

Evaluating Psychological Therapies Services: A review of outcome measures and their utility

Chris Leach^{1,3}, Mike Lucock^{2,3}, Steve Iveson² & Rachael Noble²



- ¹ The Clinical Governance Support Team, South West Yorkshire Mental Health NHS Trust
² Adult Psychological Therapies Services, South West Yorkshire Mental Health NHS Trust
³ School of Human & Health Sciences, University of Huddersfield, Huddersfield

Evaluating Psychological Therapies Services: A review of outcome measures and their utility

Chris Leach, Mike Lucock, Steve Iveson & Rachael Noble

Abstract

We describe the use of a number of outcome measures that have been used over the past eight years to evaluate adult psychological therapies services in the Wakefield & Pontefract locality of the South West Yorkshire Mental Health NHS Trust. The psychological therapies service now has a unique database of outcome measures completed by 5563 clients, which has been used nationally and internationally for service evaluation and research purposes. Internally, the database has informed the clinical service, allowing prioritising of referrals and feedback to clinicians and referral agencies on the quality of the service and appropriateness of referrals.

We describe the properties and intended use of the measures, the way results are fed back to the service and evaluate the usefulness of the measures in routine service settings. Implications for psychological therapies and other mental health services are discussed including:

- Choice of generic or specific outcome measures
- How the measures relate to one another
- Clinical usefulness of the measures, including risk assessment
- Interpretation of results, including availability of norms and how to assess change
- Practical utility, including cost

Keywords: psychological therapies; practice-based evidence; CORE-OM; BDI; BAI; IIP-32; DES

Introduction

Background

The clinical governance and clinical effectiveness agendas in the NHS emphasise the need for routine service audit and evaluation (Department of Health, 2004, p. 30). There is considerable evidence of efficacy of psychological therapies from research trials (Roth & Fonagy, 1996), which has fed into evidence-based guidelines (Department of Health, 2004). However, this evidence is often obtained from research clinic settings, where, for example, therapists are required to deliver therapy strictly according to a manual and complex clients with multiple problems are excluded. To complement this evidence, there is a need to evaluate the effectiveness of psychological therapies services in "usual service conditions" (Department of Health, 1999, page 116), where a range of therapies is provided by a range of therapists to a range of clients with a range of problems. Effectiveness research in routine clinical settings is an example of "practice-based evidence" (Barkham, Margison, Leach, Lucock, Benson, Mellor-Clark, Evans, Connell, Audin & McGrath, 2001), which complements the evidence-based guidance and provides "a framework for using local evidence to support practice" (Department of Health, 2004, p. 29). Practice-based evidence can also be used within services to feed data back to clinicians to inform their practice, and to inform clients of their progress. This feedback is an important part of any applied research, audit or evaluation and is often missed out (Parry, 1992).

Since 1997, the Adult Psychological Therapies Service (APTS) in the Wakefield & Pontefract locality of the South West Yorkshire Mental Health NHS Trust (formerly Wakefield & Pontefract Community Health NHS Trust) has been collecting outcome data for all clients referred to the service as part of the routine service delivery. The service receives about 1200 referrals a year, mainly from general practitioners, but also from the wider mental health services, psychiatry and social services and more recently from the Primary Care Liaison, Assessment, Therapy & Training (PLATT) team. The service is multi-professional and has included clinical psychologists, psychotherapists, counsellors, cognitive behaviour therapists, counselling psychologists, and an art therapist. Clients are seen on three main sites (Wakefield, Pontefract and the Specialist Psychotherapy Team based at Horbury) as well as general practice clinics.

The collection of outcome data of various sorts has been part of service delivery for many years at the Wakefield and Pontefract sites, but the appointment of research staff in 1995 allowed such data to be collected more systematically, resulting in the development of a database that now (May 2004) includes data on measures completed by 5563 clients. The database has become an archive that is useful for local service evaluation and feedback to clinicians and referrers. For example, clinicians receive feedback on all their clients annually, as well as being offered feedback on demand about individual clients and groups of clients. The data collected at referral allows prioritising of referrals based on scores on the outcome measures, risk items and comments given by clients. The database has also been used for research purposes and has attracted the attention of international researchers with whom we now collaborate on a regular basis (see, for example, Stiles, Leach, Barkham, Lucock, Iveson, Shapiro, Iveson & Hardy, 2003; Lutz, Leach, Barkham, Lucock, Stiles, Evans, Noble & Iveson, in press).

The background to the setting up of the routine collection of such data and the use of the database for research is discussed in more detail in Lucock, Leach & Iveson (1999). Our use of the outcome measures to help improve "evidence based reflective practice" is more fully described in Lucock, Leach, Iveson, Lynch, Horsefield & Hall (2003).

Outcome Measures

The choice of appropriate routine outcome measures is an important issue. Measures should be valid and reliable and measure important aspects of client change. Ideally, they should also be readily understood by clinicians and inform assessment and clinical practice. It is also important to minimise the use of measures where possible to reduce the burden on clients. In this article, we review our experience of a number of outcome measures and evaluate their usefulness in routine service delivery for both psychological therapies services and more general mental health services.

It is a challenge not only to routinely measure clinical outcome, but to feed back the results to the service and individual clinicians in a meaningful way that allows them to reflect on the service provided. Therefore, in addition to describing the measures used, we will illustrate the various ways in which the outcome data is fed back into the service.

Measures that have been routinely completed by APTS clients include the Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM: Evans, Connell, Barkham, Margison, McGrath, Mellor-Clark & Audin, 2002), the Beck Depression Inventory (BDI: Beck, Ward, Mendelson, Mock & Erbaugh, 1961), the Beck Anxiety Inventory (BAI: Beck, Epstein, Brown & Steer, 1988), the Dissociative Experiences Scale (DES-II and DES-Taxon: Carlson & Putnam, 1993) and the Inventory of Interpersonal Problems (IIP-32: Barkham, Hardy, & Startup, 1996). Partly to cut down the number of measures sent to clients and partly as a result of the prohibitive cost of using copyrighted measures, we have recently stopped sending clients the BDI and the BAI. At referral we use only the CORE-OM. The CORE-OM and IIP-32 are then used at the

four other stages, assessment, beginning of therapy, discharge and 6 month follow up. We have also introduced our own short Screener questionnaire at assessment. This is at the pilot stage, and includes items detecting issues, such as eating disorders and substance misuse, not captured by the other measures we have used. The screener also includes risk items. The DES-II has recently been replaced in the service by the shorter DES-Taxon, a subset of eight items that discriminates better than the full scale and that is more acceptable to clients. This is used at assessment only.

The APTS service

Clients referred to the service are put on a waiting list for assessment. Assessment is carried out by experienced therapists, who recommend what, if any, further therapy might be required. The clients are then put on a waiting list for therapy, which includes individual or group therapy of various sorts. Once they start therapy, they will usually stay with the same therapist until they are discharged. Clients are sent questionnaires, including a subset of the outcome measures, on up to five occasions: at referral, before assessment, before therapy, at discharge, and at six months following discharge. If the period between referral and assessment is less than six weeks, the client is not sent a set of measures at assessment, and, similarly, if the period between assessment and therapy is less than six weeks, the client is not sent a pre-therapy set of measures. Clients are told that the measures will be used for routine clinical purposes to assess their progress and that they may also be used for research purposes. They can opt not to complete the measures at all without it affecting the service provided and they can opt for any measures they complete not to be used for research purposes. Any use of the measures by research or other staff not directly involved in the service provision is done in a totally anonymous way and the research database does not include any client identifiable information. From 2002 the data has been inputted directly onto the Trust's Clinical Information System (CIS). The data is then passed on to the service in Excel files which are converted into SPSS files for analysis. Information on number of sessions, type of therapy and types of problems are recorded at discharge.

In addition to the outcome measures, the set of questionnaires includes space for client comments. The three sets sent out before therapy allow clients to add any information about their needs that they feel might be helpful to the service and the two sets after discharge allow them to comment on the quality of the service they received. The pre-therapy comments, particularly those completed at referral, can provide information that the client has not been able or willing to disclose to the referrer. These are taken into account in prioritising referrals. The comments after therapy are fed back routinely to clinicians.

The APTS Research Database

Table 1 shows numbers of clients who have completed the various measures since 1997. For example, of the 5563 clients who have completed at least one measure on one occasion, 4535 have completed the CORE-OM at referral, while 715 have completed it at discharge. The smaller numbers completing the CORE-OM and other measures on later occasions reflect both service and client issues. Because clients were not asked to complete later measures before therapy if the period following earlier completion was less than six weeks, the numbers in the Assessment and Before therapy rows are usually smaller, except for the DES-II, which was given out only at assessment. The dramatic drop in numbers of questionnaires completed after therapy and at follow-up is a result of a number of factors. First, some of the clients in the database will not have started or completed therapy, so will only have completed some of the measures before therapy began. Second, some clients will have dropped out of therapy. Third, clients often chose not to complete the questionnaires at discharge and follow up. Such huge attrition in completion of measures is typical of most psychological therapies services, where clients having come through the service may not wish to continue involvement for many different reasons. Fourth, clinicians

complete a discharge form for each client, which allows them to ask for particular clients not to be sent outcome forms if it is deemed to be clinically detrimental for the client.

Table 1. Numbers of APTS clients completing outcome measures at five stages (total n = 5563)

Stage	CORE-OM	BDI	BAI	IIP-32	DES-II	DES-Taxon	APTS Screener
1. Referral	4535	3978	1460	2555	10	-	143
2. Assessment	2046	1934	684	1913	505	64	2
3. Before therapy	923	867	351	883	2	1	-
4. Discharge	715	575	222	600	-	-	-
5. Six month follow-up	493	393	99	408	-	-	-
Pre-therapy	5228	4701	1797	3980	513	65	145
Pre-therapy & Discharge	478	397	173	397	-	-	-

The bottom two rows of Table 1 show the numbers of clients completing any of the pre-therapy measures and the numbers completing both a pre-therapy measure and a discharge measure. These are useful for tracking progress of individual clients.

The Outcome Measures and Feedback of the Data

At the request of clinicians, we provide information on the measures and the significance of scores. We provide brief background to the measure, a summary of how we have used the measure, any relevant cut-offs distinguishing levels of severity that can be used in routine clinical practice, and an indication of how the measure relates to other measures.

CORE-OM

The CORE outcome measure (CORE-OM) was developed recently in the UK (Evans et al, 2002) as part of a routine assessment system to assess clinically relevant emotional problems. It has been used predominantly in primary and secondary care settings as a generic measure to supplement more specific measures of functioning in particular areas. It is short (34 items on two sides of A4) and completed by clients to assess their feelings over the last week using a rating scale from 0 ("not at all") to 4 ("most of the time"). It produces an overall score between 0 and 4 (the average of the 34 items). It has four subscales, well-being, problems, functioning and risk. The most useful subscale is the risk subscale, which is unusual in having two items on risk to others ("I have been physically violent to others", "I have threatened or intimidated another person") as well as four items on risk to self ("I made plans to end my life", "I have thought of hurting myself", "I have thought it would be better if I were dead", "I have hurt myself physically or taken dangerous risks with my health"). We have not found the other sub-scales clinically useful.

The CORE-OM can be freely copied on a copy-left basis (i.e., don't change it and don't use it for profit). In addition, two shorter forms (19 items on one side of A4) are available and have been used in the APTS for assessing client progress session-by-session.

Table 2 shows the scores grouped into three ranges for the CORE-OM and its risk sub-scale, based on UK national data and local Wakefield data, which have fed into national norms (see Barkham et al, 2001). Clients scoring in the minimal range are defined to be below the clinical cut-off. Separate cut-offs are used for male and female clients. The table also shows the proportions of APTS clients scoring in the various ranges at referral, so that 84% of referrals (male or female) score above the clinical cut-off, with 28% of female referrals and 26% of male referrals scoring in the severe range. This allows us to conclude that APTS, like other psychological therapies services, is receiving referrals from clients with serious problems, not just the "worried well". The table also shows that well over half the clients referred score above the clinical cut-off for risk.

Table 2. Clinical cut-offs for the CORE-OM and proportions of APTS clients scoring in the various ranges at referral

Measure	Clinical Cut-offs			Percentage of Clients	
	Female	Male		Female	Male
CORE-OM (<i>n</i> = 4543)	2.50-4.00	2.50-4.00	severe	28%	26%
	1.29-2.49	1.19-2.49	moderate	56%	58%
	0.00-1.28	0.00-1.18	minimal	16%	16%
Risk sub-scale (<i>n</i> = 3727)	0.31-4.00	0.43-4.00	clinical range	59%	54%
	0.00-0.30	0.00-0.42	non-clinical	41%	46%

Fig 1. CORE-OM scores at referral 1997-2003

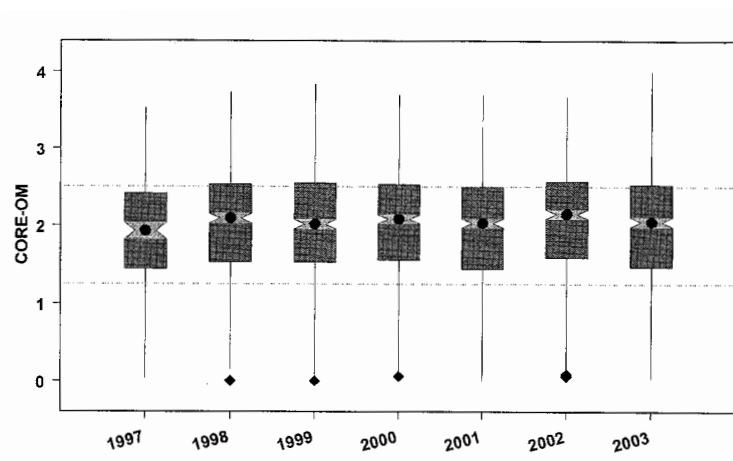


Figure 1 shows the same data in box-plot form, reported separately for each year. Each box-plot gives a summary of the CORE-OM scores for clients referred in each year. The black dots in the middle of the boxes are the average scores (medians) for each year, the boxes include the middle 50% of the data, and the whiskers include the highest and lowest 25% of the scores, with unusually large or small scores (outliers) shown as diamonds at the top or bottom of the plot. The clinical cut-off of 1.25 (the average of the male and female cut-offs) and the severe cut-off of 2.50 are shown as dotted lines. This allows us to see immediately that in each year well over 75% of referrals (the box and the upper whiskers enclosing 75% of the scores) score above the clinical cut-off. The notches in the boxes are 90% confidence intervals for the median scores, which allow a rough statistical comparison of year on year data. If the notches do not overlap when comparing two years, there is a reliable difference in average scores for the two years. With no overlaps apparent and the box-plots having very similar shapes, it is clear that there has been no variation in severity of referrals in the period 1997-2003. Although the box-plots do not allow totally accurate estimates of proportions above the cut-offs (the actual percentage given in Table 2 across all years is 84%), they do allow a quick visual comparison to be made.

Figure 2 shows box-plots of the CORE-OM scores at the five therapy stages. This shows little variation in scores before therapy starts, with over 75% of clients scoring above the clinical cut-off at referral, before assessment and before therapy and little difference between the medians. The actual percentages above the clinical cut-off at these three points are 84%, 80% and 78%, respectively. After therapy, the percentages above the clinical cut-off drop to 38% at discharge and 44% at six-months follow-up, and the figure shows reliable differences between the two post-therapy sets of scores and the three pre-therapy sets.

These plots do not track individual clients. To check on individual client progress, the so-called Jacobson plot in Figure 3 contains a dot for each of the 478 clients who have completed a CORE-OM at some point before therapy and at discharge (Jacobson & Truax, 1991). Dots appearing below the main diagonal show clients whose scores have decreased after therapy, showing some improvement. However, by the nature

Fig 2. Box-plots of CORE-OM scores at the five stages for 5563 clients

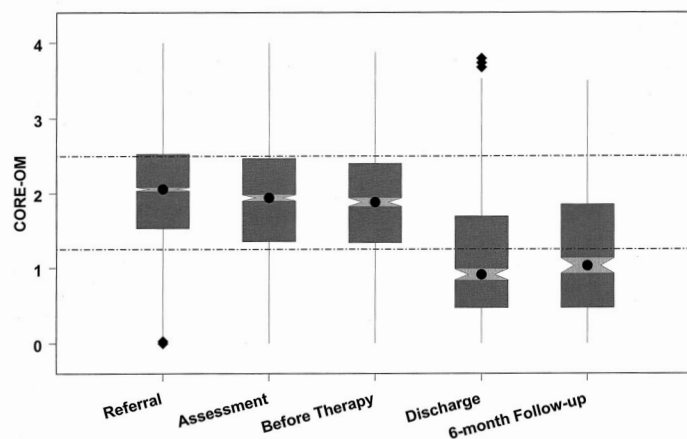
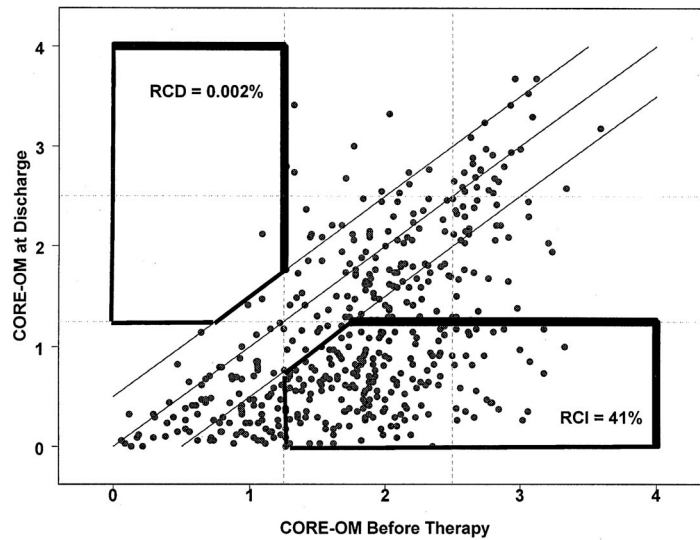


Fig 3. Jacobson plot of CORE-OM scores before therapy and at discharge for 478 clients



of all measures, some fluctuation in scores completed by the same clients on two occasions would be expected by chance, so the tram-lines on either side of the main diagonal correct for this by taking into account the reliability of the measure. Taking this into account for the CORE-OM, any client showing more than a 0.5 difference in before and after scores has a statistically reliable difference between the scores. 286 dots (60%) are below the lower tram-line, so 60% of clients show reliable improvement on the CORE-OM following therapy. 19 clients (4%) have dots above the upper tram-line, so these show reliable deterioration.

The lower box in Figure 3 encloses dots for which the pre-therapy scores were above the clinical cut-off and the discharge scores are below the cut-off, as well as being statistically different (i.e., below the lower tram-line). 194 clients (41%) are represented here and these are clients who are deemed to have shown reliable and clinical improvement following therapy. The upper box encloses any dots for which the pre-therapy scores were below the clinical cut-off and the discharge scores are above the cut-off, so these would indicate clients who have shown reliable and clinical deterioration. Only one client (0.002%) is shown here. The Jacobson plots allow us to feed back results for individual clients to clinicians as well as give a picture of the overall quality of the service. Clients figuring above the upper tramline and, in particular, in the upper box, would be of particular note.

These three figures are routinely produced for all outcome measures. They show effectively the same picture of the service, so will not be repeated when considering later measures.

The CORE-OM correlates quite highly with the BAI ($r = 0.61$), the IIP-32 ($r = 0.62$) and the DES-II ($r = 0.55$ with the DES-II total score and $r = 0.49$ with the DES-Taxon) and very highly with the BDI-I ($r = 0.85$).

The high correlation with the BDI makes it possible to translate between the two with a high degree of accuracy using either the tables provided by Leach, Lucock, Barkham, Noble & Iveson (submitted) or the following equations:

For female clients:

$$\text{BDI-I} = (2.403 \times \text{CORE} + 2.117)^{1.667} \quad (1f)$$

$$\text{CORE} = 0.309 \times \text{BDI-I}^{0.60} - 0.152 \quad (2f)$$

For male clients:

$$\text{BDI-I} = (2.289 \times \text{CORE} + 2.068)^{1.667} \quad (1m)$$

$$\text{CORE} = 0.319 \times \text{BDI-I}^{0.60} - 0.142 \quad (2m)$$

The first equation in each pair allows a BDI-I score to be predicted from a client's CORE-OM score, while the second allows a CORE-OM score to be predicted from a BDI-I score. If there is interest in translating between the CORE-OM and the more recent version of the BDI, the manual for the BDI-II (Beck, Steer & Brown, 1996) contains a table for translating between the BDI-I and the BDI-II, which can be used together with equations 1 and 2 or the tables in Leach et al (submitted).

Beck Depression Inventory (BDI-I)

The BDI is a widely used specific measure of depression. It asks clients to rate their feelings over the past week on 21 items, using a four-point rating scale from 0 to 3, with the four points spelled out separately for each item. The clinical norms are based on US data, but it has been widely used in the UK for many years. The version we have used is the original version, now known as the BDI-I. A later version, the BDI-II, is available (Beck et al, 1996), but is not yet so widely used. The main score used is the overall score, from 0 to 63, the sum of the 21 item scores. We have also found two individual items (2 and 9) useful in assessing risk to self. Item 2 asks about feelings of hopelessness and item 9 focuses on suicide ideation. We have traditionally used high scores on these items to prioritise referrals.

The BDI is copyrighted and costs over £1 per copy. Because of the large numbers of referrals and the need to send copies at five stages and repeat copies to clients who do not return them, the cost has become prohibitive to the service, so we have stopped using the BDI on a regular basis. The possibility of predicting the BDI score from the CORE-OM score has helped clinicians who have traditionally relied on the BDI to assess severity of depression and clinical change.

The clinical cut-offs for the BDI are shown in Table 3, with the percentages of APTS clients scoring in the various ranges. 75% of clients referred score in the moderate or severe clinical depression ranges. This figure includes people not referred specifically for help with depression and again points to the service being referred clients with serious problems, not the "worried well".

Beck Anxiety Inventory (BAI)

The BAI is a specific measure of anxiety completed by clients and, like the BDI, is widely used, with norms based on US data, but sufficient experience of its use abroad for these norms to apply also to the UK. Clients are asked to assess their feelings over the past week on 21 items, using a rating scale of 0 ("not at all") to 3 ("severely"). A single score is reported, the sum of the 21 item scores, ranging from 0 to 63.

It is copyrighted, so may not be copied. It also costs over £1 per copy and again the prohibitive cost of using it for 1200 referrals a year at five stages has meant that we stopped using it on a regular basis about two years ago.

The clinical cut-offs and proportions of clients scoring in the various ranges at referral are shown in Table 3. 63% of clients referred, for whatever reason, score within the

Table 3. Clinical cut-offs for the BDI-I, BAI and DES-II and proportions of APTS clients scoring in the various ranges at referral (BDI & BAI) or assessment (DES-II)

Measure	Clinical Cut-offs		Percentage of Clients	
Beck Depression Inventory (BDI) (<i>n</i> = 3978)	30-63	Severe	36%	
	17-29	Moderate	39%	
	10-16	Mild	16%	
	0- 9	Minimal	9%	
Beck Anxiety Inventory (BAI) (<i>n</i> = 1460)	26-63	Severe	35%	
	16-25	Moderate	28%	
	8-15	Mild	23%	
	0-7	Minimal	13%	
Dissociative Experiences Scale (<i>n</i> = 513)	DES II	30-100	clinical range	15%
		0-29	non-clinical	85%
DES Taxon	DES Taxon	20-100	clinical range	16%
		0-19	non-clinical	84%

“moderate” or “severe” anxiety ranges. A large proportion of clients referred score within the clinical ranges for both depression and anxiety, with the correlation between the BDI-I and the BAI being high ($r = 0.60$). Such co-morbidity is typical of NHS psychological therapies services.

Dissociative Experiences Scale

The Dissociative Experiences Scale assesses level of dissociative experience by asking clients to state what percentage of time they have particular experiences. We began using the full DES-II at assessment at the request of clinicians in one of the sites who were seeing increasing numbers of clients who experience serious dissociation. Dissociation is a significant factor in assessing suitability for therapy and type of psychological therapy indicated. On the full scale of 28 items, some of the experiences are common to most people some of the time, such as not remembering what has happened for part of a trip by car or bus. Others, such as finding yourself in a place and having no idea how you have got there, are rarely experienced by people without severe dissociative problems. The full scale (DES-II) gives a score from 0 to 100, the average of the scores for the 28 items. The DES-Taxon consists of eight rarer experiences. It therefore discriminates better between those who dissociate and those who do not. Again scores are averaged to give a score between 0 and 100. The clinical cut-off for the full DES-II is 30 and for the DES-Taxon it is 20. We used the full version initially because of concerns that using just the eight DES-Taxon items might worry clients. However, our experience so far is that those clients who do dissociate are quite prepared to fill in a shorter form based on the DES-Taxon items, and those who do not are happy to tick no to all the questions. For this reason, we have discontinued using the full version and now rely on just the eight-item DES-Taxon scale. Clients now complete this across the service at the assessment stage only.

Table 3 shows that 15% of APTS clients score above the clinical cut-off for the full DES-II, while 16% score above the cut-off for the DES-Taxon. Figure 4 is a box plot of the distribution of DES-Taxon scores, which shows most clients scoring well below the clinical cut-off of 20.

The DES-II and its Taxon subscale is not copyrighted, so may be used freely in the NHS and other setting

Inventory of Interpersonal Problems (IIP-32)

The IIP-32 is a short form developed in the UK (Barkham et al., 1996) of the much longer 127 item IIP scale originated in the US (Horowitz, Rosenberg, Baer, Ureno, & Villasenor, 1988). Two major theoretical approaches underpin common uses of the IIP. The first is the factor analytic approach embedded in the development of the short version, which posits eight factors underlying personal problems. These eight factors are the eight separate subscales of the IIP-32, which have been grouped into four bipolar factors by Barkham, Hardy & Startup (1994). The four bipolar factors were identified as problems relating to competition (hard to be assertive vs. too aggressive), socialising (hard to be sociable vs. too open), nurturance (hard to be supportive vs. too caring), and independence (hard to be involved vs. too dependent). The second theoretical approach is based on Leary's (1957) interpersonal circle and results in a more complex circumplex scale (see, for example, Alden, Wiggins & Pincus, 1990). Our use of the IIP has followed the factor analytic model. We started using it in the service at the request of clinicians following psychodynamic psychotherapy treatment models, who wanted a measure that would tap interpersonal problems.

The IIP-32 has 32 items on two sides of A4. 19 items are phrased "It is hard for me to ..." (e.g., "join in groups", "be assertive with another person"), while the remaining items are phrased "too much", such as "I fight with other people too much" and "I open up to people too much". Each item is rated by the client on a five-point scale from "not at all" (0) to "extremely" (4). An overall score is obtained as the average of all 32 item scores, giving a score from 0 to 4. In addition, eight sub-scale scores are derived to reflect difficulties in the four bipolar factors. No normative data are available, so Table 4 gives the highest scoring percentiles of APTS clients at referral, which will allow high scorers to be noted. For example, only 5% of clients score 2.62 or above on the full scale, but 25% of clients score 2 or above.

Fig 4. Box plot of scores on the Dissociative Experiences Scale (DES-Taxon) for 512 clients at assessment. The dotted line indicates the clinical cut-off

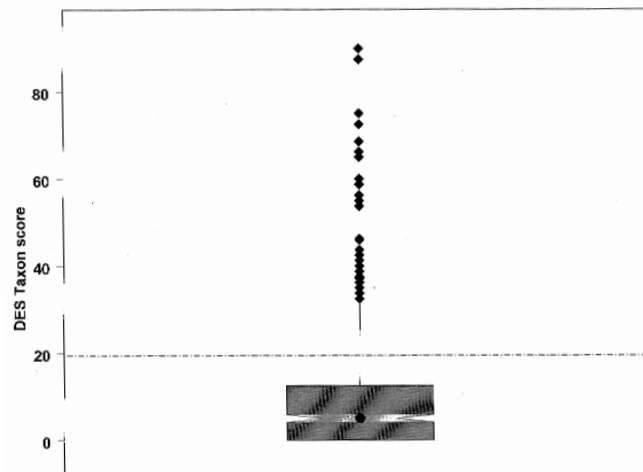


Table 4. Percentages of clients scoring in various ranges for the Inventory of Interpersonal Problems (IIP-32)

Measure	Percentage of Clients
IIP Total Score (<i>n</i> = 2555)	5% score 2.62 or above 10% score 2.40 or above 25% score 2.00 or above
IIP Subscales	
1. Hard to be Sociable	5% score 4.00 10% score 3.50 or above 25% score 2.75 or above
2. Hard to be Assertive	5% score 3.75 or above 10% score 3.25 or above 25% score 2.75 or above
3. Too Aggressive	5% score 3.75 or above 10% score 3.25 or above 25% score 2.50 or above
4. Too Open	5% score 3.25 or above 10% score 2.75 or above 25% score 2.00 or above
5. Too Caring	5% score 3.50 or above 10% score 3.00 or above 25% score 2.50 or above
6. Hard to be Supportive	5% score 3.00 or above 10% score 2.75 or above 25% score 1.75 or above
7. Hard to be Involved	5% score 3.75 or above 0% score 3.25 or above 25% score 2.25 or above
8. Too Dependent	5% score 3.25 or above 10% score 3.00 or above 25% score 2.25 or above

The IIP-32 is not copyrighted, so can be freely used in NHS settings. It correlates reasonably well with the CORE-OM ($r = 0.62$) and the BDI-I ($r = 0.63$), and moderately with the BAI ($r = 0.43$) and the DES-II ($r = 0.48$).

Discussion

We have described five measures used routinely to evaluate our psychological therapies service. Some measures, such as the CORE-OM, are used at all stages, whilst others, such as the DES-Taxon are used only at assessment. In selecting the measures we have taken into account the need for reliable and valid measures of change and the views of therapists that have led to the introduction of the IIP-32 and the DES-Taxon. We have also taken into account cost which has led us to use the CORE-OM as the main outcome measure, and abandon use of the BDI and BAI. The CORE-OM is a general measure of psychological problems whilst the BDI, BAI, DES and IIP measure more specific problems and aspects of functioning. It could be argued that the CORE-OM is therefore the best measure for the routine service evaluation and the other more specific measures could be used for more specific purposes. For example the BAI may be appropriate to evaluate anxiety management groups.

We now only use the CORE-OM at the referral stage because of concerns about asking clients to complete too many questionnaires at a time when they are not in direct contact with clinicians. It is possible clients may become distressed when completing questionnaires that raise issues for them and this can be addressed when the client has access to a therapist. The abandonment of the BDI and the use of the CORE-OM as the main outcome measure are further supported by our ability to reliably convert CORE-OM scores to BDI scores.

In order to inform assessment and reflective practice clinicians must feel part of the process and find the information understandable, interesting and helpful. There is a danger it is seen as merely an administrative exercise, an add-on to an already difficult role with little relevance to the service and clients needs. They require feedback in a form that is readily understood and information on the nature of the measures and the meaning of scores. Regarding reflective practice, when the outcome data has been fed back to clinicians, a number of further questions become apparent. For example, in Figure 3, one wonders about the characteristics of clients who significantly improve or significantly deteriorate and reasons for this change. If they deteriorate, is it due to the therapy or outside factors? When we look at the number of sessions provided, the question often raised is why are some clients seen for longer-term therapy and how is the decision made? Why do some clients opt out of the service? This has led to further off- shoot projects examining some of these questions. The use of measures to enhance assessment and evaluate clinical outcome should therefore be seen as part of a wider process of reflective practice.

The use of the measures described in this paper therefore has a number of aims in addition to service evaluation. They provide information on clinical outcome for the service as a whole, they provide information to inform the initial assessment, including risk and suitability for different forms of therapy, and they are fed back to clinicians to enhance reflective practice.

References

- Alden, L.E., Wiggins, J.S., & Pincus, A.L. 1990. Construction of circumplex scales for the Inventory of Interpersonal Problems. *Journal of Personality Assessment* 55: 521-536.
- Barkham, M., Hardy, G. E., & Startup, M. 1994. The structure, validity, and clinical relevance of the Inventory of Interpersonal Problems. *British Journal of Medical Psychology* 67: 171-185.
- Barkham, M., Hardy, G. E., & Startup, M. 1996. The development of the IIP-32: A short version of the Inventory of Interpersonal Problems. *British Journal of Clinical Psychology* 35: 21-35.
- Barkham, M., Margison, F., Leach, C., Lucock, M., Benson, E., Mellor-Clark, J., Evans, C., Connell, J., Audin, K., & McGrath, G. 2001. Service profiling and outcomes benchmarking using the CORE-OM: Towards practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology* 69: 184-196.
- Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. 1988. An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology* 56: 893-897.
- Beck, A.T., Steer, R.A., & Brown, G.K. 1996. *Manual for the Beck Depression Inventory – Second Edition (BDI-II)*. San Antonio, TX, The Psychological Corporation.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. 1961. An inventory for measuring depression. *Archives of General Psychiatry* 4: 561-571.

Carlson, E.B., & Putman, F.W. 1993. An update on the Dissociative Experiences Scale. *Dissociation* 6: 16-27.

Department of Health. 1999. National Service Framework for Mental Health. London: HMSO.

Department of Health. 2004. Organising and Delivering Psychological Therapies.. London: HMSO.

Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. 2002. Towards a standardised brief outcome measure: psychometric properties and utility. *British Journal of Psychiatry* 180: 51-60.

Horowitz, L. M., Rosenberg, S. E., Baer, B. A., Ureno, G., & Villasenor, V. S. 1988. Inventory of Interpersonal Problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology* 56: 885-892.

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.

Leach, C., Lucock, M.P., Barkham, M., Noble, R., & Iveson, S (submitted). Transforming between Beck Depression Inventory and CORE-OM scores in routine practice.

Leary, T. 1957. *Interpersonal Diagnosis of Personality*. New York, Ronald.

Lucock, M.P., Iveson, S.K., & Leach, C. 1999. Researching a psychological therapies service in the new NHS. *Clinical Psychology Forum* 132: 18-21.

Lucock, M., Leach, C., Iveson, S., Lynch, K., Horsefield, C., & Hall, P. 2003. A systematic approach to practice-based evidence in a psychological therapies service. *Clinical Psychology and Psychotherapy* 10: 389-399.

Lutz, W., Leach, C., Barkham, M., Lucock, M.P., Stiles, W.B., Evans, C., Noble, R., & Iveson, S. (in press) Predicting rate and shape of change for individual clients receiving psychological therapy: Using growth curve modelling and nearest neighbor technologies. *Journal of Consulting & Clinical Psychology*

Parry, G. 1992. Improving psychotherapy services: Applications of research, audit and evaluation. *British Journal of Clinical Psychology*, 31, 3-19.

Roth, A., & Fonagy, P. 1996. *What works for whom? A critical review of Psychotherapy Research*. London: Guilford.

Stiles, W., Leach, C., Barkham, M., Lucock, M., Iveson, S., Shapiro, D., Iveson, M., & Hardy, G. 2003. Early sudden gains in psychotherapy in routine service conditions: Practice based evidence. *Journal of Consulting and Clinical Psychology* 71: 14-21.