfehler und welcher Anteil auf die Variation in den Populationeffekten zurückzuführen ist. In einem zweiten Schritt sind diejenigen systematischen Faktoren ausfindig zu machen (z.B. durch Voruntersuchungen, theoretische Überlegungen, Clusteranalysen, konfirmative Faktorenanalysen, Strukturgleichungsmodelle u.ä.), mit denen diese Variation der Effekte erklärt werden kann. Bongartz et al. (2000) stellen zwar fest, dass sich keine diesbezüglichen Untergruppen finden ließen, formulieren jedoch leider keine entsprechenden Konsequenzen bezüglich einer Einschränkung der Interpretation. Dessen ungeachtet kommt die von uns im Vergleich AT Vorher-Nachher zum Follow-up-Zeitpunkt errechnete  $ES=0.805$  (randomisierte Studien) in die Nähe des Gesamteffekts von Psychotherapie überhaupt ( $ES=0.85$ ; Smith, Glass & Miller, 1980,  $N_{ES}=1761$ , 475 Studien). Der Signifikanztest zeigt zwar Heterogenität an, wir haben uns dennoch für die Diskussion dieses Wertes entschieden, weil die zwei anderen Hauptkriterien für die Beurteilung der Homogenität (siehe oben) für Homogenität sprechen (vgl. Tabelle 1). Der hier ermittelte Effekt im Vergleich AT versus wahre Kontrollgruppe zum Outcomezeitpunkt  $ES=0.624$  ( $r=0.298$ ,  $N=1039$ , 25 Studien) geht tendentiell in die Richtung des Effekts, den Smith et al. (1980) für psychodynamische Therapie ( $ES=0.69$ ) oder Rational-emotive Therapie ( $ES=0.68$ ) angeben. Ferner liegt er im Bereich des Effektes für die klientenzentrierte Gesprächstherapie ( $ES=0.62$ , Smith et al., 1980), adlerianische Therapie ( $ES=0.62$ , Smith et al., 1980), die Gestalttherapie ( $ES=0.64$ , Smith et al., 1980) sowie etwas höher als die Familientherapie ( $ES=0.44$ , 85 Studien, GIBLIN, SPRENKLE & SHEEHAN, 1985) und die paradoxe Behandlungsintervention ( $ES=0.42$ , 10 Studien; SHOHAM-SOLOMON & ROSENTHAL, 1987). Ferner liegt der von uns ermittelte Effekt leicht höher als Placebo ( $ES=0.56$ , Smith et al., 1980), wobei Lambert, Weber und Sykes (1993) mit  $ES=0.42$  (596 Studien, 10 Metaanalysen) einen Effekt angeben, der noch weiter darunter liegt, was dem sogenannten unspezifischen Behandlungseffekt im Sinne von Bowers und Clum (1989) entspräche (Placebo versus keine Behandlung). Der spezifische Effekt von Psychotherapie, der sich aus der Gegenüberstellung von echten Behandlungsgruppen und Placebo-Kontrollgruppen im Sinne von Bowers und Clum (1989) ergibt, beträgt nach Lambert et al. (1993)  $ES=0.48$  ( $N=250$ , 8 Metaanalysen), der reine Psychotherapieeffekt (Psychotherapie versus unbehandelte Kontrollgruppe)  $ES=0.82$  ( $r=0.38$ ,  $N=2309$ , 13 Metaanalysen, 1080 Studien). Dies entspricht einer Steigerung des Erfolgs durch das Erhalten einer psychotherapeutischen Maßnahme von 31 % auf 69 %. In dieser Berechnung ist jedoch auch die ES der Metaanalyse zu Entspannungsverfahren enthalten (Hyman et al., 1989), weshalb wir diese Daten von Lambert et al. (1993) reanalysierten. Wir ermittelten eine Hedges korrigierte  $ES=0.80$  (1024 Studien,  $r=0.38$ ). Verglichen mit dem oben angegebenen Ergebnis kann Deckungsgleichheit konstatiert werden,

wobei natürlich einzuschränken bleibt, dass dies nur als grobe Orientierung zu verstehen ist, da unser Vergleichswert aus dem Vergleich AT Vorher-Nachher zum Follow-up-Zeitpunkt stammt ( $ES=0.805$ ).

Die nächste Frage die sich stellt besteht darin, was sagt diese ES eigentlich aus und wie kann sie für die klinische Wirklichkeit interpretiert werden? Im „Handbook of Psychotherapy and Behavior Change“ von Bergin und Garfield (1994) wird darauf hingewiesen, dass eine Möglichkeit dafür die Berechnung des Binomial Effect Size Displays (BESD) sei (Rosenthal & Rubin, 1979, 1982; Rosenthal, 1983). Sie demonstrieren das an dem Beispiel, dass über alle wichtigen Metaanalysen (49 Metaanalysen; Lambert & Bergin, 1994) zur Psychotherapie die bislang vorgenommen wurden, resümiert werden kann, dass die Erfolgsresultate der Patienten der Behandlungsgruppe von denen der unbehandelten Kontrollgruppe im Outcome etwa um 10 % variieren. Dieser Varianzanteil dafür, sich in Therapie zu befinden ist äquivalent mit einer Erfolgssteigerung von 34 auf 66 % (vgl. Lambert & Bergin, 1994). Wir ermittelten in unserer Untersuchung im Rahmen der Hauptstichprobe AT vs. wahre Kontrollgruppe zum Outcomezeitpunkt, dass sich dadurch, dass sich die Patienten in der AT-Behandlungsgruppe befanden die Erfolgsrate von 35 % auf 65 % steigerte, was dem oben angegebenen Wert für die Erfolgsverbesserung von einer unbehandelten Kontrollgruppe durch die Zuweisung einer Behandlungsgruppe (34 % auf 66 %) von Lambert und Bergin (1994) sehr gut entspricht. Der Nutzen der BESD wurde vielfach kritisiert (Crow, 1991; McGraw, 1991; Strahan, 1991), was von Rosenthal (1990, 1991) zwar zurückgewiesen wurde, jedoch zu anderen Überlegungen führte. Eine weitere Möglichkeit, die Effektstärkedaten zu mehr klinisch bedeutungsvollen Aussagen zu transformieren sehen Lambert und Bergin (1994) darin, einen Cut-Off von 0.5 Standardabweichungen einzuführen. Sie demonstrieren das am Beispiel der Studie von Asay, Lambert, Christensen und Beutler (1984;  $ES=0.92$ ,  $N=2405$ ), wo nach diesem Vorgehen 66 % der Fälle als gebessert, 26 % als unverändert und 8 % als verschlechtert beschrieben werden können. Angewendet auf unsere Daten war eine Verbesserung in 68 % der Fälle zu verzeichnen, keine Veränderung in 4 % und eine Verschlechterung bei 28 % der untersuchten Fälle (Vergleich AT versus wahre Kontrollgruppe zum Outcomezeitpunkt). In der Gesamteinschätzung zeigt sich somit ein akzeptables Ergebnis. Abschließend erwähnen die Autoren auch den Einsatz des Konzeptes der Clinical Significance (Jacobson, Folette & Revensdorf, 1984), das besagt, das eine Veränderung dann als klinisch signifikant einzuschätzen ist, wenn die Veränderung den Messfehler überschreitet, gemessen durch den Reliable Change Index (absolute Größe der Veränderung geteilt durch den Standardmessfehler). Eine grafische Anwendung dieser Technik demonstrieren Lambert und Hill (1994) am Beispiel der Studie von Tingey, Burlingame, Lambert und Barlow (1989). Diesen Ansatz haben wir zum gegenwärtigen Zeitpunkt nicht auf unsere Daten angewandt.

Bei einem Vergleich dieser Ergebnisse mit dem „reinen Psychotherapieeffekt“ von Grawe, Donati und Bernauer (1994) von  $ES= 1.11$ , was einem BESD von 25 % auf 75 % Verbesserung durch Behandlung entspräche, fällt eine große Differenz auf. Auch fällt auf dass die ES von Grawe et al. (1994) bei weitem das übersteigt, was aus vielen anderen Metaanalysen ermittelt wurde (Lambert & Bergin, 1994: 15 Metaanalysen,  $ES=0.82$ ). Wichtig anzumerken ist, dass Grawe et al. (1994) zur Beurteilung nur 41 Studien zugrundelegen, weil das Ziel eigentlich die Ermittlung von Hinweisen auf eine differentielle Therapieindikation ist. Der Gesamteffekt basiert auf

1401 Einzeleffekten, was einer Messwiederholung von zirka 35 pro Studie entspricht. Grawe et al. (1994) verletzen ohnehin schon die Unabhängigkeitsforderung für Metaanalysen (Landman & Dawes, 1982; Kraemer, 1983; Bortz & Döring, 1995) gravierend, weil sie im Rahmen ihrer Prä-Post-Vergleiche die Teilstichproben doppelt zählen. Dies potenziert sich jedoch noch zusätzlich, indem jede Teilstichprobe nicht einfach, sondern 70fach gezählt wird. In jedem Standardwerk zur Metaanalyse kann nachgeschlagen werden, dass die Untersuchungseinheiten einer Metaanalyse verschiedene Studien sind und nicht die Teilergebnisse der Studien, es sei denn es läge nur ein Ergebnis pro Studie vor (Bortz & Döring, 1995; Cook, T. D., Cooper, H., Cordray, D. D., Hartmann, H., Hedges, L. V., Light, R. J., Louis, T. & Mosteller, F., 1992; Cooper, 1984; Cooper, 1998; Cooper & Hedges, 1994; Draper, Graver, Goel, Greenhouse, Hedges, Morris, Tucker & Watenaux, 1992; Durlak & Lipsey, 1991; Fricke & Treinies, 1985; Hedges & Olkin, 1985; Hunter & Schmidt, 1990; Hunter, Schmidt & Jackson, 1982; Kraemer, 1983; Wolf, 1986). Aus diesem Grunde haben wir Grawe et al.'s (1994) Daten reanalysiert und kommen in dem Vergleich von Psychotherapiegruppe und unbehandelter Kontrollgruppe auf eine  $ES=0.82$  (19 Studien,  $r=0.398$ ,  $BESD=0.30$  auf  $0.70$ ), was auch sehr gut mit den Angaben von Lambert & Bergin (1994) über 49 Metaanalysen ( $BESD: 0.34$  auf  $0.66$ ) und über 13 Metaanalysen ( $ES=0.82$ ) übereinstimmt. Grawe et al. (1994) kommen auf den reinen Psychotherapieeffekt, indem sie die Prä-Post-ES für unbehandelte Kontrollgruppen  $ES=0.1$  ( $N=111$ , Hess, unveröffentl. Diplomarbeit, zit. nach Grawe, 1992) von der Prä-Post-ES für behandelte Gruppen  $ES=1.21$  ( $N=41$ , Grawe et al., 1994) subtrahieren. Wir sehen hier bei dem Vorgehen von Grawe et al. (1994) neben dem üblichen Äpfel-Birnen-Problem auch das Karotten-und-Pflaumen-Problem, was sich u.E. nach auch nicht mehr mit dem Argument von Smith et al. (1980) entkräften lässt: „Tatsächlich vermischt der Ansatz [der Metaanalyse, Anmerkung d. Verf.] Äpfel und Orangen, wie es jemand notwendigerweise ja auch tun würde, wenn er Früchte untersuchen möchte“ (Übers. v. Verf.). Es werden u.E. nach Ergebnisse unterschiedlicher Untersuchungen auf methodisch unzulässige Art und Weise miteinander vermischt und in Beziehung gesetzt. Die Kritik an der Arbeit von Grawe et al.'s (1994) wurde an anderen Stellen differenzierter ausgeführt (Geuter, 1995; Heckrath & Dohmen, 1997; Hillecke, Krämer, Kupper, Kupper-Horster & Kächele, 1996; Kächele, 1995; Kupper, 1998; Meyer, 1994; Meyer, 1995; Rüger, 1994; Sponzel, 2000a,b,c,d; Tschuschke, Heckrath & Tress, 1997; Tschuschke, Kächele & Hölzer, 1994 u.a.), sie sollen hier nicht erneut aufgewärmt werden. Obwohl Grawe et al. (1994) zurecht an der Metaanalyse von Smith, Glass und Miller (1980) die Verwendung von Studien zur Wirksamkeit von Psychotherapie bei Angsterkrankungen/Phobien die Studentenspopulationen zugrunde legen kritisiert, nutzt er die Ergebnisse der Smith et al.-Studie (1980) als Vergleichsmaßstab für die von ihm ermittelten Ergebnisse. Was uns eigentlich Anstrengung bereitet an Grawe et al.'s (1994) Vorgehen nachzuvollziehen ist die Tatsache, wie er aus z. B. 19 Vergleichsstudien Psychoanalyse versus Verhaltenstherapie 487 Einzelvergleiche und Einzel-ES herausrechnet, was im übrigen einer Messwiederholung von über 25 pro Studie entspricht. In 4 Studien wurde die Psychoanalyse nur weniger als 10 Stunden lang durchgeführt, wobei bekannt ist dass die psychoanalytische Kurzzeittherapie mindestens 30 benötigt zur vollständigen Entfaltung ihrer Wirksamkeit, insbesondere der therapeutischen Beziehung. Auf dieses methodische Manko der zugrundegelegten Studien machen Buol und Endtner (1993), aus deren Lizentiatsarbeit (nur in der Schweiz verliehener akadem. Grad) Grawe et al. (1994)

die Datenauswertungen vollständig übernehmen, aufmerksam, nicht aber Grawe et al. (1994). Ferner treten nicht nur statistisch starke Verletzungen der Methodik auf, wenn Grawe et al. (1994) den Binomialtest bei abhängigen Messungen einsetzen mit der Entgegnung auf die diesbezügliche Kritik von Diepgen (1993), dass die Korrelationen zwischen den verschiedenen Veränderungsmaßen sehr gering seien. Die ungenaue und methodisch nicht nachvollziehbare Vorgehensweise Grawe et al.'s (1994) setzt stochastische Unabhängigkeit mit Unkorreliertheit gleich, was jedoch keinen Vergleich zulässt (Sponsel, 2000a,b,c,d; Clauss & Ebner, 1989; Bortz, 1995). Ebenfalls wurden in der Grawe et al.-Studie (1994) keine ausschließlich relevanten Zielgrößen verwendet, sondern „jedes Maß, mit dem zu irgendeinem Zeitpunkt nach Beginn der Therapie das Ausmaß der seit Therapiebeginn eingetretenen Veränderung zu erfassen versucht worden ist“ (Grawe et al., 1994). Dass kontrollierte klinische Studien mit wenigen unter Umständen sogar nur einer einzigen relevanten Zielgröße auskommen können (z.B. die Variable „Alkohol trinken“ bei der Untersuchung der Wirksamkeit einer Entwöhnungstherapie), was im übrigen ökonomisch und für die Anwendung auf die praktische Arbeit mit Patienten bedeutsam werden kann (Metzler & Krause, 1997), berücksichtigten Grawe et al. (1994) nicht. Nicht zu reden davon, dass alle Maße für VT- und Psychoanalysestudien in gleicher Gewichtung in die Analyse eingehen, obgleich es als eine Binsenweisheit, insbesondere bei der Untersuchung der differentiellen Wirksamkeit von Therapiemethoden, anzusehen ist: „man kann nicht nicht gewichten“ (Sponsel, 2000a,b,c,d). Bei den Anhängern und Kritikern von Grawe stellt sich zu gleichen Maßen mittlerweile neben der Enttäuschung auch Wut und Ärger ein, selbst von Psychologen wird tiefstes Bedauern geäußert darüber, dass „unser Grawe“ sich solche Blöße habe geben müssen (vergleiche Sponsel, 2000a,b,c,d). Die weithin unstrittige Meinung besteht darin, dass keine folgenschweren Konsequenzen bezüglich der Einstufung der Wirksamkeit bzw. dem Vorteil einzelner Therapieformen aus der Studie von Grawe et al. (1994) zu ziehen sind (Sponsel, 2000a,b,c,d).

Die methodische Qualität der in unserer Untersuchung verwendeten Studien entspricht selbstverständlich auch der jeweiligen Forschungsepoche aus der sie stammen. Da für uns jedoch ein möglichst repräsentativer Überblick über die Wirksamkeit zum Autogenen Training überhaupt entscheidend war, haben wir auch die Studien einbezogen, die auch in die Analyse von Linden (1994) und Grawe et al. (1994) einbezogen wurden, um eine Vergleichbarkeit herzustellen. Dennoch möchten wir anhand des von Petrak, Hardt, Nickel und Egle (1999) vorgeschlagenen Schemas der Checklisten zur Bewertung der wissenschaftlichen Qualität kontrollierter psychotherapeutischer Interventionsstudien (CPI) zeigen, dass sich auch methodisch hochwertige Studien in unserem Datenpool befinden. Wir haben am Beispiel von zwei Studien, die in unsere Analyse eingingen, das gegebene Cochran-Level geprüft (CPI-Summenwert:  $> 66 =$  hoch [Cochran-Level 2],  $> 33$  mäßig [Cochran-Level 3],  $\leq 33 =$  gering). In der ersten Beispielstudie (Ehlers et al., 1995) wurde ein Summenwert im CPI von 88 festgestellt, was dem Cochran-Level 2 entspricht und der zweiten Beispielstudie (Krampen, 1997) ein Summenwert im CPI von 79, was ebenfalls dem Cochran-Level von 2 entspricht. Nicht alle Studien entsprechen dem genannten Cochran-Level 2, da zum Teil auch die Untersuchungsmethodologie – insbesondere bei den Studien aus der Kreidezeit der Psychotherapieforschung (z.B. Haward, 1965; Prill, 1965) – nicht so differenziert im Rahmen der Forschungstheorie vorlag, wie bei den Studien unter aktuell gültigen

Forschungsmaßstäben (zirka ab Anfang der 90er Jahre). Die dann im Rahmen einer Metaanalyse erfolgende Mischung von Studien, die mit unterschiedlicher methodischer Qualität durchgeführt wurden, ist zu kritisieren. Dennoch glauben wir, für die Anwendung von Metaanalysen nach gegebenen aktuellen methodischen Standards und angemessenen statistischen Methoden (Fricke & Treinies, 1985) eine Lanze brechen zu können. Die Notwendigkeit von Metaanalysen besteht u.E. nach zweifellos, weil ohne diese die zirka 9000 randomisierten Studien, die pro Jahr in medizinischen Zeitschriften neu erscheinen und die damit entstehenden Informationen nicht mehr auf individueller Ebene zu verarbeiten sind (Olkin, 1995). Cochran selbst forderte, dass die Übersichten zur Wirksamkeit therapeutischer Interventionen auf Basis aller jeweils vorhandenen randomisierten kontrollierten Studien zu verfassen und regelmäßig zu aktualisieren sind (Cochran, 1972; Galandi & Antes, 2000).

Zusammenfassend kann konstatiert werden, dass das AT in mittleren bis großen klinischen Effekten auf die Hauptzielsymptomatik resultiert, die auch zum Follow-up-Zeitpunkt stabil sind und die Placebo-Effekte (Kontrollbedingungen) deutlich überschreiten. Gegenüber den anderen Behandlungsbedingungen erbringt das AT gleiche oder leicht geringere Effekte.

Im Feld der psychosomatischen Störungen zeigt das AT seine Wirksamkeit in mehr als einer Studie bei milder bis moderater essentieller Hypertonie, Spannungskopfschmerz und Migräne, koronare Herzerkrankung, Asthma bronchiale, Raynaud-Syndrom und der unspezifischen somatoforme Schmerzstörung (Neurasthenie, vegetative Dystonie). Im direkten Vergleich zu anderen psychologischen Therapien wie der progressiven Muskelrelaxation, Entspannungshypnose oder Symptom-Biofeedback zeigt das AT gleich oder leicht geringere Effekte. Zu den Darmerkrankungen, der Neurodermitis, Glaukom und Geburtsvorbereitung liegen zum Teil nur 2 oder je 1 Studie zur metaanalytischen Untersuchung vor, so dass in künftigen Untersuchungen die aktuellen Ergebnisse in diesen Störungsbereichen möglicherweise zustimmend, unter Umständen auch ablehnend beurteilt werden. Im Bereich der psychischen Störungen wirkt sich das AT positiv auf die Angstsymptomatik, Depression und Schlafstörungen aus. Abgesehen von den hier nicht parallel untersuchten verhaltenstherapeutischen Konfrontationstherapien stellt das AT seine grundsätzliche Möglichkeit als gleichwertige ergänzende, zusätzlich für die Therapiezielerreichung förderliche Therapiekomponente unter Beweis. Das AT kann daher in künftigen Therapieprogrammen im Sinne einer additiven Therapiekomponente eine wichtige Rolle spielen.

Tabelle 1. *Ergebnisse der Metaanalyse für die Bedingung „AT Vorher-Nachher-Vergleich“ für Outcome und Follow-up für randomisierte Studien*

Kennwerte der Metaanalyse	AT Prä/Post		AT Prä/Follow-up	
	Hauptzielsymptomatik	unspezifische Symptomatik	Hauptzielsymptomatik	unspezifische Symptomatik
$D^{w+}$	0.776	-	-	-
Hedges 83 (gew.)	-	0.734	0.805	0.549
r	0.362	0.345	0.374	0.266
BESD	0.32 auf 0.68	0.33 auf 0.67	0.31 auf 0.69	0.37 auf 0.63
gebessert in %	79	54	84	60
unverändert in %	0	0	5	0
verschlechtert in %	21	46	11	40
Q	47.756	27.153	43.713	6.28
$p_Q$	0.036	0.007	0.001	0.179
$V(\Delta): V(g)-V(e)$	0.00=0.14 + 0.17	0.28=0.46-0.17	0.08=0.26-0.18	0.04=0.15-0.11
$V(\Delta) \%$	0	62.2	29.8	28.5
$\frac{1}{4}$ ES	0.194	0.183	0.20	0.137
95 % Konfidenzintervall	0.673 bis 0.880	-0.313 bis 1.78	0.685 bis 0.926	0.144 bis 0.953
Signifikanz ES	s.	ns.	s.	s.
Kendalls $\tau$	-0.146	-0.263	-0.23	0.20
$p_\tau$	0.23	0.21	0.17	0.62
$N_{fs, 0.05}/N_{krit}$	1736/175	217/75	973/105	28/35
$N_{Orwd, 0.2}$	91	29	53	10
N	1530	450	1138	224
k	33	13	19	5
Anzahl Einzel-ES	117	22	61	14

Anmerkung.  $D^{w+}$  = Effektstärke für homogene Datensätze, Hedges 83 (gew.) = Effektstärke für heterogene Datensätze, r = zugehöriger Korrelationskoeffizient, BESD = Binomial Effect Size Display, gebessert in % = Prozentsatz aller Einzeleffektstärken >0.5, unverändert in % = Prozentsatz aller Einzeleffektstärken =0.5, verschlechtert in % = Prozentsatz aller Einzeleffektstärken <0.5, Q = Homogenitätstestwert,  $p_Q$  = zugehörige Wahrscheinlichkeit zum Homogenitätstestwert,  $V(\Delta)$  = Varianz der Populationseffekte,  $V(g)$  = Varianz der beobachteten Effekte,  $V(e)$  = Varianz des Stichprobenfehlers,  $V(\Delta) \%$  = Anteil der Varianz der beobachteten Effekte, die durch die Variation in den Populationseffekten erklärt wird, 95% Konfidenzintervall = Grenzen innerhalb derer der Effekt in der Population liegt, Signifikanz ES: s. = signifikant, die Grenzen des 95% Konfidenzintervalls liegen auf der negativen oder positiven Seite, ns. = nicht signifikant, die Grenzen des 95% Konfidenzintervalls schließen die Null mit ein, Kendalls  $\tau$  = Rangkorrelationskoeffizient Tau von Kendall,  $p_\tau$  = zugehörige Wahrscheinlichkeit zu Kendalls Tau,  $N_{fs, 0.05}$  = Rosenthals Fail-Safe-N mit Signifikanzgrenze 0.05,  $N_{krit}$  = kritischer Wert für Rosenthals Fail-Safe-N,  $N_{Orwd, 0.2}$  = Orwins Modifikation von Rosenthals Faile-Safe-N, N = Anzahl der untersuchten Patienten, k = Anzahl der zugrundegelegten Studien, Anzahl Einzel-ES = Menge aller Erfolgsmaße, ES = Effektstärke.



Tabelle 2. *Ergebnisse der Metaanalyse für die Bedingung „AT Vorher-Nachher-Vergleich“ für Outcome und Follow-up für nicht randomisierte Studien*

Kennwerte der Metaanalyse	AT Prä/Post		AT Prä-/Follow-up	
	Hauptzielsymptomatik	unspezifische Symptomatik	Hauptzielsymptomatik	unspezifische Symptomatik
$D^{W+}$	0.436	-	0.529	0.624
Hedges 83 (gew.)	-	0.836	-	-
r	0.213	0.387	0.256	0.299
BESD	0.39 auf 0.61	0.31 auf 0.69	0.37 auf 0.63	0.35 auf 0.65
gebessert in %	54	50	60	60
unverändert in %	0	0	0	20
verschlechtert in %	46	50	40	20
Q	14.905	19.29	8.826	5.028
$p_Q$	0.828	0.002	0.454	0.285
$V(\Delta): V(g)-V(e)$	0.00=0.04-0.14	0.19=0.31-0.11	0.00=0.08-0.01	0.00=0.12-0.14
$V(\Delta) \%$	0	63.7	0	0
$\frac{1}{4}$ ES	0.109	0.209	0.132	0.156
95 % Konfidenzintervall	0.341 bis 0.532	-0.030 bis 1.702	0.357 bis 0.701	0.363 bis 0.884
Signifikanz ES	s.	ns.	s.	s.
Kendalls $\tau$	-0.298	0.00	-0.114	0.20
$p_\tau$	0.05	1.00	0.65	0.62
$N_{fs, 0.05}/N_{krit}$	593/120	82/40	111/60	28/35
$N_{Orwd, 0.2}$	25	18	16	10
N	1720	354	538	238
k	22	6	10	5
Anzahl Einzel-ES	74	17	33	7

Anmerkung.  $D^{W+}$  = Effektstärke für homogene Datensätze, Hedges 83 (gew.) = Effektstärke für heterogene Datensätze, r = zugehöriger Korrelationskoeffizient, BESD = Binomial Effect Size Display, gebessert in % = Prozentsatz aller Einzeleffektstärken >0.5, unverändert in % = Prozentsatz aller Einzeleffektstärken =0.5, verschlechtert in % = Prozentsatz aller Einzeleffektstärken <0.5, Q = Homogenitätstestwert,  $p_Q$  = zugehörige Wahrscheinlichkeit zum Homogenitätstestwert,  $V(\Delta)$  = Varianz der Populationseffekte,  $V(g)$  = Varianz der beobachteten Effekte,  $V(e)$  = Varianz des Stichprobenfehlers,  $V(\Delta) \%$  = Anteil der Varianz der beobachteten Effekte, die durch die Variation in den Populationseffekten erklärt wird, 95% Konfidenzintervall = Grenzen innerhalb derer der Effekt in der Population liegt, Signifikanz ES: s. = signifikant, die Grenzen des 95% Konfidenzintervalls liegen auf der negativen oder positiven Seite, ns. = nicht signifikant, die Grenzen des 95% Konfidenzintervalls schließen die Null mit ein, Kendalls  $\tau$  = Rangkorrelationskoeffizient Tau von Kendall,  $p_\tau$  = zugehörige Wahrscheinlichkeit zu Kendalls Tau,  $N_{fs, 0.05}$  = Rosenthals Fail-Safe-N mit Signifikanzgrenze 0.05,  $N_{krit}$  = kritischer Wert für Rosenthals Fail-Safe-N,  $N_{Orwd, 0.2}$  = Orwins Modifikation von Rosenthals Faile-Safe-N, N = Anzahl der untersuchten Patienten, k = Anzahl der zugrundegelegten Studien, Anzahl Einzel-ES = Menge aller Erfolgsmaße, ES = Effektstärke.

Tabelle 3. *Ergebnisse der Metaanalyse für die Bedingung „AT vs. wahre Kontrollgruppe“ für Outcome und Follow-up für randomisierte Studien*

Kennwerte der Metaanalyse	AT vs. KG Outcome		AT vs. KG Follow-up	
	Hauptzielsymptomatik	unspezifische Symptomatik	Hauptzielsymptomatik	unspezifische Symptomatik
$D^{W+}$	0.624	-	-	-
Hedges 83 (gew.)	-	0.573	0.533	1.053
r	0.298	0.276	0.258	0.468
BESD	0.35 auf 0.65	0.36 auf 0.64	0.37 auf 0.63	0.27 auf 0.73
gebessert in %	68	63	50	67
unverändert in %	4	12	0	33
verschlechtert in %	28	25	50	0
Q	35.262	33.97	24.366	15.422
$p_Q$	0.065	0.00002	0.011	0.0004
$V(\Delta): V(g)-V(e)$	0.006=0.16-0.16	0.63=0.77-0.14	0.13=0.34-0.21	0.96=1.12-0.15
$V(\Delta)$ %	3.7	81.4	39	86.5
$\frac{1}{4}$ ES	0.156	0.143	0.134	0.26
95 % Konfidenzintervall	0.498 bis 0.750	-0.978 bis 2.124	-0.183 bis 1.248	-0.872 bis 2.978
Signifikanz ES	s.	ns.	ns.	ns.
Kendalls $\tau$	-0.05	-0.429	0.061	-0.333
$p_\tau$	0.71	0.138	0.78	0.60
$N_{fs, 0.05}/N_{krit}$	674/135	67/50	123/70	22/25
$N_{Orwd, 0.2}$	0	10	19	10
N	1039	372	617	106
k	25	8	12	3
Anzahl Einzel-ES	88	18	45	5

Anmerkung.  $D^{W+}$  = Effektstärke für homogene Datensätze, Hedges 83 (gew.) = Effektstärke für heterogene Datensätze, r = zugehöriger Korrelationskoeffizient, BESD = Binomial Effect Size Display, gebessert in % = Prozentsatz aller Einzeleffektstärken >0.5, unverändert in % = Prozentsatz aller Einzeleffektstärken =0.5, verschlechtert in % = Prozentsatz aller Einzeleffektstärken <0.5, Q = Homogenitätstestwert,  $p_Q$  = zugehörige Wahrscheinlichkeit zum Homogenitätstestwert,  $V(\Delta)$  = Varianz der Populationseffekte,  $V(g)$  = Varianz der beobachteten Effekte,  $V(e)$  = Varianz des Stichprobenfehlers,  $V(\Delta)$  % = Anteil der Varianz der beobachteten Effekte, die durch die Variation in den Populationseffekten erklärt wird, 95% Konfidenzintervall = Grenzen innerhalb derer der Effekt in der Population liegt, Signifikanz ES: s. = signifikant, die Grenzen des 95% Konfidenzintervalls liegen auf der negativen oder positiven Seite, ns. = nicht signifikant, die Grenzen des 95% Konfidenzintervalls schließen die Null mit ein, Kendalls  $\tau$  = Rangkorrelationskoeffizient Tau von Kendall,  $p_\tau$  = zugehörige Wahrscheinlichkeit zu Kendalls Tau,  $N_{fs, 0.05}$  = Rosenthals Fail-Safe-N mit Signifikanzgrenze 0.05,  $N_{krit}$  = kritischer Wert für Rosenthals Fail-Safe-N,  $N_{Orwd, 0.2}$  = Orwins Modifikation von Rosenthals Faile-Safe-N, N = Anzahl der untersuchten Patienten, k = Anzahl der zugrundegelegten Studien, Anzahl Einzel-ES = Menge aller Erfolgsmaße, ES = Effektstärke.

Tabelle 4. *Ergebnisse der Metaanalyse für die Bedingung „AT vs. wahre Kontrollgruppe“ für Outcome und Follow-up für nicht randomisierte Studien*

Kennwerte der Metaanalyse	AT vs. KG Outcome		AT vs. KG Follow-up	
	Hauptzielsymptomatik	unspezifische Symptomatik	Hauptzielsymptomatik	unspezifische Symptomatik
$D^{w+}$	-	-	0.314	-
Hedges 83 (gew.)	0.710	1.412	-	-
r	0.335	0.578	0.156	-
BESD	0.33 auf 0.67	0.21 auf 0.79	0.42 auf 0.58	-
gebessert in %	56	40	17	-
unverändert in %	0	20	0	-
verschlechtert in %	44	40	83	-
Q	56.181	56.33	1.65	-
$p_Q$	0.0000	0.0000	0.895	-
$V(\Delta): V(g)-V(e)$	0.31=0.45-0.14	3.26=3.49-0.23	0.00=0.03-0.12	-
$V(\Delta) \%$	69	93.5	0	-
$\frac{1}{4}$ ES	0.161	0.353	0.078	-
95 % Konfidenzintervall	-0.3888 bis 1.809	-2.128 bis 4.951	0.067 bis 0.561	-
Signifikanz ES	ns.	ns.	s.	-
Kendalls $\tau$	-0.126	0.20	0.276	-
$p_\tau$	0.498	0.62	0.437	-
$N_{fs, 0.05}/N_{krit}$	561/90	38/35	19/40	-
$N_{Orwd, 0.2}$	32	12	3	-
N	1423	224	273	-
k	16	5	6	-
Anzahl Einzel-ES	65	9	16	-

Anmerkung.  $D^{w+}$  = Effektstärke für homogene Datensätze, Hedges 83 (gew.) = Effektstärke für heterogene Datensätze, r = zugehöriger Korrelationskoeffizient, BESD = Binomial Effect Size Display, gebessert in % = Prozentsatz aller Einzeleffektstärken >0.5, unverändert in % = Prozentsatz aller Einzeleffektstärken =0.5, verschlechtert in % = Prozentsatz aller Einzeleffektstärken <0.5, Q = Homogenitätstestwert,  $p_Q$  = zugehörige Wahrscheinlichkeit zum Homogenitätstestwert,  $V(\Delta)$  = Varianz der Populationseffekte,  $V(g)$  = Varianz der beobachteten Effekte,  $V(e)$  = Varianz des Stichprobenfehlers,  $V(\Delta) \%$  = Anteil der Varianz der beobachteten Effekte, die durch die Variation in den Populationseffekten erklärt wird, 95% Konfidenzintervall = Grenzen innerhalb derer der Effekt in der Population liegt, Signifikanz ES: s. = signifikant, die Grenzen des 95% Konfidenzintervalls liegen auf der negativen oder positiven Seite, ns. = nicht signifikant, die Grenzen des 95% Konfidenzintervalls schließen die Null mit ein, Kendalls  $\tau$  = Rangkorrelationskoeffizient Tau von Kendall,  $p_\tau$  = zugehörige Wahrscheinlichkeit zu Kendalls Tau,  $N_{fs, 0.05}$  = Rosenthals Fail-Safe-N mit Signifikanzgrenze 0.05,  $N_{krit}$  = kritischer Wert für Rosenthals Fail-Safe-N,  $N_{Orwd, 0.2}$  = Orwins Modifikation von Rosenthals Faile-Safe-N, N = Anzahl der untersuchten Patienten, k = Anzahl der zugrundegelegten Studien, Anzahl Einzel-ES = Menge aller Erfolgsmaße, ES = Effektstärke.

Tabelle 5. *Ergebnisse der Metaanalyse für die Bedingung „AT vs. andere psychologische Behandlungsverfahren“ für Outcome und Follow-up für randomisierte Studien*

Kennwerte der Metaanalyse	AT vs. PT Outcome		AT vs. PT Follow-up	
	Hauptzielsymptomatik	unspezifische Symptomatik	Hauptzielsymptomatik	unspezifische Symptomatik
$D^{W+}$	-0.286	-	-0.211	0.040
Hedges 83 (gew.)	-	0.264	-	-
r	-0.142	0.131	-0.105	0.020
BESD	0.57 auf 0.43	0.43 auf 0.57	0.55 auf 0.45	0.49 auf 0.51
gebessert in %	6	29	0	0
unverändert in %	0	0	25	0
verschlechtert in %	94	71	75	100
Q	29.961	25.002	11.246	5.24
$p_Q$	0.027	0.0003	0.423	0.264
$V(\Delta): V(g)-V(e)$	0.00=0.12-0.15	0.48=0.70-0.17	0.00=0.06-0.12	0.003=0.11-0.11
$V(\Delta) \%$	0	73.1	0	2.6
$\frac{1}{4} ES$	-0.07	0.066	-0.053	0.01
95 % Konfidenzintervall	-0.403 bis -0.168	-1.088 bis 1.615	-0.336 bis -0.85	-0.213 bis 0.294
Signifikanz ES	s.	ns.	s.	ns.
Kendalls $\tau$	0.073	0.488	-0.124	0.20
$p_\tau$	0.67	0.124	0.575	0.49
$N_{fs, 0.05}/N_{krit}$	55/100	0/45	7/70	0/35
$N_{Orwd, 0.2}$	0	0	0	0
N	1207	241	1079	272
k	18	7	12	5
Anzahl Einzel-ES	62	9	69	15

Anmerkung.  $D^{W+}$  = Effektstärke für homogene Datensätze, Hedges 83 (gew.) = Effektstärke für heterogene Datensätze, r = zugehöriger Korrelationskoeffizient, BESD = Binomial Effect Size Display, gebessert in % = Prozentsatz aller Einzeleffektstärken >0.5, unverändert in % = Prozentsatz aller Einzeleffektstärken =0.5, verschlechtert in % = Prozentsatz aller Einzeleffektstärken <0.5, Q = Homogenitätstestwert,  $p_Q$  = zugehörige Wahrscheinlichkeit zum Homogenitätstestwert,  $V(\Delta)$  = Varianz der Populationseffekte,  $V(g)$  = Varianz der beobachteten Effekte,  $V(e)$  = Varianz des Stichprobenfehlers,  $V(\Delta) \%$  = Anteil der Varianz der beobachteten Effekte, die durch die Variation in den Populationseffekten erklärt wird, 95% Konfidenzintervall = Grenzen innerhalb derer der Effekt in der Population liegt, Signifikanz ES: s. = signifikant, die Grenzen des 95% Konfidenzintervalls liegen auf der negativen oder positiven Seite, ns. = nicht signifikant, die Grenzen des 95% Konfidenzintervalls schließen die Null mit ein, Kendalls  $\tau$  = Rangkorrelationskoeffizient Tau von Kendall,  $p_\tau$  = zugehörige Wahrscheinlichkeit zu Kendalls Tau,  $N_{fs, 0.05}$  = Rosenthals Fail-Safe-N mit Signifikanzgrenze 0.05,  $N_{krit}$  = kritischer Wert für Rosenthals Fail-Safe-N,  $N_{Orwd, 0.2}$  = Orwins Modifikation von Rosenthals Faile-Safe-N, N = Anzahl der untersuchten Patienten, k = Anzahl der zugrundegelegten Studien, Anzahl Einzel-ES = Menge aller Erfolgsmaße, ES = Effektstärke.

Tabelle 6 *Ergebnisse der Metaanalyse für die Bedingung „AT vs. andere psychologische Behandlungsverfahren“ für Outcome und Follow-up für nicht randomisierte Studien*

Kennwerte der Metaanalyse	AT vs. PT Outcome		AT vs. PT Follow-up	
	Hauptzielsymptomatik	unspezifische Symptomatik	Hauptzielsymptomatik	unspezifische Symptomatik
$D^{W+}$	-0.094	0.339	-0.183	0.339
Hedges 83 (gew.)	-	-	-	-
r	-0.047	0.169	-0.092	0.169
BESD	0.52 auf 0.48	0.42 auf 0.58	0.55 auf 0.45	0.42 auf 0.58
gebessert in %	0	25	0	50
unverändert in %	11	0	0	0
verschlechtert in %	89	75	100	50
Q	3.077	0.59	0.37	0.762
$p_Q$	0.929	0.899	0.996	0.383
$V(\Delta): V(g)-V(e)$	0.00=0.05-0.16	0.00=0.02-0.13	0.00=0.01-0.15	0.00=0.07-0.12
$V(\Delta) \%$	0	0	0	0
$\frac{1}{4} ES$	-0.023	0.08	-0.046	0.085
95 % Konfidenzintervall	-0.328 bis 0.140	0.023 bis 0.656	-0.477 bis 0.111	-0.093 bis 0.771
Signifikanz ES	ns.	s.	ns.	ns.
Kendalls $\tau$	0.056	-0.667	-0.276	-1.00
$p_\tau$	0.83	0.174	0.44	0.317
$N_{fs, 0.05}/N_{krit}$	0/55	10/30	0/40	0/20
$N_{Orwd, 0.2}$	0	3	0	1
N	306	166	193	92
k	9	4	6	2
Anzahl Einzel-ES	42	12	43	5

Anmerkung.  $D^{W+}$  = Effektstärke für homogene Datensätze, Hedges 83 (gew.) = Effektstärke für heterogene Datensätze, r = zugehöriger Korrelationskoeffizient, BESD = Binomial Effect Size Display, gebessert in % = Prozentsatz aller Einzeleffektstärken >0.5, unverändert in % = Prozentsatz aller Einzeleffektstärken =0.5, verschlechtert in % = Prozentsatz aller Einzeleffektstärken <0.5, Q = Homogenitätstestwert,  $p_Q$  = zugehörige Wahrscheinlichkeit zum Homogenitätstestwert,  $V(\Delta)$  = Varianz der Populationseffekte,  $V(g)$  = Varianz der beobachteten Effekte,  $V(e)$  = Varianz des Stichprobenfehlers,  $V(\Delta) \%$  = Anteil der Varianz der beobachteten Effekte, die durch die Variation in den Populationseffekten erklärt wird, 95% Konfidenzintervall = Grenzen innerhalb derer der Effekt in der Population liegt, Signifikanz ES: s. = signifikant, die Grenzen des 95% Konfidenzintervalls liegen auf der negativen oder positiven Seite, ns. = nicht signifikant, die Grenzen des 95% Konfidenzintervalls schließen die Null mit ein, Kendalls  $\tau$  = Rangkorrelationskoeffizient Tau von Kendall,  $p_\tau$  = zugehörige Wahrscheinlichkeit zu Kendalls Tau,  $N_{fs, 0.05}$  = Rosenthals Fail-Safe-N mit Signifikanzgrenze 0.05,  $N_{krit}$  = kritischer Wert für Rosenthals Fail-Safe-N,  $N_{Orwd, 0.2}$  = Orwins Modifikation von Rosenthals Faile-Safe-N, N = Anzahl der untersuchten Patienten, k = Anzahl der zugrundegelegten Studien, Anzahl Einzel-ES = Menge aller Erfolgsmaße, ES = Effektstärke.

## Literatur:

Alle Studien, die in die quantitative Analyse einbezogen wurden, sind mit einem Sternchen gekennzeichnet.

- \*Aivazyan, T. A., Zaitsev, V. P., Salenko, B. B., Yurenev, A. P. & Patrusheva, I. F. (1988): Efficacy of relaxation techniques in hypertensive patients. *Health Psychology* 7 (Suppl.), 193–200
- \*Aivazyan, T. A., Zaitsev, V. P. & Yurenev, A. P. (1988): Autogenic training in the treatment and secondary prevention of essential hypertension: Five-year follow-up. *Health Psychology* 7 (Suppl.), 201–208
- Akimenko, M. A. & Gromov, S. A. (1995): Adaptive biocontrol in the system of treating epilepsy patients. *Zurnal Nevropatologii I Psichiatrii imeni S.S. Korsakova* 95 (3), 45–48
- \*Alarcia, J., Pinard, G., Serrano, M. & Tetreault, L. (1982): Étude comparative de trois traitements du bégaiement: relaxation, désensibilisation, rééducation. *Revue de Psychologie Appliquée* 32 (1), 1–25
- Asay, T. P., Lambert, M. J., Christensen, E. R. & Beutler, L. E. (1984): A meta-analysis of mental-health treatment outcome. Unpublished manuscript, Department of Psychology, Brigham Young University
- \*Bailey, R. D. (1984): Autogenic relaxation training and sickness absence amongst student nurses in general training. *Journal of Advanced Nursing* 9, 581–587
- \*Banner, C. N. & Meadows, W. M. (1983): Examination of the effectiveness of various techniques for reducing tension. *British Journal of Clinical Psychology* 22, 183–193
- Begg, C. B. (1994): Publication bias. In H. Cooper & L. V. Hedges (eds.), *The handbook of research synthesis* (pp. 399–409). Russell Sage Foundation, New York
- \*Beitel, E. & Kröner, B. (1982): Veränderung des Selbstkonzepts durch Autogenes Training. *Zeitschrift für Klinische Psychologie* 11, 1–15
- Bentler, P. M. (1995): EQS: Structural equations program manual. Multivariate Software, Inc., Encino, CA
- Bentler, P. M. (1997): EQS: A structural equations program (Vers. 5.6, 1985–1997). Multivariate Software, Inc., Encino, CA
- Bergin, A. E. & Garfield, S. L. (Eds.). (1994): *Handbook of psychotherapy and behavior change* (4<sup>th</sup> ed.). Wiley, New York [u.a.]
- \*Berndt, H. & Maercker, W. (1985): Zur Psychotherapie beim Reizkollon. *Zeitschrift für die Gesamte Innere Medizin* 40 (4), 107–110
- Bianchi, M. S., De Punzio, C., Ferdeghini, M., Fioretti, P., Metelli, P., Neri, E. & Venticinque, M. (1994): The relationship between maternal relaxation and plasma beta-endorphin levels during parturition. *Journal of Psychosomatic Obstetrics and Gynaecology* 15, 205–210
- Binder, H. & Binder, K. (1998): *Autogenes Training – Basispsychotherapeutikum* (2. Aufl.). Deutscher Ärzteverlag, Köln
- \*Blanchard, E. B., Khramelashvili, V. V., McCoy, G. C., Aivazyan, T. A., McCaffrey, Salenko, B. B., Musso, A., Wittrock, D. A., Berger, M., Gerardi, M. & Pangburn, L. (1988): The USA-USSR collaborative cross-cultural comparison of autogenic training and thermal biofeedback in the treatment of mild hypertension. *Health Psychology* 7 (Suppl.), 175–192
- Bongartz, W., Flammer, E. & Schwonke, R. (2002): Die Effektivität der Hypnose: eine meta-analytische Studie. *Psychotherapeut* 47, 67–76

- Bortz, J. (1995): Statistik für Sozialwissenschaftler (5., vollst. überarb. und aktual. Aufl.). Springer, Berlin
- Bortz, J. & Döring, N. (1995): Forschungsmethoden und Evaluation (2., vollst. überarb. und aktual. Aufl.). Springer, Berlin
- Bowers, T. & Clum, G. (1988): Relative contributions of specific and nonspecific treatment effects: Meta-analysis of placebo-controlled behavior therapy research. *Psychological Bulletin* 103, 315–323
- Buol, C. & Endtner, K. (1993): „Doch die Verhältnisse, sie sind nicht so...“ (Brecht). Differentielle Wirkung von Psychotherapie. Eine Metaanalyse. Unveröffentlichte Lizentiatsarbeit am Psychologischen Institut der Universität Bern, eingereicht bei Prof. Klaus Grawe. *Unitobler Psych.* 1099'642
- Carpenter, W. B. (1993): Principles of mental physiology. Reprint of the 1876 edition. Routledge [u.a.], London (Original erschienen 1876: Principles of mental physiology)
- \*Carruthers, M. (1984): Health promotion by mental and physical training. *British Journal of Holistic Medicine* 1 (2), 142–147
- \*Carruthers, M. (1988): Die kardioprotektive Wirkung des Autogenen Trainings. *Zeitschrift für Allgemeine Medizin* 64, 575–579
- \*Carruthers, M. & Stetter, F. (1992): Die Beeinflussung kardiovaskulärer Risikofaktoren durch autogenes Training. *Zeitschrift für Allgemeine Medizin* 68, 154–157
- \*Charlesworth, E. A., Williams, B. J. & Baer, P. E. (1984): Stress management at the worksite for hypertension: Compliance, cost-benefit, health care and hypertension-related variables. *Psychosomatic Medicine* 46 (5), 387–397
- Clauß, G. & Ebner, H. (1989): Grundlagen (6. unveränd. Aufl.). Statistik für Soziologen, Pädagogen, Psychologen und Mediziner, Bd. 1 (hrsg. von G. Clauß & H. Ebner). Deutsch, Frankfurt a. M.
- Cochrane, A. L. (1972): Effectiveness and efficiency. Random reflections on health services. Nuffield Provincial Hospitals Trust, London
- Cohen, J. (1977): Statistical power analysis for the behavioral sciences (2nd ed.). Academic Press, New York [u.a.] (Original erschienen 1969: Statistical power analysis for the behavioral sciences)
- \*Collet, L., Cottraux, J. & Juenet, C. (1986): GSR feedback and Schultz's relaxation in tension headaches: A comparative study. *Pain* 25, 205–213
- Cook, T. D., Cooper, H., Cordray, D. D., Hartmann, H., Hedges, L. V., Light, R. J., Louis, T. & Mosteller, F. (Eds.). (1992): Meta-analysis for explanation: A casebook. Russell Sage Foundation, New York
- Cooper, H. (1984): The integrative research review: A systematic approach. Sage, Beverly Hills, CA
- Cooper, H. (1998): Synthesizing research: A guide for literature reviews (2nd ed.). Sage, Thousand Oaks
- Cooper & L. V. Hedges (Eds.). (1994): The handbook of research synthesis. Russell Sage Foundation, New York
- \*Cott, A., Parkinson, W., Fabich, M., Bedard, M. & Marlin, R. (1992): Long-term efficacy of combined relaxation: Biofeedback treatments for chronic headache. *Pain* 51 (1), 49–56
- \*Coursey, R. D., Frankel, B. L., Gaarder, K. R. & Mott, D. E. (1980): A comparison of relaxation techniques with electrosleep therapy for chronic, sleep-onset insomnia: A sleep-EEG study. *Biofeedback and Self-Regulation* 5 (1), 57–73
- Crow, E. L. (1991): Response to Rosenthal's comment "How are we doing in soft psychology?" *American Psychologist* 46, 1083

- Cucherat, M. (2000): EasyMeta 2000. Software for meta-analysis of clinical trials. Departement of clinical pharmacology cardiological hospital, Lyon, France
- \*De Rivera, J. L. G., De Montigny, C. & Remillard, G. (1977): Autogenic therapy of temporal lobe epilepsy. *Therapy in Psychosomatic Medicine* 4, 40–47
- Deter, H.-C. & Allert, G. (1983): Group therapy for asthma patients: A concept for the psychosomatic treatment of patients in a medical clinic – a controlled study. *Psychotherapy and Psychosomatics* 40, 95–105
- Diepgen, R. (1993): Eine Randbemerkung zu einer Argumentationsfigur von Grawe (1992). *Psychologische Rundschau* 44 (3), 176–177
- Draper, D., Graver, D. P., Goel, P. K., Greenhouse, J. B., Hedges, L. V., Morris, C. N., Tucker, J. R. & Watenaux, C. M. (1992): Combining information: Statistical issues and opportunities for research. National Academy Press, Washington, DC
- Durlak, J. A. & Lipsey, M. W. (1991): A practitioner's guide to meta-analysis. *American Journal of Community Psychology* 19 (3), 291–333
- \*Ehlers, A., Stangier, U. & Gieler, U. (1995): Treatment of atopic dermatitis: A comparison of psychological and dermatological approaches to relapse prevention. *Journal of Consulting and Clinical Psychology* 63 (4), 624–635
- \*Engel-Sittenfeld, P., Engel, R. R., Huber, H. P. & Zangl, K. (1980): Wirkmechanismen psychologischer Therapieverfahren bei der Behandlung chronischer Schlafstörungen. *Zeitschrift für Klinische Psychologie* 9, 34–52
- \*Farné, M. & Corallo, A. (1992): Autogenic training and signs of distress: An experimental study. *Bollettino della Società Italiana di Biologia Sperimentale* 48 (6), 413–417
- Forel, A. H. (1902): *Der Hypnotismus und die suggestive Psychotherapie* (4. umgearb. Aufl.). Enke, Stuttgart
- \*Freedman, R. R., Ianni, P. & Wenig, P. (1983): Behavioral treatment of raynaud's disease. *Journal of Consulting and Clinical Psychology* 51 (4), 539–549
- \*Freedman, R. R. (1989): Quantitative measurements of finger blood flow during behavioral treatments for Raynaud's disease. *Psychophysiology* 26 (4), 437–441
- Fricke, R. & Treinies, G. (1985): *Einführung in die Metaanalyse*. Huber, Bern
- Fricke, R. & Kreft, H. (1986): *Metaanalyse, Programmversion 03*. Braunschweig: Seminar für Pädagogik, Technische Universität Braunschweig
- Fukunishi, I., Hosaka, T., Matsumoto, T. & Hayashi, M. (1997): Liaison psychiatry and HIV infection (II): Application. *Psychiatry and Clinical Neurosciences* 51 (1), 5–8
- Galandi, D. & Antes, G. (2000): Systematische Übersichtsarbeiten und die Cochrane Collaboration. In H. J. Freyberger, G. Heuft & D. J. Ziegenhagen (Hrsg.), *Ambulante Psychotherapie: Transparenz, Effizienz, Qualitätssicherung* (S. 160–177). Schattauer, Stuttgart [u.a.]
- Geuter, U. (1995): *Psychotherapieforschung: Statistische Irrtümer?* *Psychologie Heute*, Mai, 10
- Giblin, P., Sprenkle, D. H. & Sheehan, R. (1985): Enrichment outcome research: A meta-analysis of premarital, marital and family interventions. *Journal of Marital and Family Therapy* 11, 257–271
- Glass, G. V. (1976): Primary, secondary and meta-analysis of research. *Educational Researcher* 10, 3–8
- Glass, G. V. (1977): Integrating findings: The meta-analysis of research. *Review of research in education* 5, 351–379
- Glass, G. V., McGaw, B. & Smith, M. L. (1981): *Meta-analysis in social research*. Sage, Beverly Hills, CA
- Grawe, K. (1992): *Psychotherapieforschung zu Beginn der neunziger Jahre*. *Psychologische Rundschau* 43, 132–162



- Grawe, K., Donati, R. & Bernauer, F. (1994): Psychotherapie im Wandel (2. Aufl.). Hogrefe, Göttingen
- Greenwald, A. G. (1975): Consequences of prejudice against the null hypothesis. *Psychological Bulletin* 82, 1–20
- Gurevitch, J. & Hedges, L. V. (1999): Statistical issues in ecological meta-analysis. *Ecology* 80, 1142–1149
- \*Haber, P., Moser, G., Sachs, G. & Spiess, K. (1993): Effectiveness of relaxation groups in patients with chronic respiratory tract diseases. *Wiener Klinische Wochenschrift* 105, 603–610
- Hamilton, L.C. (1992): Regression with graphics: A second course in applied statistic. Brook/Cole Publishing Co., Pacific Grove, CA
- Harlow, L. L. (1985): Behavior of some elliptical theory estimators with nonnormal data in a covariance structures framework: A Monte Carlo study. Ph.D. Thesis, University of California, Los Angeles, CA
- \*Haustein, U. F., Seikowski, K. & Weber, B. (1995): Effect of hypnosis and autogenic training on acral circulation and coping with the illness in patients with progressive scleroderma. *Hautarzt* 46, 94–101
- \*Haward, L. R. C. (1965): Reduction in stress reactivity by autogenic training. In W. Luthe (Ed.), *Autogenes Training: Correlationes psychosomaticae* (S. 96–103). Thieme, Stuttgart
- Heckrath, C. & Dohmen, P. (1997): Zu der empirischen Basis der 'hochsignifikanten Überlegenheit' der kognitiv-behavioralen gegenüber den psychoanalytischen Therapieverfahren: *Zeitschrift für psychosomatische Medizin und Psychoanalyse* 43 (2), 179–201
- Hedges, L. V. (1981): Distribution Theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics* 6, 107–128
- Hedges, L. V. (1982): Estimation of effect size from a series of independent experiments. *Psychological Bulletin* 92, 490–499
- Hedges, L. V. (1983): A random effects model for effect sizes. *Psychological Bulletin* 92, 388–395
- Hedges, L. V. (1994): Statistical considerations. In H. Cooper & L. V. Hedges (eds.), *The handbook of research synthesis* (pp. 29–38). Russell Sage Foundation, New York
- Hedges, L. V. & Olkin, I. (1985): *Statistical methods for meta-analysis*. Academic Press, New York
- Hedges, L. V. & Vevea, J. L. (1998): Fixed- and random-effects models in meta-analysis. *Psychological Methods* 3, 486–504
- \*Henry, M., de Rivera, J. L. G., Gonzalez-Martin, I. J. & Abreu, J. (1993): Improvement of respiratory function in chronic asthmatic patients with autogenic therapy. *Journal of Psychosomatic Research* 37 (3), 265–270
- \*Herbert, C. P., Gutman, G. M. (1980): Practical group autogenic training for management of stress-related disorders in family practice. In H. Wain (Ed.), *Clinical hypnosis in medicine* (pp. 109–118). Year Book Medical Publishers, Chicago
- \*Herbert, C. P. & Gutman, G. M. (1983): Practical group autogenic training for management of stress-related disorders in family practice. *Canadian Family Physician* 29, 109–117
- Hillecke, T., Krämer, B. I., Kupper, S., Kupper-Horster, C. & Kächele, H. (1996): Soweit der Stand der Dinge: Psychotherapie im Wandel! *Psychotherapie, Psychosomatik und medizinische Psychologie* 46, 96–101

- Hoffmann, B., Hoffmann, S.O. & Derra, C. (1997): Handbuch autogenes Training (13. überarb. Aufl.). Deutscher Taschenbuchverlag, München
- Hoffmann, B., Hoffmann, S.O., Derra, C. & Stetter, F. (2000): Handbuch autogenes Training (14. überarb. Aufl.). Deutscher Taschenbuchverlag, München
- Hunter, J. E. & Schmidt, F. L. (1990): Methods of meta-analysis. Sage, Newbury Park
- Hunter, J. E., Schmidt, F. L. & Jackson, G. B. (1982): Meta-analysis. Cumulating research findings across studies. Sage, Beverly Hills, CA
- Hutchings, D. F. & Reinking, R. H. (1976): Tension headaches: What form of therapy is most effective? *Biofeedback and Self-Regulation* 1 (2), 183–190
- Hyman, R. B., Feldman, H. R., Harris, R. B., Levin, R. F. & Malloy, G. B. (1989): The effects of relaxation training on clinical symptoms: A meta-analysis. *Nursing Research* 38, 216–229
- Ikemi A (1988): Psychophysiological Effects of Self-Regulation Method: EEG Frequency Analysis and Contingent Negative Variations. *Psychotherapy and Psychosomatics* 49, 230-239
- Jacobson, N. S., Follette, W. C. & Revenstorf, D. (1984): Psychotherapy outcome research : Methods for reporting variability and evaluating clinical significance. *Behavior Therapy* 15, 336–352
- \*Janssen, K. & Neutgens, J. (1986): Autogenic training and progressive relaxation in the treatment of three kinds of headache. *Behavior Research and Therapy* 24 (2), 199–208
- \*Jessup, B. A. & Neufeld, R. W. (1977): Effects of biofeedback and "autogenic relaxation" techniques on physiological and subjective responses in psychiatric patients: A preliminary analysis. *Behavior Therapy* 8, 160–167
- Kächele, H. (1995): Klaus Grawes Konfession und die psychoanalytische Profession. *Psyche* 49, 481–492
- \*Kaluza, G. & Stempel, I. (1995): Effects of self-regulation methods and visual imagery on IOP in patients with open-angle glaucoma. *Ophthalmologica* 209, 122–128
- \*Katzenstein, A., Kriegel, E. & Gaefke, I. (1974): Erfolgsuntersuchung bei einer komplexen Psychotherapie essentieller Hypertoniker. *Psychiatrie, Neurologie und Medizinische Psychologie* 26, 732–737
- \*Keefe, F. J., Surwit, R. S. & Pilon, R. N. (1980): Biofeedback, autogenic training, and progressive relaxation in the treatment of raynaud's disease: A comparative study. *Journal of Applied Behavior Analysis* 13 (1), 3–11
- Keel, P. J., Bodoky, C., Gerhard, U. & Muller, W. (1998): Comparison of integrated group therapy and group relaxation training for fibromyalgia. *Clinical Journal of Pain* 14 (3), 232–238
- Kendall, M. G. (1948): Rank correlation methods (2<sup>nd</sup> ed.). Griffin, London
- Kenny, D. A. (1999): Meta-analysis: Easy to answer (vers. 2). Internet: <http://nw3.nai.net/~dakenny/>
- König, W., di Pol, G. & Schaeffer, G. (1979): Autogenes Training: Ein Grundriss. Gustav Fischer, Jena
- Kraemer, H. C. (1983): Theory of estimation and testing of effect sizes: Use in meta-analysis. *Journal of Educational Statistics* 8, 93–101
- Krampen, G. (1991): Optimierung des Lernprozesses beim Autogenen Training bei kurzer Kurslaufzeit durch Übungsprotokolle. *Zeitschrift für Klinische Psychologie, Psychopathologie und Psychotherapie* 39, 33–45
- Krampen, G. (1992): Einführungskurse zum autogenen Training. Verlag für Angewandte Psychologie, Göttingen

- \*Krampen, G. (1997): Autogenes Training vor und begleitend zur methodenübergreifenden Einzelpsychotherapie bei depressiven Störungen. *Zeitschrift für Klinische Psychologie, Psychiatrie und Psychotherapie* 45 (2), 214–232
- \*Krampen, G. (1999): Long-term evaluation of the effectiveness of additional autogenic training in the psychotherapy of depressive disorders. *European Psychologist* 4 (1), 11–18
- \*Kröner, B., Frieg, H. & Niewendiek, U. (1982): Einsatz verschiedener Programme des Autogenen Trainings bei Prüfungsangst. *Zeitschrift für Klinische Psychologie und Psychotherapie* 30 (3), 254–266
- \*Kröner, B. & Heiß, M. (1982): Der Einsatz von Entspannungsverfahren bei chronischen Kopfschmerzen – Eine Studie über die Möglichkeiten nichtmedikamentöser Therapie. In M. P. Huber (Ed.), *Migräne* (S. 154–175). Urban & Schwarzenberg, München
- \*Kröner & Beitel (1980): Längsschnittuntersuchung über die Auswirkung des autogenen Trainings auf verschiedene Formen der subjektiv wahrgenommenen Entspannung und des Wohlbefindens. *Zeitschrift für Klinische Psychologie und Psychotherapie* 28 (2), 127–133
- Kupper, S. (1998): Partizipationsstrukturen im psychotherapeutischen Prozeß. P. Lang, Frankfurt a. M.
- \*Labbé, E. E. (1995): Treatment of childhood migraine with autogenic training and skin temperature biofeedback: A component analysis. *Headache* 35, 10–13
- \*Laberke, J. A. (1952a, 29. August): Psychotherapie: über eine psychosomatische Kombinationsbehandlung (mehrdimensionale Therapie) bei sogenannten inneren Krankheiten, Teil 1. *Münchener Medizinische Wochenschrift* 35, 1718–1724
- \*Laberke, J. A. (1952b, 5. September): Psychotherapie: über eine psychosomatische Kombinationsbehandlung (mehrdimensionale Therapie) bei sogenannten inneren Krankheiten, Teil 2. *Münchener Medizinische Wochenschrift* 36, 1809–1816
- Lambert, M. J. & Hill, C. E. (1994): Assessing psychotherapy outcomes and processes. In A. E. Bergin & S. L. Garfield (eds.), *Handbook of psychotherapy and behavior change* (4<sup>th</sup> ed., pp. 72–113). Wiley, New York [u.a.]
- Lambert, M. J. & Bergin, A. E. (1994): The effectiveness of psychotherapy. In A. E. Bergin & S. L. Garfield (eds.), *Handbook of psychotherapy and behavior change* (4<sup>th</sup> ed., pp. 143–189). Wiley, New York [u.a.]
- Lambert, M. J., Weber, F. D. & Sykes, J. D. (1993): Psychotherapy versus placebo. Poster presented at the annual meetings of the Western Psychological Association, Phoenix
- Landman, J. R. & Dawes, R. M. (1982): Psychotherapy outcome: Smith and Glass' conclusions stand up under scrutiny. *American Psychologist* 37, 504–516
- Langen, D. (1959): Autogenic training and psychosomatic medicine. In G. D. Buttows & L. Dennerstein (Eds.), *Handbook of hypnosis and psychosomatic medicine* (pp. 497–507). Elsevier/North Holland Biomedical Press, Amsterdam [u.a.]
- \*Lehrer, P. M., Atthowe, J. M. & Weber, E. S. P. (1980): Effects of progressive relaxation and autogenic training of anxiety and physiological measures, with some data on hypnotizability. In F. J. McGuigan, F. J. Simme & Wallace, J. M. (Eds.), *Stress and tension control* (pp. 171–181). Plenum Press, New York
- Linden, W. (1990): *Autogenic training: A clinical guide*. Guilford Press, New York
- Linden, W. (1994): Autogenic training: A narrative and quantitative review of clinical outcome. *Biofeedback and Self-Regulation* 19 (3), 227–264
- \*Loesch, W., Seefeldt, D. & Hecht, K. (1989): Langzeitstudie zur Effektivität psychotherapeutischer Verfahren in der Behandlung von Frühformen der primären arteriellen Hypertonie in der allgemeinmedizinisch-betriebsärztlichen Praxis. *Zeitschrift für Ärztliche Fortbildung* 83, 913–919

- \*Luborsky, L., Ancona, L., Masoni, A., Scolari, G. & Longoni, A. (1980): Behavioral versus pharmacological treatments for essential hypertension: A pilot study. *International Journal of Psychiatry in Medicine* 10 (1), 33–40
- Luthe, W. (1965): *Autogenes Training: Correlationes psychosomaticae*. Thieme, Stuttgart
- Luthe, W. (1970a): *Autogenic training. Vol IV: Research and theory*. Grune & Stratton, New York
- Luthe, W. (1970b): *Autogenic training. Vol V: Dynamics of autogenic neutralization*. Grune & Stratton, New York
- Luthe, W. (1970c): *Autogenic training. Vol VI: Treatment with autogenic neutralization*. Grune & Stratton, New York
- Luthe, W. & Schultz, J. H. (1969a): *Autogenic therapy. Vol II: Medical applications*. Grune & Stratton, New York
- Luthe, W. & Schultz, J. H. (1969b): *Autogenic therapy. Vol III: Applications in psychotherapy*. Grune & Stratton, New York
- Mann, K.F. & Stetter, F. (1982): Thermographische Befunde beim autogenen Training in Abhängigkeit von der Tagesperiodik. *Therapiewoche* 32, 2232-2238
- \*Mantovani, G., Astara, G., Lampis, B., Bianchi, A., Curreli, L., Orrù, W., Carta, M. G., Carpinello, B., Contu, P. & Rudas, N. (1996): Evaluation by multidimensional instruments of health-related quality of life of elderly cancer patients undergoing three different "psychosocial" treatment approaches: A randomized clinical trial. *Support Care Cancer* 4, 129–140
- Mardia, K. V. (1970): Measures of multivariate skewness and kurtosis with applications. *Biometrika* 57, 519–530
- Mardia, K. V. (1974): Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhya B36*, 115–128
- McCoy, G. C., Fein, S., Blanchard, E. B., Wittrock, D. A., McCaffrey, R. J. & Pangburn, L. (1988): End organ changes associated with the self-regulatory treatment of mild essential hypertension? *Biofeedback and Self-Regulation* 13 (1), 39–46
- McGraw, K. O. (1991): Problems with the BESD: A comment on Rosenthal's "How we are doing in soft psychology?" *American Psychologist* 46, 1084
- Mensen, H. (1994): *Das autogene Training. Entspannung-Gesundheit-Stressbewältigung* (16. Aufl.). Goldmann, München
- Metzler, P. & Krause, B. (1997): Methodischer Standard bei Studien zur Therapieevaluation. *Methods of psychological research online* 1997 2 (1), Pabst Science Publishers, Internet: <http://www.pabst-publishers.de/mpr/>
- Meyer, A.-E. (1994): Über die Wirksamkeit psychoanalytischer Therapie bei psychosomatischen Störungen. *Psychotherapeut* 39, 298–308
- Meyer, A.-E. (1995): Et tamen florent confessions: Schlußwort zu Grawes Replik. *Psychotherapeut* 40, 107–110
- \*Milne, B., Joachim, G. & Niedhardt, J. (1986): A stress management programme for inflammatory bowel disease patients. *Journal of Advanced Nursing* 11, 561–567
- \*Nicassio, P. & Bootzin, R. (1974): A comparison of progressive relaxation and autogenic training as treatments for insomnia. *Journal of Abnormal Psychology* 83 (3), 253–260
- Olkin, I. (1995): Statistical and theoretical considerations in meta-analysis. *Journal of Clinical Epidemiology* 48 (1), 133–146
- \*O'Moore, A. M., O'Moore, R. R., Harrison, R. F., Murphy, G. & Carruthers, M. E. (1983): Psychosomatic aspects in idiopathic infertility: Effects of treatment with autogenic training. *Journal of Psychosomatic Research* 27 (2), 145–151

- Orwin, R. G. (1983): A fail safe N for effect size in meta-analysis. *Journal for Educational Statistics* 8, 157–159
- Petrak, F., Hardt, J., Nickel, R. & Egle, U. T. (1999): Checkliste zur Bewertung der wissenschaftlichen Qualität kontrollierter psychotherapeutischer Interventionsstudien (CPI). *Psychotherapeut* 44, 390–393
- \*Polácková, J., Bocková, E. & Sedivec, V. (1982): Autogenic training: Application in secondary prevention of myocardial infarction. *Activitas Nervosa Superior (Praha)* 24 (3), 178–180
- \*Prill, H. J. (1965): Das Autogene Training in der Geburtshilfe und Gynaekologie. In W. Luthe (Ed.), *Autogenes Training: Correlationes psychosomaticae* (S. 234–246). Thieme, Stuttgart
- \*Rakov, A. L., Mandrykin, Iu. V. & Zamotaev, Iu. N. (1997): The results of autogenic training in patients with ischemic heart disease. *Voenno-Medicinskij Zurnal Taermedizin* 318 (2), 37–41
- Reed, R. & Meyer, R. G. (1974): Reduction of test anxiety via autogenic training. *Psychological Reports* 35, 649–650
- \*Reich, B. A. (1989): Non-invasive treatment of vascular and muscle contraction headache: A comparative longitudinal study. *Headache* 29, 34–41
- Rockstroh, B., Elbert, T., Birbaumer, N., Wolf, P., Düchting-Röth, A., Daum, I., Lutzenberger, W., Dichgans, J. (1993) Cortical self-regulation in patients with epilepsies. *Epilepsy Research* 14, 63-72
- Röhr, M., Lohse, H. & Ludwig, R. (1983): Statistische Verfahren. *Statistik für Soziologen, Pädagogen, Psychologen und Mediziner*, Bd. 2 (hrsg. von G. Clauß & H. Ebner). Deutscher, Frankfurt a. M.
- Rosenberg, M. S., Adams, D. C. & Gurevitch, J. (2000): *MetaWin, Version 2.0: Statistical software for meta-analysis*. Sinauer Associates, Inc., Sunderland, Massachusetts
- Rosenthal, R. (1979): The “file drawer problem” and tolerance for null results. *Psychological Bulletin* 86, 638–641
- Rosenthal, R. (1983): Assessing the statistical and social importance of the effects of psychotherapy. *Journal of Consulting and Clinical Psychology* 51, 4–13
- Rosenthal, R. (1990): How are we doing in soft psychology? *American Psychologist* 45, 775–777
- Rosenthal, R. (1991): Effect sizes: Pearson’s correlation, its display via the BESD, and alternative indices. *American Psychologist* 46, 1086
- Rosenthal, R. & Rubin, D. B. (1978): Interpersonal expectancy effects: The first 345 studies. *The Behavioral and Brain Sciences* 3, 377–415
- Rosenthal, R. & Rubin, D. B. (1979): Comparing significance levels of independent studies. *Psychological Bulletin* 86, 1165–1168
- Rosenthal, R. & Rubin, D. B. (1982): A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology* 74, 166–169
- \*Rucco, V., Feruglio, C., Genco, F. & Mosanghini, R. (1995): Autogenic training versus Erikson’s analogical technique in the treatment of fibromyalgia syndrome. *European Review for Medical and Pharmacological Sciences* 17 (1), 41–50
- Rüger, B. (1994): Kritische Anmerkungen zu den statistischen Methoden in Grawe, Donati & Bernauer: “Psychotherapie im Wandel. Von der Konfession zur Profession“ *Zeitschrift für psychosomatische Medizin* 40, 368–383
- \*Sargent, J., Solbach, P., Coyne, L., Spohn, H. & Segerson, J. (1986): Results of a controlled, experimental, outcome study of nondrug treatments for the control of migraine headaches. *Journal of Behavioral Medicine* 9 (3), 291–323
- Schwarzer, R. (1995): Meta-analysis program. Internet: <http://www.fu-berlin.de/gesund/>

- Sauer, J. & Schnetzer, M. (1978): Zum Persönlichkeitsbild des Asthmatikers und seiner Veränderung durch unterschiedliche Behandlungsmethoden im Verlauf seiner Kur. *Zeitschrift für Klinische Psychologie und Psychotherapie* 26 (2), 171–180
- \*Schrapper, D. K. & Mann, K. F. (1985): Veränderung der Befindlichkeit durch autogenes Training. *Psychotherapie und medizinische Psychologie* 35, 268–272
- Schulte, C. (1983): Das verhaltensmedizinische Gruppentraining (VMGT): Entwicklung und Evaluation in der klinischen Rehabilitation. *Verhaltensmodifikation* 3, 71–87
- Schultz, I.H. (1920): Über Schichtenbildung im hypnotischen Selbstbeobachten. *Monatsschrift für Psychiatrie* 49, 137–143
- Schultz, J. H. (1991): *Das autogene Training* (18. Aufl.). Thieme, Stuttgart (Original erschienen 1932: *Das autogene Training*)
- Schultz, J. H. & Luthe, W. (1969): *Autogenic therapy. Vol I: Autogenic methods.* Grune & Stratton, New York
- Seipp, B. (1989): *Angst und Leistung in Schule und Hochschule: eine Meta-Analyse.* Unpublizierte Dissertation, Universität Düsseldorf
- Sellers, D. J. (1974): Teaching a self-initiated control technique to individuals and a group in college. *The International Journal of Clinical and Experimental Hypnosis* 22, 38–45
- \*Sharp, C., Hurford, D. P., Allison, J., Sparks, R. & Cameron, B. P. (1997): Facilitation of internal locus of control in adolescent alcoholics through a brief biofeedback-assisted autogenic relaxation training procedure. *Journal of Substance Abuse Treatment* 14 (1), 55-60
- Simma, L. & Benzer, W. (1985): Rehabilitation nach Herzinfarkt: Autogenes Training in der ambulanten Koronarsportgruppe. *Münchener Medizinische Wochenschrift* 21, 527–531
- Shapiro, D. A. & Shapiro, D. (1982): Meta-analysis of comparative therapy outcome studies: A replication and refinement. *Psychological Bulletin* 92 (3), 581–604
- Shoham-Salomon, V. & Rosenthal, R. (1987): Paradoxical interventions: A meta-analysis. *Journal of Consulting and Clinical Psychology* 55, 22–28
- Schmid, J. E., Koch, G. G. & LaVange, L. M. (1991): An overview of statistical issues and methods of meta-analysis. *Journal of Biopharmaceutical Statistics* 1 (1), 103–120
- Smith, M. L. & Glass, G. V. (1977): Metaanalysis of psychotherapy outcome studies. *American Psychologist* 32, 752–760
- Smith, M. L., Glass, G. V. & Miller, T. I. (1980): *The benefits of psychotherapy.* The Johns Hopkins University Press, Baltimore, MD [u.a.]
- Sokal, R. R. & Rohlf, F. J. (1995): *Biometry* (3rd ed.). Freeman, San Francisco
- \*Spiess, K., Sachs, G., Buchinger, C., Röggl, G., Schnack, C. & Haber, P. (1988): Zur Auswirkung von Informations- und Entspannungsgruppen auf die Lungenfunktion und psychophysische Befindlichkeit bei Asthmapatienten. *Praxis der Klinischen Pneumologie* 42, 641–644
- \*Spinhoven, P., Linssen, C. G., Van Dyck, R. & Zitman, F. G. (1992): Autogenic training and self-hypnosis in the control of tension headache. *General Hospital Psychiatry* 14, 408–415
- Sponsel, R. (2000a): Überprüfung der schweren Vorwürfe von Heckrath & Dohmen (1997) zum direkten Wirkungsvergleich zwischen psychoanalytischer und Verhaltenstherapie durch die Forschungsgruppe Grawe et al. (1994). Teil 1. Internet Publikation für Allgemeine und Integrative Psychotherapie IP-GIPT. [http://www.sqipt.org/wisrms/ptf/grawe/g\\_hd1.htm](http://www.sqipt.org/wisrms/ptf/grawe/g_hd1.htm)

- Sponzel, R. (2000b): Überprüfung der schweren Vorwürfe von Heckrath & Dohmen (1997) zum direkten Wirkungsvergleich zwischen psychoanalytischer und Verhaltenstherapie durch die Forschungsgruppe Grawe et al. (1994). Teil 2. Internet Publikation für Allgemeine und Integrative Psychotherapie IP-GIPT. [http://www.sgipt.org/wisms/ptf/grawe/g\\_hd2.htm](http://www.sgipt.org/wisms/ptf/grawe/g_hd2.htm)
- Sponzel, R. (2000c): Überprüfung der schweren Vorwürfe von Heckrath & Dohmen (1997) zum direkten Wirkungsvergleich zwischen psychoanalytischer und Verhaltenstherapie durch die Forschungsgruppe Grawe et al. (1994). Teil 3. Internet Publikation für Allgemeine und Integrative Psychotherapie IP-GIPT. [http://www.sgipt.org/wisms/ptf/grawe/g\\_hd3.htm](http://www.sgipt.org/wisms/ptf/grawe/g_hd3.htm)
- Sponzel, R. (2000d): Überprüfung der schweren Vorwürfe von Heckrath & Dohmen (1997) zum direkten Wirkungsvergleich zwischen psychoanalytischer und Verhaltenstherapie durch die Forschungsgruppe Grawe et al. (1994). Teil 4. Internet Publikation für Allgemeine und Integrative Psychotherapie IP-GIPT. [http://www.sgipt.org/wisms/ptf/grawe/g\\_hd4.htm](http://www.sgipt.org/wisms/ptf/grawe/g_hd4.htm)
- SPSS Incorporation (1999): Statistical Package for Social Sciences for Windows (Vers. 10.0.5). SPSS Inc., Chicago
- Sterling, T. D. (1959): Publication decisions and their possible effects on inferences drawn from tests of significance – or vice versa. *Journal of the American Statistical Association* 54, 30–34
- Stetter, F. (1985): Chronobiologische Aspekte beim autogenen Training: thermometrische Befunde beim autogenen Training in Abhängigkeit von der Tagesperiodik bei vegetativ dystonen Patienten. *Zeitschrift für Psychosomatische Medizin und Psychoanalyse* 31, 172–186
- Stetter, F. (1991): Die Bedeutung der Hypnosuggestiv-Verfahren in der Psychiatrie - ein empirisch fundierter, pragmatischer Behandlungsansatz. In F. Schneider, M. Bartels, K. Foerster & H. J. Gaertner (Hrsg.), *Perspektiven der Psychiatrie* (S. 153-159). G. Fischer, Stuttgart und Jena
- Stetter, F. (1994): Gestufte Aktivhypnose, autogenes Training und zweigleisige Psychotherapie. *Fundamenta Psychiatrica* 8, 14–20
- Stetter, F. (1998): Was geschieht, ist gut. Entspannungsverfahren in der Psychotherapie. *Psychotherapeut* 43, 209–220
- Stetter, F., Mann, K. (1992): Autogenes Training. Empirisch begründetes psychotherapeutisches Verfahren in der Primärversorgung. *Deutsches Ärzteblatt* 89, 1427–1428
- Stetter, F., Günthner, A. H., Mann, K. F. & Bartels, M. (1994): Physiologische und psychische Veränderungen bei der Behandlung mit autogenem Training in der Rehabilitation von Herzinfarktpatienten. *Praxis der Klinischen Verhaltensmedizin und Rehabilitation* 25, 46–52
- Stetter, F. & Kupper, S. (1998): Autogenes Training – Qualitative Meta-Analyse kontrollierter klinischer Studien und Beziehungen zur Naturheilkunde. *Forschende Komplementärmedizin* 5, 211–223
- Stetter, F. & Kupper, S. (2002): Autogenic training: A meta-analysis of clinical outcome studies. *Applied Psychophysiology and Biofeedback* 27 (1), 45–98
- \*Stetter, F., Walter, G., Zimmermann, A., Zähres, S. & Straube, E. R. (1994): Ambulante Kurztherapie mit autogenem Training und Hypnose. *Psychotherapie, Psychosomatik und medizinische Psychologie* 44, 226–234
- Stoffelmayr, B. E., Dillavou, D. & Hunter, J. E. (1983): Premorbid functioning and outcome in schizophrenia: A cumulative analysis. *Journal of Consulting and Clinical Psychology* 51, 338–352
- Strahan, R. F. (1991): Remarks on the binomial effect size display. *American Psychologist* 46, 1083

- \*Surwit, R. S., Allen, L. M., Gilgor, R. S. & Duvic, M. (1982): The combined effect of prazosin and autogenic training on cold reactivity in raynaud's phenomenon. *Biofeedback and Self-Regulation* 7 (4), 537–544
- \*Surwit, R. S., Pilon, R. N. & Fenton, C. H. (1978): Behavioral treatment of raynaud's disease. *Journal of Behavioral Medicine* 1 (3), 323–335
- ter Kuile, M. M., Spinhoven, P., Linssen, A. C., Zitman, F. G., & Van Dyck, R. (1994): Autogenic training and cognitive self-hypnosis for the treatment of recurrent headaches in three different subject group. *Pain* 58, 331–340
- \*ter Kuile, M. M., Spinhoven, P. & Linssen, A. C. (1995): Responders and nonresponders to autogenic training and cognitive self-hypnosis: Prediction of short- and long-term success in tension-type headache patients. *Headache* 35, 630–636
- Tingey, R., Burlingame, G., Lambert, M. J. & Barlow, S. H. (1989): Assessing clinical significance: Extensions and applications. Paper presented at the Society for Psychotherapy Research, Wintergreen, VA
- Tschuschke, V., Heckrath, C. & Tress, W. (1997): Zwischen Konfusion und Makulatur: Zum Wert der Berner Psychotherapie-Studie von Grawe, Donati & Bernauer. Vandenhoeck & Ruprecht, Göttingen
- Tschuschke, V., Kächele, H. & Hölzer, M. (1994): Gibt es unterschiedlich effektive Formen von Psychotherapie? *Psychotherapeut* 39, 281–297
- \*Tsikulin, A. E., Zinkovskiy, A. K. & Volkov, V. S. (1989): L'aspect psychologique de la récupération des hypertendus. *La Santé Publique* 32 (3), 287–294
- Vaitl, D. (1993): Psychophysiologie der Entspannung. In D. Vaitl & F. Petermann (Hrsg.), *Handbuch der Entspannungsverfahren*, Bd. 1 (S. 25–63). Psychologie-Verlags-Union und Beltz, Weinheim
- Van Dyck, R., Zitman, F. G., Linssen, A. C. G. & Spinhoven, P. (1991): Autogenic training and future oriented hypnotic imagery in the treatment of tension headache: Outcome and process. *The International Journal of Clinical and Experimental Hypnosis* 39 (1), 6–23
- Wang, M. C. & Bushman, B. J. (1998): Using the normal quantile plot to explore meta-analytic data sets. *Psychological Methods* 3, 46–54
- Wolf, F. M. (1986): *Meta-analysis: Quantitative methods for research synthesis*. Sage, Beverly Hills, CA
- \*Winterfeld, H.-J., Siewert, H., Bohm, J., Frenzel, R., Aurisch, R., Ecke, A., Engelmann, U., Strangfeld, D. & Warnke, H. (1993): Autogenes Training bei hypertonen Regulationsstörungen nach aortokoronarer Venenbypass-Operation (ACVB) bei koronarer Herzkrankheit. *Innere Medizin* 48, 201–204
- \*Yurenev, A. P., Patrusheva, I. F., Aivazyan, T. A., Zaitsev, V. P. Krol, V. A., Smorodina, N. F. & Mokrievich, E. A. (1988): The use of various non-pharmacological methods in the treatment of patients in the early stages of arteriel hypertension. *Terapevticheskii Arkhiv (Moskva)* 60, 123–126
- Zimmermann-Tansella, C., Dolcetta, G., Azzini, V., Zacche, G., Bertagni, P., Siani, R. & Tansella, M. (1979): Preparation courses for childbirth in primipara: A comparison. *Journal of Psychosomatic Research* 23, 227–233
- \*Zitman, F. G., Van Dyck, R., Spinhoven, P. & Linssen, A. C. G. (1992): Hypnosis and autogenic training in the treatment of tension headaches: A two-phase constructive design study with follow-up. *Journal of Psychosomatic Research* 36 (2), 219–228