CONTENTS

SOUTH LONDON AND MAUDSLEY NHS TRUST

INTRODUCTION

This booklet has been designed to provide guidance to clinicians in administering, scoring and interpreting the Trust approved outcome measures for psychological therapies. This is part of the work in the Trust on the development of routine outcome measurement in clinical practice and the wider Patient Journey programme.

The chief purposes of establishing routine outcomes measurement in SL&M are: to evaluate the extent to which patients improve with treatment; to facilitate the planning of ongoing treatment; to support service development; and to provide feedback for clinicians that will enhance reflective practice. The measures listed below have been approved by the Trust Psychological Therapies Committee (TPTC) for use in routine outcomes monitoring for psychological therapies. Guidance information is provided for each of these measures:

Disorder	**Recommended Measure**
Generic (non-problem specific)	CORE-OM
Addictions	Maudsley Addiction Profile (MAP)
Agoraphobia	Mobility Inventory (MI)
Bipolar Disorder	Internal State Scale (ISS)
Depression	Centre for Epidemiological Studies Depression Scale (CES-D)
Eating Disorders	Eating Disorder Examination (EDE-Q)
Family/Relationship Problems/ Personality Disorder	Inventory of Interpersonal Problems (IIP – 127 items version)
Generalised Anxiety Disorder	Penn State Worry Questionnaire (PSWQ)
Obsessive Compulsive Disorder	Obsessive Compulsive Inventory (OCI)
Panic Disorder	Panic Rating Scale (PRS)
Psychosis/Schizophrenia	Psychotic Symptom Rating Scales (PSYRATS)
PTSD	Impact of Events Scale – revised (IES-R)
Social Phobia	Liebowitz Social Anxiety Scale (LSAS)
	Social Summary Rating Scale (SSRS)
Chronic Fatigue	Chalder Fatigue Questionnaire (CFQ)

All of the approved measures are patient self-report except for the MAP and the PSYRATS, which are completed by a clinician, who can be the therapist. Services are required to use one or more of these measures. However, this does not preclude either individual clinicians or services employing additional measures of their own choosing for clinical, service related or research purposes. If a service considers that none of the approved measures meets their needs, application can be made for the addition of an alternative appropriately standardised measure. Such applications should be made to the Trust Psychological Therapies Committee.

The frequency of administration during therapy should be decided, at the beginning, according to the length of the measure and the duration of therapy. Thus, it would be appropriate to give some measures every time the patient is seen, for others it would be more realistic to ask the patients to complete one every third or fourth session and for therapy with a long duration it may be more appropriate to give the measures every two months. A key priority in deciding how often measures should be administered is to increase the number of patients for whom end point data are available, including those who withdraw from therapy.

Copies of both the approved measures and the guidance will be available for downloading and printing from the Trust Intranet within the Electronic Patient Journey System. It is intended that the guidance in this booklet will be updated and revised periodicallly. Any feedback on the content and usefulness of this guidance will be gratefully received.

Philippa Garety
Trust Psychological Therapies Committee December 2003, updated June 2005

CLINICAL OUTCOMES IN ROUTINE EVALUATION-OUTCOME MEASURE (CORE-OM)

Background Information

The CORE-OM (Evans et al, 2000) is a generic self-report measure of global distress, including subjective well-being, commonly experienced problems or symptoms, social/life functioning and risk to self and others. The measure is suitable for use as an initial screening tool and for assessing response to psychological therapy across a wide range of service types. The CORE-OM is sensitive to change and has high internal and test-retest reliability (Evans et al 2000).

Description of the CORE-OM

The CORE-OM is comprising 34 items designed to measure global distress. **Eight of the items (3, 4, 7, 12, 19, 21, 31, 32) are positively framed** e.g.'I have felt optimistic about my future'. The 34 items can be separated into the following four subscales, each subscale being a dimension of global distress: Subjective Well Being (4 items), Problems/Symptoms (12 items), Functioning (including general functioning, functioning in close relationships and functioning in social relationships; 12 items) and Risk (including harm to self and harm to others; 6 items). The dimension to which an item belongs is indicated on the measure by W (Subjective Well Being), P (Problems/Symptoms), F (Functioning) or R (Risk).

Scoring the CORE-OM

Each of the 34 items is scored on a 5-point scale ranging from 0-4 (0 = not at all, 4 = most of the time). ***Scores are reversed for the positively framed items.*** Total and mean scores can be calculated for all items (global distress) and for each of the four dimensions, as follows:

Global Distress: Adding all item scores gives a score of global distress ranging from 0-136. A mean of the total score can then be calculated and will range from 0-4. If only 1 or 2 items are not complete, mean scores can be calculated by dividing the total score by the number of complete items.

Well-Being Dimension Score: Adding the 4 'W' item scores gives a total Well-Being score ranging from 0-16. A mean of the Well-Being score can then be calculated and will range from 0-4. If only 1 item is incomplete, a mean score can still be calculated.

NB The measure is problem scored: the higher the score, the more problems/distress the individual is reporting. Although this may seem counter-intuitive with regards to 'Well-Being', scoring is kept this way for consistency with the other dimensions.

Problems/Symptoms Dimension Score: Adding the 12 'P' item scores gives a total Problems/Symptoms score ranging from 0-48. A mean of the Problems/Symptoms score can then be calculated and will range from 0-4. If only 1 item is incomplete, a mean score can still be calculated.

Functioning: Adding the 12 'F' item scores gives a total Functioning score ranging from 0-48. A mean of the Functioning score can then be calculated and will range from 0-4. If only 1 item is incomplete, a mean score can still be calculated.

Risk: Adding the 6 'R' item scores gives a total Risk score ranging from 0-24. A mean of the Risk score can then be calculated and will range from 0-4. If only 1 item is incomplete, a mean score can still be calculated.

NB 'Risk' items are intended to be used as clinical indicators of risk rather than as a scale. As such, a Global Distress Score and mean can be calculated without the risk items as follows:

Global Distress score minus Risk: Adding all 28 'W', 'P' and 'F' item scores gives a score of global distress minus risk ranging from 0-112. A mean of this score can then be calculated and will range from 0-4. If only 1 or 2 items are not complete, mean scores can be calculated by dividing the total score by the number of complete 'W', 'P' and 'F' items.

Interpretation of Scores

Normative data: The Core System User Manual gives mean scores (ranging from 0-4) for clinical and non clinical populations. These are shown in the table below:

Dimension	Mean Score (SD) clinical population[1]	Mean Score (SD) non-clinical population[2]
Subjective Well Being (W)	2.37 (0.96)	0.91 (0.83)
Problems or Symptoms (P)	2.31 (0.88)	0.90 (0.72)
Functioning (F)	1.86 (0.84)	0.85 (0.65)
Risk (R)	0.63 (0.75)	0.20 (0.45)
Total minus Risk	2.12 (0.81)	0.88 (0.66)
Global Distress (Total)	1.86 (0.75)	0.76 (0.59)

[1] Clinical population n=890 from 19 NHS sites, 1 university counselling service and 1 staff support service
[2] Non-clinical population n=1106

For all dimensions, the mean scores for clinical and non-clinical populations significantly differ ($p < 0.0005$)

Cut off scores: Clinical cut off mean scores for Males and Females are given as follows:

Dimension	Males	Females
Subjective Well Being (W)	1.37	1.77
Problems or Symptoms (P)	1.44	1.62
Functioning (F)	1.29	1.30
Risk (R)	0.43	0.31
Total minus Risk	1.36	1.50
Global Distress (Total)	1.19	1.29

CORE-OM for elderly

Normative data are available for elderly people (Barkham, Culverwell, Spindler, Twigg & Connell, in press). The sample consisted of 118 elderly people (aged 65-97) presenting for mental health treatment and 214 people from a non-clinical sample (aged 65-94). Results showed that the CORE-OM is a reliable measure in both clinical and non-clinical samples for the overall mean items. Large overall differences were observed between the clinical and non-clinical elderly sample, indicating that it is equally sensitive in the older age band as it is for the working age adults. However, the norms for the elderly clinical sample were lower than the equivalent clinical norms for a working age sample. Hence, the norms for elderly CORE-OM should be used for the older population.

	Clinical population		Non-clinical population	
Dimension	Male (N=37)	Female (N=61)	Male (N=47)	Female (N=137)
Mean global distress score (SD) of CORE-OM for elderly	1.50 (0.60)	1.55 (0.57)	0.54 (0.44)	0.63 (0.33)

The clinical cut-off for the elderly is shown below:

Clinical cut-off for elderly	Male 0.95	Female 0.97

Correlations with other measures: Barkham et al (2001) report that the full CORE-OM correlates most highly with measures of symptoms, such as the Symptoms Checklist-90-Revised (SCL-90-R) and the Brief Symptom Inventory (BSI), and measures of depression such as the Beck Depression Inventory I and II (BDI and BDI-II). For all these measures Spearmans Rho > 0.80.

The CORE-OM also correlates highly with the Beck Anxiety Inventory (BAI, Spearmans Rho = 0.65) and the Inventory of Interpersonal Problems-32 item version (IIP-32, Spearmans Rho = 0.65).

Key References

Barkham M., Culverwell A., Spindler K., Twigg E., & Connell J. (in press). The CORE-OM in an older population: Psychometric status, acceptability and feasibility. *Ageing & Mental Health*.

Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., Benson, L., Connell, J., Audin, K. and McGrath, G (2001). Service Profiling and Outcomes Benchmarking: Toward Practice-Based Evidence in the Psychological Therapies. *Journal of Consulting and Clinical Psychology Vol 69, No. 2, 184-196*.

Core System User Manual. Available at http://www.coreims.co.uk

Evans, C., Connell, Barkham, Margiso, F., Mellor-Clark, J., McGrath, G., G. & Audin, K., (2002). Towards a standardised brief outcome measure. Psychometric properties & utility of the CORE-OM. British Journal of Psychiatry 180, 54-60.

Evans, C., Mellor-Clark, J., Margison, F., Barkham, M., Audin, K., Connell, J. and McGrath, G (2000). CORE: Clinical Outcomes in Routine Evaluation. *Journal of Mental Health, 9, 3, 247-255*.

THE MAUDSLEY ADDICTION PROFILE (MAP)

Background Information

The Maudsley Addiction Profile (Marsden et al, 1998) is a brief multi-dimensional instrument designed to assess treatment outcome for people with drug and/or alcohol problems. The instrument measures substance use, health risk behaviour, physical and psychological health and personal and social functioning.

Description of the MAP

The MAP has five sections. Section A is an introductory section designed to collect data management information. The remaining four sections represent the following domains:

Section B (Substance Use): This section lists eight (plus one other) substances. For each substance information is collected about the number of days in the last 30 that the substance was used, the typical amount of substance used per day when using and the main routes of substance administration (except for alcohol). Substances can be added to the list where necessary.
Section C (Health Risk Behaviour): This section has five items, three relating to injected drugs and two relating to penetrative sex (vaginal or anal) without the use of a condom.
Section D (Health Symptoms): This section has 20 items. 10 items relate to physical health symptoms (physical health sub-scale) and 10 relate to psychological symptoms (psychological health sub-scale). Five of the psychological health symptoms are anxiety related and five are depression related.
Section E (Personal/Social Functioning): This section has nine items in three areas (relationship conflict, employment and criminal behaviour).

Scoring the MAP

Section B: For each substance, score the following:

a) The number of days in the last 30 that the client has used the substance. There is a card that can be used to calculate the number of days from the client's estimate (eg 2 days per week = 9 days).
b) The amount of substance used on a typical day when using. The client's response can be written down verbatim, then scored later (in units for alcohol and in estimated weight or money for other substances).
c) The route of administration (1 = oral, 2 = snort/sniff, 3 = smoke/chase, 4 = intravenous, 9 = intramuscular).

Section C: If the client has been injecting in the last 30 days score C1-C3 as follows; if the client has engaged in non-condom penetrative sex in the past 30 days score C4 and C5 as follows:

C1) The number of days that the client has injected drugs. There is a card that can be used to calculate the number of days from the client's estimate (eg 2 days per week = 9 days).
C2) The total number of injections per typical day when using.
C3) The total number of times that the client has used a needle or syringe that has been used by someone else.
C4) The number of people the client has had penetrative sex with and not used a condom.
C5) The total number of times the client has had penetrative sex and not used a condom.

Section D:

a) Physical health sub-scale: Each of the 10 physical health items should be scored from 0-4. Adding the scores of the 10 items gives a physical health score ranging from 0-40.
b) Psychological health sub-scale: Each of the 10 psychological health items should be scored from 0-4. Adding the scores of the 10 items gives a psychological health score ranging from 0-40.

Section E:

a) E1 to E6 are paired into three items that relate to a client's relationship conflicts. Each item pair has a question about the number of days in the last 30 that the client had contact with a particular person(s) and a question about the number of days that there was conflict with that person(s). The percentage of conflict days can then be calculated for each of the three items:
   - percentage of contact days where there was conflict in a significant relationship = E2/E1X100
   - percentage of contact days where there was conflict with relatives = E4/E3X100
   - percentage of contact days where there was conflict with friends = E6/E5X100

b) E7-E9 are items related to the client's employment over the last 30 days. E7 is the number of days the client has been in paid employment. E8 is the number of days missed through sickness or unauthorised absence and E9 is the number of days unemployed. Note: the scores for items E7-E9 should add to 30.
c) E10 relates to illegal activities. For each illegal activity listed (these can be added to if necessary) there are three items to be scored as follows:
   i. Whether the client has engaged in the illegal activity over the last 30 days (T = Yes, E = No).
   ii. The number of days the client has committed the illegal activity.
   iii. The number of times the client has committed the illegal activity on a typical day.

For each illegal activity, the number of crimes can be calculated by multiplying ii and iii. These scores can be added to give the total number of all crimes in the last 30 days.

Interpretation of Scores

Reliability: The 10 item physical health subscale has a good internal reliability ($\alpha = 0.77$). The five anxiety items and five depression items of the psychological health subscales also have good internal reliability ($\alpha = 0.88$ and $\alpha = 0.81$ respectively). Three day test-retest reliability for the frequency of substance use ranged from 0.92 (heroin) to 0.77 (crack cocaine) and from 0.95 (heroin) to 0.77 (crack cocaine) for the intensity of substance use. Test-retest reliability was also high for the usual route of administration (kappa = 0.93 for heroin an 1.0 for methodone, benzodiazepines and cocaine).
For measures of health risk, health problems, relationship conflicts, employment and crime, test-retest reliability was also high, averaging 0.81.

Correlations with other measures: The physical health symptoms sub-scale items are derived from the Opiate Treatment Index (Darke et al, 1991) and the psychological health symptoms sub-scale items are derived from the Brief Symptom Inventory (BSI, Derogatis, 1975). The scores on the physical health sub-scale and the depression and anxiety items of the psychological health sub-scale correlated with medical problems, depression and anxiety items, adapted from the Addiction Severity Index (ASI, McLellan et al, 1992a, average Pearson's r = 0.72). Scores on relationship conflict measures correlated with three subscales from the Life Stressors and Social Resources Inventory (LISRES, Moos, 1988a, average Pearson's r = 0.74).

Normative data: Means for community drug users and alcohol users (n = 80 and n = 40) and inpatient drug users and alcohol users (n = 80 and n = 40) are given in the Development and User Manual (Marsden et al, 1998). These means are given for the following domains/items: percentage of days of substance use, typical amount of substance used per day, pre-treatment scores on the physical and psychological health sub-scales and relationship conflict scales, percentage of days crimes were committed and number of offences.

It is strongly recommended that the Development and User Manual (Marsden et al, 1998) is referred to.

Key References

Marsden, J., Gossop, M., Stewart, D., Best, D., Farrell, M. and Strang, J. (1998) The Maudsley Addiction Profile. A brief instrument for treatment outcome research. Development and User Manual. National Addiction Centre/Institute of Psychiatry, London. Reprint requests: J.marsden@iop.kcl.ac.uk

Marsden, J., Gossop, M., Stewart, D., Best, D., Farrell, M., Lehmann, P., Edwards, C. and Strang, J. (1998). The Maudsley Addiction Profile (MAP): a brief instrument for testing outcome. *Addiction, 93 (12), 1857-1868.*

MOBILITY INVENTORY FOR AGORAPHOBIA (MI)

Background Information

The MI (Chambless et al, 1985) is a self-report measure of agoraphobic avoidance, both when accompanied and when alone, in a variety of places and situations where a patient might experience discomfort or anxiety. The measure has good sensitivity to change and might be of use as an administration or initial screening tool, although this use might be limited due to the length of the measure.

Description of the MI

This guidance describes Parts 1 and 2 of the most recent version of the MI. These parts are specific to agoraphobic avoidance. Part 3 of the recent version relates to panic and is not included here. Part 4 asks the patient to provide the location and size of his/her safety zone, if relevant, and is an optional addition. The original version of the MI was comprised of Part 1 and a single question regarding panic frequency in the last 7 days.

Part 1 of the MI lists 26 situations (plus one optional other) in which a patient might experience anxiety or discomfort. For each situation two ratings of avoidance are made, one for when the patient is accompanied and one for when the patient is alone. One exception is the 'staying at home alone' situation, where avoidance can only be rated for when the patient is alone and not for when the patient is accompanied. This results in two sub-scales, the avoidance when accompanied sub-scale (MI-AAC; 25 items plus 1 optional item) and the avoidance when alone sub-scale (MI-AAL; 26 items plus 1 optional item).

Part 2 of the MI asks the client to circle the 5 items from part one that cause the greatest concern or impairment.

Scoring the MI

Part 1. The two sub-scales of the MI are scored as follows:

MI-AAC: Each of the 25 (plus 1 other) situations that cause discomfort or anxiety is rated for avoidance when accompanied on a scale ranging from 1-5 (1 = Never Avoid and 5 = Always avoid). Half scores can be given for each item (e.g. 3½). If an item relates to a situation that is never encountered or is irrelevant to a patient's life (e.g. if a patient has never been in an aeroplane) it should be left blank. A mean avoidance when accompanied score, ranging from 1-5, can then be calculated.

MI-AAL: Each of the 26 (plus 1 other) situations that cause discomfort or anxiety is rated for avoidance when alone on a scale ranging from 1-5 (1 = Never Avoid and 5 = Always avoid). If an item relates to a situation that is never encountered or is irrelevant to a patient's life (e.g. if a patient has never been in an aeroplane) it should be left blank. Half scores can be given for each item (e.g. 3½). A mean avoidance when alone score, ranging from 1-5, can then be calculated.

Part 2. Part 2 of the MI is not scored and should be used to provide additional individual patient information.

Interpretation of Scores

Reliability and validity: Both MI sub-scales have been found to significantly discriminate between patients with and without agoraphobia, between patients with agoraphobia and patients with social phobia and between patients with agoraphobia and non-clinical populations.

Internal consistency is excellent for both sub-scales, with Cronbach's Alpha ranging from 0.91-0.97 for the MI-AAC sub-scale and from 0.94-0.96 for the MI-AAL sub-scale. 31 day test-retest reliability has been reported as ranging from r = 0.75 to r = 0.86 for the MI-AAC sub-scale and from r = 0.89 to r = 0.90 for the MI-AAL sub-scale.

Normative data and cut off scores: Chambless et al (1985) report mean scores on the two MI sub-scales for two agoraphobic samples, a socially phobic sample and a non-patient control group. These means are given in the table below.

Sample	n	Mean score MI-AAC (SD)	Mean score MI-AAL (SD)
Agoraphobia sample 1	94	2.64 (0.90)[a]	3.35 (1.06)
2	83	2.41 (0.72)	3.30 (1.99)
Social Phobia sample	18	1.35 (0.27)	1.56 (0.41)
Controls	23	1.07 (0.08)	1.25 (0.24)

[a] Missing data for 1 subject, therefore n = 93.

Correlations with other measures: The MI-AAC has been found to correlate most highly with the agoraphobia sub-scale of the Fear Questionnaire, ranging from r = 0.44 to r = 0.63 and the MI-AAL has also been found to correlate most highly with this sub-scale of the Fear Questionnaire, with r ranging from 0.68-0.84. Both the MI-AAC and the MI-AAL have also been found to correlate with the social phobia subscale of the Fear Questionnaire (r = 0.34 and r = 0.38 respectively) and the blood/injury subscale of the Fear Questionnaire (r = 0.34 and r = 0.45 respectively).

The two MI sub-scales have also been found to correlate with the following measures: The Trait Scale of the State Trait Anxiety Inventory (r = 0.25 for the MI-AAC and r = 0.38 for the MI-AAL), the Anxiety Sensitivity Index (r = 0.39 for the MI-AAC and r = 0.42 for the MI-AAL) and the Beck Depression Inventory (r = 0.44 for the MI-AAC and r = 0.51 for the MI-AAL).

Key Reference

Chambless, D.L., Caputo, G.C., Jasin, S.E., Gracely, E.J. and Williams, C (1985). The Mobility Inventory for Agoraphobia. *Behaviour Research and Therapy, 23, 35-44*.

INTERNAL STATE SCALE (ISS)

Background Information

The Internal State Scale (ISS, Bauer et al, 1991; Glick et al. 2003) is a self-report measure that simultaneously rates manic and depressive symptoms. It is particularly useful in the assessment of patients with bipolar disorder. Based upon previous literature (Goodwin and Jamison, 1990), the ISS assesses heightened sense of activation as a core component of mania, depressed/dysphoric mood and euphoric mood states.

The drawback of the original ISS is the labour-intensive scoring of the visual analogue (VAS) format. Glick, McBride & Bauer (2003) developed a Likert version of the ISS. Each item in the ISS is divided into ten 10-unit bins (0 to 100). The new ISS was compared with the original ISS using data from the original validation sample of 86 bipolar patients and a new sample of 24 bipolar patients. Results showed that the Likert-based scoring of the ISS can be used without loss of precision when compared with the original ISS. The ISS Likert version makes automated scoring of the scale feasible.

Description of the ISS

The ISS has 16 items. Each item score will range from 0-100, in 10-unit bins. The first bin is scored as zero, the second as 10, and so on to the eleventh bin which is scored as 100. Sub-scale scores are calculated using the first 15 items. Item 16 which is used as an independent indicator of the patient's current mood state is anchored at one end by 0 (depressed and down), at the other end by 100 (manic and high), with the centre point representing a normal mood state. For each item, the patient is asked to choose a ten-unit bin based upon how they have felt in the last 24 hours. Only the zero bin (not at all/rarely) and the 100 bin (very much so/much of the time) are labelled. The first 15 items of the ISS are divided into four sub-scales: the Depression Index (DI; 2 items), the Well-Being Index (WB; 3 items), the Activation Index (ACT; 5 items) and the Perceived Conflict Index (PC; 5 items).

Depression Index (DI) score: Scores for items 7 and 9 are added to give a DI score ranging from 0-200.
Well-Being Index (WB) score: Scores for items 3, 5 and 15 are added to give a WB score ranging form 0-300.
Activation Index (ACT) score: Scores for items 6, 8, 10, 12 and 13 are added to give an ACT score ranging from 0-500.
Perceived Conflict Index (PC) score: Scores for items 1, 2, 4, 11 and 14 are added to give a PC score ranging from 0-500.

Interpretation of Scores

Reliability and validity: The initial study by Bauer et al (1991) demonstrated that the ISS discriminated between patients with mania/hypomania (who met the criteria for bipolar disorder), patients with depression (some of whom met the criteria for bipolar disorder and some of whom were classified as having recurrent major depression) and not-psychiatrically ill controls. Bauer et al (1991) found each sub-scale of the ISS to have good internal consistency (Cronbach's Alpha= 0.92 for the DI, 0.87 for the WB, 0.84 for the ACT and 0.81 for the PC). The ISS sub-scales also demonstrated good test-retest reliability for controls but not for patients whose mood state changed over time.

Comparing the original ISS with the Likert-format ISS, the within-subjects reliabilility was uniformly high across the two formats. The Likert-format ISS also showed equally good discriminant ability including kappa (Likert-form ISS VS DSM IV criteria) and receiver operator characteristic curves.

Normative data and cut off scores: Bauer et al (1991), present graphical comparison of the scores on each sub-scale for manic, depressed and control groups. However, exact mean scores are not given.

ISS Subscales as discriminators of mood epiosdes (Bauer et al. 2000)

The Well Being subscale used in conjunction with the Activation subscale is useful in discriminating between depressed, (hypo)manic, and subsyndromal/euthymic states. In this capacity, the ISS has proven

valid by discriminant function analysis, although the exact cut-off scores may vary somewhat from site to site and therefore should be standardized by each investigator. The revised algorithm for mood state discrimination (Glick et al, 2003) is as follows:

Mood State	Activation Subscale Score	Well-Being Subscale Score
(Hypo)Mania	>155	>125
Mixed State	>155	<125
Euthymia	<155	>125
Depression	<155	<125

Correlations with other measures: The ACT sub-scale of the ISS has been found to correlate with the Young Mania Rating Scale (YMRS), in the Bauer et al (1991) full sample of patients with mania, depression and controls combined (r = 0.60) and also for the mania patients alone, minus those who were severely ill with thought disorder/ delusions (r = 0.55). These particular patients were excluded because they tended to deny having any symptoms. The DI sub-scale was found to correlate with the Hamilton Rating Scale for Depression (Ham-D) in the full Bauer et al (1991) sample (r = 0.84) and also for the depression subjects alone (r = 0.73). The WB sub-scale of the ISS was also found to correlate with the Ham-D in the full sample (r = 0.73) but not for the depressed sub-group of patients alone (r = 0.20). The authors indicate that this latter finding is due to the lack of spread in scores for this sub-group: these patients' WB scores were all grouped together at the lower end of the scale.

Cooke et al (1996) found that the combined WB + ACT sub-scale score correlated significantly with the full SRMI (r = 0.58) in a sample of 20 patients with rapid-cycling bipolar disorder and also with the following sub-scales of the Self-Report Mania Inventory: energy/activity (r = 0.52), verbosity (r = 0.54), elation (r = 0.50), racing thoughts/concentration (r = 0.48). This WB+ACT combined sub-scale score was also found to correlate with the YMRS in this group of rapid-cycling bipolar disorder patients (r = 0.44).

Key References

Bauer, M.S., Crits-Cristoph, P.C., Ball, W.A., Dewees, E., McAllister, T., Alahi, P., Cacciola, J. and Whybrow, P.C. (1991). Independent Assessment of Manic and Depressive Symptoms by Self-Rating. Scale characteristics and Implications for the Study of Mania. *Archives of General Psychiatry, 48: 807-812.*

Bauer M, Vojta C, Kinosian B, Altshuler L, Glick H. The Internal State Scale: replication of its discriminating abilities in a multi-site, public sector sample. Bipolar Disorders 2000; 2: 340-346. .

Cooke, R.G., Krüger, S. and Shugar, G. (1996). Comparitive Evaluation of Two Self-Report Mania Rating Scales. *Biological Psychiatry, 40: 279-283.*

Glick HA, McBride L & Bauer MS. (2003). A manic-depressive symptom self-report in optical scanable format. *Bipolar Disorders,* 5, 366-369.

CENTRE FOR EPIDEMIOLOGICAL STUDIES DEPRESSION SCALE (CES-D)

Background Information

The CES-D (Radloff, 1977) is a self-report measure of depressive symptoms including depressed mood, feelings of worthlessness and hopelessness, loss of appetite, poor concentration and sleep disturbance. The scale is effective in measuring current levels of depressive symptoms within various populations, measuring change in depressive symptom severity over time and as an initial screening tool for depressive illness, providing appropriate follow up assessments are made.

Description of the CES-D

The CES-D has 20 items. Clients are asked to rate each item on a scale ranging from 0-3 (0 = rarely or none of the time, 3 = most or all of the time). Four of the CES-D items are positively framed. A higher score on the CES-D indicates greater depressive symptomatology.

Scoring the CES-D

Each of the items on the CES-D is scored between 0-3. **The scores for the four positively framed items are reversed (items 4, 8, 12, and 16).** Item scores are then added to give a total score ranging from 0-60.

Interpretation of Scores

Reliability and validity: Radloff (1977) reported that psychiatric patients scored significantly higher on the CES-D than non-patients. Similarly, Weissman et al, (1977) found the CES-D to discriminate between subjects with acute depression and subjects in a community sample. In this study it was also found that depressed subjects in other psychiatric groups had higher mean scores on the CES-D than the non-depressed patients in the same psychiatric groups and approaching that of the patients with acute primary depression (Mean = 38.10, SD = 9.01). The non-depressed patients in the other psychiatric groups had mean scores that were comparable to the community sample (Mean = 9.10, SD = 8.60). Moreover, Weissman et al (1977) found that for a clinical sample, the CES-D was sensitive to change with treatment, reporting that the CES-D discriminated between patients who clinicians rated as 'recovered' following treatment with medication and those who clinicians rated as 'partially responded' or 'not responded' to the treatment: mean change in CES-D score was 20 points for the patients who were deemed to have recovered and 12 points for those deemed to have partially or not responded to treatment. The CES-D has also been found to discriminate between DSM-III-R categories of mild and severe depression, but not between mild and moderate or moderate and severe depression (Fechner-Bates et al, 1994).

The CES-D has demonstrated high internal consistency in a number of different populations with alpha approximating 0.85 in community samples and 0.90 in clinical samples. Noting that only moderate test-retest reliability is to be expected for measures of symptoms that would usually change over time, test-retest reliability has been reported to range from r = 0.51-0.67 for periods ranging from 2-8 weeks and r = 0.32-0.54 for periods ranging from 3 to 12 months.

Normative data and cut off scores: Early studies identified a score of 16 or higher to indicate depressive illness. Radloff (1977) reported normative data for three all-White U.S. community samples and a clinical sample. These data are shown in the table below:

Sample	n	Mean	SD	% scoring 16
Community sample 1	2514	9.25	8.58	19
2	1060	8.17	8.23	15
3	1422	7.94	7.53	15
Clinical sample	70	24.42	13.51	70

As such, there is a body of literature investigating the sensitivity and specificity of the CES-D at predicting depressive illness using ≥ 16 as a cut off for determining cases:

- Weissman et al (1977) found that by utilising a cut off score of ≥16, sensitivity was 0.99 for acute primary depression, 0.94 for detecting depression in a sample with alcohol dependence, 0.93 for detecting depression in a sample with schizophrenia and 0.74 for detecting depression in a sample with drug dependence. However, low specificity was found for patients with remitted depression (0.56) and those with drug dependence (0.59).
- Boyd et al (1982) found that in a community sample (n = 720), sensitivity for a cut off score of ≥ 16 was 0.64 and specificity 0.94 for detecting major depression as determined by Research Diagnostic Criteria.
- Breslau (1985) found that in utilising a cut off score of ≥ 16, CES-D sensitivity was 0.75 for detecting DSM-III major depression in subjects without co-morbid DSM-III generalised anxiety and 0.67 for detecting DSM-III generalised anxiety in subjects without co-morbid DSM-III major depression. The subjects were mothers of children with chronic disabilities.
- Roberts and Vernon (1983) found that in a community sample (n = 528), sensitivity for a cut off score of ≥ 16 was 0.60 and specificity was 0.84 for detecting major depression as determined by Research Diagnostic Criteria. Similarly, Myers and Weissman found sensitivity at detecting depression to be 0.64 and specificity to be 0.94 in a community sample (n = 515).
- Fechner-Bates et al (1994) found that CES-D scores ≥16 were significantly related to DSM-III-R major depression diagnoses but also significantly related to other DSM-III-R Axis 1 diagnoses in 425 primary medical care patients.

Overall, the literature indicates the CES-D should be considered as a measure of depressive symptoms in a variety of disorders or even a general measure of distress rather than a diagnostic tool for depression. Indeed, the CES-D author notes that the measure was not developed as a diagnostic tool and as such individual scores should not be interpreted as indicting disorder by use of cut off scores. In this respect, the CES-D is similar to other self-report depression scales and has comparable validity data to the Beck Depression Inventory (BDI) and the Zung Self-Rating Depression Scale (Zung SDS).

Correlations with other measures: Note: The CES-D items were selected from the following sources: the Zung SDS, BDI, Raskin Scale, a depression checklist developed by EA Gardner (unpublished manuscript) and the Minnesota Multiphasic Personality Inventory Depression Scale (MMPI-D).

The CES-D has been found to correlate with the Sypmtom Checklist-90 (SCL-90), ranging from r = 0.73-0.89, in patients with depression, alcohol dependence, drug addiction and schizophrenia. Correlations have also been found between the CES-D and the Hamilton Depression Rating Scale (HAM-D), ranging from r = 0.49 for patients with acute depression to r = 0.85 for patients with schizophrenia. In addition the CES-D has been found to correlate with the Raskin Scale (ranging from r = 0.28 for patients with acute depression to r = 0.79 for patients with schizophrenia), the Zung SDS (r = 0.69 in a sample of elderly patients), the Bradburn Negative Affect Scale (r = 0.55 in a patient sample and r = 0.60-0.63 in non-patient samples), the Bradburn Balance (r = 0.61-0.62 in non-patient samples and r = 0.72 in a patient sample), the Lubin scale (r = 0.37-0.51 in non-patient samples and r = 0.70 in a patient sample), The Langer scale (r = 0.54 in a non-patient sample and r = 0.60 in a patient sample), Cantril Ladder (r = 0.38-0.44 in non-patient samples and r = 0.74 in a patient sample) and to negatively correlate with the Bradburn Positive Affect Scale (r = -0.21-0.25 in non-patient samples and r = -0.55 in a patient sample).

Key References

Radloff, L.S. (1977). The CES-D Scale: A Self-report depression scale for research in the general population. *Applied Psychological Measurement, 1, 385-401.*

Weissman, M.M., Sholomskas, D., Pottenger, M., Prusoff, B.A., and Locke, B.Z. (1977). Assessing depressive symptoms in five psychiatric populations: a validation study. *American Journal of Epidemiology, 106:3, 203-214.*

EATING DISORDERS EXAMINATION-QUESTIONNAIRE VERSION (EDE-Q)

Background Information

The EDE-Q (Fairburn and Beglin, 1994) is a self report version of the well-established investigator based interview, the Eating Disorders Examination (EDE; Cooper and Fairburn, 1987, Fairburn and Cooper, 1993). The measure was designed to assess the specific psychopathology of eating disorders and is useful in determining the effects of treatment. The authors have kindly agreed to the use of this scale for routine outcome monitoring in SL&M.

Description of the EDE-Q

The EDE-Q is comprised of 36 items. Fourteen of the items (items 8 and 16-28) relate specifically to the occurrence and frequency of key behavioural features of eating disorders, including binge eating, self-induced vomiting, laxative misuse and diuretic misuse. Item 8 is scored on a scale of 0-6 (0 = No days, 6 = Every day). For items 16-28 clients are either asked to state whether the behaviour occurred over the past 28 days (0 = No, 1 = Yes) or to record the number of episodes of the particular behaviour. Items 1-15 are rated for *frequency* of occurrence in the past 28 days on a scale of 0-6 (0 = No days, 6 = Every day). An exception to this is item 15 where 0 = None of the times and 6 = Every time. Items 29-36 are rated for *severity* over the past 28 days on a scale of 0-6 (0 = Not at all, 6 = Markedly). Together, items 1-15 and items 29-36 generate four sub-scales: Restraint (5 items), Eating Concern (5 items), Weight Concern (5 items) and Shape Concern (8 items). Item 11 belongs to both the Weight concern and Shape Concern sub-scales. Higher EDE-Q scores indicate greater severity of eating disorder psychopathology.

Scoring the EDE-Q *(Note: The client's weight and height should be recorded alongside the EDE-Q data.)*

The items relating to key behavioural features of eating disorders (items 8 and 16-28) do not belong to any of the sub-scales. These items are recorded individually and used as indicators of the presence and frequency of the key behaviours. Sub-scale and Global EDE-Q scores are calculated as follows:

Restraint sub-scale: A Restraint sub-scale score can be obtained by calculating a mean of the scores for items 1, 2, 3, 4, and 5. The resultant score should range from 0-6.

Eating Concern sub-scale: An Eating Concern sub-scale score can be obtained by calculating a mean of the scores for items 6, 7, 9, 15 and 34. The resultant score should range from 0-6.

Weight Concern sub-scale: A Weight Concern sub-scale score can be obtained by calculating a mean of the scores for items 11, 14, 29, 31 and 32. The resultant score should range from 0-6.

Shape Concern sub-scale: A Shape Concern sub-scale score can be obtained by calculating a mean of the scores for items 10, 11, 12, 13, 30, 33, 35 and 36. The resultant score should range from 0-6.

Global EDE-Q score: A Global EDE-Q score can be obtained by calculating a mean of the four sub-scale scores. The resultant Global score should range from 0-6. The Global EDE-Q score provides a measure of the severity of the eating disorder psychopathology. The authors recommend that the Global score should always be reported alongside data from the sub-scales and key behavioural items.

Interpretation of Scores

Reliability: Luce and Crowther (1999) provide EDE-Q reliability data, based on a sample of 139 female undergraduates. Internal consistency of the four sub-scales was high with Cronbach's Alpha ranging from 0.78 for the Eating Concern sub-scale to 0.93 for the Shape Concern sub-scale. Two week test-retest reliability was also high for all four sub-scales, ranging from r = 0.81 for the Restraint sub-scale to r = 0.94 for the Shape Concern sub-scale. The key behavioural features of eating disorders were less stable over time in this group: phi co-efficients for items relating to the occurrence of key behaviours ranged from 0.57 to 0.70. Pearson correlation co-efficients for items relating to the frequency of these behaviours ranged from 0.54 to 0.92.

Normative data and cut off scores: Normative data on the EDE-Q sub-scales have been provided in three key studies (Wilfley et al, 1997; Carter et al, 2001 and Passi et al, 2003). The Passi et al (2003) study gives data for the same group on two occasions.

The data from all three studies are given in the table below.

	Binge Eating Disorder Sample (n = 52)[1]	Control group of UK Schoolgirls (n = 808)[2]	Anorexia Nervosa Sample at Time 1 (n = 28)[3]	Anorexia Nervosa Sample at Time 2 (n = 28)[3]
Restraint	2.5 (1.5)	1.4 (1.5)	3.1 (1.9)	3.0 (1.9)
Eating Concern	3.4 (1.4)	1.0 (1.0)	2.2 (1.7)	1.8 (1.4)
Weight Concern	4.1 (1.1)	1.8 (1.7)	2.6 (1.7)	2.2 (1.8)
Shape Concern	4.8 (1.1)	2.2 (1.7)	3.4 (1.9)	3.0 (2.6)

[1] Wilfley et al, 1997; N = 6 Males and n = 46 females; Mean age= 45.4 years (SD = 9.1).

[2] Carter et al, 2001; All female; Mean age = 13.4 years (SD=0.5, range=12-14 years); Items rated based on a 14 day period rather than a 28 day period and question wording simplified due to age of subjects.

[3] Passi et al, 2003; All female; Mean age = 15.8 years (SD=1.5). Time two data: patients completed the EDE-Q for a second time. The interview version of the EDE was administered between the two questionnaire versions.

Correlations with other measures: Several studies have compared the EDE-Q with the interview version of the EDE. It is assumed that the clinician administered EDE interview is the 'gold-standard'. Fairburn and Beglin (1994) compared sub-scale scores on the EDE and the EDE-Q in a clinic sample of women with eating disorders and a female general population sample.Significant correlations were found for all sub-scales, ranging from r = 0.78 to r = 0.85 in the patient group and from r = 0.79 to r = 0.81 in the general population group. Black and Wilson (1996) reported similar results in a sample of patients with substance use disorders: sub-scale correlations between the EDE and the EDE-Q ranged from r = 0.75 to r = 0.85. In both the Fairburn and Beglin (1994) study and the Black and Wilson (1996) study, the Eating Concern sub-scale was not assessed. In a later study, Wilfley et al (1997) compared sub-scale scores on the EDE and EDE-Q, for all four sub-scales, in a sample of patients with Binge Eating Disorder. Correlations ranged from r = 0.63 to r = 0.69. The most recent study comparing the EDE and the EDE-Q sub-scales (Passi et al, 2003) found correlations ranging from r = 0.70 to r = 0.91. All of the studies above found that the EDE-Q over-estimated the EDE on each sub-scale and also found EDE and EDE-Q sub-scale scores to be significantly different. Passi et al (2003) suggest that more accurate EDE-Q scores can be obtained by providing additional information and explaining difficult concepts relating to eating disorders. As such, it has been suggested that providing subjects with additional written information and explanations and using a calendar to orient subjects to the key events over the past 28 days, could improve the accuracy of EDE-Q scores.

With regards to key behavioural features of eating disorders, Black and Wilson found significant correlations between the EDE-Q and the EDE for patients with substance use disorders (Kendall's tau-b correlation co-efficient ranging from 0.49 to 1.00 for self-induced vomiting, laxative abuse, objective bulimic episodes and strict dieting). However, Wilfley et al (1997) found that the EDE-Q lacked consistency with the EDE in measuring binge eating in a sample of patients with Binge Eating Disorder (Kendalls tau-b = 0.20, not significant). Fairburn and Beglin (1994) found agreement between the EDE-Q and the EDE for items relating to self-induced vomiting, laxative misuse and binge eating for their eating disorder patient sample and their general population sample. However, in both samples, agreement for binge eating was lower than that for the other key behaviours of eating disorders. Overall, it has been noted that the EDE-Q is an adequate substitute for the EDE for most eating disorder features, including Restraint, Weight Concern and Eating Concern. However, for complex features such as binge eating and Shape Concern, the EDE-Q is less consistent with the EDE.

Key References

Fairburn, C.G. and Beglin, S.J. (1994). The assessment of eating disorders: Interview or self-report questionnaire? *International Journal of Eating Disorders, 16: 363-370.*

Luce, K.H., and Crowther, J.H. (1999). The reliability of the Eating Disorder Examination – self-report questionnaire version (EDE-Q). *International Journal of Eating Disorders, 25:349-351.*

Passi, V.A., Bryson, S.W., Lock, J. (2003). Assessment of Eating Disorders in Adolescents with Anorexia Nervosa: Self-Report Questionnaire Versus Interview. *International Journal of Eating Disorders, 33: 45-54.*

THE INVENTORY OF INTERPERSONAL PROBLEMS (IIP-127 item version)

Background Information

The original 127 item version of the Inventory of Interpersonal problems (IIP; Horowitz et al, 1988) is a self report measure of distress arising from interpersonal sources, including distress relating to things that a patient might find hard to do (e.g. 'really care about another person's problems') and things that a patient might do too much (e.g. 'trust other people too much'). The IIP is useful in measuring change in a variety of interpersonal areas for psychiatric patients; and may be of particular use for patients with personality disorder.

There are several other versions of the IIP, including a 64-item version, a 32-item version and versions designed to specifically identify personality disorder (IIP-PD and IIP-C). The full 127-item IIP is recommended for routine outcome measurement within SL&M. Further information on the IIP-P and IIP-C can be found in Pilkonis et al (1996) and Scarpa et al (1999).

Description of the IIP

The original IIP has 127 items. The first 78 items relate to things that a patient might find it hard to do and the remaining 49 items relate to things that a patient might do too much. The IIP items are divided into 6 subscales[1]: Assertive (21 Items), Sociable (18 items), Intimate (12 items), Submissive (10 items), Responsible (12 items) and Controlling (10 items). The remaining 44 items do not belong to any sub-scale but are used to calculate total IIP scores.

Scoring the IIP

Each item is scored on a 5-point scale ranging from 0-4 (0 = not at all distressed and 4 = extremely distressed). Mean scores can be calculated for the full IIP and for each of the sub-scales, as follows:

Total IIP score: Adding all items gives a total IIP score ranging from 0-508. A mean of the score is then calculated and will range from 0-4. If items are missing a prorated score is calculated by obtaining the average of the completed items.

Assertive score: Adding items 2, 5, 6, 8, 9, 13, 14, 16, 17, 20, 22, 32, 33, 36, 41, 68, 69, 74, 83, 120 and 121 gives an Assertive sub-scale score ranging from 0-84. A mean of the score is then calculated and will range from 0-4. If items are missing a prorated score is calculated by obtaining the average of the completed Assertive sub-scale items.

Sociable score: Adding items 1, 3, 7, 10, 23, 27, 42, 67, 71, 78, 80, 87, 99, 105, 107, 18, 124, and 125 gives a Sociable sub-scale score ranging from 0-72. A mean of the score is then calculated and will range from 0-4. If items are missing a prorated score is calculated by obtaining the average of the completed Sociable subscale items.

Intimate score: Adding items 12, 15, 29, 34, 37, 38, 39, 40, 57, 61, 65, and 96 gives an Intimate sub-scale score ranging from 0-48. A mean of the score is then calculated and will range from 0-4. If items are missing a prorated score is calculated by obtaining the average of the completed Intimate sub-scale items.

Submissive score: Adding items 18, 19, 28, 50, 52, 64, 82, 112, 116 and 127 gives a Submissive sub-scale score ranging from 0-40. A mean of the score is then calculated and will range from 0-4. If items are missing a prorated score is calculated by obtaining the average of the completed Submissive sub-scale items.

Responsible score: Adding items 24, 60, 73, 75, 81, 93, 101, 110, 114, 119, 122 and 126 gives a Responsible sub-scale score ranging from 0-48. A mean of the score is then calculated and will range from 0-4. If items are missing a prorated score is calculated by obtaining the average of the completed Responsible sub-scale items.

[1]Later factor analyses have extracted 4 bi-polar factors from the IIP. Scoring can be conducted according to this factor structure. For further information see Barkham et al, 1994.

Controlling score: Adding items 4, 79, 90, 94, 98, 102, 104, 109, 111 and 113 gives a Controlling sub-scale score ranging from 0-40. A mean of the score is then calculated and will range from 0-4. If items are missing a prorated score is calculated by obtaining the average of the completed Controlling subscale items.

Note: Less typically, item scores are ipsatized in order to eliminate variance caused by a patient's overall distress. Ipsatization involves subtracting each client's total IIP score (as a mean) from each item score. Sub-scale scores can then be calculated as above using these ipsatized item scores.

Interpretation of Scores

Reliability: The IIP has high internal consistency, with alpha coefficients ranging from 0.82-0.94 for the six sub-scales. High test-retest reliability (over a 10 week waiting period for treatment) has been demonstrated for the IIP sub-scales (ranging from r = 0.80 to r = 0.87) and for the total IIP score (r = 0.98). The IIP is sensitive to change with treatment (brief dynamic psychotherapy) for psychiatric patients: no significant decrease in total IIP score was found during a 10 week waiting period and significant decreases in total IIP scores were found after 10 weeks and 20 weeks of therapy.

Normative data and cut off scores: The data below apply to normative rather than ipsative scoring methods. The table below shows mean IIP sub-scale scores for 134 psychiatric patients at first and second baseline assessments.

IIP Sub-scale	First Baseline		Second Baseline	
	Mean	SD	Mean	SD
Assertive	1.883	0.787	1.746	0.803
Sociable	1.668	0.889	1.597	0.874
Submissive	1.131	0.712	1.063	0.673
Intimate	0.989	0.720	0.950	0.671
Responsible	1.822	0.804	1.690	0.835
Controlling	0.877	0.601	0.801	0.559

Correlations with other measures: Significant correlations (ranging from r = 0.57 to r = 0.64) have been found between the total IIP score and the Symptom Check List-90-Revised (SCL-90-R). The IIP sub-scales also correlate with the SCL-90-R (ranging from r = 0.33 for the Controlling sub-scale to r = 0.56 for the Sociable sub-scale). However, some of these correlations are quite low and it is evident that the SCL-90-R correlates best with the total IIP score.

The UCLA Loneliness Scale (Russel et al, 1980) correlates most highly with the IIP Sociable sub-scale (r = 0.73) and the Rathus Assertiveness Scale (Rathus, 1973) correlates most highly with the IIP Assertive sub-scale (r = 0.64), demonstrating that the IIP sub-scales can assess specific constructs.

Key References

Barkham, M., Hardy, G.E., and Startup, M. (1994). The structure, validity and clinical relevance of the Inventory of Interpersonal Problems. *British Journal of Medical Psychology, 67, 171-185.*

Horowitz, L.M., Rosenberg, S.E., Baer, B.A., Ureño, G. and Villaseñor, V.S. (1988). Inventory of Interpersonal Problems: Psychometric Properties and Clinical Applications. *Journal of Consulting and clinical Psychology, 56, No. 6, 885-892.*

Pilkonis, P.A., Kim, Y., Proietti, J.M. and Barkham, M (1996). Scales for personality disorders developed from the Inventory of Interpersonal Problems. *Journal of Personality Disorders, 10 (4): 355-369.*

Scarpa, A., Luscher, K.A., Kadee, J.S., Pilkonis, P.A. Kim, Y. and Williams, W.C. (1999). Screening for personality disorders in a non-clinical population. *Journal of Personality Disorders, 13 (4): 345-360.*

PENN STATE WORRY QUESTIONNAIRE (PSWQ)

Background Information

The PSWQ (Meyer et al, 1990) is a self-report measure of an individual's tendency to worry. A general cognitive style of excessive, uncontrollable worry is assessed rather than the specific content of thoughts. This cognitive style is seen as characteristic of Generalised Anxiety Disorder (GAD). The PSWQ has been found to be sensitive to change for cognitively oriented therapy in a sample of GAD patients.

Description of the PSWQ

The PSWQ comprises 16 items which clients rate on a scale of 1-5 (1 = not at all typical or characteristic, 5 = very typical or characteristic). Five of the PSWQ items are positively framed. A higher score on the PSWQ indicates greater worry.

Scoring the PSWQ

Each of the items on the PSWQ is scored between 1-5. **The scores for the positively framed items are reversed (items 1, 3, 8, 10 and 11).** Item scores are then added to give a total score ranging from 16-80.

Interpretation of Scores

Reliability: The PSWQ literature shows that the measure has good internal consistency ranging from $\alpha = 0.86\text{-}0.93$ in anxiety disorder patients and ranging from $\alpha = 0.91\text{-}0.95$ in non-clinical samples.

The measure also has good test-retest reliability, ranging from $r = 0.74$ to $r = 0.93$ across periods ranging from 2-10 weeks.

The PSWQ significantly discriminates between GAD and other anxiety disorders: subjects with GAD score more highly on the measure than subjects with other anxiety disorders and controls (Meyer et al, 1990; Brown et al, 1992). The PSWQ has also been found to discriminate between subjects who meet all, some or none of the DSM-III-R criteria for GAD (Meyer et al, 1990).

Normative data and cut off scores: In various non-clinical populations, mean scores on the PSWQ range from 42.2 (SD = 11.5; Gillis et al, 1995) to 53.7 (SD=14.2; Meyer et al, 1990). Gillis et al (1995) present the percentile scores on the PSWQ for their non-clinical sample. The 50th percentile score was 41, the 80th was 51 and the 90th was 57. In the Meyer et al (1990) non-clinical studies, mean scores for females were higher than those for males: mean scores ranged from 48.0 to 53.3 for females and 43.0 to 50.1 for males.

In a clinical population (n = 34), Meyer et al (1990) report a mean score for males of 67.8 (SD = 7.9) and for females of 68.4 (SD = 6.3). In a larger clinical population (n = 50), Brown et al (1992) report a mean score for males of 67.10 (SD = 9.43) and for females of 69.30 (SD = 9.86).

Although there are no published clinical cut off scores, data suggest that non-anxious subjects tend to score ≤ 40, subjects with anxiety disorders other than GAD tend to score between 40-60 and GAD subjects tend to score ≥ 60.

Correlations with other measures: In non-clinical populations, the PSWQ has been found to significantly correlate with the following measures of worry and anxiety: The Student Worry Scale ($r = 0.59$), the Worry Domains Questionnaire ($r = 0.67$), the Cognitive subscale of the Cognitive Somatic Anxiety Questionnaire ($r = 0.70$), the Somatic subscale of the Cognitive Somatic Anxiety Questionnaire ($r = 0.55$), the State Trait Anxiety Inventory (STAI)-Trait Scale ($r = 0.64$), the STAI-State Scale ($r = 0.49$), the Test Anxiety Inventory Emotionality Subscale ($r = 0.58$) and Worry Subscale ($r = 0.40$).

Also, in non-clinical populations, the PSWQ has been found to significantly correlate with the BDI (r = 0.36) and the Self-Handicapping Scale (r = 0.33).

In clinical populations the PSWQ has been found to significantly correlate with the Self Analysis Questionnaire Tension Subscale (r = 0.36) and a significant negative correlation has been found with the Emotional Control Questionnaire (r = -0.53).

In clinical populations, the PSWQ has also been found to correlate weakly (between 0.02 and 0.18) with the Hamilton Anxiety Rating Scale, the STAI-Trait Scale and the Zung Self-Rating Scale and between –0.10 and 0.04 with the Hamilton Rating Scale for Depression and the BDI.

Key References

Brown, T.A., Antony, M.M. and Barlow, H.B. (1992). Psychometric Properties of the Penn State Worry Questionnaire in a Clinical Anxiety Sample. *Behaviour Research and Therapy, No. 1, 33-37.*

Meyer, T.J., Miller, M.L., Metzger, R.L. and Borkovec, T.D. (1990). Development and Validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy, Vol 28, No.6, 487-495.*

OBSESSIVE COMPULSIVE INVENTORY (OCI)

Background Information

The Obsessive Compulsive Inventory (OCI; Foa et al, 1998) is a self-report measure of the frequency of a broad range of obsessions and compulsions and associated distress. The measure can be used to aid the diagnosis and determine the severity of OCD.

Description of the OCI

The OCI has 42 items. For each item a rating is made for frequency and also for associated distress over the past month. The items are divided into seven sub-scales: Washing (8 items), Checking (9 items), Doubting (3 items), Ordering (5 items), Obsessing (8 items), Hoarding (3 items) and Mental Neutralizing (6 items).

Scoring the OCI

Each of the 42 items is scored for frequency on a five-point scale, ranging from 0-4 (0 = Never and 4 = Almost Always). Each item is also scored for distress on a five-point scale, ranging from 0-4 (0 = Not at all and 4 = Extremely). Total and sub-scale scores for frequency and distress can be calculated as follows:

Total frequency and distress scores: A *Total frequency score* (ranging from 0-168) can be calculated by adding the frequency scores for all 42 items. A *Total distress score* (ranging from 0-168) can be calculated by adding the distress scores for all 42 items.

Washing sub-scale frequency and distress scores: A *Washing sub-scale frequency score* can be calculated by adding the frequency scores for items 2, 4, 8, 21, 22, 27, 38 and 42 and then calculating a mean (ranging from 0-4). A *Washing sub-scale distress score* can be calculated by adding the distress scores for items 2, 4, 8, 21, 22, 27, 38 and 42 and then calculating a mean (ranging from 0-4).

Checking sub-scale frequency and distress scores: A *Checking sub-scale frequency score* can be calculated by adding the frequency scores for items 3, 7, 9, 10, 19, 24, 31, 32 and 40 and then calculating a mean (ranging from 0-4). A *Checking sub-scale distress score* can be calculated by adding the distress scores for items 3, 7, 9, 10, 19, 24, 31, 32 and 40 and then calculating a mean (ranging from 0-4).

Doubting sub-scale frequency and distress scores: A *Doubting sub-scale frequency score* can be calculated by adding the frequency scores for items 26, 37 and 41 and then calculating a mean score (ranging from 0-4). A *Doubting sub-scale distress score* can be calculated by adding the distress scores for items 26, 37 and 41 and then calculating a mean (ranging from 0-4).

Ordering sub-scale frequency and distress scores: An *Ordering sub-scale frequency score* can be calculated by adding the frequency scores for items 14, 15, 23, 29 and 35 and then calculating a mean score (ranging from 0-4). An *Ordering sub-scale distress score* can be calculated by adding the distress scores for items 14, 15, 23, 29 and 35 and then calculating a mean (ranging from 0-4).

Obsessing sub-scale frequency and distress scores: An *Obsessing sub-scale frequency score* can be calculated by adding the frequency scores for items 1, 12, 13, 17, 20, 28, 30 and 33 and then calculating a mean score (ranging from 0-4). An *Obsessing sub-scale distress score* can be calculated by adding the distress scores for items 1, 12, 13, 17, 20, 28, 30 and 33 and then calculating a mean (ranging from 0-4).

Hoarding sub-scale frequency and distress scores: A *Hoarding sub-scale frequency score* can be calculated by adding the frequency scores for items 6, 11 and 34 and then calculating a mean (ranging from 0-4). A *Hoarding sub-scale distress score* can be calculated by adding the distress scores for items 6, 11 and 34 and then calculating a mean (ranging from 0-4).

Mental Neutralizing sub-scale frequency and distress scores: A *Mental Neutralizing sub-scale frequency score* can be calculated by adding the frequency scores for items 5, 16, 18, 25, 36 and 39 and then calculating a mean (ranging from 0-4). A *Mental Neutralizing sub-scale distress score* can be calculated by adding the distress scores for items 5, 16, 18, 25, 36 and 39 and then calculating a mean (ranging from 0-4).

Interpretation of Scores

Reliability and validity: Excellent internal consistency has been found for the OCI Total frequency (0.93) and Total distress (0.92) scores in a sample of patients with OCD (Foa et al, 1998). Internal consistency ranged from 0.72-0.96 for sub-scale frequency ratings and 0.68-0.94 for sub-scale distress ratings. Foa et al (1998) also found the OCI Total scores to have high test-retest reliability in an OCD sample ($r = 0.84$ for Total frequency and $r = 0.87$ for Total distress) and in non-patient controls ($r = 0.90$ for Total frequency and $r = 0.89$ for Total distress). Sub-scale scores also demonstrated high test-retest reliability in an OCD sample (ranging from $r = 0.79$-0.95 for sub-scale frequency scores and $r = 0.77$-0.97 for sub-scale distress scores) and in non-patient controls (ranging from $r = 0.82$-0.90 for sub-scale frequency scores and $r = 0.68$-0.89 for sub-scale distress scores).

The Total frequency and Total distress OCI scores also discriminated between those with OCD and those with PTSD, Generalised Social Phobia or no anxiety. This was also found for all sub-scales except Hoarding.

Normative data and cut off scores: Foa et al (1998) give mean scores for the following groups: OCD patients, PTSD patients, Generalised Social Phobia patients and controls on all sub-scales and total scores. The table below displays these means for the OCD group and the control group only.

	Frequency Score Mean (SD)		Distress Score Mean (SD)	
	OCD Group	Control Group	OCD Group	Control Group
Washing sub-scale	1.44 (1.4)	0.76 (0.7)	1.44 (1.3)	0.55 (0.7)
Checking sub-scale	1.51 (0.9)	0.72 (0.5)	1.51 (0.9)	0.52 (0.5)
Doubting sub-scale	2.01 (1.1)	0.78 (0.8)	1.84 (1.3)	0.65 (0.8)
Ordering sub-scale	1.87 (1.1)	1.08 (0.8)	1.87 (1.2)	0.80 (0.8)
Obsessing sub-scale	1.67 (0.8)	0.69 (0.6)	1.79 (1.1)	0.60 (0.6)
Hoarding sub-scale	1.22 (1.1)	1.52 (0.9)	1.24 (1.3)	1.06 (0.8)
Mental Neutralizing sub-scale	1.49 (0.9)	0.64 (0.6)	1.38 (1.0)	0.41 (0.5)
Total score	66.36 (29.4)	34.15 (21.2)	66.33 (31.9)	25.25 (20.8)

Correlations with other measures: Foa et al (1998) found the OCI Total distress score and Total frequency scores to significantly correlate with the following measures of obsessions/compulsions in a group of OCD patients and a control group:

	OCI Total frequency		OCI Total distress	
	OCD Group	Control Group	OCD Group	Control Group
Y-BOCS[1]	0.43	0.31	0.23	0.55
Y-BOCS obsessions scale	0.31	0.52	0.14*	0.47
Y-BOCS compulsions scale	0.41	0.29	0.25	0.44
Maudsley Obsessive Compulsive Inventory (MOCI)	0.75	0.72	0.68	0.66
Compulsion Activity Checklist (CAC)	0.67	——-	0.65	——

[1] Yale-Brown Obsessive-Compulsive Scale, interview version for OCD group, questionnaire version for control group.
* Not significant.

In the OCD group, a significant correlation was found between OCI Total distress and Total frequency and the BAI and between OCI Total distress and the HAM-D. In the control group, both OCI Total distress and Total frequency significantly correlated with the STAI-Trait Scale, the BAI and the BDI. Foa et al (1998) also give correlations for the OCI Washing, Checking and Doubting sub-scales and the corresponding sub-scales of the MOCI and CAC.

Key References

Foa, B.E., Kozak, M.J., Salkovskis, P.M., Coles, M.E. and Amir, N. (1998). The validation of a New Obsessive-Compulsive Disorder Scale: The Obsessive-Compulsive Inventory. *Psychological Assessment, 10:3, 206-214.*

PANIC RATING SCALE (PRS)

Background Information

The Panic Rating Scale (PRS; Clark et al, 1994) is a self-report measure designed to assess the frequency of panic attacks, panic related distress/disability and avoidance associated with panic disorder.

Description of the PRS

The PRS has 3 items. The first item relates to the frequency of panic attacks over the past two weeks and is rated on a five-point scale ranging from 0-4 (0 = No panic attacks, 4 = One or more panic attacks per day).

The second item relates to the degree of perceived disturbance and/or disability relating to the panic attacks and is rated on a nine-point scale ranging from 0-8 (0 = Not at all disturbing and/or disabling, 8 = Very disturbing or disabling). The final item on the PRS relates to the level of avoidance of particular situations over the past two weeks (or the need to be accompanied in particular situations), due to fear of panic attacks or panic symptoms.

This final item is rated on a nine-point scale ranging from 0-8 (0 = Never avoid, 8 = Always avoid). To ensure that clients have an understanding of panic attacks when completing the measure, a definition is provided.

Scoring the PRS

Each PRS item is scored separately. Thus three scores will be obtained for the PRS:

- A Panic Frequency score ranging from 0-4.
- A Panic-related Distress/Disability score ranging from 0-8.
- An Associated Avoidance score ranging from 0-8.

Interpretation of Scores

There have been two notable Randomised Controlled Trials of treatment for Panic Disorder (Clark et al, 1994; Clark et al, 1999). Both of these RCTs have used items one and two of the PRS as measures of treatment outcome.

The third PRS item was added to the measure by clinicians at the Centre for Anxiety Disorders and Trauma, South London and Maudsley NHS Trust, who have found it to be clinically useful (personal communication, Professor D.M. Clark, 2003). As yet, there is no published data on the third PRS item. The data presented here relate to PRS items one and two only.

Reliability and validity: Items one and two of the PRS have both been found to be sensitive to change with treatment for panic disorder and to discriminate between treatments that vary in effectiveness. For example, Clark et al (1994) found significantly reduced scores for panic disorder patients receiving cognitive therapy, applied relaxation and treatment with imipramine in comparison to patients on a waiting list. In addition, the PRS items discriminated between the effectiveness of the three therapeutic approaches. A later study (Clark et al, 1999) also found PRS items one and two to be sensitive to change for brief cognitive therapy for panic disorder (five sessions) in addition to full-length cognitive therapy (12 sessions).

Normative data and cut off scores: The PRS data from the two RCTs (Clark et al, 1994; Clark et al, 1999) are given in the tables below:

PRS data from Clark et al (1994)[1]

PRS Item	Assessment	Panic disorder patients receiving cognitive therapy Mean (SD)	Panic disorder patients receiving applied relaxation therapy Mean (SD)	Panic disorder patients receiving imipramine Mean (SD)
Item 1:	Pretreatment	2.6 (1.0)	2.8 (1.0)	3.0 (0.9)
Panic Frequency	3mths	0.3 (0.7)	1.4 (1.4)	1.1 (1.3)
	12mths	0.4 (0.7)	1.6 (1.5)	0.6 (0.9)
	15mths	0.5 (0.7)	1.3 (1.6)	0.9 (1.4)
Item 2:	Pretreatment	5.2 (2.0)	6.1 (1.7)	6.0 (2.2)
Panic-related	3mths	1.2 (1.1)	3.0 (1.8)	3.0 (2.5)
distress/disability	12mths	1.4 (1.2)	2.9 (2.0)	1.8 (2.0)
	15mths	1.1 (1.6)	2.8 (2.0)	2.6 (2.3)

[1] All treatments were administered in 12 sessions over 3 months with up to three booster sessions in the following three months. Imipramine was gradually withdrawn after 6 months. The data above are based on 64 patients with panic disorder, randomly assigned to the three treatment groups and one waiting list group. Of these 64 patients, 19 per cent had no agoraphobic avoidance, 48 per cent had mild agoraphobic avoidance and 33 per cent had moderate agoraphobic avoidance.

PRS data from Clark et al (1999)[2]

PRS Item	Assessment	Panic Disorder Patients receiving Cognitive Therapy Mean (SD)	Panic Disorder Patients receiving Brief Cognitive Therapy Mean (SD)
Item 1:	Pretreatment	2.8 (1.0)	2.8 (0.7)
Panic Frequency	Posttreatment	0.4 (0.8)	0.5 (0.9)
	3mths	0.2 (0.6)	0.4 (0.8)
	12mths	0.6 (0.9)	0.4 (0.5)
Item 2:	Pretreatment	5.6 (1.7)	5.1 (1.6)
Panic-related	Posttreatment	0.7 (1.0)	1.1 (0.8)
distress/disability	3mths	0.8 (1.1)	0.9 (0.9)
	12mths	0.9 (1.2)	1.0 (1.2)

[2] Cognitive Therapy was administered in 12 weekly sessions and Brief Cognitive Therapy was administered in five sessions over 3 months. Both groups received a mean of 1.5 one-hour booster sessions in the three months following treatment. The data above are based on 42 patients with panic disorder, randomly assigned to the two treatment groups and one waiting list group. Of these 42 patients, 15 per cent had no agoraphobic avoidance, 63 per cent had mild agoraphobic avoidance and 22 per cent had moderate agoraphobic avoidance.

Key References

Clark, D.M., Salkovskis, P.M., Hackman, Middleton, H., Anastasiades, P. and Gelder, M.G. (1994). A comparison of cognitive therapy, applied relaxation and imipramine in the treatment of panic disorder. *British Journal of Psychiatry, 164, 759-769.*

Clark, D.M., Salkovskis, P.M., Hackman, A., Wells, A., Ludgate, J. and Gelder, M. (1999). Brief cognitive therapy for panic disorder: a randomised controlled trial. *Journal of Consulting and Clinical Psychology, 67, 583-589.*

PSYCHOTIC SYMPTOM RATING SCALES (PSYRATS)

Background Information

The Psychotic Symptom Rating Scales (PSYRATS; Haddock et al, 1999) are two scales: one scale for auditory hallucinations and one scale for delusions. Each scale measures the presence and severity of a number of different dimensions of the psychotic symptom. The scales are appropriate for measuring outcome to treatment for positive psychotic symptoms in schizophrenia/ psychosis in terms of changes in symptom severity and can also be used to assess how symptom dimensions co-vary over time/with treatment. The PSYRATS are to be completed by a clinician using a structured interview.

Description of the PSYRATS

The PSYRATS is comprised of two scales, the auditory hallucinations sub-scale (AH) and the delusions sub-scale (DS). The AH has 11 items which assess the following dimensions of auditory hallucinations: frequency, duration, location, loudness, beliefs about the origin of voices, amount of negative content, degree of negative content, amount of distress, intensity of distress, disruption to life caused by voices and controllability of voices. The DS has six items which assess the following dimensions of delusions: amount of preoccupation with delusions, duration of preoccupation with delusions, conviction, amount of distress, intensity of distress and disruption to life caused by beliefs.

Each subscale is prefaced with a general instruction sheet that includes relevant patient details. In general the PSYRATS items are designed to rate the patient's experiences over the last week. Exceptions to this are items 4 and 5 on the AH subscale and item 3 on the DS subscale. These items are rated according to the patient's beliefs/experience at the time of interview (or for item 4 on the AH subscale, at the last time the voices were heard).

In addition, the AH sub-scale has two supplementary items, intended to be scored separately from the subscale. The first supplementary item records the number of different voices heard by the patient over the last week. The second supplementary item records the form of the voices (whether voices are heard in the first, second, third person or as single words or phrases without pronouns) and the frequency of these voice forms.

Scoring the PSYRATS

For both sub-scales, each item is scored between 0-4. Whilst looking at individual item scores can give clinicians a clear picture of progress for individual patients, the PSYRATS authors (Haddock et al, 1999) acknowledge that for the purpose of outcome monitoring it may be useful to have global dimensional symptom scores in addition to individual item scores for a target symptom.

Adding the 11 AH item scores gives a total score for AH. This total AH score (T-AH) will range from 0-44. Adding the six DS item scores gives a total score for DS. This total DS score (T-DS) will range from 0-24.

Interpretation of Scores

Reliability: The PSYRATS has good inter-rater reliability (Haddock et al, 1999). For the AH sub-scale, inter-rater reliability co-efficient estimates ranged from 0.79 (item 10)-1.00 (items 4 and 8). For the DS sub-scale, inter-rater reliability co-efficient estimates ranged from 0.88 (item 6)-1.00 (item 3).

Normative data and cut off scores: There are no published cut off scores for the PSYRATS. Haddock et al (1999) present median item and sub-scale total scores for 71 patients with schizophrenia (n = 52) or schizoaffective disorder (n = 19). These are shown in the table below:

Median PSYRATS scores for patients with schizophrenia or schizoaffective disorder, from Haddock et al, (1999)

Subscale	Item(s)	Median	Range
AH sub-scale			
	Frequency, Duration, Location and Beliefs re-origin (items 1, 2, 3, and 5)	3	(1-4)
	Loudness (item 4)	2	(1-4)
	Amount and Degree of Negative content (items 6 and 7)	3	(0-4)
	Amount of Distress (item 8)	3	(0-4)
	Intensity of Distress (item 9)	2	(0-4)
	Disruption (item 10)	2	(0-3)
	Control (item 11)	3	(0-4)
	T-AH	**28**	**(14-39)**
DS sub-scale			
	Amount of Pre-occupation (item 1)	3	(1-4)
	Duration of Pre-occupation (item 2)	2	(1-4)
	Conviction (item 3)	3	(1-4)
	Amount of Distress (item 4)	3	(0-4)
	Intensity of Distress (item 5)	2	(0-4)
	Disruption (item 6)	2	(0-4)
	T-DS	**15**	**(5-22)**

Correlations with other measures: Haddock et al (1999) compared scores on the PSYRATS with those on the modified version of the Psychiatric Assessment Scale (KGV; Krawiecka et al, 1977; modified by Lancashire, 1994).

Significant relationships were found between the AH control item and both the total modified KGV score ($r = 0.40$) and the KGV hallucinations/delusions score ($r = 0.40$). The T-AH score was found to correlate with the KGV hallucinations items ($r = 0.33$).

Significant associations were also found between the DS disruption item and both the total modified KGV score ($r = 0.36$) and the KGV hallucination/delusions score ($r = 0.40$) and between the DS duration of pre-occupation item and the KGV affective symptoms score ($r = 0.34$). The T-DS score was found to correlate significantly with the total modified KGV score ($r = 0.35$), the KGV delusions item ($r = 0.38$) and the KGV hallucinations/delusions score ($r = 0.34$).

Key Reference

Haddock, G., McCarron, J., Tarrier, N. and Faragher, E.B. (1999). Scales to measure dimensions of hallucinations and delusions: the psychotic symptom rating scales (PSYRATS). *Psychological Medicine, 29, 879-889.*

IMPACT OF EVENT SCALE-REVISED (IES-R)

Background Information

The Impact of Event Scale-Revised (IES-R, Weiss and Marmar, 1996) is a self-report assessment of psychological response to a specific traumatic stressor or stressful life event. The IES-R differs from the original IES (Horowitz et al, 1979) in two ways: first, items have been added to tap the hyperarousal symptoms of PTSD and minor modifications have been made to original IES items and second, items are no longer rated on an unequal four-point scale of 0, 1, 3, and 5 but on a five-point scale of 0-4.

Description of the IES-R

The IES-R has 22 items, divided into three sub-scales. The first sub-scale relates to intrusive thoughts, images or feelings and dissociative-like re-experiencing (Intrusion subscale, 8 items). The second sub-scale relates to attempts to dampen or avoid experiences associated with the traumatic event, including associated numbing of feelings (Avoidance sub-scale, 8 items). The third sub-scale relates to the hyperarousal symptoms that can be part of the psychological response to traumatic events and are evidenced in PTSD (Hyperarousal sub-scale, 6 items).

Scoring the IES-R

Each item is scored on a five-point scale ranging from 0-4 (0 = not at all distressed or bothered and 4 = extremely distressed or bothered). Mean scores can be calculated for the full IES-R and for each of the sub-scales as follows:

Total IES-R score: A Total IES-R score can be obtained by calculating a mean of all item scores. The resultant score should range from 0-4.

Intrusion sub-scale score: An Intrusion sub-scale score can be obtained by calculating a mean of the scores for items 1, 2, 3, 6, 9, 14, 16 and 20. The resultant score should range from 0-4.

Avoidance sub-scale score: An Avoidance sub-scale score can be obtained by calculating a mean of the scores for items 5, 7, 8, 11, 12, 13, 17 and 22. The resultant score should range from 0-4.

Hyperarousal sub-scale score: A Hyperarousal sub-scale score can be obtained by calculating a mean of the scores for items 4, 10, 15, 18, 19 and 21. The resultant score should range from 0-4.

Interpretation of Scores

There are few data available on the IES-R. Also, the initial psychometric data provided on the measure (Weiss and Marmar, 1996; Marmar et al, 1996) were based on the first revision of the IES-R. In this first revision, the hyperarousal items were added to the IES but ratings were still made on an unequal four-point scale of 0, 1, 3, and 5 rather than the current five-point scale described above. Thus, these data are not given here.

Since the publication of the IES-R (Weiss and Marmar, 1996), researchers have differed in the version of the scale that they use. Some researchers still use the first revision of the IES-R, with the unequal four-point rating scale, due to the fact that psychometric data are available. Other researchers have used the current version of the IES-R that is published by Weiss and Marmar (1996) and described here. The normative data provided below are from Walsh and Clarke (2003), who used the current rating scale of 0-4.

Normative data and cut off scores: Walsh and Clarke (2003) provide data on the IES-R for 126 employees of a UK NHS Community Trust working in the areas of primary care, learning disabilities, mental health or older adults. The subjects had all reported an incident of either verbal or physical aggression. IES-R scores for these subjects are given in the table overleaf.

	Mean	Standard Deviation	Range (Possible range 0-4)
Intrusion sub-scale score	0.36	0.63	0-3.29
Avoidance sub-scale score	0.38	0.57	0-3.00
Hyperarousal sub-scale score	0.49	0.61	0-3.14
Total IES-R score	0.41	0.55	0-3.00

Normative data on this version of the IES-R are currently not available for patients with PTSD.

Key References

Marmar, C.R., Weiss, D.S., Metzler, T.J., Ronfeldt, H.M. and Foreman, C. (1996). Stress Responses of Emergency Service Personnel to the Loma Prieta Earthquake Interstate 880 Freeway collapse and Control Traumatic Incidents. *Journal of Traumatic Stress, 9 (1): 63-85.*

Walsh, B.R. and Clarke, E. (2003). Post-trauma symptoms in health workers following physical and verbal aggression. *Work and Stress, 17 (2): 170-181.*

Weiss, D.S. and Marmar, C.R. (1996). The Impact of Event Scale-Revised. In Wilson, J.P. and Keane, T.M. (Eds.). Assessing Psychological Trauma and PTSD. New York, Guildford, 399-411.

LIEBOWITZ SOCIAL ANXIETY SCALE (LSAS) AND SHEEHAN DISABILITY RATINGS

Background Information

The original Liebowitz Social Anxiety Scale (LSAS; Liebowitz, 1987) is a clinician-administered measure of fear and avoidance across a number of different social situations. A self-report version of the measure has also been developed and the information given here is for the self-report version. The self-report LSAS has been used to assess severity of social phobic fear and avoidance and is sensitive to change with treatment (Baker et al, 2002).

Description of the LSAS

As with the original LSAS, the self-report LSAS has 24 items. For each item a rating is made for the amount of fear or anxiety experienced and also for associated avoidance. The items are divided into two sub-scales. The first sub-scale measures difficulty with social interaction (*Social Interaction sub-scale*) and has 11 items. The second sub-scale measures difficulty with performance (*Performance sub-scale*) and has 13 items.

Scoring the LSAS

Each of the 24 items is scored for fear/anxiety on a four-point scale, ranging from 0-3 (0 = None and 3 = Severe). Each item is also scored for avoidance on a four-point scale, ranging from 0-3 (0 = Never and 4 = Usually-68-100%). Total and sub-scale scores for both fear and avoidance can be calculated as follows:

Total fear and avoidance scores: A *Total fear score* (ranging from 0-72) can be calculated by adding the fear/anxiety scores for all 24 items. A *Total avoidance score* (ranging from 0-72) can be calculated by adding the avoidance scores for all 24 items. A *Total LSAS* score, ranging from 0-144, can be calculated by adding the *Total fear* and *Total avoidance* scores. This *Total LSAS* score is most commonly used in research studies and routine clinical practice. Sub-scale scores can be calculated, but for routine outcome monitoring, the Total score is sufficient. Clinicians may want to look at individual items to target treatment interventions.

Social Interaction sub-scale fear and avoidance scores: A *Social Interaction sub-scale fear score*, ranging from 0-33, can be calculated by adding the fear/anxiety scores for items 5, 7, 10, 11, 12, 15, 18, 19, 22, 23 and 24. A *Social Interaction sub-scale avoidance score,* ranging from 0-33, can be calculated by adding the avoidance scores for items 5, 7, 10, 11, 12, 15, 18, 19, 22, 23 and 24.

Performance sub-scale fear and avoidance scores: A *Performance sub-scale fear score*, ranging from 0-39, can be calculated by adding the fear/anxiety scores for items 1, 2, 3, 4, 6, 8, 9, 13, 14, 16, 17, 20 and 21. A *Performance sub-scale avoidance score,* ranging from 0-39, can be calculated by adding the avoidance scores for items 1, 2, 3, 4, 6, 8, 9, 13, 14, 16, 17, 20 and 21.

Interpretation of Scores

Reliability: Baker et al (2002) found all of the self-report LSAS scales to have internal consistency equal to or greater than $\alpha = 0.79$ in a sample of 175 patients diagnosed with social phobia (approximately half with the generalised subtype of social phobia and half with the non-generalised sub-type). Fresco et al (2001) found internal consistency for the self-report LSAS scales to range from $\alpha = 0.82$-0.95 in a sample of 99 patients diagnosed with social phobia and from $\alpha = 0.73$-0.94 in a sample of 53 non-anxious controls. In their sample of patients with social phobia, Baker et al (2002) reported 12 week test-retest reliability to be 0.83 for the self-report LSAS total score, 0.79 and 0.83 for the Total fear and Total avoidance scores respectively, 0.80 and 0.85 for the Social Interaction sub-scale fear and avoidance scores respectively and 0.53 and 0.75 for the Performance sub-scale fear and avoidance scores respectively.

Normative data and cut off scores: Both Baker et al (2002) and Fresco et al (2001) give mean scores for all self-report LSAS scales. The data from both studies are given in the table overleaf.

Correlations with other measures: All scales of the self-report LSAS have been found to significantly correlate with the following measures in patients with social phobia: the total, social and agoraphobia scales of the Social Phobia Anxiety Inventory (SPAI; Turner et al, 1989), ranging from 0.48-0.83; the Negative Self Statements scale of Self Statements during Public Speaking (SSPS; Hofmann and DiBartolo, 2000), ranging from 0.38-0.44; the Social Phobia Scale (SPS; Mattick and Clarke, 1998),

Mean self-report LSAS scores from Baker et al (2002) and Fresco et al (2001).

	Patients with social phobia (n = 175)[1]	Patients with social phobia (n = 99)[2]	Non-anxious controls (n = 53)[2]
	Mean (SD)	Mean (SD)	Mean (SD)
Social Interaction sub-scale fear score	18.9 (7.3)	19.4 (5.9)	3.3 (3.4)
Social Interaction sub-scale avoidance score	16.6 (7.9)	18.7 (6.7)	2.9 (3.4)
Performance sub-scale fear score	19.5 (6.1)	19.2 (6.1)	4.2 (4.1)
Performance sub-scale avoidance score	16.6 (7.3)	17.1 (6.9)	3.1 (3.3)
Total Fear score	37.2 (12.9)	38.7 (11.3)	7.5 (7.2)
Total Avoidance score	33.2 (14.4)	35.9 (12.7)	6.0 (6.2)
Total self-report LSAS score	69.1 (25.5)	74.5 (23.3)	13.5 (12.7)

[1] From Baker et al (2002). Approx. half of these patients had a diagnosis of generalised sub-type social phobia and approx. half had a diagnosis of non-generalised subtype social phobia.
[2] From Fresco et al (2001).

ranging from 0.48-0.66; the Social Interaction Anxiety Scale (SIAS; Mattick and Clarke, 1998), ranging from 0.56-0.77; the social phobia sub-scale of the Fear Questionnaire (FQ; Marks and Matthews, 1979), ranging from 0.47-0.64 and the Beck Depression Inventory (BDI; Beck et al, 1979), ranging form 0.44-0.48. Also, in patients with social phobia, the following scales of the self-report LSAS have been found to significantly correlate with the Anxiety Sensitivity Index (ASI; Reiss et al, 1986): the LSAS total score and the Performance sub-scale fear and avoidance scores (0.14, 0.17 and 0.19 respectively). In non-anxious controls, all scales of the self-report LSAS have been found to significantly correlate with the SPS (ranging from 0.47-0.60) and the SIAS (ranging from 0.62-0.72).

Sheehan Disability Ratings

The Sheehan Disability Ratings are three self-report items designed to measure disability/impairment related to psychiatric symptoms. The three items are related to impairment at Work, impairment with Social Life/Leisure Activities and impairment with Family Life/Home Responsibilities. Each item is rated on a discretised analogue scale, ranging from 0-10 (0 = Not at all, 10 = Very severely). Item scores are interpreted individually, with higher scores representing greater disability in that domain. A total disability score can also be calculated by adding the three item scores. Sheehan et al (1996) cite mean scores for each item in two groups of patients with social phobia and two groups of patients with panic disorder. These means are given in the table below.

	Social Phobia Patients		Panic Disorder Patients	
	(n = 576)	(n = 76)	(n = 77)	(n = 101)
Work	6.67	5.76	4.58	5.42
Social Life/Leisure Activities	7.17	6.56	4.94	6.14
Family Life/Home Responsibilities	4.57	3.80	3.10	5.18

The Sheehan Disability Ratings are sensitive to treatment effects for patients with social phobia and patients with panic disorder (Sheehan et al. 1996).

Key References

Baker, S.L., Heinrichs, N., Kim, H-J. and Hofmann, S.G. (2002). The Liebowitz Social Anxiety Scale as a self-report instrument: a premliminary psychometric analysis. *Behaviour Research and Therapy, 40, 701-715.*

Fresco, D.M., Coles, M.E., Heimberg, R.G., Leibowitz, M.R., Hami, S., Stein, M.B. and Goetz, D. (2001). The Liebowitz Social Anxiety Scale: a comparison of the psychometric properties of self-report and clinician administered formats. *Psychological Medicine, 31, 1025-1035.*

Liebowitz, M.R. (1987). Social Phobia. *Anxiety, 22: Modern Problems in Pharmacopsychiatry, 141-173.*

Sheehan, D.V., Harnett-Sheehan, K. and Raj, B.A. (1996). The measurement of disability. *International Clinical Psychopharmacology, 11, Suppl. 3, 89-95.*

SOCIAL SUMMARY RATING SCALE

Background Information

The Social Summary Rating Scale (SSRS) is a self-report measure designed to assess social anxiety and related constructs. The scale is an adaptation of the Social Phobia Weekly Summary Scale (SPWSS; Clark et al, 2003).

Description of the SSRS

The SSRS has eight items. The first six items form the current version of the SPWSS and relate to social anxiety, social avoidance, self-focused versus external attention in general social situations, self-focused versus external attention in feared social situations, anticipatory processing and post-event rumination.

The final two items of the SSRS are additional items to assess anxiety-related distress and anxiety-related interference/impairment. All eight items are rated on a nine-point scale ranging from 0-8, with higher scores indicating greater social phobia psychopathology.

Scoring the SSRS

Three scores are calculated from the SSRS: an SPWSS score, and separate scores for Anxiety-related Distress and Anxiety-related Interference/impairment.

SPWSS score: An SPWSS score can be obtained by calculating a mean of the scores for the first six items of the SSRS (items a-f). The resultant score should range from 0-8.

Anxiety-related Distress: The score for seventh item of the SSRS (item g) is the Anxiety-related Distress Score and should range from 0-8.

Anxiety-related Interference: The score for final item of the SSRS (item h) is the Anxiety-related Interference score and should range from 0-8.

Interpretation of Scores

Items g and h have been added to the measure by clinicians at the Centre for Anxiety Disorders and Trauma, South London and Maudsley NHS Trust. As yet, there are no published data on these two items.

The remaining six items form the current version of the SPWSS. Published data are not yet available on this version of the SPWSS. However, a study featuring this measure is due to be submitted for publication early in 2004. It has been noted that the six-item SPWSS is sensitive to treatment effects and has good internal consistency (Cronbach's Alpha = 0.75; personal communication, Professor D.M. Clark, 2003).

Normative data and cut off scores: Again, there are no published normative data on the six-item SPWSS or the final two SSRS items. Data are available on a previous five-item version of the SPWSS (Clark et al, 2003) which only differs from the six-item version in that it has one item (rather than two items) relating to self-focused versus external attention.

A mean score for the current six-item SPWSS should be calculated for routine outcome monitoring. Thus, the data on the five-item version are not provided here.

Key Reference

For information on the five-item SPWSS:

Clark, D.M., Ehlers, A., McManus, F., Hackman, A., Fennell, M., Campbell, H., Flower, T., Davenport, D. and Louis, B. (2003). Cognitive Therapy versus Fluoxetine in Generalized Social Phobia: A Randomized Placebo-Controlled Trial. *Journal of Consulting and Clinical Psychology, 71:1058-1067.*

Chalder Fatigue Questionnaire

Background Information

The CFQ (Chalder et al 1993) is a self-report questionnaire enquiring about the various physical and mental fatigue symptoms. It was designed to measure fatigue in hospital and community populations, and is recommended for use in chronic fatigue syndrome. The CFQ has been found to be sensitive to change in patients who received cognitive behaviour therapy.

Description of the CFQ

The CFQ consists of 11 physical and mental fatigue items rated on: less than usual, no more than usual, more than usual, or much more than usual. Scores on these 11 items give a total Fatigue score.

At the end of the questionnaire the questions about duration of tiredness and the percentage of time feeling tired are not generally scored, but are used to provide additional information about the patient. The final item asks respondants to give a reason for their tiredness in an open ended question "Why do you think you are feeling tired?".

Scoring the CFQ

A total score can be obtained by adding all 11items on the main scale. There is evidence for two constructs, physical fatigue (first 7 items) and mental fatigue (last 4 items). Scoring is normally bimodal (values of 0, 0, 1, 1 are assigned to columns). The range of bimodal scoring is from 0-11. This method has the advantage of eliminating errors due to "end users" and "middle users", and is often used to dichotomise probable "normals" and "cases". Alternatively, a much less frequently adopted Likert system (0, 1, 2, 3) can be used. The range for the Likert scoring is from 0-33. In both scoring systems, higher scores indicate greater fatigue.

Interpretation of scores

Split halves reliability are 0.85 to 0.86 for part 1 and part 2. The value of Cronbach's alpha of the 11-item scale is 0.89.

Normative data and cut off scores

Fatigue is normally distributed in the population. Pawlikowska et al (1994) show graphically the frequency distribution of fatigue and the relationship between fatigue and psychological distress as measured by the GHQ-12. The two constructs are moderately correlated (r = 0.62, 95% CI 0.61-0.63).

For the bimodal scoring, a score of ≥ 4 is severe fatigue. A score of ≥ 4 for six months or more indicates caseness for chronic fatigue.

In a cognitive behaviour therapy treatment study for chronic fatigue syndrome (Deale et al. 1997), patients had the following CFQ scores (bimodal scoring system):

	Cognitive behaviour therapy (N = 30)	Relaxation group (N = 30)
Pre-treatment	10.2 (SD 1.3)	9.5 (SD 2.6)
Post-treatment	7.2 (SD 4.0)	7.5 (SD 4.1)
6-month follow-up	4.1 (SD 4.0)	7.2 (SD 4.0)

Key reference

Chalder, T., Berelowitz, G., Pawlikowska, T., Watts, L., Wessley, S., Wright, D. & Wallace, E. (1993). Development of a fatigue scale. *Journal of Psychosomatic Research, Vol 37, 147-153*.

Deale, A., Chalder, T., Marks, I. & Wessley, S. (1997). Cognitive behaviour therapy for chronic fatigue syndrome: a randomised controlled trial. *American Journal of Psychiatry, Vol 154, 408 -414*.

Pawlikowska, T., Chalder, T., Hirsch, S.R., Wallace, P., Wright, D.J.M., Wessely, S. (1994) A Population Based Study of Fatigue and Psychological Distress. **British Medical Journal**. 308; 763-766.

ACKNOWLEDGMENTS

Project Development and Review

Trust Psychological Therapies Committee

Professor Philippa Garety (Chair)
Fran Bristow
Dr. June Brown
Sarah Burleigh
Professor David Clark
Dr. Ingrid Doherty
Dr. Caron Gaw
Hilary Harries
Dr. Brian Jacobs
Dr. Dominic Lam
Dr. Adrienne Little
Dr. Stirling Moorey
Dr. Siobhan Murphy
Professor Jan Scott
Ian Simpson
Roger Smith

Project Workers

Helen Coelho
Caroline Cross

Patient Journey Project Managers

Matthew Broadbent
Carmel de Silva

Trust Clinical Outcomes Group

Consultation on measures and guidance

Professor Anke Ehlers
Dr. Daniel Freeman
Professor Elizabeth Kuipers
Professor Dianne Chambless
Professor Peter Cooper
Professor Chris Fairburn
Dr. Emmanuelle Peters
Dr. Melanie Shepherd
Padmal de Silva
Dr. Janet Feigenbaum
Dr. Gillian Haddock

Contact for Further Information and Comments:

Professor Philippa Garety
philippa.garety@slam.nhs.uk