Another effectiveness study

The discussion about empirically supported (validated) treatments

Peter Wilhelm

25.4.2018
Overview of Today’s Lecture

• A more convincing effectiveness study
  – Evaluating the effectiveness of psychotherapy for patients with panic disorder and agoraphobia (Hahlweg et al. 2001)

• The attempt to identify empirically supported treatments and its discussion
Short- and long-term effectiveness of an empirically supported treatment for Agoraphobia

Hahlweg et al. (2001)

Research Questions

- Examining transportability of HDE by comparing effectiveness and clinical significance at:
  - Post treatment
  - 1 year follow up
- Investigating generalizability of results by examining differences between:
  - Clinics
  - Therapists with different levels of experience
- Exploring potential predictors of outcome
- Comparing Dropouts with Completers
Patients

- Patients had primary diagnosis of panic disorder with agoraphobia DSM-III-R
  - Exclusion Criterion: alcohol and substance abuse, psychosis, medical problems

- 692 patients applied for treatment
  - in clinics of the Christoph-Dornier Stiftung (Foundation) (CDS) (Marburg, Braunschweig und Dresden)

- 416 patients completed treatment and had at least Pre Post data
  - 67 % women
  - Ø 35.6 years
  - Duration of disease Ø 8.4 years

- 95% patients received already treatment for panic disorder with agoraphobia
  - 19% Antidepressent medication
  - 37% Anxiolytica
  - Only 35% without medication
Therapists

- 52 diploma psychologists
  - Training in CBT
    - Broad range of experience (1 to 60 patients treated)
  - Supervision
Treatment

- High-density cognitive-behavioral in vivo exposure (HDE)
  - Confrontation with fear inducing stimuli for several hours a day
  - Patients often stay for 2 or more weeks in the clinic

- Three phases:
  - 1. Psychological and medical examination
    - 4 to 6 sessions
  - 2. Diagnostic Feedback und cognitive preparation
    - Patient is given 1 or 2 weeks to decide whether to participate
  - 3. HDE
    - Individual adjustment of treatment to patient’s needs
      - In the beginning: Confrontation with most difficult situation with support of therapist
      - Later self initiated confrontation
      - Duration of treatment 4 to 10 days
      - Up to 12 h confrontation per day

- Ø 36.2 sessions (a 50-min) (SD = 17)
Design

- Pre
- Post
  - 6 weeks after termination
- Follow up
  - 1 year
Measures

- Diagnostic Interview (DIPS)

- Beck Anxiety Inventory (BAI) $\rightarrow$ Anxiety symptoms

- Agoraphobic Cognition Questionnaire (ACQ) $\rightarrow$ Anxiety and agoraphobic cognitions (Heart attack)

- Body Sensation Questionnaire (BSQ) $\rightarrow$ physical symptoms

- Mobility Inventory (MI) $\rightarrow$ avoidance behavior

- Beck Depression Inventory (BDI) $\rightarrow$ Depressive Symptoms

- Symptom Checklist-90-Revised (SCL-90-R) $\rightarrow$ Mental health symptoms and distresses
1. Comparison between completers and dropouts

- 692 Patients applied for treatment
  - 90 Dropouts (13%) after cognitive preparation
  - 59 Dropouts (8.5%) during exposition
  - Dropouts due to other reasons

- 416 Completers (1 year follow up)

- 3 sig. differences between completers and dropouts (ANOVA and chi²-Test):
  - Dropouts had:
    - longer duration of disease
    - higher pre-treatment-BDI-Scores
    - lower education
2. Effectiveness of treatment: Pre – Post

- **Big pre-follow-up effect sizes (IGES)**
  - $\bar{d} = 1.23$

- **Reliable Change (RC)**
  - $(\text{Pre – Post score / SD diff.})$
  - $\bar{\text{Improvement}} = 81\%$

- **Clinical Significance (CS)**
  - Dysf. → Funct. (Norms)
  - $\bar{\text{Improved}} = 55\%$

<table>
<thead>
<tr>
<th>Variable</th>
<th>IGES</th>
<th>RC</th>
<th>CS</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAI</td>
<td>1.16</td>
<td>8</td>
<td>46</td>
</tr>
<tr>
<td>ACQ</td>
<td>1.09</td>
<td>2</td>
<td>38</td>
</tr>
<tr>
<td>BSQ</td>
<td>1.22</td>
<td>4</td>
<td>53</td>
</tr>
<tr>
<td>MIA</td>
<td>1.82</td>
<td>3</td>
<td>68</td>
</tr>
<tr>
<td>MIB</td>
<td>1.43</td>
<td>3</td>
<td>71</td>
</tr>
<tr>
<td>SCL–GSI</td>
<td>0.99</td>
<td>6</td>
<td>53</td>
</tr>
<tr>
<td>BDI</td>
<td>0.93</td>
<td>6</td>
<td>60</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>1.23</strong></td>
<td><strong>5</strong></td>
<td><strong>55</strong></td>
</tr>
</tbody>
</table>

IGES = intragroup effect sizes (dz), RC = reliable change (%), CS = clinical significance (%), FU = Follow-Up (1 Jahr), D = Detoriation (%), I = Improvement (%)
### 2. Effectiveness of treatment: Pre – Follow up

- **Big pre-follow-up effect sizes** (IGES)
  - $\bar{d} = 1.24$

- **Reliable Change** (RC)
  - $(\text{Pre} - \text{Post score} / \text{SD diff.})$
  - $\bar{\text{Improvement}} = 79\%$

- **Clinical Significance** (CS)
  - Dysf. $\rightarrow$ Funct. (Norms)
  - $\bar{59\%}$ Improved

<table>
<thead>
<tr>
<th>Variable</th>
<th>FU</th>
<th>D</th>
<th>I</th>
<th>FU</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAI</td>
<td>1.16</td>
<td>10</td>
<td>72</td>
<td>49</td>
</tr>
<tr>
<td>ACQ</td>
<td>1.18</td>
<td>4</td>
<td>72</td>
<td>44</td>
</tr>
<tr>
<td>BSQ</td>
<td>1.33</td>
<td>8</td>
<td>77</td>
<td>70</td>
</tr>
<tr>
<td>MIA</td>
<td>1.70</td>
<td>4</td>
<td>92</td>
<td>70</td>
</tr>
<tr>
<td>MIB</td>
<td>1.30</td>
<td>5</td>
<td>83</td>
<td>64</td>
</tr>
<tr>
<td>SCL–GSI</td>
<td>1.06</td>
<td>6</td>
<td>86</td>
<td>55</td>
</tr>
<tr>
<td>BDI</td>
<td>0.92</td>
<td>7</td>
<td>70</td>
<td>61</td>
</tr>
<tr>
<td>Average</td>
<td>1.24</td>
<td>6</td>
<td>79</td>
<td>59</td>
</tr>
</tbody>
</table>

IGES = intragroup effect sizes (dz), RC = reliable change (%), CS = clinical significance (%), FU = Follow-Up (1 Jahr), D = Detoriation (%), I = Improvement (%)
3. Patients’ Satisfaction at follow up

- Patients and Therapists evaluated improvement on a 7 point scale

- Improvement (much better or very much better)
  - Patients 78%
  - Therapists 77%

- no change
  - Patients 4.4%
  - Therapists 5.6%

- detioration
  - Patients 4%
  - Therapists 1.8%
4. Generalizability of results

- No difference between clinics
  - Equal implementation

- No difference between less and more experienced therapists
  - Novice therapists can successfully treat patients
5. Predictors of Post outcome

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>$r$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDI</td>
<td>.30***</td>
<td>.30***</td>
</tr>
<tr>
<td>Age</td>
<td>- .21***</td>
<td>.20**</td>
</tr>
<tr>
<td>No. of sessions</td>
<td>.12*</td>
<td>.08</td>
</tr>
<tr>
<td>Duration of disorder</td>
<td>- .13*</td>
<td>- .02</td>
</tr>
</tbody>
</table>

Note. $R = .36; R^2 = .13$. BDI = Beck Depression Inventory.  
* $p < .05$.  ** $p < .01$.  *** $p < .000$. 
Discussion: Clinical Representativeness

- Evaluation of study according to criteria of clinically representativeness (Shadish et al., 1997)

  - Treatment
    - conducted in a non-university setting (+)
    - payed by patients health insurance (+)
    - implementation was not monitored (+)

  - Patients
    - referred through usual clinical routes not by experimenter (+)
    - were heterogeneous in personal characteristics (+)
    - But homogeneous with regard to the primary diagnosis (PDAG) (-)

  - Therapists
    - did not use a treatment manual (+)
    - free to use a variety of procedures (+)
    - not restricted to a fixed number of sessions (+)
    - only 50% experienced, professional therapists with regular caseloads (-)
Discussion

- Pre-Post and Pre-Follow up results show high effectiveness of HDE treatment
  - Limitations: No control group, but
    - spontaneous remission in PDAP patients unlikely
    - effect sizes comparable to findings in efficacy studies -> weakens alternative explanations
  - Critic: 21% drop outs: No intent to treat analysis was performed -> Completer analysis overestimates effect sizes

- Assessment: Standardized self reports
  - independent blind assessor ratings should be included; however
    - Expensive and difficult to implement in ongoing clinical setting
    - Meta analysis showed that effect-sizes of clinical ratings exceeded those from self reports

- Replicability of findings (Uhle, 2006) (N = 379 patients, Clinic of CDS in Münster)
  - 1 year follow up: $\bar{\delta}$ Cohens $d$ (Pre – 1y-FU) = 1.26
  - 7 year follow up: $\bar{\delta}$ Cohens $d$ (Pre – 7y-FU) = 1.22
  - Demonstrated that improvement was reliable and stable in the long run

Discussion

- Dropout Analysis -> HDE needs to be improved
  - to prevent patients dropping out before or during exposure
    - especially vulnerable to drop out: patients with high levels of depression, low education, and long duration of disease

- Big sample of patients and therapists
  - HDE can be successfully trained
  - Therapist with low experience but frequent supervision can successfully treat patients

- Generalizability of findings limited to Institutions
  - where anxiety patients are inpatients
  - where therapists get frequent supervision
  - which are empirically oriented, and support research
  -> likely generalizable to settings and treatment conditions in CDS clinics
Conclusions

- Well conducted effectiveness study.

- Despite limitations conclusions can be drawn by comparing results with those from efficacy studies

- HDE can be successfully transferred into clinical setting

- HDE can be successfully trained

- Results are generalizable to settings and treatment conditions in CDS clinics

- HDE needs to be improved for those patients who have high risk for drop out
  - Depression, low education, long duration of disease
• Attempts to implement evidence-based practice in the health care systems
  • Empirically Supported Treatments
Background
Evidence-Based Treatment movement


Premises (Chambless & Ollendick, 2001):

▪ Patient care can be enhanced by acquisition and use of up-to-date empirical knowledge
▪ Difficult for clinicians to keep up with newly emerging information pertinent to their practice
▪ If they do not, their knowledge and clinical performance will deteriorate over the years after their training
▪ Clinicians need summaries of evidence provided by expert reviews and instructions on how to access this information during their routine practice.
Background of the attempt to identify Empirically Supported Treatments (ESTs)

- Many RCT studies have provided evidence that PT is efficacious
- “Nethertheless, a perception existed in many corners of the health delivery system that psychological treatments for particular disorders were either ineffective or inferior to pharmacological treatment.” (APA, 2006, p. 272)

  -> Promotion of PT

- 1993 APA Division 12 (Clinical Psychology) appointed Task Force on Promotion and Dissemination of Psychological Procedures
- 1995 Report of Task Force
  - Criteria for identification of empirically supported treatments (ETSs) (initially called empirically validated treatments)
  - Initial list of treatments
    - Preliminary list of ESTs that should be extended and updated
    - distribution and promotion of ESTs for training programs in clinical psychology
Criteria for Empirically Supported Treatments: Well Established Treatments (APA, 1995; Chambless & Hollon, 1998)

I. At least two good between-group design experiments (RCTs) must demonstrate efficacy in one or more of the following ways:
   • A. Superiority to pill- or psychotherapy placebo, or to another treatment
   • B. Equivalence to already established treatment with adequate sample sizes

OR

II. A series of single-case design experiments (2 * N ≥ 3) must demonstrate efficacy with:
   • A. Use of good experimental design and
   • B. Comparison of intervention to another treatment

III. Experiments must be conducted with treatment manuals or equivalent clear description of treatment

IV. Characteristics of samples must be specified

V. Effects must be demonstrated by at least two different investigators or teams
Criteria for Empirically Supported Treatments: Probably Efficacious Treatments (APA, 1995; Chambless & Hollon; 1998)

I. Two experiments
   • must show that the treatment is superior to waiting-list control group
OR

II. One or more experiments
   • must meet well-established criteria IA or IB, III, and IV,
   • but independent replication (V) is not met
OR

III. A small series of single-case design experiments (N ≥ 3)
   • must meet well-established-treatment criteria

Experimental treatments
   • Treatment not yet tested in trials meeting task force criteria for methodology
Initial List of Well Established Treatments (APA, 1995)

- Beck's cognitive therapy for depression (Dobson, 1989)
- Behavior modification for developmentally disabled individuals (Scotti et al., 1991)
- Behavior modification for enuresis and encopresis (Kupfersmid, 1989; Wright & Walker, 1978)
- Behavior therapy for headache and for irritable bowel syndrome (Blanchard et al., 1987; Blanchard et al., 1980)
- Behavior therapy for female orgasmic dysfunction and male erectile dysfunction (LoPiccolo & Stock, 1986; Auerbach & Kilmann, 1977)
- Behavioral marital therapy (Azrin, Bersalel et al., 1980; Jacobson & Follette, 1985)
- Cognitive behavior therapy for chronic pain (Keefe et al., 1992)
- Cognitive behavior therapy for panic disorder with and without agoraphobia (Barlow et al., 1989; Clark et al.)
- Cognitive behavior therapy for generalized anxiety disorder (Butler et al., 1991; Borkovec et al., 1987; Chambless & Gillis, 1993)
Initial List of Well Established Treatments (APA, 1995)

- Exposure treatment for phobias (agoraphobia, social phobia, simple phobia) and PTSD (Mattick et al., 1990; Trull et al., 1988; Foa et al., 1991)
- Exposure and response prevention for obsessive-compulsive disorder (Marks & O'Sullivan, 1988; Steketee et al., 1982)
- Family education programs for schizophrenia (Hogarty et al., 1986; Falloon et al., 1985)
- Group cognitive behavioral therapy for social phobia (Heimberg et al., 1990; Mattick & Peters, 1988)
- Interpersonal therapy for bulimia (Fairburn et al., 1993; Wilfley et al., 1993)
- Klerman and Weissman's interpersonal therapy for depression (DiMascio et al., 1979; Elkin et al., 1989)
- Parent training programs for children with oppositional behavior (Wells & Egan, 1988; Walter & Gilmore, 1973)
- Systematic desensitization for simple phobia (Kazdin & Wilcoxin, 1976)
- Token economy programs (Liberman, 1972)
Lists of ESTs and potentially efficacious treatments

- APA (1995)
  - 25 ESTs (18) and potentially efficacious treatments (7)
- Chambless & Hollon (1998)
  - 71 ESTs and potentially efficacious treatments
- Chambless & Ollendick (2001)
  - 108 ESTs and potentially efficacious treatments for adults
  - 37 for children
Limitations of ESTs-Lists  
(Chambless & Ollendick, 2001)

- For none of the lists a complete review of the psychotherapy literature was performed
- No information about treatments which are not efficacious

- That a treatment does not appear on any lists could have several meanings:
  - treatment has was not efficacious in research trials
  - treatment has not been examined in research trials
  - treatment was not reviewed
Underlying features of ESTs (Westen et al., 2004)

- Treatments are typically designed for a single DSM disorder
  - Patients are screened to maximize homogeneity of diagnosis and minimize co-occurring conditions
    - minimize within-group variability.
- Treatments are manualized and of brief and fixed duration
  - minimize within-group variability.
- Outcome assessment focuses primarily (though not necessarily exclusively) on focus symptoms
  - aimed at maximizing the internal validity
  - draw relatively unambiguous conclusions about cause and effect
Implicit assumptions of ESTs (Westen et al., 2004)

- ESTs focus on brief treatments (< 20 sessions)
  
  - pragmatic considerations
    - cost effective
    - Research design: The longer therapy, the more variability within experimental conditions -> more difficult to draw causal conclusions -> Applies also to long term follow up assessment
  
  - Assumes that improvement can be achieved with brief treatments
    - Long term effects for many disorders questionable
    - High relapse rates
Implicit assumptions of ESTs (Westen et al., 2004)

- Most patients have one primary problem or can be treated as if they do so
  - DSM criteria are essential for ESTs but can be questioned
  - Many diseases have high comorbidity
    - Different pathological features may interact with each other
    - Assumption of ESTs: comorbidity is random or additive
      -> can be ignored or sequentially treated with different EST manuals
  - Personality factors are not take into account
    -> focus on symptoms rather than processes
Critic of ESTs (Westen et al., 2004)

- **Manualization**
  - Initially developed and restricted for research purposes:
    - Artificially maximize variation between groups
    - Minimizing variation within groups
  - Manualized ESTs
    - Cook book approach
    - Restrict responsiveness of therapist and flexibility of treatment
Fundamental criticism about attempt to identify ESTs (Chambless & Ollendick, 2001)

EST research is based on treatment manuals

- Use of manuals to train therapists will lead to decrements in the quality of psychotherapy

Reply Chambless & Ollendick (2001)

- no study that investigates whether manual-trained therapists would be less effective than therapists trained without manuals

- 2 studies in which therapists trained in ESTs were compared under two conditions:
  - therapists were free to design individually tailored treatments (all within a general cognitive-behavioral framework)
  - therapists operated under the standard EST guidelines.
  - Results: Standardized treatment proved equivalent or superior to individually tailored treatment.
Critic of ESTs (Westen et al., 2004)

- Selection of treatments for RCTs
  - Selection of treatments that are easy to study
  - Selection of treatment that receive funding
    - Treatments that are based on major paradigms
Fundamental criticism about attempt to identify ESTs (Chambless & Ollendick, 2001)

No difference in efficacy among various forms of psychotherapy

-> identification of ESTs is unnecessary

- Chambless & Ollendick (2001) found considerable evidence of specificity, even within cognitive and behavioral methods
- However, evidence of specificity is not uniform
  - E.g. for adult depression, no treatment has been clearly demonstrated to be superior to another.
- Meta-analyses indicate that question of specificity is complex
  - may depend on the target, problem, and clients’ age
Quantitative research is not an appropriate paradigm for psychotherapy research. Qualitative research or clinical observation should be the evidence source.

Reply Chambless & Ollendick (2001)

- Evidence reviews cannot address this question
- Fundamentally different views of science
  -> schism in paradigms

- No matter how large or consistent the body of evidence found for identified ESTs, findings will be dismissed as irrelevant by those with fundamentally different views.
Concerns about the strength of treatment and quality of studies (Tolin et al., 2015, Tab. 2)

- Inadequate attention to null or negative findings
  - Systematic reviews that take all research evidence into account rather than individual studies

- Inadequate attention to long-term outcomes
  - Long-term efficacy (indicator of sustainability) in addition to short-term efficacy

- Reliance on statistical, rather than clinical significance
  - Reporting clinical significance in addition to statistical significance
    - (different indicators of clinical significance: reliable change; clinically significant change: post score below diagnostic cut off or within normal range; good end state functioning: not only symptoms but also indicators of social functioning in the normal range) -> Meaning depends on context and kind of disease
  - Separating strength of effect from strength of evidence

- Potentially significant variability in study quality
  - Grading quality of studies
  - Taking quality of studies into account for the evaluation of evidence
Concerns about selecting among multiple treatment options (Tolin et al., 2015, Tab. 2)

- Within a given EST category, there is little basis for choosing one over another
  -> Quantitative information about treatment strength should be reported
  -> Cost effectiveness of treatments should be considered

- Lack of clarity about whether empirical support translates to a recommendation
  -> Specific recommendations based on clinical outcomes and quality of the available research
Concerns about the relevance of findings (Tolin et al., 2015, Tab. 2)

- Inadequate attention to functional outcomes
  - Include functional or other health-related outcomes in addition to symptom outcomes
    - E.g. quality of life

- Inadequate attention to effectiveness in non-research settings or with diverse populations
  - in routine praxis
    - Patients are more heterogeneous and complex (comorbidities)
    - Therapist might be less trained and motivated
  - Address generalization of research findings to nonresearch settings and diverse populations
Concern about unclear active treatment ingredients and the proliferation of manuals for specific diagnoses (Tolin et al., 2015, Tab. 2)

- Listing of packaged treatments rather than empirically supported principles of change
- Evaluate and encourage dismantling research to identify empirically supported principles of change (e.g. Jacobson et al. 1996 study on efficacy of behavior activation in depression)

- Emphasis on specific psychiatric diagnoses (DSM-5)
- De-emphasize diagnoses and emphasize syndromes/mechanisms of psychopathology
  - Overcoming categorical diagnoses derived from DSM-5
  - focus on distinct, empirically derived syndromes of psychopathology (mild to severe) -> Understanding the core dimensions of psychopathology -> develop treatments that target those dimensions
New criteria to identify Empirically Supported Treatments and to develop treatment guidelines (Tolin et al., 2015)
Procedure for evaluating ESTs and the development of guidelines (Tolin et al., 2015, Tab. 3, p. 11)

Table 3. Summary of the proposed Division 12 procedure for evaluating empirically supported treatments

<table>
<thead>
<tr>
<th>Step</th>
<th>Process</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>Systematic review</td>
<td>- Treatment is nominated</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Existing systematic review is evaluated according to:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- PICOTS (population, intervention, comparison, outcomes, timeline, setting)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Risk of bias (low, unclear, high)</td>
</tr>
<tr>
<td>Step 2</td>
<td>Committee-based evidence review</td>
<td>- GRADE (Grading of Recommendations Assessment, Development, and Evaluation) recommendation by committee: very strong, strong, weak</td>
</tr>
</tbody>
</table>
Factors that might decrease quality of evidence according to Grading of Recommendations Assessment Development and Evaluation (GRADE) (Guyatt et al., 2008, p. 996)

- Study limitations
- Inconsistency of results
- Indirectness of evidence
  - E.g. Study 1 compares treatments A vs placebo, study 2 compares B vs placebo.
- Imprecision
  - E.g. Studies with small samples
- Publication bias
- Factors that might increase quality of evidence
- Large magnitude of effect
- Plausible confounding, which would reduce a demonstrated effect
- Dose-response gradient
Evaluation of risk of bias of single studies according to Cochrane criteria (example from Bighelli et al. 2016)

<table>
<thead>
<tr>
<th></th>
<th>Random sequence generation (selection bias)</th>
<th>Allocation concealment (selection bias)</th>
<th>Blinding of participants and personnel (performance bias)</th>
<th>Blinding of outcome assessment (detection bias)</th>
<th>Incomplete outcome data (attrition bias)</th>
<th>Selective reporting (reporting bias)</th>
<th>Other bias</th>
</tr>
</thead>
</table>
Evaluation of risk of bias of single studies according to Cochrane criteria (example from Bighelli et al. 2016)

Figure 3. 'Risk of bias' graph: review authors’ judgements about each risk of bias item presented as percentages across all included studies.
Evaluating evidence according to Grading of Recommendations Assessment Development and Evaluation (GRADE) recommendations (Tolin et al., 2015, adapted from Guyatt et al., 2008)

Very strong recommendation

1. There is a wide range of studies included in the analyses with no major limitations.
2. There is little variation between studies.
3. The summary estimate has a narrow confidence interval.

All of the following:

- There is high-quality evidence that the treatment produces a clinically meaningful effect on symptoms of the disorder being treated.
- There is high-quality evidence that the treatment produces a clinically meaningful effect on functional outcomes.
- There is high-quality evidence that the treatment produces a clinically meaningful effect on symptoms and/or functional outcomes at least 3 months after treatment discontinuation.
- At least one well-conducted study has demonstrated effectiveness in nonresearch settings.
Evaluating evidence according to GRADE recommendations (Guyatt et al., 2008)

Strong recommendation

At least one of the following:

1. There are only a few studies, and some have limitations but not major flaws.
2. There is some variation between studies, or the confidence interval of the summary estimate is wide.

- There is moderate- to high-quality evidence that the treatment produces a clinically meaningful effect on symptoms of the disorder being treated.
- There is moderate- to high-quality evidence that the treatment produces a clinically meaningful effect on functional outcomes.
### Evaluating evidence according to GRADE recommendations (Guyatt et al., 2008)

<table>
<thead>
<tr>
<th>Weak recommendation</th>
<th>Any of the following:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The studies have major flaws.</td>
<td>● There is only low- or very low-quality evidence that the treatment produces a clinically meaningful effect on symptoms of the disorder being treated.</td>
</tr>
<tr>
<td>2. There is important variation between studies.</td>
<td>There is only low- or very low-quality evidence that the treatment produces a clinically meaningful effect on symptoms of the disorder being treated as well as on functional outcomes.</td>
</tr>
<tr>
<td>3. The confidence interval of the summary estimate is very wide.</td>
<td>● There is moderate- to high-quality evidence that the effect of the treatment, although statistically significant, may not be of a magnitude that is clinically meaningful.</td>
</tr>
</tbody>
</table>
Organizations that promote ESTs and adjusted evaluation standards

- National Institute of Clinical Excellence (NICE), U.K. (NICE; Baker & Kleijnen, 2000)
  - ensures that clinicians practice specific and accepted empirically based interventions for different psychological conditions

- American Psychiatric Association (APA)

- Leitlinien der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. (AWMF)

- **Ia** Evidence from a meta-analysis across at least 3 RCTs
- **Ib** Evidence from one RCT or a meta-analysis across less than 3 studies

- **IIa** Evidence from two controlled studies (Non – RCTs)
- **IIb** Evidence from one controlled study, quasi-experimentel study

- **III** Evidence from non experimental observational studies, case studies

- **IV** Evidence from case reports or expert recommendations
Further Approaches to Evidence Based Treatments

- Task Force on empirically based principles of change (Costangue & Beutler, 2006)

- American Psychiatric Association: Clinical Practice Guidelines
  https://www.psychiatry.org/psychiatrists/practice/clinical-practice-guidelines

- National Institute for Health Care and Excellence, U.K. (NICE) Guidelines
  https://www.nice.org.uk/guidance/published?type=guidelines

- Cochrane Reviews
  http://www.cochranelibrary.com/

- Leitlinien der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e.V. (AWMF)
  http://www.awmf.org/leitlinien/aktuelle-leitlinien.html