Evaluation in Psychotherapy Research

Prof Anthony W Bateman
Visiting Professor in Psychotherapy
Copenhagen University
The Current Story of the Psychotherapy Evidence Base
Psychotherapy works

- Therapy vs no Therapy effect size: $d = .80$
  - Over 80% of treated patients better than those with no treatment
- Number Needed to Treat = 3
  - Aspirin as a prophylaxis for heart attacks (NNT = 129)
- Superior to almost all interventions in cardiology, geriatric medicine, asthma, flu vaccine, cataract surgery
- Psychotherapy mostly as effective as medication
  - Considerable evidence for combined treatments
Problems with Effect Size

- A treatment could obtain a moderate effect size
  - by producing a very large effect for a small subset of patients
  - Or a moderate but incomplete reduction in symptoms for many

- Other useful metrics are
  - percent recovered
  - percent improved

- Six disorders: major depressive disorder, panic disorder, generalized anxiety disorder, obsessive-compulsive disorder, bulimia nervosa, and PTSD
  - Across disorders, treatment-versus-control effect sizes tend to be moderate to large
  - But on average, only roughly half of patients who complete treatments in such trials improve.
% improved (not effect sizes)  
What can be said with certainty?

- Overall, good evidence for the efficacy of psychological therapies

- In relation to major mental health conditions:
  - can achieve symptom reduction & in some cases freedom from symptoms
  - can improve social adjustment and work relationships
Some commonly-used therapies less well researched than others

More evidence for CBT than psychodynamic therapy BUT


Little evidence on 'eclectic' therapy

Good evidence for some little-used therapies e.g. Interpersonal Psychotherapy (IPT)

Absence of evidence limit scope and strength of EBP statements
The size of the database varies across diagnostic categories

Cochrane RCT base

Number of studies

<table>
<thead>
<tr>
<th>Condition</th>
<th>Number of Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phobias</td>
<td>200</td>
</tr>
<tr>
<td>Panic</td>
<td>150</td>
</tr>
<tr>
<td>OCD</td>
<td>100</td>
</tr>
<tr>
<td>Eating</td>
<td>75</td>
</tr>
<tr>
<td>PTSD</td>
<td>50</td>
</tr>
<tr>
<td>Bipolar</td>
<td>25</td>
</tr>
</tbody>
</table>
Research Design and Random assignment
Definition of Evidence Based Medicine

- “Evidence based medicine is the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients” (Sackett et al., 1996)

- Most EBM definitions are motivational, persuasive and essentialist rather than reportive, stipulative and operational

- EBM is synonymous with the experimental establishment of causality between treatment and outcome ➔ the RCT
RCT’s: Misleading Hierarchies?

- Inappropriateness – e.g. TC?
- Choice of control
  - No treatment, Wait-list, Attention, TAU/PAU
- Active treatment control
- Generalisability of the results
  - Patients - ?under-represented groups
  - Treatment – dose, timing, duration
  - Setting – quality, clinic, university
- Assessment of benefit
- Assessment of harms
RCT’s: Misleading Hierarchies?
(Rawlins, Harveian Oration, R.C.Phys. 2008)

- Experimentation vs Observation
  - Randomised controlled trials (RCT’s) inappropriately elevated above observational studies

- Hierarchy ‘illusory’
  - RCT’s have advantages and disadvantages
  - …so do observational studies

- Need for appraisal of all evidence
  - …and exercise of judgment
  - Wait-list control
Limitations of RCT’s (Rawlins 2008)

Resources

- Money, time and energy
- 153 Pharmaceutical RCT’s 2005-2006:
  - Median cost was £3,202,000
  - Interquartile range of £1,929,000 to £6,568,000
- Recent proposals suggest a reduction of 40% to 60% is possible (Eisenstein et al. 2005 & 2008)
Parker (2009, editorial, BJPych, 194, 1–3)

- Depression is a non-specific domain diagnosis
  - Population displaying heterogeneous characteristics (cf Dyspnoea as a criterion for RCT, bronchodilator trialled on participants with varied respiratory conditions)

- Trial incentives may lead to ‘up-rating’ of those with less substantive disorders to meet criteria
  - Recruitment increasingly weighted to milder, briefer self-limiting forms (Walsh et al., 2002)

- Remission status: Difficulty separating base-function and state depression
# Inclusion/Exclusion Criteria

<table>
<thead>
<tr>
<th>Exclusions</th>
<th>MBT</th>
<th>DBT</th>
<th>SFT</th>
<th>TFP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychotic/Bipolar</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Opiate dependence</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Organic</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Men/Women</td>
<td>Included</td>
<td>Men mostly excluded</td>
<td>Included</td>
<td>Included</td>
</tr>
<tr>
<td>Antisocial PD</td>
<td>Included</td>
<td>Excluded</td>
<td>Excluded</td>
<td>Excluded</td>
</tr>
<tr>
<td>DID</td>
<td>Included</td>
<td>?Included</td>
<td>Excluded</td>
<td>Excluded</td>
</tr>
</tbody>
</table>
Research Design and Random assignment – control groups
Comparisons between psychodynamic psychotherapy and an inactive comparator

- K=28 (68%)
- K=12 (30%)
- K=1 (3%)
Linear relationship between study quality and effect size for psychotherapies (regardless of modality)

Quality Score by Year of Publication of RCTs of CBT & PDT

Thoma et al., Am J Psychiatry, 169(10) (2012), 22-30
Secular trends in ESs for ESTs: Effect size of CBT in 27 trials for youth depression

Comparisons between psychodynamic psychotherapy and an active comparator in adequate studies.
Recovery from depressive symptoms (BDI) and no auxiliary treatment

Adequate work ability (Work Ability Index, WAI36) and no auxiliary treatment.
Manual-based treatment and what sort of therapy do therapists deliver in routine settings?
Manuals in Psychotherapy

Positive
- Enhance internal validity and treatment integrity
- Reduce confounds e.g. dose, contact, format
- Facilitate training and allow multi-centre deliver

Negative
- Reduce creativity and reactivity and flexibility
- Lack clinical sensitivity and lose moment to moment sensitivity

Compromise
- Manual must offer *flexibility* within *fidelity*
What sort of therapy do therapists deliver in routine settings?

- we don’t know.....

- but it probably isn’t the same as:

  - the therapy specified in textbooks
  - therapy as delivered in research
    - which forms the basis for any claims for the efficacy of psychological therapy
Therapy adherence/competence in routine practice
Brosan et al. (2006)

- 24 CBT therapists
  - all had received ‘some’ training in CBT

- submitted mid-treatment tape session
  - rated on Cognitive Therapy Rating Scale (CTRS)
Therapy adherence/competence in routine practice

Quality of therapy highly variable, and unrelated to

- years of experience
- frequency of supervision
- accreditation

was related to formal post-qualification training in CBT

reasonable to assume that this variability in quality is usual
Self-evaluation of competence

Brosan et al.

- 22 CBT therapists, each submit tape of one session

- rated on CTS by:
  - therapist (self-rating)
  - independent expert rater

- on basis of expert rater, therapists coded as:
  - competent (N=12)
  - not competent (N=10)
Self-evaluation compared to expert evaluation

- correlation of 0.57 between self-ratings and expert ratings - BUT

- self-ratings are higher than expert ratings
- compared to competent therapists
  - less competent therapists show a significant tendency to overestimate their competence
Self-evaluation and competence (2)
Lafferty et al. 1989

- 30 trainee therapists, mixed therapeutic orientation (60% psychodynamic, 30% eclectic)

- identified “more” or “less” effective therapists
  - residualised change scores for 2 randomly selected clients, indicating improvement or worsening

- less effective therapists
  - showed poorer self-evaluation skills
  - compared to observer, rated their clients as
    - more involved in therapy
    - making more progress
57 clients referred for trial of CBT for OCD

- all had prior (failed) psychological intervention

40% reported previous treatment with CBT/ BT

- 50% did not recall receiving core elements which characterise CBT/ BT (e.g. exposure)

- some indication that those who had CBT/ BT had better specific improvements in OCD
Does Competence make a difference?
Competence and outcome in IPT

- 11 therapists treating 35 patients in NIMH trial

- all therapists selected for competence
  - reduces variance attributable to therapists:
Selection, training and supervision in the NIMH trial (Roth, Pilling and Turner)

**Selection of therapists**

- 2-27 years experience, average 11.4 years, prior experience of treating at least 10 depressed clients
- All candidates screened for competence using CV, interview and video of treatment sessions

**Training**

- IPT training (from Weissman) 5 days
- CBT training (from Beck) 1-2 weeks

**Supervision and monitoring**

- Monthly, plus call-back if red-line
Ratings of competence

Therapist Strategy Rating Form & Process Rating Form – evaluates:

- therapist accuracy in identifying problem areas
- strategies for bringing about change
- quality of application of IPT techniques

- includes ratings of generic therapeutic skills e.g.
  - alliance
  - maintaining session focus
Competence and outcome

- correlation of 0.56 between supervisor skill rating and patient-rated outcome

- median split of therapists to “high” and “low” in competence
  - greater patient change in more skilled group

- therapist performance contributes 23% outcome variance in patient-rated change beyond initial patient factors (e.g. level of functioning)
Competence and outcome in IPT
Frank et al. (1991)

- IPT for depression, with 3 year maintenance phase

- therapist stratified into high and low competence (median split)

median survival time to relapse

- “high” competence therapists: 2 years
- “low” competence therapists: 5 months

- but a fairly clear interaction between patient ‘difficulty’ and therapist ability to adhere to IPT protocol
Links between specific techniques and outcome

- few studies link *specific* techniques to outcome
- (DeRubies et al) - studies of CBT for severe depression

deploy ‘concrete’ CBT competences early in treatment
- i.e. those associated with pragmatic, structuring aspects of the therapy

- better outcomes when therapists:

  focus on specific (concrete) beliefs and behaviours early in treatment
“Concrete” CBT competences (concerned with ‘pragmatic aspects of therapy)

- set and followed agenda
- reviewed homework
- assigned homework
- asked patient to report cognitions verbatim
- asked for specific examples of beliefs
- labelled cognitive errors
- examined evidence concerning beliefs
- practised rational responses with patient
- assigned/ reviewed self-monitoring
- asked patient to record thoughts
Cognitive therapy – “abstract” competences

some (but weaker) evidence of association between abstract competences and outcome

• encouraged independence
• explained rationale for cognitive therapy
• explored personal meaning of thoughts
• explored underlying assumptions
• encouraged distancing of beliefs
• recognised adaptive/functional value of beliefs
• negotiated content of session with client
Adherence and treatment outcome

- how many poor outcomes relate to poor adherence?
  - not knowing enough about the therapy to apply it properly

- how many poor outcomes ”relate to an excess of adherence?
  - problems in making the method tractable, relevant and acceptable…
Adherence and client motivation in PCT

Huppert et al. 2006

Panic Disorder Severity % change

low adherence        high adherence

low motivation clients  high motivation clients
Therapists often like to travel “off-piste”

Schulte and Eifert (2002)

- studies of the application of manualised CBT for anxiety disorders
- therapists make surprisingly large numbers of changes to their initial plans
  - usually response to sense of pessimism or a lack of control over therapeutic direction
- significant negative correlation (-0.49) between outcome and the frequency of changes of method
Adherence and competence

- some evidence for the benefit of adherence and competence
  - adherence alone is not enough

- competence involves the judicious manipulation of adherence
  - flexible (but not over-flexible) in technique
  - capacity for alliance management

- metacompentence also important
  - ability to employ procedural knowledge
The Impact of Therapists and Therapeutic Alliance on Treatment Outcome
Therapists co-opted or recruited for the trial were randomly assigned to a 3 day training in MBT-OP or SCM-OP with continued supervision.

- All 11 therapists had
  - minimum of 2 years' experience of treating patients in general psychiatric services following their generic training
  - minimum of 1 year’s experience treating patients with personality disorder
  - did not differ in their years of psychiatric experience (mean [SD]: MBT-OP, 6.16 [1.6]; SCM-OP, 6.8 [2.3] years)

VS.
Reducing the Harmful Effects of Psychotherapy: The work of Lambert (2009)

- Across studies the rate of observed deterioration in psychotherapy was 10-25% with young people.
- Some therapists have rates of deterioration of around 50% and their treatment is NEVER associated with recovery.
- Introduction of outcome tracking (session by session monitoring):
  - Early warning when patient goes off trajectory.
- Therapists randomized to feedback vs no-feedback:
  - Deterioration reduced by 50%.
  - Recovery improves by 50%.
  - Average therapy is shorter.
  - Patients who show early negative response receive longer and more effective treatment.
Impact of individual therapists in routine practice
Okiishi et al. 2006 (J Clin Psychol 62:9, 1157)

- 6,499 patients seen by 71 therapists

- therapists had to see at least 15 clients
  - on average saw 92

- number of sessions: range 1-203; mean 8.7

- therapists saw equivalent range of clients in terms of disturbance & presentation

- HLM used to compare ‘trajectories’ (recovery curves) of patients using OQ45
Clients of Some Therapists Improve Faster or Slower Than Others

![Graph showing the relationship between Score on OQ 45 and Session number.]
Slope of Improvement Across Therapists unaffected by:

- therapist experience
- gender
- type of training
  - counselling psychology, clinical psychology, social work, marital/family therapist, psychiatry
- orientation
  - CBT, humanistic, psychodynamic
# Outcomes for Best and Worst Performing Therapists

<table>
<thead>
<tr>
<th></th>
<th>recovered</th>
<th>improved</th>
<th>deteriorated</th>
</tr>
</thead>
<tbody>
<tr>
<td>top 10% therapists</td>
<td>22.4%</td>
<td>21.5%</td>
<td>5.2%</td>
</tr>
<tr>
<td>bottom 10% therapists</td>
<td>10.6%</td>
<td>17.4%</td>
<td>10.5%</td>
</tr>
</tbody>
</table>
Odds of a self-harming in MBT by therapist

Therapist x Time interaction: $p<0.05$
Incidence of Harmful Effects

- estimates are that 5-10% of therapy clients deteriorate
  - across all orientations, client groups, modalities
  - in RCTs of ‘empirically supported treatments’

- rates higher than in control groups
  - e.g. NIMH reanalysis (Ogles et al. 1995)
  - 13/162 (8%) deteriorated, all in active treatments

- in Lambert’s work therapists tend to be poor at:
  - predicting who will do badly
  - recognising failing therapies
Therapist Technique


- Therapists may overcompensate for patient's poor outcome by giving more complex explanations to patients. Higher complexity does not necessarily lead to better outcomes.

- Therapists who give over-complex explanations engage in developing a poorer outcome by not matching intervention to mentalizing capacity and so inducing harm.
Do no harm... outcomes informed care

- Most therapists see themselves as better than average:
  
  Dew & Riemer (2003, 16th Annual Research Conference, University of South Florida)
  
  - 143 counselors asked to grade their job performance on scale from A+ to F
  - 66% rate themselves as A or better
  - None rated themselves as below average

- Outcomes informed care may be a critical way of linking the EBP approach and practice based evidence
Harm and psychological therapy

- we intervene with the best intentions and assume our impacts are:
  - beneficial at best
  - neutral at worst

- but if treatments are effective (they can change people) they also have power to harm
Ways of harming

inappropriate treatment choices

• critical incident stress debriefing
• grief counselling for normal bereavement
• psychodynamic therapy for schizophrenia

sub-optimal therapy

• couples therapy that benefits one partner at the cost of the other

misapplied therapy

• failure to adequately treat
• applying “technique” in the absence of a good alliance

mistakes

• therapist error
Minimising harm by maximising effectiveness

- research trials demonstrate efficacy of clearly-specified therapies in relation to a range of mental health conditions
- could be argued that this represents a demonstration of best practice, in terms of:
  - training
  - monitoring
  - supervision

helpful to try to do the right thing in the right way
Minimising harm by maximising effectiveness

- in routine settings therapists often adapt these therapies in uncertain ways

- if adherence and competence of delivery make a difference, it makes sense to enhance this
  
  - competence frameworks can help to bridge research and practice
Thank you for listening

For further info
anthony.bateman@ucl.ac.uk