Using Large Practice-Based Data Sets: The Dodo Verdict and Responsive Regulation

William B. Stiles
Department of Psychology, Miami University
Oxford, Ohio, USA

Presented at the
SPR-Europe Research Methods Workshop
Bern, Switzerland, September, 2007

Overview

1. Practice-based evidence
2. The CORE measures and two samples
3. Issue 1: Which theoretical approach?
4. Issue 2: How many sessions?

Randomized trials versus practice-based research

- Internal versus external validity
- Our study addressed comparative *effectiveness* of treatments.
  - as routinely delivered,
  - using practitioners' versions of treatments,
  - with typical clients,
  - data collected primarily for clinical and administrative use, rather than for research.
- Risks (threats to internal validity) such as
  - selection biases associated with lack of randomization.
  - no assurance that treatments were delivered in a standard way.
- Balanced by greater realism, or external validity.
  - This is what is actually happening.

Some characteristics of Practice-Based Evidence

- Continuing data collection rather than discrete studies
  - Incremental improvements in design and implementation
- Research often starts with the data rather than with hypotheses.

The CORE Assessment Program

- Collaboration with Michael Barkham, John Mellor-Clark, others.
- Data gathered for clinical and administrative reasons.
  - CORE IMS provides training and feedback
  - Currently used in about 300 services
- Three forms
  - CORE Assessment
  - CORE Outcome Measure
  - CORE End of Therapy
- Copyleft (free if you don't change it and acknowledge the source)

Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM)

- 34 items in 4 domains:
  - subjective well-being;
  - symptoms (anxiety, depression, physical problems, trauma)
  - functioning (general functioning, close relationships, social relationships)
  - risk (risk to self, risk to others)
- half low intensity (e.g. 'I feel anxious/nervous')
- half high intensity (e.g. 'I feel panic/terror').

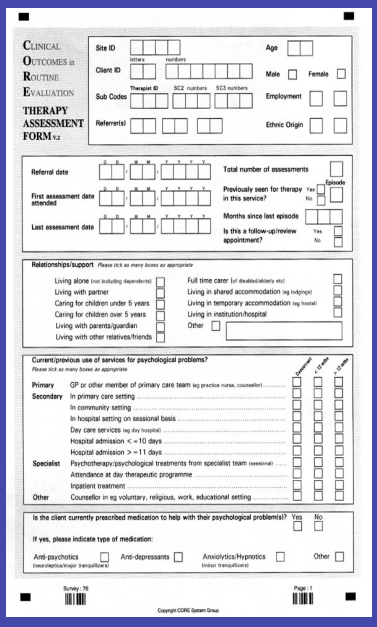
- Items scored on 0-4 scale
  - *Not at all*, *Only occasionally*, *Sometimes*, *Often*, *All or most of the time*

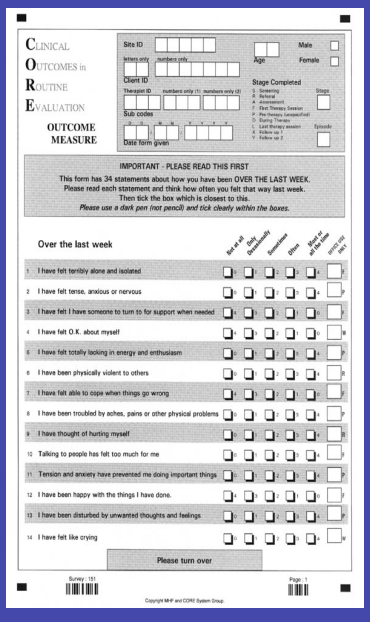
- CORE Clinical Scores= **mean** of all completed items **times 10**.
  - scores can range from 0 to 40.
  - clinically meaningful differences represented by whole numbers.
- Internal consistency alpha = .94 ; one month test-retest = .88

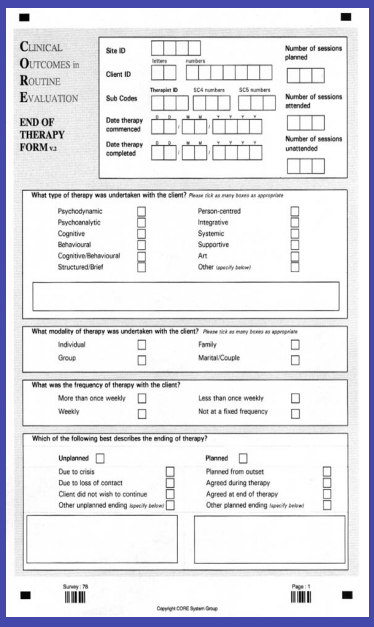
CORE Assessment and EOT Forms

- **Assessment Form** (completed by therapist after intake):
- referral information, client demographics,
- nature, severity, and duration of presenting problems in 14 categories:
  - depression, anxiety, psychosis, personality problems, cognitive/learning difficulties, eating disorder, physical problems, addictions, trauma/abuse, bereavement, self-esteem, interpersonal problems, living/welfare and work/academic.

- **End of Therapy Form** (completed by therapist)
- which approaches were used (11 categories + Other):
  - psychodynamic, psychoanalytic, cognitive, behavioural, cognitive/behavioural, structured/brief, person-centred, integrative, systemic, supportive, art, and other.
- number of sessions attended, whether end was planned, other information ...





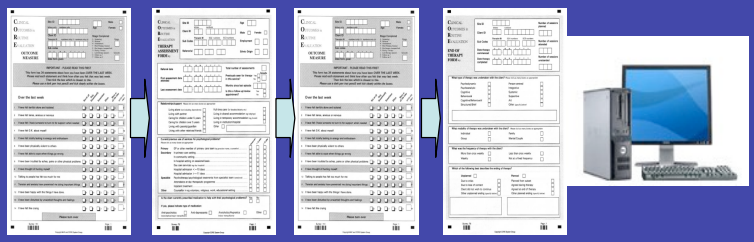


The CORE Practice Research Network

CORE System Trust provide free CORE System tools to support the development and growth of practice-based evidence

CORE National Research Database helps develop practice-based evidence, benchmarks and service quality management guidelines

The CORE System Methodology



CORE System Tools

CORE-PC

CORE IMS provide CORE implementation training, CORE-PC, and data management support

CORE Network meets to explore and share good practice

Two CORE Data Sets

Data Set 1

- Collected 1999 – 2001
- N = 12,571 clients
  - 69.8% female
  - Mean age = 36.9
- N = 421 therapists
- 58 NHS services
  - primary care, secondary, tertiary, other

Data Set 2 (CORE NRD)

- Collected 2002 – 2005
- N = 33,587 clients
  - 69.4% female
  - Mean age = 38.5
- N = 637 therapists
- 34 NHS services
  - all primary care

Issue 1: Comparison of Three Major Approaches to Psychotherapy

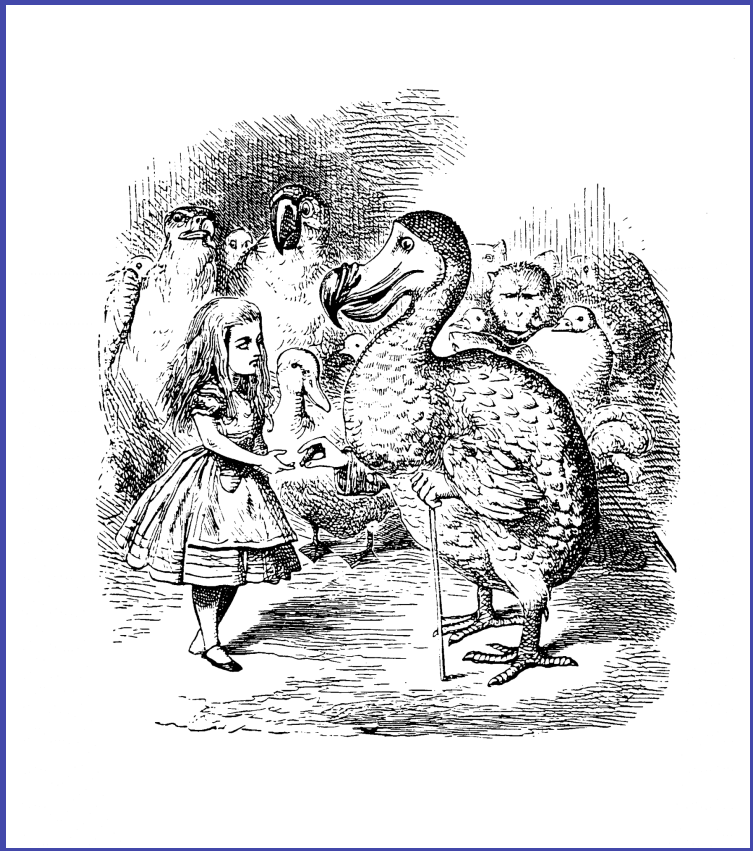
- Cognitive-behavioral therapy (CBT)
- Person-centered therapy (PCT)
  Psychodynamic/psychoanalytic therapy (PDT)

- CBT, PCT, and PDT are distinct in terms of
  - usual repertoires of interventions.
  - assumptions about psychopathology.
- Each encompasses a range of techniques
  - a family of treatments, not a single protocol.
- Each is considered by its practitioners as widely applicable across disorders and populations.

- Reporting two samples—one now published, one new.

Dodo Verdict (Carroll, 1865)

- "*Everybody* has won, and *all* must have prizes"
  - First applied to psychotherapy 70 years ago by Rosenzweig (1936)
  - Repeatedly, comparisons suggest *bone fide* therapies tend to be similarly effective (though this is debated).
- Equivalence Paradox
  - Equivalent outcomes despite different psychotherapeutic theories and techniques
  - Researchers don't believe it and keep trying.



Counterpoint: Dominance of CBT

- Much evidence for efficacy and effectiveness of CBT.
- Fewer studies of PCT and PDT, though available evidence shows they too are effective.
- The overwhelming quantity of published research gives CBT greater credibility.
  - Majority of approaches on APA Div 12 list of empirically supported treatments are in CBT family.
  - Layard initiative in UK emphasizes CBT

Study Design

- Samples from routine practice
- At start and end of treatment, clients completed self-report symptom inventories (CORE-OM).
- At end of treatment, therapists indicated which treatment approach(es) they had used.
- Select clients who received CBT, PCT, or PDT

Procedure

- Pre-treatment CORE-OM (all clients attending for psychological assessment or therapy)
  - screening or assessment (77.1%)
  - immediately before the first session (22.9%).
- Clients allocated to treatments and therapists following site's normal procedures
- Therapist Assessment Form after intake.
- Post-treatment CORE-OM given at last session
  - timing depended on site's administrative procedures
- End of Therapy Form at discharge or when client stopped coming.
- Completed measures sent to the University of Leeds for processing with no client identifiers.

Targeted Approaches

- CBT = *cognitive, behavioral,* and/or *cognitive/behavioral*
- PCT = *person-centered*
- PDT = *psychodynamic* and/or *psychoanalytic.*

- Pure treatments:
  one and only one targeted approach
  CBT, PCT, or PDT.
- Diluted/enhanced treatments:
  - one targeted approach plus one additional treatment
    - (i.e., one of: *structured/brief, integrative, systemic, supportive, art,* or *other*),
  - abbreviated CBT+1, PCT+1, and PDT+1

Selection of Clients for CBT-PCT-PDT Comparison

	Sample 1	Sample 2
Therapist returned initial CORE Assessment Form	12,571	33,587
Primary/secondary/tertiary (S1) or primary care only (S2)	10,351	33,587
Pre-treatment CORE-OM but no post-treatment CORE-OM – includes those who did not attend any sessions and those still in treatment	(5,444)	(14,945)
Pre- & post-treatment CORE-OM *AND* End of Therapy form	3,051	12,162
Met specifications for one of the six groups	1,309	5,613

Sample Characteristics

Sample 1 (2001)

- N = 1,309 clients
  - 70.7% female
  - Mean age = 40.3
- N = 251+ therapists
- 58 NHS services
  - primary and secondary care

Sample 2 (2005)

- N = 5,613 clients
  - 70.7% female
  - Mean age = 40.7
- N = 399 therapists
- 32 NHS services
  - all primary care

Overall Outcomes of Treatment

- Very substantial gains on CORE-OM

	Sample 1	Sample 2
Pre-treatment	17.41 (6.52)	17.60 (6.33)
Post-treatment	8.50 (6.27)	8.77 (6.43)
Difference	8.91 (6.81)	8.83 (6.64)
Effect size	1.36	1.39

  - (ES = diff/pre-treat sd) Comparable to efficacy trials

- Pre-therapy means not signif. different across groups
  - Sample 1: $F(5, 1303) = 0.66$, $p = 0.654$, partial $\eta^2 = 0.003$
  - Sample 2: $F(5, 5607) = 1.90$, $p = 0.091$, partial $\eta^2 = 0.002$,
  - **(partial eta squared = effect variance divided by effect plus error variance)**

Dodo Result

- Repeated-measures (pre- vs post-) ANOVA with two fixed factors:
  - treatment approach (CBT vs PCT vs PDT)
  - and degree of purity (pure vs "+1")
- Results
- 1. A very large within-clients main effect of treatment
  - Sample 1: F(1, 1303) =1905.70, p < .001, partial η^2 = .594.
  - Sample 2: F(1, 5607) = 6805.63, p <.001, partial η^2 = .548.
  - Improvement across treatment accounted for a large proportion of the variation in CORE-OM scores.
- 2. Differential treatment effect (treatment x pre-post interaction) was significant but very small in sample 1, nonsignficant in sample 2.
  - Sample 1: F(2, 1303) = 3.94, p = .020, partial η^2 = .006
  - Sample 2: *F*(2, 5607) = 0.81, *p* = .446, partial η^2 < .001
- Mean improvement accounted for 100+ times as much of the variance in CORE-OM scores as did differential effects.

Results (cont'd)

- 3. Treatment purity (pure vs "+1") was marginally significant but very small (purity by pre-post interaction)
  - Sample 1: F(1, 1303) = 4.02, p = .045, partial η^2 = .003
  - Sample 2: F(1, 5607) = 3.23, p = .073, partial η^2 = .001
  - The diluted/enhanced ("+1") groups averaged slightly better outcomes
- 4. Three-way interaction was not significant
  - Sample 1: F(2, 1303) = 1.40, p = .248, partial η^2 < .001
  - Sample 2: F(2, 5607) = 0.58, p = .561, partial η^2 < .001
  - would have indicated that purity was differentially important for the therapies,

Sample 1 CORE-OM clinical scores: Means, pre-post differences, effect sizes

Treat-ment	n	Pre-therapy		Post-therapy		Pre-Post Difference		
		Mean	sd	Mean	sd	Mean	sd	Effect size
CBT	298	16.9	7.0	8.1	6.4	8.9	7.1	1.27
PCT	332	17.6	6.6	8.9	6.1	8.7	6.6	1.32
PDT	122	17.6	6.3	9.9	6.8	7.7	6.4	1.23
CBT+1	181	17.9	6.3	7.9	5.6	10.0	6.5	1.59
PCT+1	249	17.3	6.4	7.8	6.3	9.5	6.9	1.48
PDT+1	127	17.3	5.9	9.1	6.3	8.2	6.9	1.38

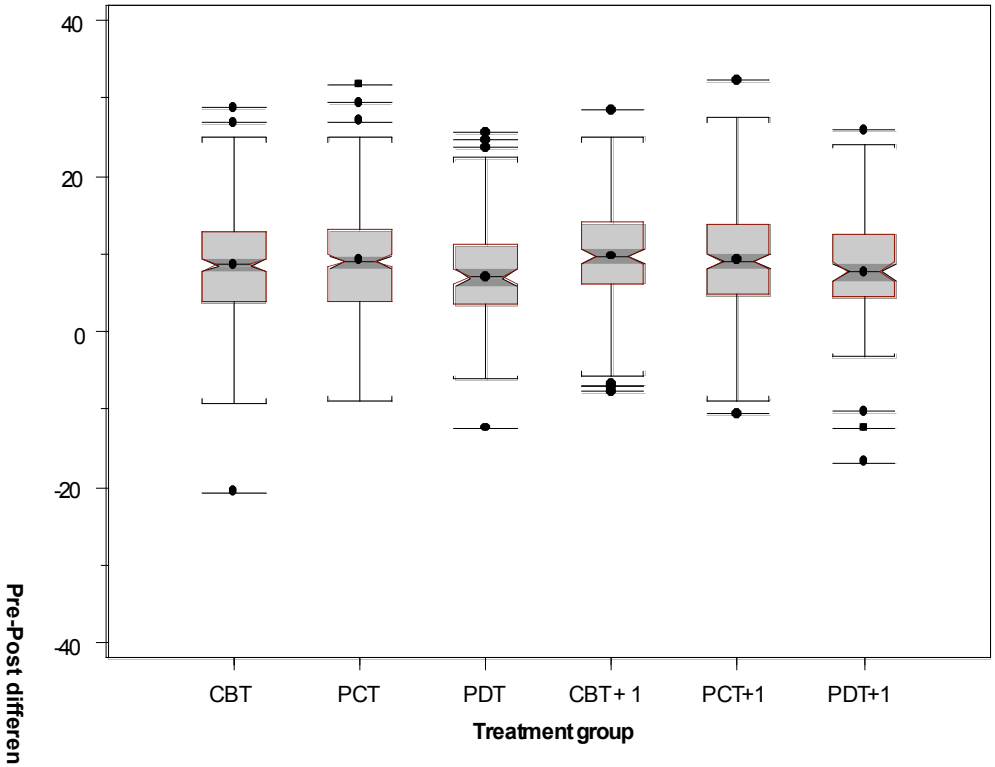
Sample 2 CORE-OM clinical scores: Means, pre-post differences, effect sizes

Treat-ment	n	Pre-therapy		Post-therapy		Pre-Post Difference		
		Mean	SD	Mean	SD	Mean	SD	Effect size
CBT	1,045	17.3	6.7	8.6	6.5	8.7	6.8	1.38
PCT	1,709	17.7	6.4	8.9	6.9	8.8	7.0	1.39
PDT	261	17.7	6.7	9.5	6.9	8.2	7.1	1.29
CBT+1	1,035	17.3	6.0	8.4	5.7	8.9	6.2	1.40
PCT+1	1,033	17.9	6.4	8.9	6.4	9.0	6.4	1.43
PDT+1	530	17.7	5.6	8.8	6.2	9.0	6.2	1.42

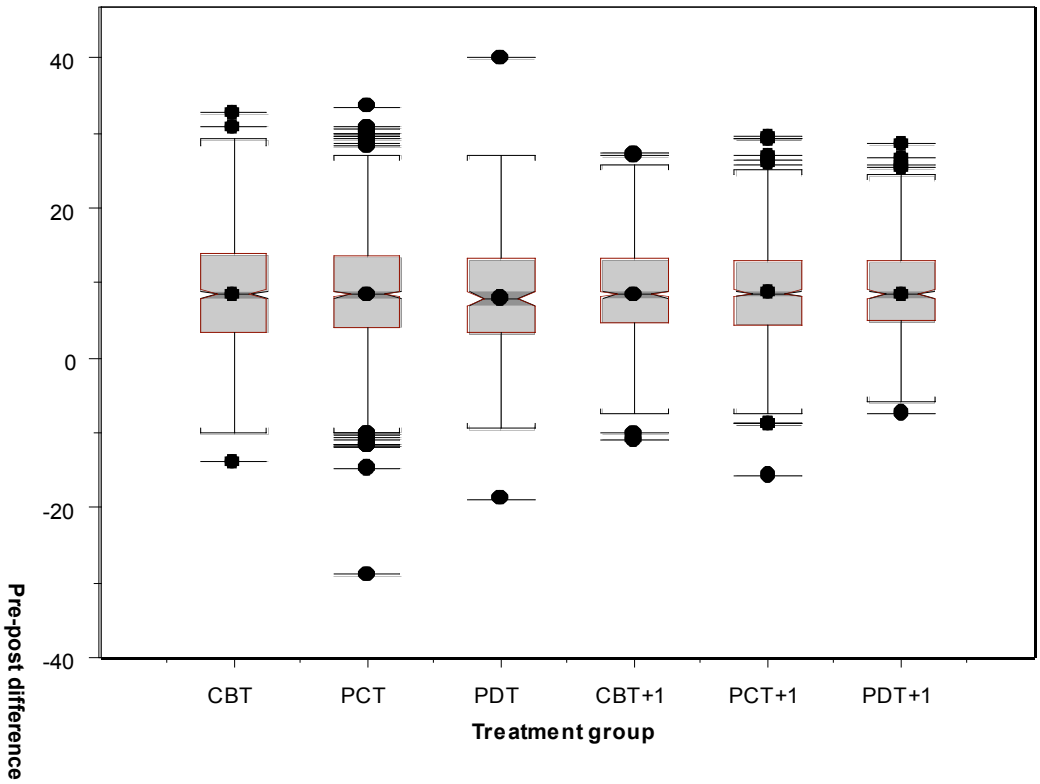
Notched box plots

- The notch shows the 95% confidence interval around the median.
- The boxes show the middle 50% of the distribution.
- The whiskers show the range,
- except that observations falling 1.5 times the interquartile range or more away from the top or bottom of the box are considered outliers and are shown separately.

Study 1 (2001 sample, N = 1,309)



Study 2 (2005 sample, N = 5,613)

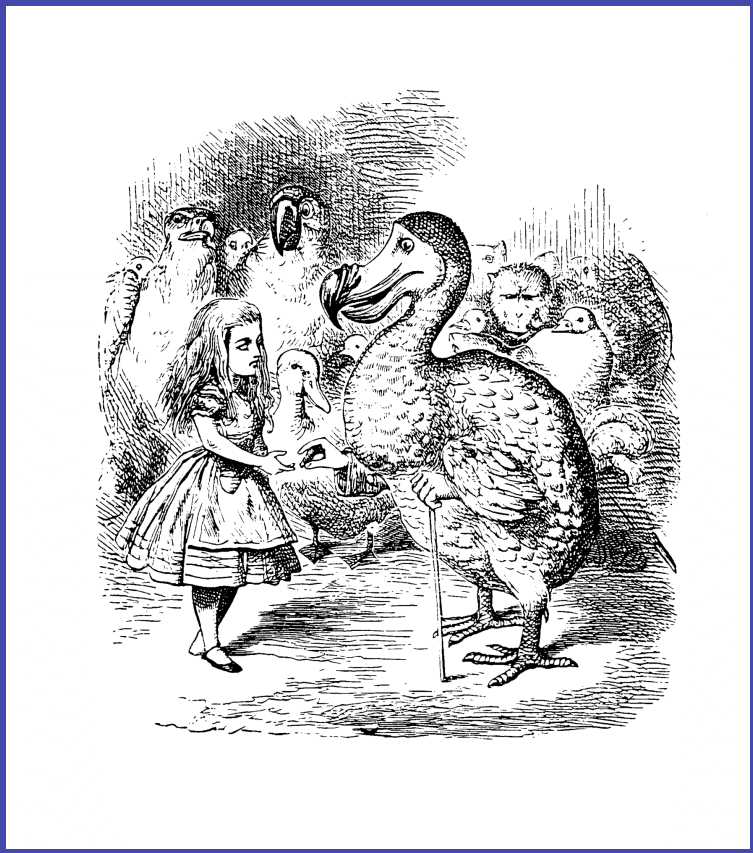


Conclusions

- 1. Support for the Dodo verdict.
- 2. Dilution of treatments didn't impair effectiveness.
- 3. Results may be of particular interest to practitioners of PCT and PDT.
  - Comparable effectiveness to CBT in routine practice may have been unappreciated.

Qualifications, Limitations, and Alternative Hypotheses

- Limited specification of treatments
  - Were treatments faithful and pure? Were they alike?
- Non-random assignment of clients
  - Did CBT get the most difficult clients?
- Absence of a control group
  - Would clients have improved naturally?
- Incomplete data
  - Biased reporting?
- Restriction to one self-report measure
- Investigator allegiance



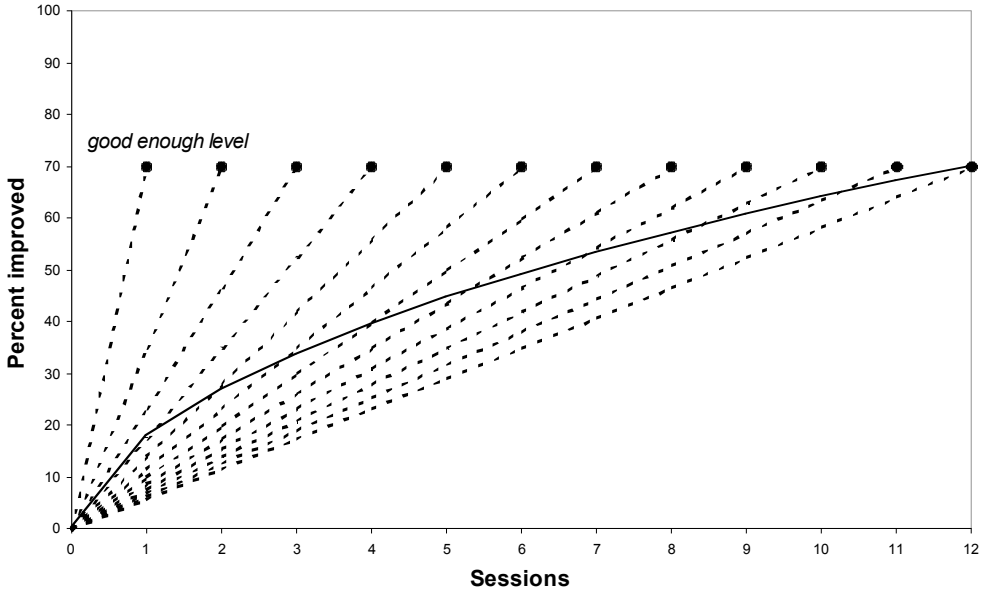
Issue 2: How Many Sessions Should Clients Receive?

- Dose-effect model
  - Asks what *dose* of therapy clients require.
  - Dose is denominated in sessions.
  - More sessions is presumed to represent a stronger dose.
  - Dose-effect curve is negatively accelerated
- Responsive regulation of treatment duration
  - Regulation of dose to fit requirements
  - Selective adaptation of treatment approach (clinical judgment, timing, choice of techniques)
  - Acceleration of processes to fit externally-imposed limits on time and resources.
  - Treatment continues until client reaches a *good enough level.*

The Good Enough Level (GEL)

- Level of improvement and treatment duration are mutually regulated.
- Treatments end when clients have improved to a level that is *good enough*.
- GEL and hence treatment duration my vary with characteristics of the clients, therapists, and settings
  - nature and severity of problems, personalities, available resources, etc.
- The GEL is a manifestation of *responsiveness* (behavior influenced by emerging context).
  - adjusting the length of therapy, degree of focus, effort, etc.
- Clients and therapists may agree to end treatment
  - when goals have been reached or acceptable progress has been made.
- Imposing time limits may accelerate the rate of improvement.

Clients end treatment at GEL: Theoretical account of the negatively-accelerating dose-response curve



Reliable and Clinically Significant Improvement (RCSI)

- Based on Jacobson and Truax (1991).
- Clients had achieved RCSI if they
  - (a) entered treatment in a dysfunctional state and left treatment in a normal state,
  - (b) having changed to a degree that was probably not due to measurement error.

- (A) *clinical cutoff*, dividing the dysfunctional from the normal populations
  - Study 1: cutoff = 11.9 for men and 12.9 for women based on earlier studies (Evans et al., 2002).
  - Study 2: cutoff = 10, based on comparing clinical populations with a systematic general population sample (Connell et al., 2007).

- (B) *reliable change index*, a pre-post difference that, when divided by the standard error of the difference, is equal to 1.96
  - Study 1: reliable change index = 4.8.
  - Study 2: reliable change index = 4.5.

RCSI: Methodological Considerations

- Clinical cutoff depends on normal comparison sample.
- Study 1: Cutoffs of 11.9 and 12.9 based on comparison with university students and convenience sample
  - M = 6.9 for men and 8.1 for women (Evans et al., 2002).
- Study 2: Cutoff of 10 based on comparison with general population sample (using formulas given by Jacobson and Truax, 1991)
  - M = 4.8, with no gender differences (Connell et al., 2007).
- We included only clients who began treatment above cutoff.
  - Clients who began below the clinical threshold could not move from the clinical to the normal population.
- Reliable change index (RCI) depends on standard error of difference, which depends on the SD of the pre-post difference, and the reliability of the CORE-OM.
- For example, in study 2, SD_{diff} = 6.51
  - Using internal consistency reliability = .94 yields RCI = 4.5.
  - Using one-month test-retest reliability = .88 yields RCI = 6.4.

Clinical cutoff =

$$\frac{mean_{clin}sd_{norm} + mean_{norm}sd_{clin}}{sd_{norm} + sd_{clin}}$$

Reliable Change index:

$$RCI = 1.96\,\mathrm{sd}\sqrt{2}\sqrt{1-r}$$

$$\mathrm{RC} = \frac{x_2 - x_1}{S_{diff}}$$

$$S_{diff} = \sqrt{2(S_E)^2}\,.$$

$$SE = \mathrm{sd}\sqrt{1-r}$$

Selection of Clients for Dose-Response Comparison

	Sample 1	Sample 2
Therapist returned initial CORE Assessment Form	12,571	33,587
Primary care	6,610	33,587
Ending planned AND Valid pre- and post-treatment CORE-OM	1,956	11,352
Number of sessions reported AND ≤ 12 (S1) or ≤ 20 (S2) AND began treatment above cutoff: CORE-OM ≥ 11.9 / 12.9 (S1) CORE-OM ≥ 10 (S2)	1,472	9,703

Sample Characteristics

Sample 1 (2001)

- N = 1,472 clients
  - 73.1% female
  - Mean age = 40.0
- N = ???+ therapists
- ?? NHS services
  - primary and secondary care

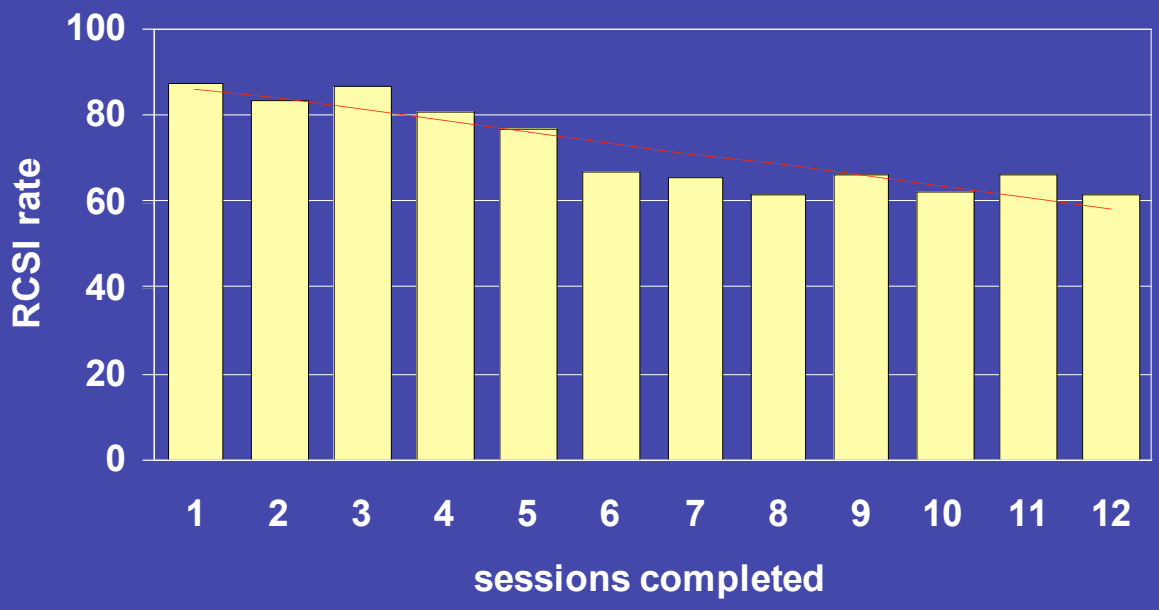
Sample 2 (2005)

- N = 9,703 clients
  - 70.7% female
  - Mean age = 40.7
- N = 445 therapists
- 33 NHS services
  - all primary care

Reliable and clinically significant improvement (RCSI) as a function of number of sessions attended: Sample 1 (2001)

Sessions attended	N above cutoff at intake	Clients who achieved RCSI	
		n	percent
1	8	7	87.5
2	84	70	83.3
3	120	104	86.7
4	172	139	80.8
5	193	149	77.2
6	536	358	66.8
7	70	46	65.7
8	79	49	62.0
9	62	41	66.1
10	77	48	62.3
11	24	16	66.7
12	47	29	61.7
Total	1472	1056	71.7

Reliable and clinically significant improvement (RCSI) as a function of number of sessions attended: Sample 1 (2001)

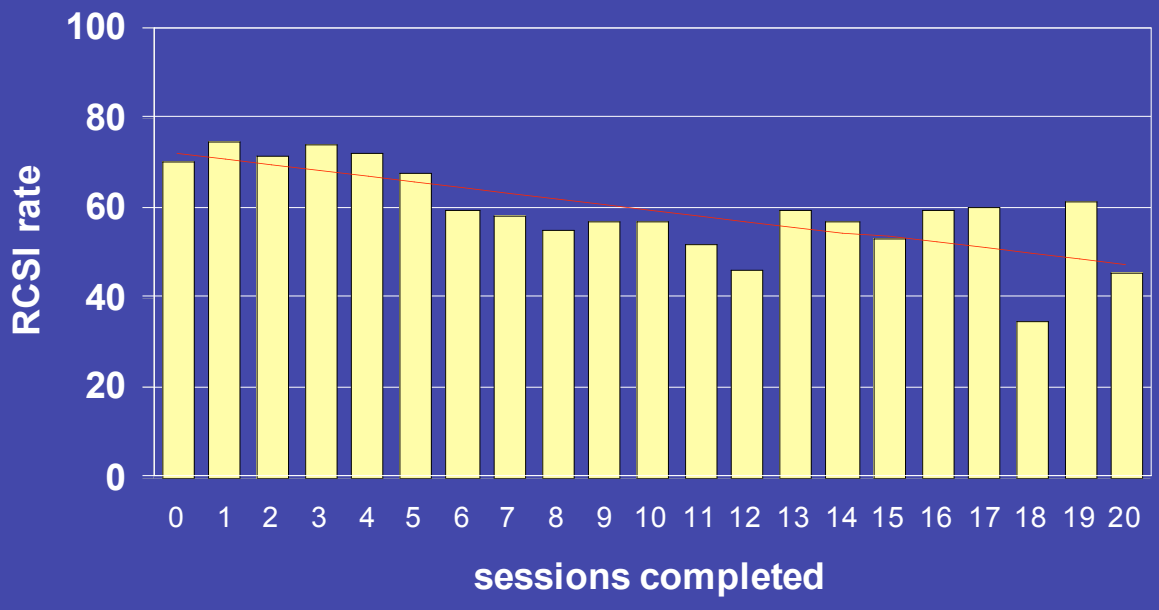


Reliable and clinically significant improvement (RCSI) as a function of number of sessions attended:

Sample 2 (2005)

Sessions attended	N above cutoff	RCSI	
		N	Percent
0	17	12	70.6
1	67	50	74.6
2	413	297	71.9
3	747	554	74.2
4	1030	747	72.5
5	1274	861	67.6
6	2674	1587	59.3
7	905	527	58.2
8	958	527	55.0
9	373	213	57.1
10	326	185	56.7
11	213	111	52.1
12	428	197	46.0
13	59	35	59.3
14	44	25	56.8
15	49	26	53.1
16	37	22	59.5
17	25	15	60.0
18	29	10	34.5
19	13	8	61.5
20	22	10	45.5
Total	9703	6019	62.0

Reliable and clinically significant improvement (RCSI) as a function of number of sessions attended: Sample 2 (2005)



Improvement slightly negatively correlated with number of sessions

- Correlations calculated across categories (not across clients)
- Number of sessions negatively correlated with RCSI rate:
  - Study 1: $r = -.91$, $p < .0001$, $n = 12$ categories.
  - Study 2: $r = -.75$, $p < .001$, $n = 21$ categories.

Conclusions - 1

- Outcomes did not improve with larger doses
  - Clients who had 1 or 2 sessions improved as much (on average) as clients who had 15 or 16 sessions.
  - Seems paradoxical and surprising if treatment is considered as an independent variable in an experimental manipulation.
  - But clinically sensible if clients and therapists are considered as responsively ending treatment when a GEL has been reached.
    - clients change at different rates and achieve a satisfactory level of gains at different treatment durations.

Conclusions - 2

- The slight decline in RCSI rates across dose (sessions) suggests that a client's GEL is influenced by costs.
  - Clients and therapists satisfied with less as more time and effort is required.
- Responsive regulation could help explain the Dodo verdict.
  - If outcome is regulated by GEL, treatments may tend to have similar mean outcomes.
  - Therapists responsively use whatever tools their approach provides.
  - Presumes that all approaches have some effective tools, which can be adapted.