

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/236058464>

Everyday Evidence: Outcomes of Psychotherapies in Swedish Public Health Services

Article *in* Psychotherapy Theory Research Practice Training · March 2013

DOI: 10.1037/a0031386 · Source: PubMed

CITATIONS

6

READS

99

4 authors, including:



Andrzej Werbart
Stockholm University
71 PUBLICATIONS 407 CITATIONS

[SEE PROFILE](#)



Lars Levin
Stockholm University
2 PUBLICATIONS 9 CITATIONS

[SEE PROFILE](#)



Rolf Sandell
Linköping University
82 PUBLICATIONS 1,071 CITATIONS

[SEE PROFILE](#)

Everyday Evidence: Outcomes of Psychotherapies in Swedish Public Health Services

Andrzej Werbart and Lars Levin
Stockholm University and Stockholm County Council,
Stockholm, Sweden

Håkan Andersson
Stockholm University

Rolf Sandell
Linköping University and Lund University

This naturalistic study presents outcomes for three therapy types practiced in psychiatric public health care in Sweden. Data were collected over a 3-year period at 13 outpatient psychiatric care services participating in the online Quality Assurance of Psychotherapy in Sweden (QAPS) system. Of the 1,498 registered patients, 14% never started psychotherapy, 17% dropped out from treatment, and 36% dropped out from data collection. Outcome measures included symptom severity, quality of life, and self-rated health. Outcomes were studied for 180 patients who received cognitive-behavioral, psychodynamic, or integrative/eclectic therapy after control for dropout representativeness. Among treatment completers, patients with different pretreatment characteristics seem to have received different treatments. Patients showed significant improvements, and all therapy types had generally good outcomes in terms of symptom reduction and clinical recovery. Overall, the psychotherapy delivered by the Swedish public health services included in this study is beneficial for the majority of patients who complete treatment. Multilevel regression modeling revealed no significant effect for therapy type for three different outcome measures. Neither did treatment duration have any significant effect. The analysis did not demonstrate any significant therapist effects on the three outcome measures. The results must be interpreted with caution, as there was large attrition and incomplete data, nonrandom assignment to treatment, no treatment integrity control, and lack of long-term follow-up.

Keywords: psychotherapy, effectiveness, naturalistic design, routine clinical practice, therapy types

The current focus on generating lists of evidence-based psychotherapies ultimately implies the simplistic question, “Which form(s) of psychotherapy is (are) the best?” However, as a recent statement from the American Psychological Association (2012) makes clear, there are various sources of evidence, and a multitude of research methods are needed to build an empirical knowledge base applicable to routine psychotherapy practice. Different research paradigms stress distinct methodological aspects, for instance regarding validity. Naturalistic studies focus on external

validity, whereas experimental studies try to maximize internal validity. Incompatibilities between the prerequisites of experimental design in randomized controlled trials (RCT) and clinical practice have been pointed out frequently (Braslow et al., 2005; Sanson-Fisher, Bonevski, Green, & D’Este, 2007), leading to the suggestion of applying pluralistic methodologies (Black, 1996; Morrison, Bradley, & Westen, 2003; Westen, Novotny, & Thompson-Brenner, 2004; Norcross, Beutler, & Levant, 2006; Wachtel, 2010). Ideally, RCTs should be succeeded by effectiveness studies, which aim at the generalizability of results to practice.

The effectiveness of psychotherapy as practiced in the community is an important matter for politicians, public health-service managers, and above all for ordinary patients and their therapists. There is also a need for the empirical evaluation of psychotherapies to make them part of the larger health care system (Kendall, 1998). Knowledge of how therapy works in routine clinical practice may help us optimize treatments, provide individually adjusted interventions, and reduce the hiatus between research and practice (Kazdin, 2008; Norcross, 2002). Finally, it can provide clients with research-based information about effective treatments and help them make adequate choices (Høglend, 1999). The evidence base for psychotherapy can be enriched by naturalistic effectiveness studies, addressing such epidemiological and “market-relevant” questions as: Who is in psychotherapy in routine public service

Andrzej Werbart, Department of Psychology, Stockholm University, Stockholm, Sweden, and Child and Adolescent Psychiatry, Stockholm County Council, Stockholm, Sweden; Lars Levin, Outpatient Psychiatric Clinic West, Northern Stockholm Psychiatry, Stockholm County Council and Department of Psychology, Stockholm University; Håkan Andersson, Department of Psychology, Stockholm University; Rolf Sandell, Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden, and Lund University, Department of Psychology, Lund, Sweden.

This study was supported by grants from the Research and Development Committee, Stockholm County Council, Sweden.

Correspondence concerning this article should be addressed to Andrzej Werbart, PhD, Department of Psychology, Stockholm University, SE-106 91 Stockholm, Sweden. E-mail: andrzej.werbart@sll.se or andrzej@werbart.se

settings? What types of psychotherapy are practiced? What is the outcome of the various psychotherapies, and how does outcome differ among treatments?

This type of questions has been raised, among others, in the U.K. in CORE (*Clinical Outcomes in Routine Evaluation*) studies. Stiles, Barkham, Twigg, Mellor-Clark, and Cooper (2006, *Stiles, Barkham, Mellor-Clark, & Connell, 2008a*) have found that theoretically different psychotherapy types appear to be equally effective in National Health Service settings. Nowadays, there are several systems for routine outcome monitoring in regular clinical practice, such as the *Treatment Outcome Package* (TOP; Kraus, Seligman, & Jordan, 2005) in U.S., the *AKtive Interne Qualitätsicherung* (web-AKQUASI; Kordy & Bauer, 2003; Kordy & Gallas, 2007; Kordy, Hannover, & Richard, 2001; Percevic, Gallas, Arikan, Mössner, & Kordy, 2008) in Germany, or the *Routine Outcome Monitoring* (ROM; de Beurs et al., 2011) in the Netherlands. A system for *Quality Assurance of Psychotherapy in Sweden* (QAPS) was another attempt to address these questions.

The general aim of the present study is to replicate and complement the U.K. studies in the Swedish public health services and, using several well-established outcome measures, widen the basis for everyday evidence. Our main objective is to study outcomes for patients in different types of psychotherapy. To address the issue of representativity, we first investigate differences between patients who remained in treatment and patients who dropped out of treatment or data collection, present patients' pretreatment characteristics, and describe the distribution of patients among different types of psychotherapy. Given this distribution, we then compare pretreatment characteristics of patients assigned to different types of psychotherapy. Finally, we compare three different outcome measures, of symptoms, quality of life, and self-rated health, presumably capturing different aspects of change in different types of psychotherapy.

Method

Setting

Data were collected over a 3-year period between January 2007 and February 2010 from patients who received cognitive-behavioral, psychodynamic, or integrative/eclectic psychotherapy and their therapists at 13 outpatient services, primarily under the auspices of the Stockholm County Council. The included clinics participated in QAPS, a routine evaluation system with a core battery of well-established instruments, which are theory-neutral and use online data entry. The clinics had to offer psychotherapy treatment, apply for the participation in the QAPS quality register, and provide resources approved by the clinic leadership for a local coordinator and for data collection. The patient questionnaire consists of sociodemographic data and self-rating scales aimed at covering different aspects of outcome. The therapist questionnaire contains data about the therapist and the treatment as well as different assessments of the patient. The questionnaires were administered before and after treatment, rendering four possible data sets: patient data pre- and posttreatment, and therapist data pre- and posttreatment. Both patients and therapists completed an informed consent form.

Treatment Types

At termination, the therapists had to specify the frequency, number of sessions, and type of the treatment delivered, using the predefined response categories "cognitive," "cognitive-behavioral," "psychodynamic," "psychoanalysis," "integrative/eclectic," "systemic," "creative (art, dance, music)," or "other." Based on therapists' self-reports, the treatments were eventually classified as belonging to one of three categories: CBT (cognitive or cognitive-behavioral), PDT (psychodynamic), or INT (integrative or eclectic). No case of psychoanalysis was indicated in the material. The remaining therapy types (systemic, expressive, combination of therapies or other), as well as missing indication, were treated as a separate category (0.5% of cases), not included in further analyses.

Attrition

We distinguish among four main forms of attrition: nonstarters (patients who were only assessed for psychotherapy, those referred to other services, or who never showed up for psychotherapy after the initial contact), treatment dropouts (the treatment was discontinued before the planned termination), incomplete data at termination (completed therapy but missing posttreatment questionnaires), and ongoing therapies (no end-data available). Not starting treatment was indicated in the therapist pretreatment questionnaire, whereas premature termination was indicated in the therapist post-treatment questionnaire. The interruption of the treatment had to have taken place after the completion of the intake interviews and mutual agreement to commence therapy had been reached. This coding was based on therapist judgment, which is one of the two methods that have demonstrated adequate agreement in classifying participants as premature terminators (Hatchett & Park, 2003). The participants' possibilities and willingness to complete questionnaires was in many cases affected by reorganization, closure, or privatization of services, owing to structural changes introduced in Stockholm County psychiatry during the data collection period. The attrition from admission to data analysis is presented in Figure 1.

The QAPS database comprised pretreatment questionnaires from 1,498 patients, of whom 204 (13.6%) never started psychotherapy after initial assessment, 260 (17.4%) dropped out of treatment, and 535 (35.7%) were incomplete data collections. At the closure of data collection, 311 of the remaining 499 patients were in ongoing therapy. Consequently, complete data (all four data sets) were available for 188 patients (12.6%) who had finished treatment. Our effectiveness analysis included all patients with enough data to perform pre/post treatment analysis and for whom psychotherapy type was PDT, CBT, or INT ($n = 180$), reaching a "core sample" of 12.0% of the total sample.

Core Sample

The 180 patients in the core sample were treated at three specialized psychotherapy units ($n = 101$; 56.1%), one primary-care center ($n = 54$; 30.0%), one young adults psychotherapy unit ($n = 17$; 9.4%), and four outpatient psychiatric care clinics ($n = 8$; 4.4%) (see Table 1). The average age of the patients was 36 at treatment start (range, 15–70; $SD = 10.9$), and 74% were female.

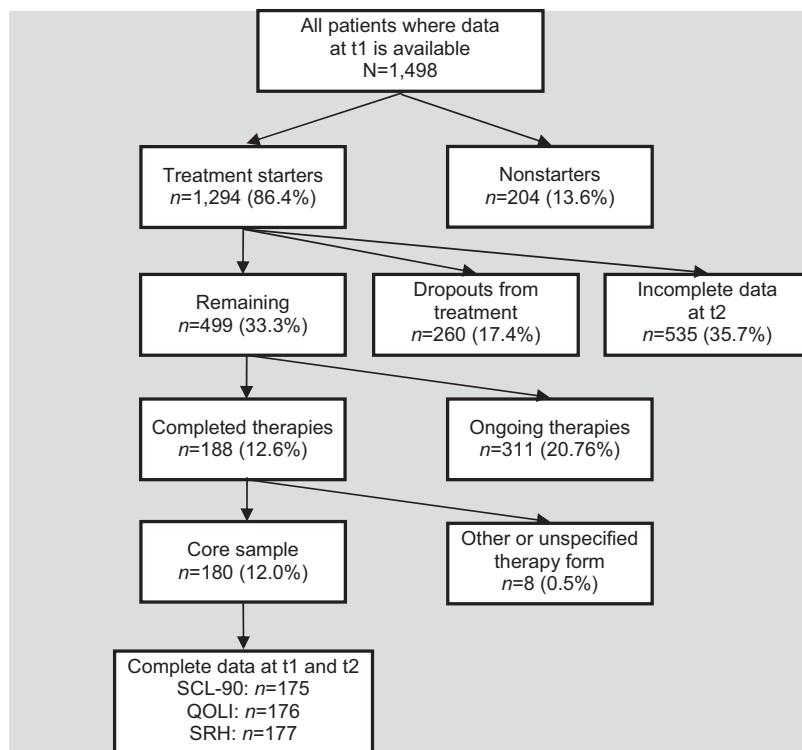


Figure 1. Flow-chart of attrition from admission to complete data collection. Pretreatment data = t1, posttreatment = t2.

The patients were treated by 75 therapists; 94 (52.2%) of patients by therapists with 3 years (halftime) advanced postgraduate psychotherapy training and the Swedish National Board of Health and Welfare license to practice as a psychotherapist, 81 (45.0%) by clinicians with basic 1.5-year (halftime) training in psychotherapy meant to make the clinicians sufficiently competent to provide psychotherapy under supervision, and 5 (2.8%) by therapists whose level of education had not been indicated. Licensed therapists had access to supervision in difficult cases, whereas therapists with basic training were supervised on a regular basis. No further data on the therapists' experience were available.

The therapists' levels of psychotherapy training did not differ significantly between therapy types [χ^2 (4, 180) = 5.215, p =

.266]. The majority, 90.4%, of patients met their therapist once a week, 3.6% less frequently, and 6.0% twice a week. Each therapist had between 1 and 23 patients (Mdn = 1); 7 therapists had 5 or more patients, while 68 therapists had 4 or fewer. The average age of the therapists was 45 at the start of treatments (range, 29–68; SD = 12.9), and 87% were female.

Data Collection

Patients were asked to complete the pretreatment questionnaire either before their first visit, during psychological screening, or after the first therapy session. In most cases, the questionnaires were administered online; paper-and-pencil versions were sometimes used. Patients were allocated to treatments and therapists following routine procedure at each unit. The posttreatment QAPS questionnaires were administered in connection with the last but one or last therapy session. Nonresponders received at least one reminder by mail after 2–3 weeks. Therapists completed pretreatment QAPS questionnaires after the initial assessment and post-treatment questionnaires after the last therapy session. Anonymized data were stored on an independent IT management system.

Self-Rated Outcome Measures

In addition to extensive sociodemographic data, three instruments included in the QAPS questionnaire were applied to assess pretreatment characteristics and measure outcomes in the sample.

Symptom Check List-90 (SCL-90; Derogatis, 1994) reports psychiatric symptoms experienced over the last seven days. Items

Table 1
Clinic Characteristics and Distribution of Patients in the Core Sample

Clinic ID	Clinic characteristics	Frequency	%
7	Primary-care center	54	30.0
1	Specialized psychotherapy unit	51	28.3
2	Specialized psychotherapy unit	37	20.6
8	Young adults psychotherapy unit	17	9.4
6	Specialized psychotherapy unit	13	7.2
4	Outpatient mental care clinic	4	2.2
3	Outpatient mental care clinic	2	1.1
12	Outpatient mental care clinic	1	0.6
13	Outpatient mental care clinic	1	0.6
Total		180	100.0

are rated on 5-point Likert scales ranging from 0 ("not at all") to 4 ("very much"). The Global Severity Index (GSI) was used for the analyses in this study. The SCL-90 demonstrates adequate reliability (Derogatis, Rickels, & Rock, 1976; Fridell, Cesarec, Johansson, & Malling Thorsen, 2002).

The Quality of Life Inventory (QOLI; Frisch, Cornell, Villanueva, & Retzlaff, 1992; Frisch et al., 2005) is a measure of life satisfaction, well-being, and positive mental health. The 17 areas of life are rated in terms of their importance (from 0, "not at all important" to 2, "extremely important") and in terms of satisfaction with the area (from -3, "very dissatisfied" to +3, "very satisfied"). The overall QOLI score is obtained by averaging all weighted satisfaction ratings that have nonzero importance ratings. The reported test-retest and internal consistency reliabilities were high (Frisch et al., 1992; Paunović & Öst, 2004; Öst, personal communication, 28 May, 2010).

The correlation between SCL-90 and CORE-OM (all items) was reported as $r = .88$, whereas the correlation with CORE-OM subset "well-being" had an $r = .68$ (Evans et al., 2002). This indicates that symptom severity and life satisfaction can be viewed as alternative outcome measures (Frisch et al., 2005; Perry & Bond, 2009). In our material, we found a significant moderate correlation of $r(184) = -0.519$, $p = .000$, between GSI and QOLI pretreatment (all cases included), the negative correlation being an issue of scaling, with GSI higher scores indicating more pathology and QOLI higher scores indicating greater life satisfaction.

Self-Rated Health (SRH, Björner et al., 1996) is a single-item measure where the respondent rates his or her present subjective mental and somatic health on a 7-point Likert scale ranging from 1 ("very bad") to 7 ("very good"). SRH has been shown to be a good predictor of health problems (Nilsson & Orth-Gomér, 2000) and to have good test-retest reliability (Lorig et al., 1996).

Therapist Assessments

The therapist questionnaire requested information about the therapists, the patients, and the treatments. The therapists' assessment of the patient comprised data on the nature, severity, and duration of the patient's presenting problems, using 18 categories: depression, anxiety, psychosis, stress/exhaustion, aggressive acting-out, sexual acting-out, eating disorder, addiction, physical problems, self-harm, suicidal attempts, self-esteem, interpersonal problems, work/academic, criminal acting-out and changes in temper, the patient's experiences of bereavement/loss and trauma/abuse, risk assessment of being dangerous for oneself and for others (extended version of the CORE categories; Stiles et al., 2006, 2008a), as well as Global Assessment of Functioning (GAF; American Psychiatric Association, 2000). Furthermore, the therapists indicated the patient's *DSM-IV* or ICD-10 diagnoses following routine clinical procedure, but without a systematic use of any structured diagnostic instrument.

Statistical Analysis

Patients were included in the analyses only if they had data at both baseline and posttreatment. Owing to the various types of attrition and missing data, the sample size varies for different statistical analyses and for each specific measure.

Following Stiles et al. (2008a), for within-group comparisons, we used Cohen's d (Cohen, 1992) as well as reliable and clinically

significant change (Jacobson, Roberts, Berns, & McGlinchey, 1999; Jacobson & Truax, 1991) as effectiveness measures. Cohen's d is defined as mean pre/post difference divided by the pooled standard deviation. Confidence intervals for effect sizes were set to 95% and calculated following the Wilson score method without continuity correction (Newcombe, 1998).

Reliable change and clinically significant change were calculated separately for the three outcome measures. Reliable change (RC) is achieved if the reliable change index is ≥ 1.96 ($p < .05$). For movement into a functional distribution, the cutoff between clinical and nonclinical range was determined in accordance with Jacobson & Truax (1991; Jacobson et al., 1999) criterion (c), as recommended when the distributions of the functional and dysfunctional population overlap. For clinically significant change (conditional stimulus [CS]), the patients have to achieve both reliable change and move out of the clinical distribution into the functional distribution.

To calculate the RC index of GSI, we used a test-retest reliability of 0.94 over a 2-week period based on a nonpatient sample (Edwards, Yarvis, Mueller, Zingale, & Wagman, 1978; Tingey, Lambert, Burlingame, & Hansen, 1996). Comparing the QAPS sample and Swedish norms (Fridell et al., 2002; $M = 0.55$; $SD = 0.46$), the cutoff for GSI was calculated as 0.85. A test-retest reliability of 0.80, as reported for a nonclinical sample over a period of 2 to 3 weeks (Frisch et al., 1992) was used to calculate the RC index of QOLI. The clinical cutoff for QOLI was calculated as 1.87, using a Swedish nonclinical norm group of 100 individuals, drawn from 1,000 individuals randomly selected from the whole population in the Stockholm ($M = 2.92$; $SD = 1.08$; Paunović & Öst, 2004; Öst, personal communication, 28 May, 2010). A test-retest reliability of 0.92, reported by Lorig et al. (1996), was used to calculate the RC index of SRH. The clinical cutoff for SRH was calculated as 4.32, using a Swedish nonclinical norm group of 207 individuals in the age span 20 to 64 years ($M = 5.15$; $SD = 1.36$; Lazar, personal communication, 8 November 2004, cf., Philips, Wennberg, Werbart, & Schubert, 2006).

For between-groups comparisons, we used multilevel regression modeling and Pearson's chi-square where applicable. The multilevel regression analyses were performed using the statistical package Mplus (version 6.12; Muthén & Muthén, 1998–2010). All other analyses were performed using SPSS v. Nineteen software.

Results

As preliminaries to the study of treatment outcomes in the core sample, we present attrition groups and compare patient pretreatment characteristics by therapy type.

Patient Pretreatment Characteristics by Attrition Group

The mean age of treatment starters was 32.2 years (range, 14–78; $SD = 11.0$), and 74% were female. There were significant differences for nonstarters regarding age ($M = 34.5$) ($t = 2.708$, $df = 1472$, $p = .007$) and gender (66% female) [$\chi^2 (1, 1492) = 5.775$, $p = .016$].

Among treatment starters, sociodemographic data were compared for dropouts from treatment, cases with incomplete data at termination (patient and/or therapist questionnaires lacking), and

remaining patients. There was a significant difference regarding age between remaining ($M = 32.3$) and dropouts from treatment ($M = 30.2$) [$F(2, 1269) = 6.291, p = .002$, partial $\eta^2 = 0.010$]. There was also a significant difference regarding gender [$\chi^2 (2, 1289) = 12.780, p = .002$], as revealed in pairwise comparisons between remaining (78.9% female) and cases with incomplete data at termination (69.1% female). No significant differences were found in educational level [$\chi^2 (4, 1282) = 8.196, p = .085$] or in occupational status [$\chi^2 (6, 1286) = 6.961, p = .324$]. There was significant variation regarding previous psychiatric contacts [$\chi^2 (4, 1273) = 13.204, p = .010$]. However, post hoc tests did not reveal any significant differences in pairwise comparisons of dropouts from treatment, cases with incomplete data at termination, and remaining patients.

Regarding patients' self-assessments at baseline (see Table 2), there was nominally a statistically significant difference in GSI, but post hoc analysis (Bonferroni) revealed only a tendency ($p = .064$) toward higher mean GSI level for dropouts from treatment than cases of incomplete data at termination. There were no significant differences between the groups regarding QOLI and SRH. This implies that the remaining patients were comparable with dropout groups in terms of self-rated psychological distress.

Distribution of Patients Between Psychotherapy Types

In the core sample ($n = 180$), the most common psychotherapy type was PDT ($n = 118$; 65.6%), followed by CBT ($n = 31$; 17.2%) and INT ($n = 31$; 17.2%). A comparable distribution of these three psychotherapy types was found among all cases where psychotherapy type had been indicated ($n = 328$): the most common psychotherapy was PDT ($n = 206$; 62.8%), followed by CBT ($n = 63$; 19.2.0%) and INT ($n = 59$; 18.0%).

Patient Pretreatment Characteristics by Therapy Type

For the three treatment groups in the core sample, a significant difference was found regarding age [$F(2, 179) = 4.558, p = .012$, partial $\eta^2 = 0.049$]. Post hoc pairwise tests (Bonferroni) revealed that INT patients were significantly younger than PDT patients

Table 2
Patient Self-Ratings at Baseline by Attrition Group (All Patients Where Data Are Available)

Patient Self-Ratings	Remaining	Dropouts from treatment	Incomplete data at termination	F	p
GSI, SCL-90 ($n = 1,258$)	487	254	517		
Mean	1.37	1.41	1.29	3.061	.047*
SD	0.03	0.04	0.03		
QOLI ($n = 1,278$)	493	258	527		
Mean	0.23	0.14	0.33	1.312	.270
SD	0.07	0.10	0.07		
SRH ($n = 1,279$)	495	254	530		
Mean	3.45	3.50	3.53	0.426	.653
SD	0.06	0.08	0.06		

Note. n varies owing to the varying number of respondents who completed each specific instrument.

* $p < .05$.

($p = .010$). INT patients' mean age was 26.1, as compared with PDT patients' mean age 32.5 years (while CBT patients' mean age was 32.3). No significant differences were found regarding gender [$\chi^2 (2, 180) = 2.292, p = .318$].

There was a significant difference regarding education level [$\chi^2 (4, 180) = 12.414, p = .015$]. Pairwise comparisons revealed that the difference was between INT patients and the other two groups, INT patients having slightly lower education level than PDT patients ($p = .002$). There was also a tendency in the same direction for INT versus CBT patients ($p = .067$). These differences in educational level were obviously correlated with age differences. There were no significant differences regarding occupational status [$\chi^2 (6, 180) = 16.306, p = .178$] or previous psychiatric contacts [$\chi^2 (4, 179) = 1.959, p = .743$].

The patients' clinical data at baseline by therapy type are presented in Table 3. There were no significant differences between the therapy types in pretreatment levels of GSI, QOLI, and SRH. We also studied the type and number of presenting problems indicated by the therapists of the 180 patients in the core sample in the three therapy types. The distribution by therapy type was fairly similar, even if there were some small differences. As in the study by Stiles et al. (2008a), patients in CBT were less likely to be reported as presenting interpersonal problems (67.7%), compared with patients in PDT (92.4%) and INT (96.8%). Also, problematic self-esteem was less frequently reported for CBT patients (71.0%) compared with PDT (94.1%) and INT (96.8%). CBT therapists also reported less frequently anxiety, aggression, sexual acting-out, self-harm, stress, and mood changes in their patients. On average, CBT therapists indicated their patients reported fewer presenting problems ($n = 6$) than PDT and INT ($n = 8$) therapists. In comparison with the U.K. sample, Swedish patients seem to be described as depressed and anxious to a larger extent.

Missing data on psychiatric diagnoses differed significantly between therapists in the three therapy types. Out of the 180 in the core sample, psychiatric diagnoses were lacking for 19.4% of CBT patients, followed by 36.4% of PDT patients and as many as 58.1% for INT patients (see Table 4). Therefore, the distribution of diagnoses must be treated with caution. This considered, anxiety disorders were significantly more frequent in the CBT group (even if the CBT therapists more seldom indicated anxiety as the patients' presenting problem), whereas mood disorders tended to be more frequent in the PDT group. Between 10.7% and 15.4% of patients did not fulfill criteria for any psychiatric diagnosis.

The frequency of patients with medication differed significantly between therapy types (see Table 4). Pairwise comparisons revealed that CBT patients were more often ($p < .05$) on medication (54.8%) than PDT patients (39.0%) and INT patients (25.8%).

Outcomes of Treatment by Therapy Type

Table 5 shows the mean pre/post treatment scores, standard deviations, and within-group effect sizes for each of the three outcome measures and each treatment group in the core sample. According to these analyses, all three therapy types showed good treatment effects. However, although the CIs overlap, they suggest that INT patients have improved slightly more than patients in the other therapy types in terms of all outcome measures, while the QOLI and the SRH seem to have increased least for CBT patients.

Table 3
Patient Self-Ratings at Baseline by Therapy Type for the Core Sample

Patient Self-Ratings	Therapy type			<i>F</i>	<i>p</i>
	CBT	PDT	INT		
GSI, SCL-90 (<i>n</i> = 177)	30	116	31	2.486	.086
Mean	1.26	1.19	1.46		
SD	0.63	0.57	0.73		
QOLI (<i>n</i> = 178)	30	117	31	0.660	.518
Mean	0.73	0.41	0.58		
SD	1.66	1.44	1.11		
SRH (<i>n</i> = 177)	30	116	31	0.448	.640
Mean	3.73	3.51	3.45		
SD	1.46	1.25	1.23		

Note. *n* varies owing to the varying number of respondents who completed each specific instrument.

CBT patients averaged 18 sessions, PDT patients 31, and INT patients 23. This difference was significant [$F(2, 178) = 4.06, p = .019$, partial $\eta^2 = 0.044$] for CBT and PDT patients (Scheffé, $p < .05$). However, the intensity of treatments, measured as the frequency of sessions, did not differ significantly between psychotherapy types [$\chi^2 (6, 180) = 10.461, p = .107$]. All CBT treatments had sessions weekly; 4.6% of PDT treatments were less frequent, and 9.3% were twice a week; all INT patients but one (3.2%) had weekly sessions.

Next, we tested whether the three therapies showed different treatment outcomes. Acknowledging the nonindependent nature of the data with patients being nested within therapists, multilevel regression analyses were used to ensure that our standard errors were not biased (Hox, 2010). In these models, level 1 was the patient level and level 2 was the therapist level. To estimate the amount of variance explained by the therapist, intraclass correlations (ICCs) were estimated in three random-intercept models, controlling for pretest scores for each outcome variable, respec-

tively (i.e., estimating the ICCs for the residual change scores). For the GSI, ICC = .03; QOLI, ICC = .02; SRH, ICC = .07 ($p > .10$). As patients were not randomly assigned to treatment type, ICCs for pretest scores were calculated. A significant ICC would indicate that patients' mean levels of pretreatment scores would differ between therapists. For the GSI pretest, ICC = 0.07; QOLI, ICC = 0.13; SRH, ICC = 0.05 ($p > .10$). Although the intraclass correlations were generally small for both the residual change scores and the pretest scores, clustered data can still potentially bias significance tests. Therefore, we chose to use multilevel regression models with therapist as a level 2 random effect (random intercept only).

Given that there was no natural control group, effect coding was used instead of dummy coding for therapy type (INT was used as the reference category and coded -1 in both effect variables). With effect coding, the regression coefficient represents the effect of the predictor in relation to the grand mean of the outcome. In addition to therapy type and pretest scores, the logarithm of number of sessions (timelog) and the logarithm of the age of the patient (agelog) were included as covariates. Number of sessions, pretest scores, and patients' age were centered on the grand mean. Table 6 summarizes the results of these analyses, where standardized β is used as a measure of effect size. Note that to calculate the effect of the reference category (INT), one has to add the two regression coefficients for CBT and PDT and multiply this sum by -1. Pretest scores were significantly related to their respective posttest scores. The effects of therapy type were not significant for any of the three outcomes, meaning that the effect of therapy on the three outcomes did not differ between therapy types. Neither was the effect of treatment duration significant. Finally, higher patient age predicted lower SRH residuals.

Table 7 shows the frequency and percentage of patients with reliable change, movement into the functional distribution, clinical significance, and deterioration by therapy types and in total core sample, and outcome measure.

Table 4
Number of Patients in Each Primary DSM-IV Diagnosis, Comorbidity With Axis II and Medication (Core Sample)

Diagnoses and Medication	Therapy type						χ^2	<i>p</i>
	CBT		PDT		INT			
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%		
Core sample (<i>n</i> = 180)	31		118		31			
No DSM-IV diagnosis	3	12.0	8	10.7	2	15.4	0.370	.831
Mood disorders	7	28.0	36	48.0	4	30.8	4.639	.098
Bipolar disorders	0		2	2.7	1	7.7	2.023	.364
Anxiety disorders	13	52.0	11	14.7	3	23.1	14.313	.001***
Maladaptive stress reaction	0		6	8.0	1	7.7	2.214	.331
Somatoform disorders	0		0		1	7.7	7.900	.019*
Eating disorders	0		0		1	7.7	7.900	.019*
Sleep disorders	1	4.0	1	1.3	0		0.983	.612
Interpersonal/relational problems	1	4.0	4	5.3	0		0.772	.680
Other states/disorders	0		4	5.3	0		2.151	.341
Personality disorders	4	16.0	4	5.3	1	7.7	2.912	.233
Co-morbid personality disorders	4	16.0	3	4.0	1	7.7	4.113	.128
Missing diagnostic data	6	19.4	43	36.4	18	58.1	10.029	.007**
Medication	17	54.8	46	39.0	8	25.8	10.884	.028*

Note. Distribution of diagnoses is calculated as percentage of the total number of diagnosed cases in each therapy type.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 5

Clinical Scores for Treatment Groups: Pre- and Posttherapy Means, Within-Group Effect Sizes With 95% Confidence Intervals (Core Sample)

Patient Self-Ratings by Therapy Type	Pretherapy			Posttherapy			Within-groups effect sizes	
	n	Mean	SD	n	Mean	SD	Cohen's <i>d</i>	±95% CI
GSI, SCL-90								
CBT	29	1.30	0.60	29	0.77	0.58	0.89	0.36–1.39
PDT	115	1.19	0.57	115	0.77	0.59	0.72	0.47–1.00
INT	31	1.46	0.73	31	0.72	0.53	1.17	0.58–1.46
Total	175	1.25	0.61	175	0.76	0.58	0.83	0.60–1.01
QOLI								
CBT	29	0.64	1.63	29	1.38	1.53	0.47	0.06–0.96
PDT	116	0.41	1.45	116	1.56	1.52	0.77	0.53–1.06
INT	31	0.58	1.11	31	1.91	1.29	1.10	0.65–1.75
Total	176	0.48	1.42	176	1.59	1.49	0.77	0.57–1.00
SRH								
CBT	30	3.73	1.46	30	5.03	1.16	0.99	0.42–1.36
PDT	116	3.51	1.25	116	4.96	1.21	1.18	0.91–1.42
INT	31	3.45	1.23	31	5.29	1.16	1.54	1.00–1.98
Total	177	3.54	1.28	177	5.03	1.19	1.20	0.96–1.37

Note. n varies owing to the varying number of respondents who completed each specific instrument.

Reliable deterioration was in no treatment group or on no outcome measure more than 4%, whereas reliable improvement varied between 34% and 65%. The clinical dysfunctional group pretreatment consisted of 72% of the patients with GSI data available, 85% of the patients with QOLI data available, and 76% of the patients with SRH data available. Movement from the dysfunctional to the functional distribution varied between 38% and 84%, with somewhat lower proportions for the QOLI. Clinically significant improvement (both reliable change and movement across the clinical cutoff) varied between 14% and 55%, again with lower proportions for the QOLI. No consistent differences between the treatment groups were found.

Discussion

The present study, like the extensive body of previous research (Lambert, Garfield, & Bergin, 2004), showed good outcomes for different psychotherapy types. The within-group effect sizes were comparable with the results from other naturalistic studies (e.g., Minami et al., 2008; Johansson, 2009), and about half of the patients showed reliable change on all outcome measures. Multilevel regression modeling revealed no significant differences between therapy types for the three outcome measures. Neither did treatment duration have any significant effect, despite significant differences in number of sessions between the three therapy types.

Table 6
Multilevel Modeling Estimates Predicting Therapy Outcomes

Fixed effects	GSI		QOLI		SRH	
	Coefficient (SE)	β	Coefficient (SE)	β	Coefficient (SE)	β
Intercept	0.73 (0.05)**		1.58 (0.13)**		5.01 (0.11)**	
CBT ^a	0.00 (0.07)	.00	-0.24 (0.18)	-.09	-0.01 (0.16)	-.00
PDT ^a	0.07 (0.05)	.09	0.04 (0.13)	.02	-0.09 (0.12)	-.06
Timelog	0.12 (0.10)	.08	0.16 (0.27)	.04	-0.05 (0.23)	-.02
Agelog	0.47 (0.27)	.12	-1.01 (0.68)	-.10	-1.62 (0.62)*	-.19
Pretest score	0.56 (0.06)**	.59	0.61 (0.07)**	.59	0.41 (0.06)**	.45
Random effects	Variance	Wald Z	Variance	Wald Z	Variance	Wald Z
Therapist variance	0.005	0.39	0.02	0.22	0.06	0.84
Patient variance	0.21	8.45**	1.41	8.36**	0.99	8.18**

Note. CBT = Cognitive behavioral therapy; INT = Integrative/eclectic therapy; PDT = Psychodynamic therapy; Timelog = the logarithm of number of sessions; Agelog = the logarithm of patients' age.

^a INT is used as the reference category, but because effect coding of the therapy types was used, CBT and PDT represent main effects in relation to the grand mean.

* $p < .01$. ** $p < .001$.

Table 7
Frequency and Percentage of Patient Reliable Change, Movement Into the Functional Distribution, Clinical Significance, and Deterioration by Therapy Types and Outcome Measure (Core Sample)

Outcomes by Outcome Measure	PDT		CBT		INT		Total	
	n	%	n	%	n	%	n	%
GSI								
n	115		29		31		175	
RC	56	49	14	48	20	65	90	51
Functional distribution	76	66	19	66	22	71	117	67
CS	36	31	10	34	15	48	61	35
Deterioration	5	4	0	0	0	0	5	3
QOLI								
n	115		29		31		175	
RC	53	46	10	34	19	61	82	47
Functional distribution	53	46	11	38	16	52	80	46
CS	31	27	4	14	12	39	47	27
Deterioration	2	2	0	0	1	3	3	2
SRH								
n	116		30		31		177	
RC	47	41	13	43	19	61	79	45
Functional distribution	82	71	22	73	26	84	130	73
CS	43	37	12	40	17	55	72	41
Deterioration	0	0	0	0	0	0	0	0

Note. RC = reliable improvement; Functional distribution = crossing the cutoff between clinical and nonclinical population; CS = reliable change and crossing the cutoff between clinical and nonclinical population; Deterioration = reliable deterioration.

These results may imply that the participating units had offered different patients different types and duration of treatments based on patients' individual needs, a factor that should eliminate differences in outcome (Watzke et al., 2010). It may also be that the therapists adapted their interventions and the number of sessions to different patient needs, thus resulting in the lack of treatment differentiation (Perepletchikova, Treat, & Kazdin, 2007). Furthermore, the lack of effects of therapy type and treatment duration may be valid at termination but no longer at long-term follow-up. In the Helsinki Psychotherapy Study (Knekt et al., 2008), short-term therapies were more effective than long-term psychodynamic psychotherapy during the first year of follow-up. During the second year of follow-up, no significant differences were found between the short-term and long-term therapies, and after 3 years of follow-up, long-term psychodynamic psychotherapy was more effective. In the Munich Psychotherapy Study (Huber, Zimmermann, Henrich, & Klug, 2012), no differences in Beck Depression Inventory (BDI) mean scores were found at termination and 1-year follow-up of CBT, PDT, and psychoanalytic therapy. However, patients in psychoanalytic therapy, but not other therapy types, continued to improve on BDI up to 3-year follow-up, which was interpreted as a dose effect.

In the present study, we used three outcome measures, one focusing on severity of symptoms, the second on quality of life, and the third on self-rated health. Comparing GSI, QOLI, and SRH, the total pre/post treatment effect sizes were fairly similar, even if there were some differences between outcome measures and therapy types. This implies that the completed therapies resulted not only in symptomatic change but also in increased life satisfaction. When we apply the criterion of reliable and clinically significant change, the picture is fairly similar, but when looking at

the more generous criterion of movement into the functional distribution, more patients improved in terms of SRH and GSI than in terms of QOLI.

Research suggests that different psychotherapy types address different types of problems (Ambühl & Orlinsky, 1997; Philips, 2009). Furthermore, different kinds of change require different amounts of time (Grande et al., 2009; Huber, Henrich, Gastner, & Klug, 2012), and improvements in self-reported distress precede improvement in life satisfaction (Perry & Bond, 2009). Lambert and Ogles (2004) suggested that >50 sessions are necessary for >75% of patients to recover, whereas Kopta, Howard, Lowry, and Beutler (1994) concluded that acute distress symptoms need a smaller dosage than chronic distress and especially characterological symptoms. In the present study, increased life satisfaction seemed to be more difficult to achieve than symptomatic improvement. It may also be the case that changes in life satisfaction continue after termination and would be observed only at follow-up. In our sample, the patients treated in the CBT group received on average relatively short treatments and showed large effect sizes in terms of GSI and SRH, but small effect size in terms of QOLI. As the present study was limited to treatment effectiveness at termination, long-term follow-up would be necessary to study the effects of interaction of therapy type and treatment duration on different dimensions of therapeutic change.

At the therapist level, our analysis did not demonstrate any significant therapist effects. In fact, therapist variance was remarkably low, except on the SRH. In a naturalistic study of managed care, Wampold and Brown (2005) estimated variability in outcomes attributable to therapists to equal 5%, when the initial level of symptom severity was taken into account. In the present study,

42 of the 75 therapists (56%) treated only one patient each, which may have biased the estimate of therapist variance.

Looking at the pretreatment differences in the three therapy types, INT patients were younger and less educated. Some minor differences could be observed in the distribution of presenting problems by treatment group. CBT therapists reported fewer presenting problems per patient and fewer interpersonal problems (as in the U.K. sample; Stiles et al., 2008a), less problematic self-esteem, anxiety, aggression, sexual acting-out, self-harm, and mood changes. On the other hand, CBT therapists were more prone to diagnose their patients, and they used diagnoses in the anxiety spectrum more frequently. PDT therapists used mood disorder diagnosis more frequently, while INT therapists had most missing diagnostic data. In addition, pretreatment medication was more frequent in the CBT group as compared with PDT and INT. In our view, such systematic differences in the patients' diagnoses and in the therapists' inclination to use diagnostic categories reflect the routine psychotherapy practice, but of course set limits for comparing outcomes between therapy types. The proportion of patients in psychotherapy without psychiatric diagnoses (12% in the core sample) is probably common in routine clinical practice. In a Swedish naturalistic study of psychodynamic psychotherapy, 45% of outpatients lacked psychiatric diagnosis, either owing to the underuse of diagnostics or owing to the patients' status (Wilczek, Weinryb, Barbier, Gustavsson, & Åberg, 2000).

Overall, our study replicated the main findings of Stiles et al. (2008a). Psychotherapy in Swedish routine practice was effective for the majority of patients who complete treatment. The theoretically different psychotherapy approaches (CBT, PDT, and INT) all had good outcomes. However, there were some differences in our sample compared with the U.K. study (Stiles et al., 2008a). The most common psychotherapy type in this Swedish sample was PDT, followed by CBT and INT, whereas PDT was the least frequent therapy type and person-centered therapy (PCT), which is unusual in Sweden, was the most common therapy type in the CORE sample. Furthermore, patients in QAPS attended more psychotherapy sessions ($M = 18, 31$, and 23 for CBT, PDT, and INT, respectively) than the CORE sample ($M = 8$ for PDT and 7 for PCT and CBT). This difference may be due to different organizational frames for delivering psychotherapy to the population in Sweden and in the U.K. The QAPS sample was treated at specialized outpatient psychiatric care services, whereas the CORE sample came from primary-care services. The overall within-group effect sizes were lower than those presented by Stiles et al. (2008a) (CORE-OM $d = 1.39$). There was also a difference in clinical effectiveness in terms of reliable and clinically significant improvement measured by GSI in our sample (35%) compared with the reliable and clinically significant improvement measured by CORE-OM (58%). This difference may be due to use of different outcome measures, different pretreatment levels of problems, or greater effectiveness of the therapists in U.K. National Health Service.

Limitations

The strength of this study is the clinically representative patient sample, therapist sample, and therapy types. Nevertheless, several limitations with naturalistic "real-life" studies are applicable here.

Of the 1,294 treatment starters in the QAPS database, only 180 patients (13.9%) could be included in the study of effectiveness. As a comparison, the two CORE studies (Stiles et al., 2006, 2008a) comprised selected samples of 12.6% (1,309 of 10,351) and 16.7% (5,613 of 33,587), respectively, of the total samples. Thus, our study had a much smaller total sample but a comparable proportion of patients with outcome data. Nonstarters, 13.6% in our study, constitute a group that is rarely presented and analyzed in psychotherapy research (Barrett, Chua, Crits-Christoph, Gibbons, & Thompson, 2008; Defife, Conklin, & Smith, 2010). In the present study, 17.4% of patients dropped out of treatment. This dropout rate is comparable with the weighted dropout rate across 669 studies of 19.7%, reported in a recent meta-analysis (Swift & Greenberg, 2012). Incomplete data collection undermines most outcome studies, but is not as well investigated as dropouts. In a study of young adults in routine care settings in Sweden, 52.5% of patients had incomplete data at termination (Falkenström, 2010), as compared with 35.7% in our study. The 499 remaining patients did not differ significantly from the 535 dropouts from data collection, except in terms of gender and number of previous psychiatric contacts. On the other hand, Clark, Fairburn, and Wessely (2008) argued that patients who complete pre- and posttreatment measures are more likely to have improved than patients who fail to do so. It is likely that the attrition rates and reasons were similar in the three treatment groups, however. Because of the large attrition, incomplete data, and the possibility of selective reporting, our conclusions primarily apply to the patients included in the analyses.

As a consequence of insufficient diagnostics and nonreporting of the use of established diagnostic instruments, no definite conclusions can be made on the diagnostic differences between the groups. However, no systematic differences were found in terms of symptom severity, self-rated health, or quality of life.

Nonrandom assignment and nonequivalence of treatment groups is another possible limitation of practice-based studies. Although there were no significant pretreatment differences on the instruments included, there may have been some systematic differences not detected with the measures used in this study. Differences between the groups in numbers, diagnoses, medication, and treatment duration restrict the comparability of the three therapy types.

The present study was restricted to self-report measures, and observer-rated measures were not included. The QAPS self-report instruments may be subject to distortions. Furthermore, like in most practice-based studies, statistical power was insufficient for a meaningful examination of moderator variables and the clinical judgments of patient suitability for different treatments (Fonagy, 2010). The lack of long-term follow-up is another limitation of the present study.

Limited specification of treatments and therapist responsiveness is a common problem in practice-based studies. There has been no control regarding the treatment integrity or therapists' adherence. Thus, the three therapy types may be a matter of "trademarks" rather than substantial differences in the therapists' approaches. Furthermore, there was no control group.

A dropout rate from treatment of 17.4% is obviously a problem in routine practice, as is also the 3% of patients who deteriorated in terms of symptom severity. Additionally, we have to learn more about why patients (and therapists) do not start therapy after the initial assessment. We concur with the suggestion formulated by

Stiles, Barkham, Mellor-Clark, and Connell (2008b) that noncompletion in routine clinical practice is more often attributable to personal, institutional, social, and economic conditions than to the theoretical approach the therapist uses. Indeed, in a study of potential predictors of treatment attendance and discontinuation among the same QAPS sample, organizational factors predicted both starting and continuing in treatment (Werbart & Wang, 2012).

Conclusions

Our results suggest that psychotherapy delivered in psychiatric routine care settings is efficient for a majority of the patients who complete treatment. Treatment completers showed significant improvements, and all psychotherapy types had good effects in terms of symptom reduction and clinical recovery. Although clinical significance findings suggested some differences among treatments, these may have been confounded by pretreatment differences among patients. More sensitive multilevel analyses indicated no outcome differences between psychotherapy types. Finally, are the effects reported true treatment effects? Effect sizes were comparable with within-group effect sizes usually reported in both randomized controlled studies and naturalistic studies and exceed the upper limit of the confidence interval for change occurring in psychiatric patients participating in control groups not receiving specific treatment. This supports our conclusion that the results reported are true treatment effects.

References

- Ambühl, H., & Orlinsky, D. E. (1997). Zum Einfluss der theoretischen Orientierung auf die psychotherapeutische Praxis [How does the theoretical orientation influence the psychotherapeutic work?]. *Psychotherapeut*, 42, 290–298. doi:10.1007/s002780050078
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: American Psychiatric Association.
- American Psychological Association. (2012). Resolution on the recognition of psychotherapy effectiveness—Approved August 2012. Washington, DC: American Psychiatric Association. Retrieved from <http://www.apa.org/news/press/releases/2012/08/resolution-psychotherapy.aspx>. Retrieved August 10, 2012.
- Barrett, M. S., Chua, W. J., Crits-Christoph, P., Gibbons, M. B., & Thompson, D. (2008). Early withdrawal from mental health treatment: Implications for psychotherapy practice. *Psychotherapy: Theory, Research, Practice, Training*, 45, 247–267. doi:10.1037/0033-3204.45.2.247
- Bjørner, J. B., Søndergaard Kristensen, T., Orth-Gomér, K., Tibblin, G., Sullivan, M., & Westerholm, P. (1996). *Self-rated health: A useful concept in research, prevention and clinical medicine* (No 96:9). Stockholm, Sweden: Swedish Council for Planning and Coordination of Research (FRN).
- Black, N. (1996). Why we need observational studies to evaluate the effectiveness of health care. *British Medical Journal*, 312, 1215–1218. doi:10.1136/bmj.312.7040.1215
- Braslow, J. T., Duan, N., Starks, S. L., Polo, A., Bromley, E., & Wells, K. B. (2005). Generalizability of studies on mental health treatment and outcomes 1981 to 1996. *Psychiatric Services*, 56, 1261–1268. doi:10.1176/appi.ps.56.10.1261
- Clark, D. M., Fairburn, C. G., & Wessely, S. (2008). Psychological treatment outcomes in routine NHS services: A commentary on Stiles et al. (2007). *Psychological Medicine*, 38, 629–634. doi:10.1017/S0033291707001869
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159. doi:10.1037/0033-2909.112.1.155
- de Beurs, E., den Hollander-Gijsman, M. E., van Rood, Y. R., van der Wee, N. J. A., Giltay, E. J., van Noorden, M. S., . . . Zitman, F. G. (2011). Routine outcome monitoring in the Netherlands: Practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. *Clinical Psychology and Psychotherapy*, 18, 1–12. doi:10.1002/cpp.696
- Defife, J. A., Conklin, C. Z., & Smith, J. M. (2010). Psychotherapy appointment non-shows: Rates and reasons. *Psychotherapy: Theory, Research, Practice, Training*, 47, 413–417. doi:10.1037/a0021168
- Derogatis, L. R. (1994). *Symptom checklist-90-R: Administration, scoring and procedures manual* (3rd ed., revised). Minneapolis, MN: National Computer Systems.
- Derogatis, L. R., Rickels, K., & Rock, A. F. (1976). The SCL-90 and the MMPI: A step in the validation of a new self-report scale. *British Journal of Psychiatry*, 128, 280–289. doi:10.1192/bjp.128.3.280
- Edwards, D. W., Yarvis, R. M., Mueller, D. P., Zingale, H. C., & Wagman, W. J. (1978). Test-taking and the stability of adjustment scales: Can we assess patient deterioration? *Evaluation Quarterly*, 2, 275–291. doi:10.1177/0193841X7800200206
- Evans, C., Connell, J., Barkham, M., Margison, F., Mellor-Clark, J., McGrath, G., & Audin, K. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry*, 180, 51–60. doi:10.1192/bjp.180.1.51
- Falkenström, F. (2010). Does psychotherapy for young adults in routine practice show similar results as therapy in randomized clinical trials? *Psychotherapy Research*, 20, 181–192. doi:10.1080/10503300903170954
- Fonagy, P. (2010). Psychotherapy research: Do we know what works for whom? *British Journal of Psychiatry*, 197, 83–85. doi:10.1192/bjp.bp.110.079657
- Fridell, M., Cesarec, Z., Johansson, M., & Malling Thorsen, S. (2002). *Svensk normering, standardisering och validering av symtomskalet SCL-90 [Swedish norms, standardisation, and validation of the symptom scale SCL-90]*. Västervik, Sweden: National Board of Institutional Care (SiS).
- Frisch, M. B., Clark, M. P., Rouse, S. V., Rudd, M. D., Paweleck, J. K., Greenstone, A., & Kopplin, D. A. (2005). Predictive and treatment validity of life satisfaction and the quality of life inventory. *Assessment*, 12, 66–78. doi:10.1177/1073191104268006
- Frisch, M. B., Cornell, J., Villanueva, M., & Retzlaff, P. J. (1992). Clinical validation of the quality of life inventory: A measure of life satisfaction for use in treatment planning and outcome assessment. *Psychological Assessment*, 4, 92–101. doi:10.1037/1040-3590.4.1.92
- Grande, T., Dilg, R., Jakobsen, T., Keller, W., Krawietz, B., Langer, M., . . . Rudolf, G. (2009). Structural change as a predictor of long-term follow-up outcome. *Psychotherapy Research*, 19, 344–357. doi:10.1080/10503300902914147
- Hatchett, G. T., & Park, H. L. (2003). Comparison of four operational definitions of premature termination. *Psychotherapy: Theory, Research, Practice, Training*, 40, 226–231. doi:10.1037/0033-3204.40.3.226
- Høglend, P. (1999). Psychotherapy research: New findings and implications for training and practice. *Journal of Psychotherapy Research and Practice*, 8, 257–263. PMCID: PMC3330564.
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. New York: Routledge.
- Huber, D., Henrich, G., Gastner, J., & Klug, G. (2012). Must all have prizes? The Munich Psychotherapy Study. In A. R. Levy, J. S. Ablon & H. Kächele (Eds.), *Psychodynamic psychotherapy research: Evidence-based practice and practice-based evidence* (pp. 51–69). New York: Springer. doi:10.1007/978-1-60761-792-1_4
- Huber, D., Zimmermann, J., Henrich, G., & Klug, G. (2012). Comparison of cognitive-behaviour therapy with psychoanalytic and psychodynamic therapy for depressed patients: A three-year follow-up study. *Zeitschrift*

- für Psychosomatische Medizin und Psychotherapie, 58, 299–316. doi: [PMID:22987495](#)
- Jacobson, N. S., Roberts, L. J., Berns, S. B., & McGlinchey, J. B. (1999). Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. *Journal of Consulting and Clinical Psychology*, 67, 300–307. doi: [10.1037/0022-006X.67.3.300](#)
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12–19. doi: [10.1037/0022-006X.59.1.12](#)
- Johansson, H. (2009). Do patients improve in general psychiatric outpatient care? Problem severity among patients and the effectiveness of a psychiatric outpatient unit. *Nordic Journal of Psychiatry*, 63, 171–177. doi: [10.1080/08039480802571051](#)
- Kazdin, A. E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist*, 63, 146–159. doi: [10.1037/0003-066X.63.3.146](#)
- Kendall, P. C. (1998). Empirically supported psychological therapies. *Journal of Consulting and Clinical Psychology*, 66, 3–6. doi: [10.1037/0022-006X.66.1.3](#)
- Knekt, P., Lindfors, O., Häkkinen, T., Välikoski, M., Virtala, E., Laaksonen, M. A., ... Renlund, C.; the Helsinki Psychotherapy Study Group. (2008). Randomized trial on the effectiveness of long- and short-term psychodynamic psychotherapy and solution-focused therapy on psychiatric symptoms during a 3-year follow-up. *Psychological Medicine*, 38, 689–703. doi: [10.1017/S003329170700164X](#)
- Kopta, S. M., Howard, K. I., Lowry, J. L., & Beutler, L. E. (1994). Patterns of symptom recovery in psychotherapy. *Journal of Consulting and Clinical Psychology*, 62, 1009–1016. doi: [10.1037/0022-006X.62.5.1009](#)
- Kordy, H., & Bauer, S. (2003). The Stuttgart-Heidelberg model of active feedback driven quality management: Means for the optimization of psychotherapy provision. *International Journal of Clinical and Health Psychology*, 3, 615–631.
- Kordy, H., & Gallas, C. (2007). Dokumentation und Qualitätssicherung in der Psychotherapie. In B. Strauss, X. Hohagen & F. Caspar (Hrsg.), *Lehrbuch Psychotherapie* (pp. 974–998). Göttingen: Hogrefe.
- Kordy, H., Hannöver, W., & Richard, M. (2001). Computer-assisted feedback-driven quality management for psychotherapy: The Stuttgart-Heidelberg Model. *Journal of Consulting and Clinical Psychology*, 69, 173–183. doi: [10.1037/0022-006X.69.2.173](#)
- Kraus, D. R., Seligman, D. A., & Jordan, J. R. (2005). Validation of a behavioral health treatment outcome and assessment tool designed for naturalistic settings: The Treatment Outcome Package. *Journal of Clinical Psychology*, 61, 285–314. doi: [10.1002/jclp.20084](#)
- Lambert, M. J., Garfield, S. L., & Bergin, A. E. (2004). Overview, trends and future issues. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th ed., pp. 805–821). New York: Wiley.
- Lambert, M. J., & Ogles, B. M. (2004). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th ed., pp. 139–193). New York: Wiley.
- Lorig, K., Stewart, A., Ritter, P., Gonzalez, V., Laurent, D., & Lynch, J. (1996). *Outcome measures for health education and other health care interventions*. Thousand Oaks, CA: Sage.
- Minami, T., Wampold, B. E., Serlin, R. C., Hamilton, E. G., Brown, G. S., & Kircher, J. C. (2008). Benchmarking the effectiveness of psychotherapy treatment for adult depression in a managed care environment: A preliminary study. *Journal of Consulting and Clinical Psychology*, 76, 116–124. doi: [10.1037/0022-006X.76.1.116](#)
- Morrison, K. H., Bradley, R., & Westen, D. (2003). The external validity of controlled clinical trials of psychotherapy for depression and anxiety: A naturalistic study. *Psychology and Psychotherapy: Theory, Research and Practice*, 76, 109–132. doi: [10.1348/147608303765951168](#)
- Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.
- Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Statistics in Medicine*, 17, 857–872. doi: [10.1002/\(SICI\)1097-0258\(19980430\)17:8>857::AID-SIM777<3.0.CO;2-E](#)
- Nilsson, P., & Orth-Gomér, K. (Eds.) (2000). *Self-rated health in European perspective* (No 2000:2). Stockholm, Sweden: Swedish Council for Planning and Coordination of Research (FRN).
- Norcross, J. C. (2002). *Psychotherapy relationships that work: Therapists contributions and responsiveness to patients*. Oxford: Oxford University Press.
- Norcross, J. C., Beutler, L. E., & Levant, R. F. (Eds.). (2006). *Evidence-based practice in mental health: Debate and dialogue on the fundamental questions*. Washington, DC: American Psychological Association. doi: [10.1037/11265-000](#)
- Paunović, N., & Öst, L.-G. (2004). Clinical validation of the Swedish version of the quality of life inventory in crime victims with posttraumatic stress disorder and a nonclinical sample. *Journal of Psychopathology and Behavioral Assessment*, 26, 15–21. doi: [10.1023/B:JOBA.0000007452.65270.13](#)
- Percevic, R., Gallas, C., Arikan, L., Mössner, M., & Kordy, H. (2008). Web-AKQUASI: Ergebnismonitoring und Qualitätssicherung in der Psychotherapie. Retrieved from <http://www.klinikum.uni-heidelberg.de/Qualitaetssicherung-und-Evaluation.7354.0.html?&L=1>. Retrieved August 10, 2012.
- Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology*, 75, 829–841. doi: [10.1037/0022-006X.75.6.829](#)
- Perry, J. C., & Bond, M. (2009). The sequence of recovery in long-term dynamic psychotherapy. *Journal of Nervous and Mental Disease*, 197, 930–937. doi: [10.1097/NMD.0b013e3181c29a0f](#)
- Philips, B. (2009). Comparing apples and oranges: How do patient characteristics and treatment goals vary between different forms of psychotherapy? *Psychology and Psychotherapy: Theory, Research and Practice*, 82, 323–336. doi: [10.1348/147608309X431491](#)
- Philips, B., Wennberg, P., Werbart, A., & Schubert, J. (2006). Young adults in psychoanalytic psychotherapy: Patient characteristics and therapy outcome. *Psychology and Psychotherapy: Theory Research and Practice*, 79, 89–106. doi: [10.1348/147608305X52649](#)
- Sanson-Fisher, R. W., Bonevski, B., Green, L. W., & D'Este, C. (2007). Limitations of the randomized controlled trial in evaluating population-based health interventions. *American Journal of Preventing Medicine*, 33, 155–161. doi: [10.1016/j.amepre.2007.04.007](#)
- Stiles, W. B., Barkham, M., Mellor-Clark, J., & Connell, J. (2008a). Effectiveness of cognitive-behavioral, person-centred, and psychodynamic therapies in UK primary-care routine practice: Replication in a larger sample. *Psychological Medicine*, 38, 677–688. doi: [10.1017/S0033291707001511](#)
- Stiles, W. B., Barkham, M., Mellor-Clark, J., & Connell, J. (2008b). Routine psychological treatment and the Dodo verdict: A rejoinder to Clark et al. (2007) *Psychological Medicine*, 38, 905–910. doi: [10.1017/S0033291708002717](#)
- Stiles, W. B., Barkham, M., Twigg, E., Mellor-Clark, J., & Cooper, M. (2006). Effectiveness of cognitive-behavioral, person-centred and psychodynamic therapies as practised in UK National Health Service settings. *Psychological Medicine*, 36, 555–566. doi: [10.1017/S0033291706007136](#)
- Swift, J. K., & Greenberg, R. P. (2012). Premature discontinuation in adult psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology*, 80, 547–559. doi: [10.1037/a0028226](#)

- Tingey, R., Lambert, M. J., Burlingame, G. M., & Hansen, N. B. (1996). Assessing clinical significance: Proposed extensions to method. *Psychotherapy Research*, 6, 109–123. doi:10.1080/10503309612331331638
- Wachtel, P. L. (2010). Beyond “ESTs”: Problematic assumptions in the pursuit of evidence-based practice. *Psychoanalytic Psychology*, 27, 251–272. doi:10.1037/a0020532
- Wampold, B. E., & Brown, G. S. (2005). Estimating variability in outcomes attributable to therapists: A naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology*, 73, 914–923. doi:10.1037/0022-006X.73.5.914
- Watzke, B., Rüddel, H., Jürgensen, R., Koch, U., Kriston, L., Grothgar, B., & Schultz, H. (2010). Effectiveness of systematic treatment selection for psychodynamic and cognitive-behavioral therapy: Randomised controlled trial in routine mental healthcare. *British Journal of Psychiatry*, 197, 96–105. doi:10.1192/bjp.bp.109.072835
- Werbart, A., & Wang, M. (2012). Predictors of not starting and dropping out from psychotherapy in Swedish public service settings. *Nordic Psychology*, 64, 128–146. doi:10.1080/19012276.2012.726817
- Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2004). The empirical status of empirically supported psychotherapies: Assumptions, findings, and reporting in controlled clinical trials. *Psychological Bulletin*, 130, 631–663. doi:10.1037/0033-2909.130.4.631
- Wilczek, A., Weinryb, R. M., Barbier, J. P., Gustavsson, J. P., & Åsberg, M. (2000). The core conflictual relationship theme (CCRT) and psychopathology in patients selected for dynamic psychotherapy. *Psychotherapy Research*, 10, 100–113. doi:10.1093/ptr/10.1.100

Received October 30, 2012

Accepted November 5, 2012 ■