**AMHOCN**

**Australian Mental Health Outcomes and Classification Network**

**'Sharing Information to Improve Outcomes'**

**An Australian Government funded initiative**

# Review of standardised measures used in the National Outcomes and Casemix Collection (NOCC)

Jane Pirkis, Philip Burgess, Pia Kirk, Sarity Dodson and Tim Coombs

Version 1.1

May 2005

**national mental health strategy**

A joint Australian, State and
Territory Government Initiative

# Table of contents

# Executive Summary

## Background

The continued improvement of the quality and effectiveness of the treatment of people with a mental illness is one of the major objectives of the National Mental Health Strategy.  The Strategy recognises that this objective can only be achieved through the development of sound information to support planning and service delivery.  It has therefore fostered the use of routine outcome measurement, and all Australian states/territories have signed Information Development Agreements that require them to submit routinely-collected outcome and casemix data to the Australian Government.

The outcome data are collected via 10 standardised instruments that form part of what is known as the National Outcomes and Casemix Collection.  Specifically, these measures are:
- the Health of the Nation Outcome Scales (HoNOS);
- the Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA);
- the Health of the Nation Outcome Scales 65+ (HoNOS65+);
- the Life Skills Profile 16 (LSP-16);
- the Resource Utilisation Groups – Activities of Daily Living Scale (RUG-ADL);
- the Children's Global Assessment Scale (CGAS);
- the Mental Health Inventory (MHI);
- the Behaviour and Symptom Identification Scale 32 (BASIS-32®);
- the Kessler-10 Plus (K-10+); and
- the Strengths and Difficulties Questionnaire (SDQ).

The current report provides a critical review of these standardised measures, providing descriptive and evaluative information on each.

## Method

Articles and reports relating to the above instruments were retrieved through searches of electronic databases, the Internet and article reference lists.  Relevant descriptive and evaluative information was extracted from each article and report.  Specifically, the following categories of information were sought:
- Background
- Purpose
- Availability
- Description
- Versions
- Psychometric properties:
  - Validity (Content, Construct, Concurrent and Predictive)
  - Reliability (Test-retest and Inter-rater)
  - Sensitivity to change
  - Feasibility and utility

## Results

### HoNOS

The HoNOS was developed as an instrument that could be routinely used by clinicians in the United Kingdom to measure outcomes for consumers with a mental illness.  It can be regarded as a general measure of severity of symptoms for this group.  It comprises 12 items that collectively cover the sorts of problems that may be experienced by people with a mental illness.  Each item is rated from 0 (no problem) to 4 (very severe problem), resulting in individual item scores, subscale scores and a total score.

In tests of its psychometric properties, the HoNOS has been shown to have reasonably good content, construct, concurrent and predictive validity, and fair to moderate test-retest and inter-rater reliability. It is reasonably sensitive to change, at least for consumers in inpatient settings. Certain items have consistently been identified as problematic, particularly Item 8 (Other mental and behavioural problems), Item 11 (Problems with living conditions) and Item 12 (Problems with occupation and activities). The HoNOS has been shown to be generally acceptable to clinicians, but there is debate as to whether it will be feasible to use it as a routine outcome measure unless it also has some clinical utility, and its introduction is accompanied by appropriate resourcing, adequate infrastructure, regular feedback and ongoing training.

### HoNOSCA

Like its parent instrument, the HoNOS, the HoNOSCA arose out of the Health of the Nation Strategy in the United Kingdom. It was designed specifically as a brief measure that was relevant for routine use in child and adolescent mental health services. Structurally, it bears a strong resemblance to the HoNOS. It comprises 15 items that collapse into two sections, one comprising four subscales and the other one. Each item is scored on a scale of 0 (no problem) to 4 (severe problem).

Compared with the HoNOS, fewer studies have considered the HoNOSCA's psychometric properties. However, the studies that have been conducted provide a fairly consistent picture. Its content validity has not been adequately assessed, but it would appear to have reasonable construct and predictive validity. Tests of its concurrent validity indicate that it performs well against other clinician-rated measures and can discriminate between groups of consumers based on their clinical and/or treatment profiles. Its test-retest reliability and inter-rater reliability appear generally sound. It appears to be sensitive to change, as judged by the movement of scores over time, and, more particularly, its performance against other 'gold standards' that indicate improvement, deterioration or stability. Clinicians seem positive about its brevity and ease of use, clinical utility and ability to be incorporated into clinical practice (given adequate time and resources), but are wary about its applicability to children aged under five and consumers with particular disorders (e.g., anxorexia and somatoform disorders).

### HoNOS65+

The HoNOS65+ was developed in response to the observation that some modifications to its parent instrument, the HoNOS, might make it more appropriate for use with older consumers. The HoNOS65+ has essentially the same structure as the HoNOS, comprising 12 items which are each rated from 0 (no problem) to 4 (severe to very severe problem), yielding individual item scores, subscale scores and a total score. The key difference between the two instruments is the more detailed glossary developed for the HoNOS65+, which better reflects the reasons that older consumers come into contact with mental health services and makes it a discernibly different instrument.

Less work has been done on establishing the psychometric properties of the HoNOS65+ than the HoNOS. However, the studies that have been undertaken indicate that the HoNOS65+ has reasonably good (concurrent) validity and reliability, and is sensitive to change. Having said this, certain items perform better than others when assessed against relevant criteria. The HoNOS65+ appears to be acceptable to clinicians, but its successful implementation as a routine outcome measure is likely to require timely feedback to staff, minimal paperwork burden for clinicians, suitable infrastructure, appropriate data management and analysis and reporting that meets the requirements of clinicians.

### LSP-16

The LSP-16 is derived from the LSP-39. It is designed to measure the level of functioning and adaptation of people with a mental illness living in the community. It consists of 16 items that address issues faced when adapting to life in the community. Each item is rated on a four-point

scale ranging from 0 to 3 with high scores indicating higher disability, resulting in an individual item scores, subscale scores and a total score.

Taken together, studies examining the psychometric properties of the LSP suggest that it has: moderately good content, construct, concurrent and predictive validity; high test-retest and adequate inter-rater reliability; and good sensitivity to change. It also appears to demonstrate feasibility and utility in clinical settings. Having said this, relatively few studies have examined the LSP-16 specifically; far more have been concerned with the LSP-39. Additional analyses are needed to strengthen the evidence base regarding the shorter version.

### RUG-ADL

The RUG-ADL is a component of the RUG-III, a casemix classification system developed in the United States for use in long term care facilities for the elderly. The RUG-ADL is clinician-administered and measures consumer dependency or functional status related to four items (bed mobility, toileting, transfer and eating).

Psychometric testing has tended to focus on the RUG-III (or its predecessors) as a whole, rather than on the ADL component, and has emphasised its use as a casemix tool rather than an outcome measure. In terms of content validity, the RUG has been criticised for inadequately dealing with mental illness in general and dementia in particular (although the RUG-III represents an improvement in this regard), failing to account for the care demands associated with cognitive impairment, and measuring actual rather than potential performance. The construct and concurrent validity of the RUG have not been adequately tested, but it shows good predictive validity. Similarly, the test-retest reliability of the RUG has not been assessed, but its inter-rater reliability is sound. No studies have considered the sensitivity of the RUG to change in functional status. There has been little formal examination of the feasibility and utility of the RUG, but published commentaries have tended to be positive in this regard.

### CGAS

The CGAS is a clinician-administered instrument that provides an assessment of levels of functioning in children and adolescents (aged 4 to 16). It provides a single global rating, ranging from 1 (severe dysfunction) to 100 (superior functioning).

In the main, tests of the CGAS's psychometric properties have concentrated on concurrent validity and inter-rater reliability, where it has performed well. There is also evidence, albeit from a smaller number of studies, that the instrument has adequate content validity, predictive validity, test-retest reliability and sensitivity to change. Published commentaries have generally regarded the CGAS as having adequate feasibility and utility, although concerns have been expressed about its vulnerability to rater manipulation, its lack of any overarching organising principles, and its accuracy.

### MHI

The MHI is a self-report instrument developed for use in the RAND Health Insurance Experiment to assess psychological distress and wellbeing of people in the general population. Its full-length version comprises 38 items. Most items are rated on a scale from 1 to 6 based on frequency or intensity where higher scores reflect more frequent occurrence of favourable mental health symptoms.

In tests of its psychometric properties, the MHI has been shown to have adequate to good content, construct, concurrent, predictive validity, test-retest and inter-rater reliability. The MHI has demonstrated sensitivity to change as well as feasibility and utility.

*BASIS-32®*

The BASIS-32® is a consumer-rated measure that was originally developed to assess outcomes among inpatients with mental health problems. Subsequent studies have confirmed its utility as an outcome measure for use across a range of mental health settings. The instrument comprises 32 items which collectively measure symptoms and behavioural distress in people with a mental illness. Each item is rated from 0 (no difficulty) to 4 (extreme difficulty), resulting in 32 individual scores, five subscale scores and a single total score.

The BASIS-32® has been shown to have adequate validity and reliability, and to be sensitive to change during treatment (although the Impulsive and addictive behaviours subscale and the Psychosis subscale perform less well in these areas than do the other three subscales). Arguably, the instrument is also regarded as demonstrating adequate feasibility and utility. Having said this, it should be noted that the developers of the instrument have since prepared a revised version, the BASIS-24®, which only retains three of the original items.

*K-10+*

The K-10+ is a version of the K-10, which was developed for use as a measure of non-specific psychological distress. It was originally designed for use in the United States National Health Interview Survey, but was deliberately constructed in a manner that would allow it to have utility in clinical settings as well.

The K-10 is a 10-item self-report measure which asks the consumer about symptoms of depression and anxiety in the past four weeks. The K-10+ includes an additional four items that quantify the level of disruption and disability resulting from the problems identified in the first 10 items.

The K-10 is extremely widely used, both as a measure of mental health status in general population surveys and as an outcome measure in primary care settings, suggesting that it is well-regarded by the mental health field. The published studies on the psychometric properties of the K-10 are not extensive, but the instrument appears to have adequate to good content, construct and concurrent validity, and test-retest reliability. Its predictive validity and sensitivity to change require further exploration. So too do its feasibility and utility, although its brevity and widespread use in a range of settings augur well in this regard.

*SDQ*

The SDQ is a brief screening tool that describes children/adolescents' behaviours, emotions and relationships. Parent-rated, teacher-rated and self-report versions of the SDQ are available. The core instrument comprises 25 items that depict a positive or negative attribute and 'roll up' into five scales: Emotional symptoms; Conduct problems; Hyperactivity-inattention; Peer problems; and Prosocial behaviours. Extended versions of the SDQ include these items plus an impact supplement.

Extensive testing of the construct and concurrent validity and test-retest and inter-rater reliability of the SDQ has been undertaken, and the instrument is strong in terms of these psychometric properties. Less work has been done in the areas of content validity, predictive validity, sensitivity to change and feasibility and utility, but those studies that do provide information in this regard indicate that the instrument performs well on these dimensions as well.

## Conclusions

The current review provides evidence that the NOCC suite constitutes a group of measures that can collectively assess outcomes for different groups, from different perspectives, on a range of mental health-related constructs. Where they have been tested, they appear to perform adequately or better in terms of validity, reliability, sensitivity to change, and feasibility and utility.

To this extent, they can be regarded as appropriate for the purposes of monitoring outcomes for consumers, with a view to improving the quality and effectiveness of treatment.

# Chapter 1:  Background

## Routine outcome measurement and casemix development in public sector mental health services:  The policy context

The continued improvement of the quality and effectiveness of the treatment of people with a mental illness is one of the major objectives of the National Mental Health Strategy.  The Strategy recognises that this objective can only be achieved through the development of sound information to support service planning and delivery.  It has therefore fostered the use of routine outcome measurement to monitor services' impacts for consumers, and the development casemix systems to understand the role of provider variation in differences between services' costs and outcomes.

Under the First National Mental Health Plan (1992-1997),[1] considerable groundwork was undertaken in this regard.  Instruments that might have utility for routine outcome measurement were reviewed[2] and field-tested,[3] and a major casemix development project, known as the Mental Health Classification and Service Costs (MH-CASC) project, was undertaken.[4, 5] Under the Second National Mental Health Plan (1998-2003),[6] the systematic implementation of routine outcome measurement in all public sector mental health services became a priority, and an ambitious plan was put in place to develop information infrastructure in all public mental health services to support and encourage good clinical practice, regularly inform about consumer outcomes, inform judgements about value for money, and produce national and State/Territory data as a by-product.[7] The new National Mental Health Plan (2003-2008)[8] continues to provide a framework for the collection of outcomes and casemix data, supporting the comprehensive implementation and further development of routine consumer outcome measurement in mental health, and the reform of public sector funding models to better reflect need via continued development of mental health casemix classifications.

## A collaborative approach

State/territory governments and the Australian Government are collaborating in a coherent national approach to implement this policy framework 'on the ground'.

### *The role of the state/territories*

All states/territory jurisdictions prepared Information Development Plans and signed Information Development Agreements requiring them to provide the Australian Government with de-identified, consumer-level data comprising for 'outcomes dataset' and the 'casemix dataset', specified in the *National Mental Health Information Priorities* document.[7] Together, these datasets are referred to as the National Outcomes and Casemix Collection (NOCC).  The NOCC dataset is described in more detail below.

The Information Development Agreements also required states/territories to submit the consumer-level components of the National Minimum Data Set (NMDS) – Mental Health Care, as described in the most current version of the National Health Data Dictionary.

### *The role of the Australian Government*

For its part, the Australian Government has established three Expert Groups (Adult, Child and Adolescent and Older Persons) to advise on the implementation and use of routine outcome data in mental health services.  In addition, it has provided resources to support the training of the mental health workforce in the use of outcome measures, and setting up arrangements to receive, process, analyse and report on the NOCC and NMDS data submitted by states/territories.

***The role of the Australian Mental Health Outcomes and Classification Network (AMHOCN)***

The latter arrangements have been established by the Australian Government, under the banner 'Australian Mental Health Outcomes and Classification Network' (AMHOCN). AMHOCN is a consortium of three groups, and has been funded to provide national leadership in the development of outcomes and casemix concepts in mental health. It is undertaking a work program with three components, each of which is the remit of one of the consortium members: a data management component (Strategic Data Pty Ltd, Victoria); an analysis and reporting component (The University of Queensland, Queensland); and a training and service development component (Institute of Psychiatry, New South Wales). The three components of AMHOCN were each contracted in late 2003, and first came together as a consortium in December of that year.

## The National Outcomes and Casemix Collection (NOCC)

The NOCC dataset comprises 10 standardised instruments (see Table 1), and four additional measures (see Table 2). Selection of these instruments was guided by the developmental work conducted under the first National Mental Health Plan (described above).

As Table 1 shows, the standardised instruments represent a mix of clinician-rated, consumer-rated and consumer- and parent-rated measures, and the specific instrument(s) used at a given data collection occasion depend on the age group of the consumer (adults, older persons, children and adolescents). Although not shown in Table 1, the type of episode (inpatient, ambulatory, residential) and the reason for collection (admission – new referral, admission – admitted from other treatment setting, admission – other, review – 3-month review, review – other, discharge – no further care, discharge – change of treatment setting, discharge – death, discharge – other) also influence the specific instrument(s) used. The contextual rules governing the administration of specific instruments are described in detail elsewhere.[9]

As indicated in Table 2, the four additional measures are factors influencing health status, focus of care, mental health legal status, and principal and additional diagnoses. Again, the protocol dictating their collection is governed by consumer age group (shown in Table 2), and factors related to the type of episode and reason for collection (not shown in Table 2).

**Table 1:  Standardised instruments included in the NOCC dataset**

| | | Adults | Older persons | Children and adolescents |
|---|---|---|---|---|
| **Clinician-rated** | Health of the Nation Outcome Scales (HoNOS) | √ | | |
| | Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA) | | | √ |
| | Health of the Nation Outcome Scales 65+ (HoNOS65+) | | √ | |
| | Life Skills Profile 16 (LSP-16) | √ | √ | |
| | Resource Utilisation Groups – Activities of Daily Living Scale (RUG-ADL) | | √ | |
| | Children's Global Assessment Scale (CGAS) | | | √ |
| **Consumer-rated** | Mental Health Inventory (MHI) or Behaviour and Symptom Identification Scale 32 (BASIS-32®) or Kessler-10 Plus (K-10+) | √ | √ | |
| **Consumer- and parent-rated** | Strengths and Difficulties Questionnaire (SDQ) | | | √ |

Source: Department of Health and Ageing (2003)[9]


**Table 2:  Additional measures included in the NOCC dataset**

| | | Adults | Older persons | Children and adolescents |
|---|---|---|---|---|
| **Clinician-rated** | Factors Influencing Health Status (FIHS) | | | √ |
| | Focus of Care | √ | √ | |
| | Mental Health Legal Status | √ | √ | √ |
| | Principal and Additional Diagnoses | √ | √ | √ |

Source: Department of Health and Ageing (2003)[9]

**Purpose and scope of the current review**

There is no up-to-date, comprehensive review of the standardised instruments that underpin the NOCC dataset. The original review of instruments was extremely valuable in homing in on those with the greatest potential for use in routine outcome measurement, and played a large part in shaping the specific measures that are in use today.[2] However, it was conducted in 1994, and much has been written about many of these measures in the intervening ten year period. In addition, some of the measures that are included in the current suite were not covered in the original review.

The current review critically considers the standardised instruments in the NOCC suite of measures. It provides descriptive and evaluative information on each measure, in a manner that will hopefully prove useful to clinicians and managers involved in routine mental health outcome measurement in Australia, as well as others with a more general interest in the area.

The additional measures that do not constitute standardised instruments are out of scope of the review, primarily because of the dearth of published literature relating to them.

**Report structure**

The remainder of this report describes the review. Chapter 2 provides a description of the method by which the review was conducted. Chapters 3 to 11 are organised around the instruments, and present the salient findings from the review. Chapter 12 summarises these findings, and discusses them in the context of current efforts in the area of routine outcome measurement in Australia.

# Chapter 2: Method

## Article retrieval

Searches of the electronic databases MEDLINE and PSYCINFO were conducted from their respective years of inception to September 2004. The search was retrieved articles using the following search terms:

- MENTAL HEALTH or PSYCHIATR*
- OUTCOME MEASURE* or ROUTINE OUTCOME MEASURE*;
- HEALTH OF THE NATION OUTCOME SCALES or HONOS;
- HEALTH OF THE NATION OUTCOME SCALES 65+ or HONOS65+;
- HEALTH OF THE NATION OUTCOME SCALES FOR CHILDREN AND ADOLESCENTS or HONOSCA;
- LIFE SKILLS PROFILE or LSP;
- RESOURCE UTILISATION GROUPS – ACTIVITIES OF DAILY LIVING or RUG-ADL;
- CHILDREN'S GLOBAL ASSESSMENT SCALE or CGAS;
- MENTAL HEALTH INVENTORY or MHI;
- BEHAVIOUR AND SYMPTOM IDENTIFICATION SCALE or BASIS-32®;
- KESSLER 10 PLUS or K-10+; and
- STRENGTHS AND DIFFICULTIES QUESTIONNAIRE or SDQ.

Potentially relevant peer-reviewed journal articles were retrieved by this means, and their reference lists scanned for further pertinent articles. Efforts were also made to retrieve government and other reports, both from within Australia and overseas, largely by conducting Internet searches using the above terms. Greatest weight was given to the peer-reviewed articles for two reasons. Firstly, it was possible to be confident that they had undergone some academic checking for scientific merit. Secondly, this approach created a relatively 'level playing field' for all instruments. It is acknowledged, however, that the relative standing of the given journal was not taken into account, and the individual studies were not systematically rated for quality (although consideration was given to the strength of their design).

In addition, the review primarily concerned itself with articles (and reports) that involved explicit testing of the psychometric properties of a given instrument (e.g., a study that examined the validity and reliability of the HoNOS). Articles that described the use of a given instrument in a study of some other kind (e.g., a study that used the BASIS-32® as an outcome measure in assessing the relative merits of two different types of treatment) were given less weight. This decision was made on the grounds that the latter type of study, by design, implicitly accepted the psychometric value of the given instrument and to use the findings as evidence for the psychometric robustness of that instrument would create a somewhat circular argument.

## Critical appraisal of NOCC instruments

Evidence from the above articles and reports was used to critically appraise each of the instruments in the NOCC dataset. The critical appraisal exercise was guided by a checklist that drew on the work of Greenhalgh et al,[10] Green and Gracely[11] and McDowell and Newell.[12]

Specifically, the checklist elicited descriptive and evaluative information on each instrument. The descriptive information covered the background and purpose of the given instrument, and provided a description of its structure, administration and scoring. The evaluative information considered the validity, reliability, sensitivity to change and feasibility/utility of the instrument. Each of these dimensions is described in more detail below.

### Background

The development of each instrument was described in terms of why it was perceived to have been needed and the rationale behind its design. Where possible, information was provided on the settings or application areas in which it has been used.

### Purpose

Consideration was given to what the given instrument aims to measure, and whether it was designed as an outcome measure, or simply as a measure of health status.

### Availability

Information was provided regarding whether the given instrument was in the public domain or commercially available.

### Description

Each instrument was described in terms of its main domains, and number of items and subscales. The response format was described, as was the scoring method. Consideration was given to whose perspective it captures, and how it is administered. Comment was also provided, wherever possible, on the training requirements associated with the instrument.

### Versions

The different versions of each instrument were described. A broad definition of the term 'version' was used, which included any alternatives to the original instrument including. In scope were:
- Abbreviated versions of the original instrument (e.g., versions where original items had been omitted);
- Extended versions of the original instrument (e.g., versions where additional items had been included);
- Modified versions of the original instrument (e.g., versions where the structure or content of items had been changed, or where the rating period under consideration had been lengthened or shortened);
- Versions that are completed by different classes of informants; and
- Non-English language versions of the original instrument.

All major versions of any given instrument were considered, irrespective of whether the alterations to the original version had been made by the original developers or by another party (including stakeholders in the NOCC process).

### Psychometric properties: Validity

Judgements were made about the validity of each instrument, on the basis of the available evidence. Validity was defined as 'the extent to which the instrument measures what it intends to measure'.[10] The types of validity that were explicitly considered were:

- **Content validity**, which refers to the instrument's comprehensiveness (i.e., how adequately the sampling of items reflects its aims), and is commonly ascertained by asking consumers, clinicians and experts in the field to review the content of the instrument;[10]

- **Construct validity**, which involves conceptually defining the construct to be measured by the instrument, and assessing the internal structure of its components and the theoretical relationship of its item and subscale scores;[10] and

- **Criterion validity**, which assesses the extent to which the instrument correlates with a 'gold standard' or more established measure of the same theme. Criterion validity can be split into:

  - **Concurrent validity**, which pits the instrument against a comparable measure or measures at the same point in time; and

  - **Predictive validity**, which assesses its ability to predict a future outcome, such as resource use (particularly important in the context of casemix development) or treatment response.[10]

### *Psychometric properties: Reliability*

Each instrument was assessed in terms of its reliability. Reliability can be viewed as the extent to which a given instrument gives stable, consistent results.[10] Alternatively, it can be considered as the inverse of the degree of error obtained from any measurement. Two types of reliability were considered:

- **Test-retest reliability**, or the degree of agreement when the same instrument is applied to the same consumer by the same rater at two different points in time;[10] and

- **Inter-rater reliability**, or the degree of agreement when the same instrument is applied to the same consumer by different raters at the same point in time.[10]

The level of reliability of an instrument is traditionally measured by a kappa value. Kappas of $\leq 0.20$ are regarded as poor, 0.21-0.40 as fair, 0.41-0.60 as moderate, 0.61-0.80 as good, and $\geq 0.81$ as very good.[13, 14]

### *Psychometric properties: Sensitivity to change*

Sensitivity to change is related to both validity and reliability. Indeed, it has been argued that it is the most stringent test of validity because it compounds the error component of initial and follow-up scores on any given instrument.[15] In other words, an instrument that is both valid and reliable, and which demonstrates change over time, can be regarded as being sensitive to change. Consideration was given to the sensitivity to change of each instrument.

### *Psychometric properties: Feasibility and utility*

The feasibility of implementing the given instrument as a routine measure of outcome was assessed, as was its acceptability to and utility for consumers, carers, clinicians and managers. Note that information on feasibility and utility excluded information on the instrument's comprehensiveness, covered under the aegis of content validity (see above).

# Chapter 3:  Health of the Nation Outcome Scales (HoNOS)

## Background

The HoNOS was developed by Wing and colleagues from the College Research Unit of the Royal College of Psychiatrists in the United Kingdom, as a means of assessing the extent to which the Government's Health of the Nation target 'to improve significantly the health and social functioning of mentally ill people' was being met.  Specifically, it was designed as an instrument that could be used routinely by clinicians in the United Kingdom's National Health Service to measure outcomes for consumers with a mental illness.[16, 17]

## Purpose

The HoNOS can be regarded as a general measure of mental health and social functioning in people with a mental illness.[16]

## Availability

The HoNOS is in the public domain, and can be used at no cost.[18]

## Description

The HoNOS is a clinician-administered instrument comprising 12 items:
- Item 1: Overactive, aggressive, disruptive or agitated behaviour;
- Item 2: Non-accidental self-injury;
- Item 3: Problem drinking or drug taking;
- Item 4: Cognitive problems;
- Item 5: Physical illness or disability problems;
- Item 6: Problems associated with hallucinations and delusions;
- Item 7: Problems with depressed mood;
- Item 8: Other mental and behavioural problems;
- Item 9: Problems with relationships;
- Item 10: Problems with activities of daily living;
- Item 11: Problems with living conditions; and
- Item 12: Problems with occupation and activities.[99]

Collectively, the items cover the sorts of problems that may be experienced by people with a mental illness.  The items 'roll up' into four subscales:
- Behaviour (Items 1-3);
- Impairment (Items 4-5);
- Symptoms (Items 6-8); and
- Social (Items 9-12).

Each item is rated on a five-point scale (0 = no problem; 1 = minor problem; 2 = mild problem; 3 = moderately severe problem; 4 = very severe problem), resulting in individual item scores, subscale scores and a total score.  In assigning ratings, the clinician makes use of a glossary which details the meaning of each point on the item being rated.[19]

The clinician rates the consumer on each of the items in terms of their assessment of the consumer's situation over the recent period.  The specific instruction is to 'rate the most severe problem that occurred in the period rated' for all items except Item 11 and Item 12 where the instruction is to 'rate the usual [situation]'.  For the purposes of the NOCC collection, this modification has been extended to the other items in the Social subscale as well (i.e., Items 9 and 10); in training, clinicians have been instructed to 'rate the usual or typical [situation]'.[20]

For the purposes of the NOCC collection, the period is the standard two weeks at almost all collection occasions. The exception is discharge from an inpatient setting, where the rating period is three days in recognition of the brevity of episodes in such settings.

The clinician is expected to draw on all relevant and useful information to make their ratings, and will typically use information obtained from interview with or direct observation of the consumer, medical records, and consultation with his/her family, carer(s) and case worker(s).[21] The HoNOS can be administered in less than five minutes, once the clinician becomes familiar with the items.[22] The HoNOS is designed to be used by any mental health clinician – psychiatrists, psychologists, nurses, social workers etc – providing they have been appropriately trained.[23] One day training is recommended initially, with a half day retraining every two years.[22] Various training packages and resources (e.g., vignettes) have been developed.[17, 24]

## Versions

A number of iterations of the HoNOS were developed and field tested by the original developers, culminating in the version that is in current use.[16, 17] Since its original development, the HoNOS has been translated into a number of languages, including Danish,[25] French,[26, 27] Italian,[28, 29] German[30] and Norwegian.[31]

As noted above, two versions of the HoNOS are used in the NOCC collection. These vary only by rating period. The main version, which is used at almost all collection occasions, has a rating period of two weeks (as prescribed in the original instrument). The alternative version, which is used at discharge from an inpatient setting, has a rating period of three days, in recognition of the brevity of episodes in such settings.

Studies that have examined the psychometric properties of the HoNOS (see below) have focused on the two-week version of the HoNOS. Most, but not all, have considered the English-language version.

## Psychometric properties

Relatively extensive testing has been undertaken of the psychometric properties of the HoNOS.

### *Content validity*

Several attempts have been made to explore the content validity of the HoNOS. Shergill et al,[32] Orrell et al[33] and McClelland et al[34] asked consumer/carer advocacy groups and mental health professionals to comment on whether the HoNOS items reflected areas of concern for them. In the main, respondents in these studies were positive, suggesting that the HoNOS was appropriate, well-designed and thorough, and highlights consumers' problems quickly, indicating changes in their mental health status over time.[32-34]

However, there were some reservations about specific items. Respondents were concerned about the restriction imposed by the rater being forced to indicate only one problem in Item 8 (Other mental and behavioural problems),[32, 34] and expressed concerns about the ability of Item 6 (Problems associated with hallucinations and delusions) to given an accurate picture of the symptoms and role performance of a person with schizophrenia.[33] They also felt that the social items (Items 10, 11 and 12) were problematic because the information needed to rate them is complex and/or not always available.[33, 34]

More generally, respondents in these studies noted that, for some items, anchor points and their associated terminology are subjective and might require clarification.[32, 33] They also commented on difficulties with knowing which item to use for rating some symptoms, such as elated mood. In addition, they observed some omissions, such as the instrument's inability to take into account factors such as culture, poverty, abuse, safety and risk, bereavement and medication compliance.[32, 33] In addition, they noted some limitations that were introduced by human error in

the use of the HoNOS, specifically warning that the guidelines contained in the glossary are not always adhered to, and that ratings are influenced by rapport with and knowledge of the consumer, which is likely to increase over time.[34] Some respondents went so far as to say that the HoNOS is a blunt instrument and may be open to misinterpretation.[34]

## *Construct validity*

A number of studies have examined the internal consistency of the HoNOS, as measured by Cronbach's alpha (derived from the mean inter-item correlation and the total number of items).[35] In these studies, Cronbach's alpha has ranged from 0.59 to 0.76, indicating that the HoNOS has a moderately high level of internal consistency and low levels of item redundancy, supporting its use as a meaningful summary of severity of symptoms.[3, 16, 32-34, 36, 37] Having said this, Trauer has argued that the HoNOS should not be regarded as unidimensional, measuring a single, underlying construct of mental health status. Instead, it should be viewed in the context of its original intention, namely to provide a broad coverage of the problems typically experienced by consumers of mental health services.[36, 38]

McClelland et al[34] examined the relative contribution of each of the HoNOS items to the total score, and found that Item 7 (Problems with depressed mood), Item 8 (Other mental and behavioural problems) and Item 9 (Problems with relationships) had the greatest weight, contributing 15%, 19% and 14% to the total, respectively. By contrast, Item 11 (Problems with living conditions) and Item 12 (Problems with occupation and activities) contributed only 3% each.

Preston,[39] Trauer[36] and McClelland[34] examined the subscale structure of the HoNOS, in separate studies that employed confirmatory factor analysis and principal components analysis. In his study, Preston found that the four factor model defined by the original subscales had good fit, but that the contribution of individual items to their respective subscales varied in two separate mental health services, indicating differentiation in construct interpretation. Trauer's examination of the subscales revealed a poorer fit than Preston's, leading him to propose an alternative structure comprising: Behaviour (Items 1 and 3); Impairment (Items 4 and 5); Hallucinations and delusions (Item 6); Social Problems (Items 9, 10, 11 and 12); and Depression (Items 2, 7, 8 and 9). McClelland's study also identified alternative factors: Severity of illness (Items 1, 3, 6, 7, 9, 10, 11 and 12); Suicidality/depression (Items 2 and 4); Physical illness (Item 5); and Other mental and behavioural problems (Item 8).

## *Concurrent validity*

Numerous studies have considered the concurrent validity of the HoNOS, assessing its individual item scores, subscale scores and total scores on the HoNOS in terms of their correlation with relevant scores on more established instruments that have been shown to validly measure related constructs. In the main, the HoNOS has been shown to perform well against clinician-rated instrument such as the Role Functioning Scale (RFS),[16] the Brief Psychiatric Rating Scale (BPRS),[16, 32-34] the GAF scales for symptoms (GAF-s) and disability (GAF-d) from the Global Assessment Scale (GAS),[32-34, 40-42] the Life Skills Profile (LSP),[40] the Manchester Audit Tool (MAT),[43] the Behaviour Rating Scale from the Clifton Assessment Procedures for the Elderly (CAPE-BRS),[32] the Clinical Dementia Rating,[32] the Mini-Mental State Examination (MMSE),[32] Schedules for Clinical Assessment in Neuropsychiatry (SCAN),[15, 42] the Broad Rating Schedule (BRS),[42] the Disability Assessment Schedule (DAS),[42] the Scale for Assessment of Negative Symptoms (SANS),[42] Location of Community Support Scale (LOCSS),[33] the Social Behaviour Scale (SBS),[15, 33] the Hamilton Rating Scale for Depression (Ham-D),[44] and the Positive and Negative Symptoms Scale (PANSS).[44] There are some exceptions, with low correlations being found between the HoNOS and the BPRS in one study[45] and the Beck Depression Inventory (BDI) in another.[37] As a general rule, the strength of correlation has been found to be higher when research workers rather than case managers or other clinical workers conduct the assessment,[15, 42] Not surprisingly, higher correlations have also been found in studies where consideration has been given to the theoretical relationship between the specific item or subscale and the comparison instrument of interest.[44]

By contrast, the HoNOS has shown poor or mixed performance against consumer-rated instruments such as the Symptom Checklist 90-Revised (SCL-90-R),[45, 46] the Social Adjustment Scale (SAS),[45] the Medical Outcome Study Short Form 36 (SF-36),[46] the Camberwell Assessment of Need Short Appraisal Schedule (CANSAS),[47] an instrument adapted from the Quality of Life Index for Mental Health (QLI-MH),[40] the Quality of Life assessment (QoL),[32] the Avon Mental Health Measure (AVON),[48] the Outcome of Problems of Users of Services (OPUS),[48] and even a self-rating version of the HoNOS with a similar question structure.[49] As with the clinician-rated measures, there are exceptions to the general rule, but even where studies have reported correlations between the HoNOS and consumer-rated measures – e.g., the CANSAS,[50] the SF-36,[33, 44] the General Health Questionnaire (GHQ)[33] and the Comprehensive Quality of Life scale (ComQOL)[44] – they tend to vary across domains and be lower than those between the HoNOS and clinician-rated measures. These findings are not surprising, given that poorer correspondence is typically found between instruments that rely on information from informants of different classes than those which rely on information from informants of the same class, since different informants have access to different information.

Another frequently-used method of examining the concurrent validity of the HoNOS has been to consider its ability to discriminate between consumer groups differentiated on a range of treatment- and service-based indicators. Several studies have found high total scores on the HoNOS to be associated with diagnoses of drug and alcohol, psychotic and bipolar disorders, high scores on items relating to hallucinations/delusions and social and cognitive problems to be associated with a diagnosis of schizophrenia, high scores on items relating to aggressive behaviour, drinking/drug taking and anxiety to be associated with a diagnosis of mania, and high scores on items relating to suicidal thoughts/behaviours, physical illness and depressed mood to be associated with a diagnosis of depression.[25, 26, 34, 41, 43] Similarly, a number of studies have found that the HoNOS can discriminate between consumers with differing levels of need or disability, as indicated by their current or expected location of treatment – e.g., those receiving standard case management versus those assertive case management,[51] those in residential/nursing home, day patient, outpatient and inpatient settings,[32, 33, 44], and those in long-stay settings with low, medium and high expectations of discharge.[52]

### Predictive validity

Several studies have examined the predictive validity of the HoNOS. Most have found the instrument to have reasonably good predictive validity, explaining a significant proportion of the variance in resource use (e.g., as measured by service contacts, length of stay and costs) and treatment outcome (e.g., as measured by readmission rates, retention in the community, treatment response and death).[40, 44, 53-55] There have been exceptions to this rule, however, with some studies finding no correspondence or only a small association between the HoNOS and resource use.[56, 57]

### Test-retest reliability

Comparatively few studies have examined the test-retest reliability of the HoNOS, but those that have generally report fair to moderate overall reliability scores.[32, 33, 46] Particularly low reliability scores have been reported for Item 1 (Overactive, aggressive, disruptive or agitated behaviour), Item 3 (Problem drinking or drug taking), Item 7 (Problems with depressed mood), and Item 10 (Problems with activities of daily living).

### Inter-rater reliability

Far more studies have assessed the inter-rater reliability of the HoNOS. Most of these have found that the overall agreement between pairs of raters is fair to moderate,[15, 32, 46] or even moderate to good,[16, 33, 42, 44] but that agreement is poor on particular items. Items identified as problematic include Item 4 (Cognitive problems),[15] Item 7 (Problems with depressed mood),[15] Item 8 (Other mental and behavioural problems),[15, 16] Item 9 (Problems with relationships),[33] Item 11 (Problems with living conditions)[33, 58] and Item 12 (Problems with occupation and activities).[15,

[16, 58] Factors associated with improved inter-rater reliability include training, familiarity with consumers and setting.[46 58]

### Sensitivity to change

The ability of the HoNOS to detect genuine improvement or deterioration (or indeed stability) in symptoms has been assessed in a number of studies. Typically, these studies have examined the extent to which the direction and magnitude of movement in the total HoNOS score and/or in individual item scores correlates with some external measure of change.

The simplest of these studies have examined change in HoNOS over time in given settings, hypothesising that there should be a decrease in severity as the consumer nears the end of an episode. In general, these studies have found decreases of the greatest magnitude in inpatient settings and of lesser magnitude in community settings,[23, 34, 58, 59] suggesting that the HoNOS may be more sensitive to change in the former environment. Having said this, there is some evidence that there may be an interaction between setting, diagnosis and severity, and that the HoNOS may be able to detect change in the community for those with depression and anxiety.[43] Particular items may also interact with setting, with one study that took in the range of inpatient and community settings finding that scores on all items except Item 11 (Problems with living conditions) showed decreases over time,[34] and another that concentrated on a community setting only finding that only Items 7 (Problems with depressed mood), 8 (Other mental and behavioural problems) and 9 (Problems with relationships) had sufficient relevance and variability to change over time.[59]

Other studies have used clinician or consumer judgement as the 'gold standard' against which to judge whether change has occurred and, if so, whether the HoNOS is capable of detecting it. In separate studies, Taylor and Wilkinson[60] and Gallagher and Teesson[51] found correlations between changes in consumers' HoNOS scores and clinical judgements about whether they had improved, remained stable or deteriorated made by general practitioners (GPs) and case managers, respectively. Likewise, Hunter et al[48] found that significant decreases in total HoNOS scores between initial and repeat ratings corresponded with consumers' self report of their goals having been met.

Still other studies have compared the HoNOS's dynamic properties and capacity to detect change against other, more established measures of outcome. Using these criteria, McClelland et al[34] found the HoNOS to perform commensurately with the GAS and the BPRS. Sharma et al[61] found it performed well against the CGI, although the correlations were greatest for those with extreme improvement or deterioration. Ashaye et al[55] found the HoNOS was correlated with the CAPE-BRS and two quality of life scales in elderly consumers, particularly those with dementia and depression. By contrast, Bebbington et al[15] found the HoNOS performed poorly by comparison with the SCAN and the SBS.

A final approach to examining sensitivity to change has involved assessing whether improvements in HoNOS scores are observed for consumers who receive evidence-based therapies and therefore would be expected to show reductions in symptom severity. Bech et al,[25] for example, hypothesised that consumers who received lithium and/or electro-convulsive therapy would show greater improvement on the HoNOS than consumers who did not, and found this to be the case, at least for the Behaviour and Symptoms subscales.

### Feasibility and utility

There has been considerable debate about the feasibility and utility of the HoNOS. The most positive authors have suggested although no instrument will fulfil all needs, the HoNOS is a comprehensive, user-friendly tool that is likely to have utility in routine outcome measurement (possibly with minor modifications).[16, 26, 34, 37, 44, 51, 62] Less enthusiastic published commentaries have argued that while it is acceptable to clinicians and feasible to administer during routine outcome measurement, it is of limited value in informing care planning, restricting its application to being a pure research instrument.[41, 61, 63-65] This has been countered by the view that the

HoNOS alone would not be expected to guide day-to-day clinical practice, but could make a valuable contribution in conjunction with other pieces of evidence that normally form the basis for clinical judgements.[17]

Audits of the extent to which the HoNOS is being used in particular settings have generally lent support to the former view. Glover and Sinclair-Smith,[66] for example, surveyed mental health care provider trusts in Britain in 2000, and discovered that 60% had implemented (or at least piloted) routine outcome measurement, with the majority using the HoNOS among the outcome measures of choice. They predicted that full coverage would be achieved within four years. Likewise, James and Kehoe[67] audited a random sample of consumers of a district service in the United Kingdom, and found that 77% had HoNOS scores recorded in their care plans. The latter finding was corroborated in a similar study in the United Kingdom conducted by Broadbent, who found that the HoNOS was completed for the vast majority of consumers on an electronic case register.[53]

Evidence from studies that have elicited the views of clinicians about their experiences with using the HoNOS has been more mixed. In separate studies conducted in the United Kingdom, James and Kehoe,[67] Broadbent,[53] and Milne et al[68] found that clinicians were relatively positive about the HoNOS, viewing it as potentially useful, but insisting that its ongoing use would depend on appropriate resourcing, adequate infrastructure, regular feedback and ongoing training. By contrast, Gilbody surveyed United Kingdom psychiatrists and found that only 26% were asked to collect the HoNOS by their trusts, and many questioned the usefulness of outcome measures, believed that the infrastructure was not in place for routine outcome measurement, and did not find routine outcome measurement informed their clinical decisions.[65] In field trials conducted at five sites in Victoria, Trauer found that clinicians at one site were extremely positive about the HoNOS, whereas those at the other four were more ambivalent, believing that it contributed only minimally to their treatment practices.[69]

## Summary

The HoNOS was developed as an instrument that could be routinely used by clinicians in the United Kingdom to measure outcomes for consumers with a mental illness. It can be regarded as a general measure of severity of symptoms for this group. It comprises 12 items that collectively cover the sorts of problems that may be experienced by people with a mental illness. Each item is rated from 0 (no problem) to 4 (very severe problem), resulting in individual item scores, subscale scores and a total score.

In tests of its psychometric properties, the HoNOS has been shown to have reasonably good content, construct, concurrent and predictive validity, and fair to moderate test-retest and inter-rater reliability. It is reasonably sensitive to change, at least for consumers in inpatient settings. Certain items have consistently been identified as problematic, particularly Item 8 (Other mental and behavioural problems), Item 11 (Problems with living conditions) and Item 12 (Problems with occupation and activities). The HoNOS has been shown to be generally acceptable to clinicians, but there is debate as to whether it will be feasible to use it as a routine outcome measure unless it also has some clinical utility, and its introduction is accompanied by appropriate resourcing, adequate infrastructure, regular feedback and ongoing training.

# Chapter 4: Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA)

## Background

Like the HoNOS, the HoNOSCA arose out of the Health of the Nation strategy in the United Kingdom. The HoNOSCA was developed by the Department of Child and Adolescent Psychiatry at the University of Manchester, in conjunction with the College Research Unit from the Royal College of Psychiatrists, and with the assistance of a multi-disciplinary steering committee. It was welcomed as having the potential to fill a gap identified by several reviews[70, 71] – namely that existing instruments were too lengthy or too specific to be useful for measuring global outcomes for children and adolescents in routine clinical practice.[72, 73]

Several imperatives drove the development of the HoNOSCA. It had to be brief, and share a similar structure with its parent instrument, the HoNOS. It had to be applicable to the consumer profile of child and adolescent mental health services, in terms of both age range and clinical characteristics. It had to provide a quantitative measure of severity that demonstrated sound psychometric properties, including validity, reliability and sensitivity to change. And it had to be acceptable and useful to clinicians, managers and funders.[72]

## Purpose

The HoNOSCA can be regarded as a broad measure of the range of behavioural, symptomatic, social and impairment domains in children and adolescents.[72]

## Availability

Like the HoNOS, the HoNOS is in the public domain, and can be used free of charge.[18]

## Description

The HoNOSCA is a clinician-administered instrument comprising 15 items:
- Item 1: Problems with disruptive, antisocial or aggressive behaviour;
- Item 2: Problems with over-activity, attention or concentration;
- Item 3: Non-accidental self-injury;
- Item 4: Problems with alcohol, substance or solvent misuse;
- Item 5: Problems with scholastic or language skills;
- Item 6: Physical illness or disability problems;
- Item 7: Problems associated with hallucinations, delusions or abnormal perceptions;
- Item 8: Problems with non-organic somatic symptoms;
- Item 9: Problems with emotional and related symptoms;
- Item 10: Problems with peer relationships;
- Item 11: Problems with self-care and independence;
- Item 12: Problems with family life and relationships;
- Item 13: Poor school attendance;
- Item 14: Problems with knowledge or understanding about the nature of the child or adolescent's difficulties; and
- Item 15: Problems with lack of information about services or management of the child or adolescent's difficulties.[72, 73]

Items 1-13 comprise Section A, and collectively cover typical problems experienced by children and adolescents presenting to mental health services. These items can be further grouped into four subscales:
- Behaviour (Items 1-4);
- Impairment (Items 5 and 6);

- Symptom (Items 7-9); and
- Social (Items 10-13)[74]

Items 14 and 15 comprise Section B. This section covers lack of information about difficulties and services, usually on the part of the parent of the child or adolescent.[72, 73]

The clinician is asked to draw on all available information, and to rate each item on a scale of 0 (no problem) to 4 (severe problem) for the period under consideration. For the purposes of the NOCC collection, the period is the standard two weeks at almost all collection occasions. The exception is discharge from an inpatient setting, where the rating period is three days in recognition of the brevity of episodes in such settings.

For Items 1-9, the clinician is asked to 'rate the most severe problem that occurred in the period rated'. Items 10-13 require 'a more general rating'.[75] A comprehensive glossary provides descriptions of the anchor points associated with each item.[76] Once completed, the HoNOSCA is scored in a manner that provides individual item scores, subscale scores and a total score. The total score comprises the items in Section A – Items 1-13 – only.[72, 73]

Like the HoNOS, the HoNOSCA is designed to be used by clinicians of any discipline, subject to their receiving adequate training. Various training resources have been developed to assist in this regard, and these have been distributed widely both within the United Kingdom and internationally.[73] Reports on the typical time taken by trained clinicians to complete the HoNOSCA vary from two to 20 minutes, with an average of 5-10 minutes.[72-74, 77, 78]

## Versions

As noted above, two versions of the HoNOSCA are used in the NOCC collection. These vary only by rating period. The main version, which is used at almost all collection occasions, has a rating period of two weeks (as prescribed in the original instrument). The alternative version, which is used at discharge from an inpatient setting, has a rating period of three days, in recognition of the brevity of episodes in such settings.

Studies that have examined the psychometric properties of the HoNOSCA (see below) have focused on the two-week version of the instrument.

## Psychometric properties

The psychometric properties of the HoNOSCA have been examined, although not as extensively as those of the HoNOS.

### Content validity

As yet, there are no published studies that provide evidence about the content validity of the HoNOSCA, although there are several that have considered the related concepts of its feasibility and utility (see below).

### Construct validity

The construct validity of the HoNOSCA was examined in the original field work that underpinned its development. Specifically, Gowers et al[72, 73] examined the internal structure of the HoNOSCA, in terms of both individual items and subscales. They considered the correlations between the individual items and found them to be low, which they took as evidence that each item carried independent weight. They then examined the factor structure of the HoNOSCA, and found that it generally mirrored the instrument's subscales. Brann,[79] by contrast, also examined the factor structure of the HoNOSCA and produced preliminary evidence for a different set of factors – externalising, internalising, developmental and emerging marginalised. Neither Gowers et al nor Brann found support for the instrument's sections. Indeed, Brann has argued that the total score

and the individual item scores provide the main useful information, and that the subscale scores and section scores should be treated with caution.

Gowers et al[72, 73] also considered the extent to which the total HoNOSCA score accurately reflected clinical severity, essentially arguing that high total scores should more frequently be associated with high scores on a few items than on mild to moderate scores on a number of items. They found that the total score increased as a linear function of high individual item scores, a finding confirmed by Brann et al[80] in a subsequent study.

### *Concurrent validity*

Several studies have examined the concurrent validity of the HoNOSCA, weighing up its performance against other measures. Studies that have examined the correlation between the total HoNOSCA score and scores on other clinician-rated measures have typically reported moderate correlations (r=0.6 or above). This was the case in a study by Yates et al,[77] which compared the HoNOSCA with the Children's Global Assessment Scale (CGAS) and Paddington Complexity Scale (PCS), and in a study by Bilenberg where the 'gold standard' was the Global Assessment of Psychosocial Disability.[81]

Studies that have evaluated the HoNOSCA against parent- and child/adolescent-rated instruments have typically produced lower correlations. Yates et al[77] found significant but at best only modest correlations between the HoNOSCA and the Behaviour Check List (BCL), the Strengths and Difficulties Questionnaire (SDQ), the Child Health Related Quality of Life Questionnaires (CHRQOL) and the Modified Harter Self-Esteem Questionnaire (Harter). Gowers et al found overall low levels of agreement between the HoNOSCA and the HoNOSCA-SR (a consumer-rated version of the instrument applicable to adolescents only) at an individual level, although some groups (e.g., outpatients with eating disorders) provided an exception to this rule.[82] As noted in the equivalent subsection on the HoNOS (see Chapter 3, above), these findings are to be expected, given that instruments that rely on information from different classes of informants are likely to demonstrate lower levels of correspondence than those that rely on informants from the same class.

Other studies have used a different methodology to examine concurrent validity, assessing the ability of the HoNOSCA to discriminate between groups of consumers based on their clinical and/or treatment profile. Gowers et al[72, 73] and Yates et al[77] found that the HoNOSCA could distinguish between consumers in inpatient and outpatient settings and between consumers presenting to clinics with different areas of focus, respectively, with the HoNOSCA profiles according with expectations. Manderson and McCune[74] and Brann et al[80] found that the HoNOSCA yielded coherent results by sex and age. So, for example, boys typically scored higher than girls on Item 1 (Problems with disruptive, antisocial or aggressive behaviour) but lower on Item 9 (Problems with emotional and related symptoms). Likewise, children aged under five scored higher than their older counterparts on Item 5 (Problems with scholastic or language skills) but lower on Item 3 (Non-accidental self-injury). Brann et al[80] also reported that the HoNOSCA behaved in an intuitive manner when they considered diagnosis, noting, for example, that consumers with attention deficit and conduct disorders scored highest on Items 1 and 2 (Problems with disruptive, antisocial or aggressive behaviour, and Problems with over-activity, attention or concentration). Along similar lines, Bilenberg[81] found that high HoNOSCA scores were associated with comorbidity.

### *Predictive validity*

Few studies have considered the predictive validity of the HoNOSCA. The exception is work by Brann,[79] which found that HoNOSCA scores at community assessment could discriminate between adolescents who later received treatment from intensive adolescent outreach teams and their counterparts who went on to other forms of community care.

### Test-retest reliability

There are few published studies on the test-retest reliability of the HoNOSCA, and those which do exist are arguably studies of the sensitivity to change (or lack of change) of the instrument (see below), since they cover considerable time periods and consider stability in relation to other measures. Garralda et al[78] examined the test-retest reliability of the instrument over a six month period, for consumers for whom clinicians indicated there had been no change on a global rating scale, and reported a figure of 0.69. Similarly, Brann[79] reported correlations of 0.80 over three months and 0.76 over five months when he examined the instrument's test-retest reliability, again in a group of consumers who were judged not to have changed over the given period.

### Inter-rater reliability

Several studies have considered the inter-rater reliability of the HoNOSCA, generally with reasonably positive results. These studies have consistently found that the majority of items in Section A demonstrate good or very good reliability, although there is less agreement about which items perform poorly. For example, Brann et al[80] reported a particularly low intra-class correlation (0.06) for Item 10 (Problems with peer relationships), but Gowers et al[72, 73] found that this item, while not performing as well as some others, achieved an intra-class correlation of 0.77.

There is also debate about the inter-rater reliability of Section B. In the original field work associated with the development of the HoNOSCA, Gowers et al[72, 73] found that the two Items comprising this section each had good inter-rater reliability: Item 14 (Problems with knowledge or understanding about the nature of the child or adolescent's difficulties) had an intra-class correlation of 0.73 and Item 15 (problems with lack of information about services or management of the child or adolescent's difficulties) one of 0.78. By contrast, the equivalent figures in a later study by Garralda et al[83] were 0.27 and 0.03, respectively.

### Sensitivity to change

Three different approaches have been taken to assessing the ability of the HoNOSCA to detect change, with individual studies often using a combination of these approaches. The first and methodologically weakest approach involves simply determining whether HoNOSCA scores change over time, with no reference to whether this reflects real change. In the original field work associated with the development of the HoNOS, for example, Gowers et al noted that 'the HoNOSCA demonstrated satisfactory sensitivity to change, with a mean overall reduction in total scores of 38% between rating points, on average nearly three months apart'.[72, 73] Manderson and McCune[74] made a similar observation in their study. These and other studies have considered factors that are predictive of particular patterns of change on the HoNOSCA, and have identified diagnosis and related symptomatology,[74, 78, 81] initial case severity,[78] and setting.[82]

The second approach examines the correspondence between change as assessed by the HoNOSCA and change as defined by the difference between scores on other measures. Collectively, studies by Gowers et al,[82] Garralda et al[78] and Bilenberg[81] have reported changes in HoNOSCA scores that are comparable in direction and magnitude with other clinician-rated measures, such as the CGAS and the Global Assessment of Psychological Disability (GAPD), and, to a lesser extent with parent- and/or consumer-rated measures such as the HoNOSCA-SR, the BCL and the SDQ.

The third approach uses global judgements of outcome as the 'gold standard'. Typically, these require clinicians (and sometimes parents and referrers) to give their opinion regarding whether the consumer has improved, deteriorated or remained stable over time, via some sort of Likert scale. Studies by Gowers et al,[72, 73] Garralda et al,[78] Brann et al[79, 80] and Bilenberg[81] have all reported close correspondence between change (or lack of change) recorded on the HoNOSCA and such global judgements of outcome.

*Feasibility and utility*

In considering the feasibility and utility of the HoNOSCA in routine outcome measurement, several studies have explicitly sought the views of clinicians.[72-74, 81, 83] Clinicians in these studies have been positive about the HoNOSCA's brevity and ease of use, its clinical utility, and its ability to be incorporated into routine practice (given adequate time and resources). The main concerns raised by clinicians about the instrument have been its applicability to children aged under five, its emphasis on child/adolescent symptoms and functioning, and its failure to take into account context. Some have also questioned whether it may be less useful in the case of particular disorders (e.g., anxorexia and somatoform disorders).

These and other studies have further considered feasibility and utility by examining the behaviour of services and individual clinicians. For example, Gowers et al[72, 73] reported that in their original field study none of the sites dropped out, 71% of consumers who were rated at Time 1 were also rated at Time 2, and that compliance did not vary by the setting (inpatient versus outpatient) or discipline of the clinician. They continued to report optimal completion rates in their later work.[73]

## Summary

Like its parent instrument, the HoNOS, the HoNOSCA arose out of the Health of the Nation Strategy in the United Kingdom. It was designed specifically as a brief measure that was relevant for routine use in child and adolescent mental health services. Structurally, it bears a strong resemblance to the HoNOS. It comprises 15 items that collapse into two sections, one comprising four subscales and the other one. Each item is scored on a scale of 0 (no problem) to 4 (severe problem).

Compared with the HoNOS, fewer studies have considered the HoNOSCA's psychometric properties. However, the studies that have been conducted provide a fairly consistent picture. Its content validity has not been adequately assessed, but it would appear to have reasonable construct and predictive validity. Tests of its concurrent validity indicate that it performs well against other clinician-rated measures and can discriminate between groups of consumers based on their clinical and/or treatment profiles. Its test-retest reliability and inter-rater reliability appear generally sound. It appears to be sensitive to change, as judged by the movement of scores over time, and, more particularly, its performance against other 'gold standards' that indicate improvement, deterioration or stability. Clinicians seem positive about its brevity and ease of use, clinical utility and ability to be incorporated into clinical practice (given adequate time and resources), but are wary about its applicability to children aged under five and consumers with particular disorders (e.g., anxorexia and somatoform disorders).

# Chapter 5: Health of the Nation Outcome Scales 65+ (HoNOS65+)

## Background

Two of the studies described in Chapter 3 examined the psychometric properties of the HoNOS when it was used with older consumers.[32, 55] These studies concluded that although the HoNOS performed well, some modifications might make it more appropriate for use with this group. As a consequence, Burns and colleagues[84] from the College Research Unit of the Royal College of Psychiatrists began a process of modifying the general adult scale for older people. This process involved developing, piloting and evaluating the modified instrument, which became known as the HONOS65+.

## Purpose

The HoNOSCA65+ constitutes a general measure of mental health and social functioning in older people with a mental illness.[84]

## Availability

The HoNOS65+ is subject to the same free availability at the HoNOS.[18]

## Description

The HoNOS65+ is a clinician-administered instrument that has essentially the same structure as the HoNOS, but includes a more detailed glossary that better reflects the reasons that older consumers come into contact with mental health services.[a] So, for example, the glossary acknowledges the behavioural disturbances associated with dementia, and captures some of the unique characteristics of depression in older people (e.g., guilt and passive aspects of suicidal ideation).[85-87] Despite the structure remaining the same, these glossary changes introduce significant semantic changes that make the HoNOS65+ a discernibly different instrument.

The clinician is asked to rate the consumer on each of 12 problems or symptoms in the recent period. Each problem or symptom is captured by a single item:
- Item 1: Behavioural disturbance (e.g., overactive, aggressive, disruptive or agitated behaviour, uncooperative or resistive behaviour);
- Item 2: Non-accidental self-injury;
- Item 3: Problem drinking or drug taking;
- Item 4: Cognitive problems;
- Item 5: Physical illness or disability problems;
- Item 6: Problems associated with hallucinations and delusions;
- Item 7: Problems with depressive symptoms;
- Item 8: Other mental and behavioural problems;
- Item 9: Problems with relationships;
- Item 10: Problems with activities of daily living;
- Item 11: Problems with living conditions; and
- Item 12: Problems with occupation and activities.[99]

As with the HoNOS, these items can be aggregated into four subscales:
- Behaviour (Items 1-3);
- Impairment (Items 4-5);
- Symptoms (Items 6-8); and

---

[a] It should be noted that the HoNOS65+ glossary used in Australia and New Zealand differs from that used in the United Kingdom.

- Social (Items 9-12).[85, 86]

Each item is rated on a five-point scale, with anchor points that more or less mirror the HoNOS (0 = no problem; 1 = minor problem requiring no action; 2 = mild problem, but definitely present; 3 = moderately severe problem; 4 = severe to very severe problem). Individual item scores, subscale scores and total scores can be calculated.[84]

For each item, the clinician makes an assessment of the consumer's situation over the recent period. The overarching instruction is to 'rate the most severe problem that occurred in the period rated', but for Items 11 and 12 this is modified to 'rate the usual [situation]'. Under the NOCC protocol, Items 9 and 10 are also treated this way, with clinicians being trained to 'rate the usual or typical situation'.[20]

For the purposes of the NOCC collection, the period is the standard two weeks at all collection occasions, except discharge from an inpatient setting, where the rating period is three days.

Like the HoNOS, the HoNOS65+ is designed to be completed by clinicians of any discipline, who draw on relevant information from their clinical assessment of the consumer, medical records, and consultation with the consumer's family and carers.[85] A clinician who is familiar with the items on the HoNOS65+ can administer the instrument in less than five minutes; others may take up to 15 minutes.[85, 88]

## Versions

As noted above, two versions of the HoNOS65+ are used in the NOCC collection. These vary only by rating period. The main version, which is used at almost all collection occasions, has a rating period of two weeks (as prescribed in the original instrument). The alternative version, which is used at discharge from an inpatient setting, has a rating period of three days, in recognition of the brevity of episodes in such settings.

Studies that have examined the psychometric properties of the HoNOS65+ (see below) have focused on the two-week version of the instrument.

## Psychometric properties

Less work has been done on establishing the psychometric properties of the HoNOS95+ than the HoNOS.

### *Content validity*

During the initial development of the HoNOS65+, Burns et al[84] asked psychiatrists, nurses, psychologists and other mental health professionals working with older consumers to review the content of the HoNOS. As noted above, this process resulted in modifications to the glossary to address their concerns regarding the comprehensiveness of the instrument for older consumers.[85] Since this time, ongoing issues have been noted, and further refinements to the glossary have been made.[86, 87, 89] It is fair to say, however, that there has been no systematic assessment of the content validity of the HoNOS65+ since its development phase.

### *Construct validity*

There is a paucity of evidence on the construct validity of the HoNOS65+. There are no published studies on the internal consistency of the instrument, on the relative contribution of each item to the total score, or on the robustness of the subscale structure. The only relevant data come from the original pilot work by Burns et al,[85] where a factor analysis revealed that four factors accounted for 57.4% of the variance in total HoNOS65+ scores. The factors represented the following domains: disability/disturbance (Items 1, 4, 9, 10 and 12), psychiatric symptoms

(Items 2, 7 and 8), physical illness (Items 5 and 6) and psychosocial disturbance (Items 3 and 11).

### Concurrent validity

Separate studies by Burns et al,[85] Mozley et al,[90] Spear et al,[88] and Bagley et al[91] have examined the concurrent validity of the HoNOS65+. Most commonly, these studies have examined the correlations between the HoNOS65+ and more established clinician-rated measures that assess similar domains. Reasonable correlations have been observed between the HoNOS65+ total score and the Mini-Mental State Examination (MMSE),[85, 88, 90] the Crighton Royal Behaviour Rating Scale (CRBRS),[85] and the Barthel Activities of Daily Living Index.(Barthel).[85]

As a general rule, however, stronger correlations have been observed between specific HoNOS65+ items and other instruments:
- Item 4 (Cognitive problems) with the MMSE;[85, 88]
- Item 6 (Problems associated with hallucinations and delusions), Item 7 (Problems with depressive symptoms), Item 8 (Other mental and behavioural problems) and Item 9 (Problems with relationships) with the Brief Psychiatric Rating Scale (BPRS);[85]
- Item 4 (Cognitive problems), Item 5 (Physical illness or disability problems) and Item 12 (Problems with occupation and activities) with the Barthel[85];
- Item 1 (Behavioural disturbance), Item 4 (Cognitive problems), Item 5 (Physical illness or disability problems), Item 7 (Problems with depressive symptoms), Item 8 (Other mental and behavioural problems), Item 10 (Problems with activities of daily living), Item 11 (Problems with living conditions) and Item 12 (Problems with occupation and activities) with the CRCRS;[85] and
- Item 1 (Behavioural disturbance), Item 4 (Cognitive problems), Item 9 (Problems with relationships) with the Brief Agitation Rating Scale (BARS).[88]

There are exceptions, however. Equivocal findings have been reported regarding the relationship between HoNOS65+ Item 7 (Problems with depressive symptoms) and the Geriatric Depression Scale (GDS). The original pilot found the correlations between the relevant HoNOS65+ item and individual items on the GDS were good, but that there was no significant correlation between it and the total GDS score.[85] Later studies have produced conflicting results, with one finding a good correlation between Item 7 and the GDS[88] and the other finding that the former detected only a minority of the consumers identified as depressed by the latter.[91]

In addition to the above studies that have examined the performance of the HoNOS65+ against other instruments, there is a small group of studies that have investigated its ability to discriminate between different consumer groups. In the original pilot study, Burns et al found the instrument was able to discriminate between consumers with dementia and those with functional psychiatric disorders, with the former scoring higher on Item 1 (Behavioural disturbance), Item 4 (Cognitive problems) and Item 10 (Problems with activities of daily living), and the latter scoring higher on Item 2 (Non-accidental self injury), Item 7 (Problems with depressive symptoms), Item 8 (Other mental and behavioural problems).[85] The later study by Spear et al[88] reported similar findings, demonstrating that consumers with dementia generally had higher HoNOS65+ scores than those with mood disorders, but had lower scores on the symptoms subscale.

### Predictive validity

To date, the predictive validity of the HoNOS65+ has not been adequately assessed.

### Test-retest reliability

As yet, there are no published assessments of the test-retest reliability of the HoNOS65+.

*Inter-rater reliability*

By contrast, several studies have examined the inter-rater reliability of the HoNOS65+. Burns et al[85] and Spear et al[88] both found inter-rater reliability to be good to very good for most items, whereas Allen et al were less positive.[89] In Burns et al's study, only Item 2 (Non-accidental self-injury), Item 10 (Problems with activities of daily living), Item 11 (Problems with living conditions) and Item 12 (Problems with occupation and activities) did not consistently perform well. In Spear et al's study, Item 4 (Cognitive problems), Item 5 (Physical illness or disability problems) and Item 9 (Problems with relationships) demonstrated only poor to moderate inter-rater reliability. In Allen et al's study, there were problems with a broader range of items, largely related to difficulties in interpretation of the glossary.

*Sensitivity to change*

Only one study has considered the sensitivity to change of the HoNOS65+. Spear et al[88] found that consumers showed improvement on all HoNOS65+ subscales and on the HoNOS65+ total score between assessment and discharge from inpatient and community services. More particularly, they found that the discharge HoNOS65+ score and the change in HoNOS65+ scores showed moderate but significant correlations with the Clinician's Interview Based Impression of Change Scale (CIBIC+).

*Feasibility and utility*

Several studies have examined different aspects of the feasibility and utility of the HoNOS65+, including its acceptability to clinicians. In the original pilot of the HoNOS65+, Burns et al[85] asked raters whether or not they would find the instrument helpful in working with individual consumers; 39% indicated it would be very useful and 50% that it would be of some use. Spear et al[88] reported similar findings. In both studies, almost all respondents reported that it was easy to administer.

Feasibility and utility have also been considered in terms of uptake, both at a national level and at a service level. Reilly et al[92] conducted a survey of old age psychiatrists across the United Kingdom, and found that 18% reported that the HoNOS65+ was being used in their service. Spear et al examined the proportion of episodes of care at which the HoNOS65+ was administered within a single service, and found completion rates of 96%.

Other studies have examined the feasibility and utility of the HoNOS65+ in a more general way, considering issues that have arisen during implementation. Allen et al,[89] for example, observed that clinical leadership and timely feedback to staff were crucial, as was minimising the paperwork burden for clinicians and clarifying analysis and reporting issues. In a similar vein, MacDonald[93] argued that suitable infrastructure must be in place, that the data must be put together in an appropriate manner (e.g., the time period for analysis must be determined, and missing data should be dealt with), and that analysis and reporting should be guided by the requirements of clinicians.

## Summary

The HoNOS65+ was developed in response to the observation that some modifications to its parent instrument, the HoNOS, might make it more appropriate for use with older consumers. The HoNOS65+ has essentially the same structure as the HoNOS, comprising 12 items which are each rated from 0 (no problem) to 4 (severe to very severe problem), yielding individual item scores, subscale scores and a total score. The key difference between the two instruments is the more detailed glossary developed for the HoNOS65+, which better reflects the reasons that older consumers come into contact with mental health services and makes it a discernibly different instrument.

Less work has been done on establishing the psychometric properties of the HoNOS65+ than the HoNOS.  However, the studies that have been undertaken indicate that the HoNOS65+ has reasonably good (concurrent) validity and reliability, and is sensitive to change.  Having said this, certain items perform better than others when assessed against relevant criteria.  The HoNOS65+ appears to be acceptable to clinicians, but its successful implementation as a routine outcome measure is likely to require timely feedback to staff, minimal paperwork burden for clinicians, suitable infrastructure, appropriate data management and analysis and reporting that meets the requirements of clinicians.

# Chapter 6:  Life Skills Profile 16 (LSP-16)

## Background

The Life Skills Profile (LSP-16) was developed as a short-form version of the LSP-39.  The original instrument was produced by Rosen and colleagues in Sydney to measure constructs relevant to survival and adaptation in the community for individuals with schizophrenia and chronic mental illness.[94, 95]  The LSP-16 was derived for use in the Australian Mental Health Classification and Service Costs (MH-CASC) Project[4, 5, 96] to reduce the rating burden on clinicians.  The original developers were commissioned by the MH-CASC team to create the shortened version.  Several other versions of the instrument have since been developed (see below).

## Purpose

The LSP-16 can be regarded as a measure of the level of disability experienced by people with a mental illness, particularly those living in the community.[18]

## Availability

The rights of all versions of the LSP are held by the authorship group, but permission has been given to the Australian Government for the free use of the LSP by mental health services throughout Australia.

## Description

The Life Skills Profile (LSP) is a clinician-administered instrument comprising the following 16 items:
- Item 1: Does this person generally have any difficulty with initiating and responding to conversation?
- Item 2: Does this person generally withdraw from social contact?
- Item 3: Does this person generally show warmth to others?
- Item 4: Is this person generally well groomed (e.g., neatly dressed, hair combed)?
- Item 5: Does this person wear clean clothes generally, or ensure that they are cleaned if dirty?
- Item 6: Does this person generally neglect her or his physical health?
- Item 7: Is this person violent to others?
- Item 8: Does this person generally make and/or keep up friendships?
- Item 9: Does this person generally maintain an adequate diet?
- Item 10: Does this person generally look after and take her or his own prescribed medication (or attend for prescribed injections on time) without reminding?
- Item 11: Is this person willing to take psychiatric medication when prescribed by a doctor?
- Item 12: Does this person co-operate with health services (e.g., doctors and/or other health workers)?
- Item 13: Does this person generally have problems (e.g., friction, avoidance) living with others in the household?
- Item 14: Does this person behave offensively (includes sexual)?
- Item 15: Does this person behave irresponsibly?
- Item 16: What sort of work is this person generally capable of (even if unemployed, retired or doing unpaid domestic duties)?[96]

The above items aggregate into four subscales, namely:
- Withdrawal (Items 1, 2, 3, 8);
- Self-care (Items 4, 5, 6, 9, 16);
- Compliance (Items 10, 11, 12); and

- Anti-social behaviour (Items 7, 13, 14, 15).[96]

It should be noted here that the original LSP-39 comprised five subscales: Communication, Social Contact, Non-turbulence, Self-care and Responsibility.[94] An alternative set of subscales was proposed by Trauer et al[97] – Bizarre, Withdrawal, Self-care, Compliance and Anti-social behaviour – following a study of the psychometric properties of the instrument (see below). When the LSP-16 was developed for the MH-CASC project, the Bizarre subscale was excluded on the grounds that equivalent information could be gleaned from the Health of the Nation Outcome Scales (HoNOS) (which was included in the MH-CASC Project's data collection protocol, and is now part of the NOCC suite of measures), and the four most highly-correlated items in each of the remaining subscales were retained.[4, 5, 96]

The clinician rates the consumer's general functioning (not during crisis or when he/she is becoming ill) on each of the 16 items in terms of his or her behaviour over the preceding three month period, drawing on direct observation and information from other individuals in contact with the consumer (including family, friends, carers and mental health professionals).[98] The clinician rates each item on a four-point scale ranging from 0 (no problem) through 1 (slight problem) and 2 (moderate problem) to 3 (extreme problem). Individual item scores, subscale scores and a total score can then be calculated. As noted, this scoring system represents a reversal from the LSP-39, where a higher score denotes better functioning to be consistent with the strengths emphasis of contemporary recovery approaches.

The LSP-16 has no training requirements (although familiarisation training and clarification of individual items are available through NOCC) and may be administered by clinicians of any discipline, providing they have had moderate contact with the consumer. Clinicians who are familiar with the instrument can typically administer it in five minutes or less.

## Versions

As noted above, several versions of the LSP exist. The original instrument, the LSP-39, was abbreviated into the LSP-16, which is the version being used under the NOCC protocol. In addition, several other versions have been developed, the most recent being the LSP-20.[99] Some of these versions have been translated into other languages, including Italian,[100] Spanish,[101, 102] French.[103] and Japanese.

The LSP-39 can be regarded as a measure of strengths, since high scores indicate better functioning. By contrast, the LSP-16 can be viewed as a measure of impairment, since high scores indicate poorer functioning. The LSP-20, which represents a cut-down version of the LSP-39 and an extension of the LSP-16 can be scored in either direction.[99]

In the remainder of this chapter, primary emphasis is given to the LSP-16, as this is the version included among the NOCC measures. Lesser weight is given to descriptive and evaluative information about the alternative versions of the LSP. Having said this, studies of the alternative versions (particularly the original LSP-39) are cited where relevant, particularly if evidence is lacking regarding the LSP-16.

## Psychometric properties

The LSP has been subjected to reasonably comprehensive testing of its psychometric properties, although it must be acknowledged that more work has been done on the LSP-39 than on the LSP-16.

### Content validity

Several studies have commented on the content validity of the LSP. The majority of these have been concerned with the LSP-39, but to the extent that the content of the LSP-16 is a cut-down version of the larger instrument, their findings are relevant here. Studies conducted by Andrews

et al[2] and Stedman et al[3] generally elicited a positive responses from public sector mental health service providers regarding the LSP-39 as a measure of disability. Specifically, the majority viewed the instrument as comprehensive and able to provide an accurate picture of the consumer's level of functioning.

However, these and other authors have reported some issues with the instrument. One criticism is that it lacks relevance in particular settings. For example, in long-term inpatient settings, consumers often do not have the opportunity to exhibit some of the skills assessed by certain items on the LSP-39 (and included in the LSP-16), such as Item 10 (Does this person generally look after and take her or his own prescribed medication without reminding?).[104, 105] In Stedman et al's[3] study, some respondents even argued that the LSP-39 lacked relevance in the community setting, for which it was specifically designed, because of the restrictiveness of the response options.

### Construct validity

A number of studies have examined the internal consistency of the LSP-39, as measured by Cronbach's alpha. In these studies, the internal consistency has been reported as moderately high with subscale and total score correlations ranging from 0.64 to 0.88[3, 94, 95, 97] and 0.93 to 0.94,[3, 97, 106] respectively. The Communication subscale has been shown to have the poorest internal consistency.[95, 97]

In the original development work that underpinned the LSP-39, Rosen et al[94] on the LSP-39, conducted a principal components analysis to determine the components that should be retained. Five components were retained, together accounting for more than half (53.8%) of the total variance. These five components became defined as the original five subscales: Communication, Social contact, Non-turbulence, Self-care and Responsibility.

As noted above, Trauer et al[97] carried out further psychometric testing on the LSP-39. Replicating Rosen et al's method, they also identified five components that accounted for just over half (50.7%) of the overall variance. However, they explored the exact nature of these components further, using confirmatory factor analysis, and recommended that the factor structure be modified. Specifically, they proposed the alternative subscale structure described earlier – Bizarre, Withdrawal, Self-care, Compliance and Anti-social behaviour – since this model fitted the data better. The correlations with the original subscales were 0.85, 0.94, 0.98, 0.97 and 0.90 respectively. Andrews et al[107] suggested the five subscales could be further divided into two dimensions, which they described as 'general impairment' and 'difficulty'.

### Concurrent validity

Collectively, numerous studies have considered the concurrent validity of the LSP-39, LSP-20 and LSP-16 by examining the correlation of individual, subscale and total scores to that of some other instrument that is viewed as an acceptable measure of the behaviour being examined:

- The LSP-39 has been shown to perform well against the Health of the Nation Outcome Scales (HoNOS),[3, 40, 108] the Katz Adjustment Scale (KAS),[95] the Multnomah Communities Ability Scale (MCAS),[106] the Strauss-Carpenter Levels of Functioning Scale (LOF),[106] the Global Assessment of Functioning Scale (GAF),[40, 106, 109] the Role Functioning Scale (RFS),[3] the Quality of Life Scale (QOL),[110] the Interviewer-rated Quality of Life Scale (IQL),[108] the Social Behaviour Schedule (SBS),[109] the Resource Associated Functional Level Scale (RAFLS)[97] and the Global Assessment Scale (GAS).[108] However, it has demonstrated poor or mixed performance against the BASIS-32®,[3] the Mental Health Inventory (MHI),[3] the Short Form-36 (SF-36),[3] the General Wellbeing Scale (GWB),[110, 111] the Brief Psychiatric Rating Scales (BPRS),[95, 97, 108] the Dysexecutive Questionnaire (DEX),[108, 109] Cantril's Ladder[108] and the Affect Balance Scale (ABS).[108] It should be noted that several of the latter are self-report measures, so the poor levels of correspondence may reflect the expected gap between clinician and self-report ratings.

In addition, some (e.g., the BPRS) are symptomatic scales, and it would be expected that symptoms might vary in intensity and not co-vary with the more robust and enduring, though often slower, changes in functioning or disability.

- The LSP-20 has been shown to produce comparable results to the LSP-39, and to the Positive and Negative Syndrome Scale (PANSS).[99]

- Of greatest relevance to the current discussion is the performance of the LSP-16. In developing the shorter instrument, Buckingham et al[96] found an 85% to 90% concordance between the top four items of each subscale with the full subscale of the LSP-39. Likewise, Rosen et al[99] found the LSP-16 performed well against its parent instrument. Trauer et al[112] also found the LSP-16 correlated well with the HoNOS, but that it showed poor or mixed performance against the Behaviour and Symptom Identification Scale (BASIS-32®).

An alternative method of assessing concurrent validity has been to consider its ability to discriminate between particular groups of consumers. Rosen et al[94] found that low scores on the Responsibility and Non-turbulence scales of the original LSP-39 were characteristically scored by younger people. Other studies have found that total scores on the various versions of the LSP can distinguish between consumers based on their locus of accommodation and/or care, with higher levels of disability being at least moderately associated with those experiencing frequent changes in accommodation (as opposed to those in stable living environments) and/or living in long-term residential care settings or in the family home (relative to those living independently or semi-independently).[107, 111, 113-116] Still other studies have found the LSP to be able to discriminate between consumers on the basis of their levels of social functioning, as measured by factors like unstable employment, low-grade accommodation (e.g., hostel or refuge), welfare dependency, police contact and complaints by neighbours.[94, 107, 117]

### Predictive validity

Several studies have examined the predictive validity of the LSP-16 and LSP-39. In general, these studies have shown that the different versions of the instrument can predict outcomes relating to retention in the community,[118] hospital readmission,[107, 117] change in locus of care,[113] length of inpatient stay,[119, 120] and overall costs[111, 120] An exception to this rule was a study by Parker et al,[40] which failed to support the predictive validity of the LSP-39.

### Test-retest reliability

The few studies that have examined the test-retest reliability of the LSP have reported a high overall reliability score, albeit for the LSP-39 only.[2, 3, 95] Parker et al,[95] for example, established high test-retest reliability for case workers, residential carers and parents, each of whom were asked to rate the same person with relatively stable chronic schizophrenia at two points in time (one month apart).

### Inter-rater reliability

A number of studies have assessed the inter-rater reliability of both the LSP-39 and the LSP-16. These studies found the overall agreement between pairs of raters on the LSP-39 to be fair to moderate[2, 3, 97] or moderate to good.[94, 95, 97] Some studies have found that raters of similar backgrounds are more likely to show high correlations between their respective ratings,[95] whereas others have found the background of raters to have little bearing on levels of agreement.[94]

### Sensitivity to change

A number of studies have examined the degree to which changes in LSP scores correlate with some external measure of change. Several of these studies have compared the LSP's ability to

detect change against other, more established measures of outcome. Stedman et al,[3] for example, found significant associations between changes on the LSP and changes on the Global Change Ratings Scale, the Modified Clinical Global Impressions Scale (CGI), the Role Functioning Scale (RFS) and the HoNOS. The latter finding was reproduced by Parker et al[40] who also found a moderate associated between the LSP and the Global Assessment of Functioning (GAF).

Other studies have examined changes in LSP scores for different consumer groups that would be expected to show greater or lesser degrees of improvement depending on their treatment circumstances. Typically, these studies have found, as hypothesised, that the LSP demonstrates greater levels of improvement in those who participate in intensive case management than in those who undergo routine case management.[121-126] However, there have been some exceptions to this rule.[127, 128]

Still other studies have used self-reported improvement or deterioration as the 'gold standard' against which to assess the sensitivity to change of the LSP. Stedman et al,[3] for example, conducted an analysis of LSP-39 scores over time for groups showing differing levels and directions of self-reported change. This study found that LSP scores worsened in the group who reported a decline in their levels of functioning, but there was no association between LSP change score and self-reported change for any other group.

### *Feasibility and utility*

Published commentaries have reported that the overall feasibility and utility of the LSP-16 is moderately high.[40] However, only the work of Stedman et al[3] has really put this assertion to the test. These authors elicited the views of service providers about their experience with using the LSP, via a purpose-designed utility questionnaire. The majority of service providers had little difficulty with the language and viewed the questions as relevant, useful and effective in measuring outcomes for consumers. Respondents in public sector psychiatric settings, in particular, rated the LSP more highly than other observer-rated measures.[3, 129]

## Summary

The LSP-16 is derived from the LSP-39. It is designed to measure the level of functioning and adaptation of people with a mental illness living in the community. It consists of 16 items that address issues faced when adapting to life in the community. Each item is rated on a four-point scale ranging from 0 to 3 with high scores indicating higher disability, resulting in an individual item scores, subscale scores and a total score.

Taken together, studies examining the psychometric properties of the LSP suggest that it has: moderately good content, construct, concurrent and predictive validity; high test-retest and adequate inter-rater reliability; and good sensitivity to change. It also appears to demonstrate feasibility and utility in clinical settings. Having said this, relatively few studies have examined the LSP-16 specifically; far more have been concerned with the LSP-39. Additional analyses are needed to strengthen the evidence base regarding the shorter version.

# Chapter 7: Resource Utilisation Groups – Activities of Daily Living (RUG-ADL)

## Background

The Resource Utilization Groups – Activities of Daily Living (RUG-ADL) is a component of the RUG-III, a casemix classification system developed in the United States by Fries et al[130] for use in long term care facilities for the elderly.[131] In such facilities, diagnosis is a poor predictor of resource use, and better predictors are factors like objectives of care, functional deficits, rehabilitation potential, and behavioural problems.[132]. In particular, basic activities of daily living, such as toileting, bathing, eating and mobility are strong predictors of resource use and consequently their measurement forms a key element of the RUG system.[133] The RUG-III involves allocating consumers to one of seven cost-ranked clinical groups (i.e., Rehabilitation, Extensive Services, Special Care, Clinically Complex, Impaired Cognition, Behavioural Problems and Reduced Physical Function) on the basis of clinical characteristics or services received. Consumers are further subgrouped on the basis of their RUG-ADL score, care received, and the presence or absence of depression.[134-136]

## Purpose

The ADL component of RUG-III aims to measure consumer dependency or functional status, with a focus on assistance required for toileting, bathing, eating and mobility.[131-133]

## Availability

The RUG-III (and consequently the RUG-ADL) is readily available and has no licence costs attached.[18]

## Description

In the context of NOCC, the RUG-ADL is completed for consumers aged 65 or older. It is clinician-administered, and comprises only four items. Ratings are based on the consumer's poorest performance during 'the period rated'. The duration of this period is unspecified, but NOCC training materials interpret it as 'current status'. Consumers are given a score for each item and these are then summed to provide a total score. The specific items are:

- Bed Mobility: ability to move in bed after the transfer into bed has been completed
    1. Independent or Supervision only: the patient is able to readjust position in bed, and perform own pressure area relief, through spontaneous movement around bed or with prompting from carer. No hands-on assistance is required. May be independent with the use of a device.
    3. Limited assistance: the patient is able to readjust position in bed, and perform own pressure area relief, with the assistance of one person
    4. Other than 2 person physical assistance: the patient requires the use of a hoist or other assisting device to readjust position in bed and physically assist pressure relief. Still requires the assistance of only one person for task.
    5. 2-person physical assistance: requires two assistants to readjust position, and perform own pressure area relief.

- Toileting: includes mobilising to the toilet, adjustment of clothing before and after toileting and maintaining perineal hygiene without the incidence of incontinence or soiling of clothes. If the person cares for the catheter or other device independently and is independent on all other tasks, rate 1.
    1. Independent or Supervision only: the patient is able to mobilise to the toilet, adjust clothing, cleans self, has no incontinence or soiling of clothing. All tasks performed

independently or with prompting from carer.  No hands on assistance required.  May be independent with the use of device.
3. Limited assistance: the patient requires hands on assistance of one person for one or more of the tasks
4. Other than 2-person physical assistance: the patient requires the use of a catheter, uridome or urinal, or a colostomy, bedpan or commode chair, or insertion of enema or suppository.  Requires the assistance of one person for the management of the device.
5. 2-person physical assist: the patient requires two assistants to perform any step of the task

- Transfer: includes the transfer in and out of bed, bed to chair, in and out of shower or tub.
  1. Independent of Supervision only: the patient is able to perform all transfers independently or with prompting from carer.  No hands-on assistance required.  May be independent with the use of a device.
  3. Limited assistance: the patient requires hands-on assistance of one person to perform any transfer of the day or night
  4. Other than 2-person physical assistance: the patient requires the use of device for any of the transfers performed in the day or night.
  5. 2-person physical assist: the patient requires two persons to perform any transfer of the day or night.

- Eating: Includes the tasks of cutting food, bringing food to the mouth and the chewing and swallowing of food.  Does not include preparation of the meal.
  1. Independent or Supervision only: the patient is able to cut, chew and swallow food, independently or with supervision, once meal has been presented in the customary fashion.  No hands on assistance required.  If individual relies on parenteral or gastrostomy feeding which he or she administers him or her self then rate 1.
  2. Limited assistance: the patient requires hands on assistance of one person to assist in bringing food to the mouth, or requires food to be modified (soft or staged diet)
  3. Extensive assistance/Total dependence/Tube fed: the patient needs to be fed meal by assistant, or if the individual does not eat or drink full meals by mouth but relies on parenteral or gastrostomy feeding and does not administer feeds by him or her self.[130]

## Versions

As noted, the RUG-ADL is a component of the RUG-III.  The RUG-III is the culmination of several earlier versions of the same instrument.  The RUG-III deals with mental health and dementia care more effectively than the original RUG and the RUG-II.

In the remainder of this chapter, primary emphasis is given to the RUG-ADL, rather than the full RUG-III (or earlier versions of the RUG system), wherever possible.  However, it should be noted that the majority of studies that have considered the psychometric properties of the instrument have not restricted their analyses to the RUG-ADL alone.

## Psychometric properties

Psychometric testing has tended to focus on the RUG-III (or its predecessors) as a whole, rather than on the ADL component, and has emphasised its use as a casemix tool rather than an outcome measure.

### Content validity

The RUG-II received significant criticism for its inadequacies in dealing with mental illness (particularly dementia),[131, 137] and this was specifically addressed by the RUG-III.  Beyond this, there have been few concerted empirical efforts to explore the content validity of the RUG-ADL (or indeed the RUG-III), although some concerns have emerged through published discussion

papers and commentaries. Most of these concerns relate to the instrument's use as a casemix tool (its primary purpose), rather than as an outcome measure.

One criticism is that because cognitive ability is not directly assessed, the RUG-ADL fails to account for the significant demands associated with the care of cognitively impaired consumers.[131, 137] RUG-III developers have countered that additional resource use can be explained by ADL deficits, irrespective of the cause of these deficits. However, research has indicated that within given ADL domains, cognitively impaired consumers experience particular difficulties that may impede care tasks, and it has been recommended that an additional assessment of cognitive impairment may be necessary to estimate the care needs of consumers with dementia more accurately.[131]

Another concern raised is that what the RUG-ADL measures is the activities performed by the consumer, as opposed to the activities that he or she may be capable of. The index assumes that staff will allow consumers to perform activities to their maximum capacity, but in some settings and circumstances it may be more efficient for the staff member to take over a given activity. This will artificially inflate consumers' levels of dependency as measured by the RUG-ADL.[138]

### Construct validity

No studies have attempted to test the construct validity of the RUG-ADL, or RUG system more broadly.

### Concurrent validity

There is a dearth of studies examining the concurrent validity of the RUG-ADL, or indeed the RUG-III. A single early study compared the RUG-II to the alternative Diagnosis Related Groups (DRG) system in the United States, and this showed that the RUG-II explained significantly more of the variance in resource use in long term care settings for the elderly than did DRGs (53% compared with 30-35%).[135]

### Predictive validity

As a casemix measure, the RUG system is designed to predict resource use. Consequently, the majority of the psychometric validation of the various versions of the RUG has been concerned with predictive validity.

A number of studies have examined the predictive validity of the RUG-III system, with positive results in various countries:[135]

- In the United States, the system explained 55.5% of the variance for wage-weighted nursing and therapy costs, and 41.2% of nursing staff costs alone.[130]
- In Japanese psychiatric facilities, a per day model was used to explain 33.3% of the variance in resource use.[139]
- In England and Wales, Carpenter et al[135] found that variance explanation differed by ward type, and was highest for consumers whose length of stay was greater than 10 days. Acute (57%), acute/rehabilitation (59%) and rehabilitation wards (47.7%) displayed the highest variance explanation, while rehabilitation/long stay and long stay wards showed lower explanation (36.6% and 23.7%). The ADL score alone explained 18.3% of the variance overall; 32.1% in acute wards, and 19.3% in long stay wards.
- In Australia, the RUG-III was shown to explain 44% of the variance in costs, but was shown to be a poor predictor of length of stay. Consequently, the ADL component of the RUG system was added to a major clinical grouping system (the National Non-Acute Inpatient Classification Project (NAIP) classification). Testing of this system revealed an explanation of 28% of the variance in cost per day.[140]

One study, examining the RUG-II rather than the RUG-III, considered the ADL index specifically. Ljunggren and Brandt[136] found the RUG-ADL to be significant predictor of length of stay, although this finding was largely explained by the fact that a single, sizeable group with very low dependency scores on the RUG-ADL had particularly short lengths of stay. The RUG-ADL was also predictive of discharge status.

### Test-retest reliability

No published studies exist that examine the test-retest reliability of the RUG-ADL (or indeed the RUG-III).

### Inter-rater reliability

Carpenter et al[135] also examined the inter-rater reliability of the RUG-III, comparing the ratings by two nurses of 27 consumers. The same RUG-III clinical grouping was allocated in 85% of the cases, and the same subgrouping (based on the RUG-ADL) in 74% of cases. In those cases where different allocations were made, the differences were accounted for only by one or two questions.

### Sensitivity to change

No studies have explicitly considered the ability of the RUG-ADL to detect change in functional ability, which can occur rapidly in elderly populations.[136]

### Feasibility and utility

Although there has been little formal testing of the feasibility and utility of the RUG system in general, or of the RUG-ADL in particular, various authors have commented positively in this regard. Carpenter et al,[135] Lombardo et al[137] and Hirdes et al,[139] for example, are sanguine about the consumer-focused structure of the RUG system, noting that it provides the basis for improved equity of resource allocation based on consumers' needs, and eliminates incentives for providers to avoid the admission of consumers with complex care requirements. There remain some concerns with its use in dementia care, but evaluation of its use within Japanese psychiatry facilities suggest that it can be successfully administered within this context.[139] Considering the RUG-ADL specifically, Mitty[141] has commented that its structure makes it ideal for classification and payment systems which can be easily audited.

## Summary

The RUG-ADL is a component of the RUG-III, a casemix classification system developed in the United States for use in long term care facilities for the elderly. The RUG-ADL is clinician-administered and measures consumer dependency or functional status related to four items (bed mobility, toileting, transfer and eating).

Psychometric testing has tended to focus on the RUG-III (or its predecessors) as a whole, rather than on the ADL component, and has emphasised its use as a casemix tool rather than an outcome measure. In terms of content validity, the RUG has been criticised for inadequately dealing with mental illness in general and dementia in particular (although the RUG-III represents an improvement in this regard), failing to account for the care demands associated with cognitive impairment, and measuring actual rather than potential performance. The construct and concurrent validity of the RUG have not been adequately tested, but it shows good predictive validity. Similarly, the test-retest reliability of the RUG has not been assessed, but its inter-rater reliability is sound. No studies have considered the sensitivity of the RUG to change in functional status. There has been little formal examination of the feasibility and utility of the RUG, but published commentaries have tended to be positive in this regard.

# Chapter 8:  Children's Global Assessment Scale (CGAS)

## Background

The Children's Global Assessment Scale (CGAS) was adapted from the Global Assessment Scale (GAS) for adults[142] by Shaffer and colleagues of the Department of Psychiatry, Columbia University, United States,[143] to provide an assessment of functioning for those aged 4 to 16.

## Purpose

The CGAS offers a means of establishing levels of dysfunction for children and adolescents.  It is designed to be used in conjunction with diagnostic measures, and aims to provide more detailed information upon which to develop management plans.[144]  It also enables comparisons to be made between different groups of young consumers, and offers a means to evaluate improvement or deterioration in functioning following treatment or over time.[145]

## Availability

The CGAS is readily available, and incurs no cost to the user.[18]

## Description

The CGAS is clinician-administered, and provides a single global rating of a child or adolescent's lowest level of functioning over the previous two weeks.  Ratings range from 1 (severe dysfunction) to 100 (superior functioning), and the threshold of psychopathology is suggested to sit between 61 and 71.[146, 147]

In making their rating, clinicians draw on a thorough psychiatric assessment that incorporates an understanding of the consumer's social and psychological functioning.[148]  They rely heavily on clinical judgement and are therefore presumed to have clinical expertise and training in the use of psychometric measures.  No formal training materials are provided with the CGAS, although AMHOCN has developed training materials as part of its standard package.

Clinicians are instructed to select the 'lowest level which describes his/her (the consumer's) functioning on a hypothetical continuum of health-illness'.  Ratings are made regardless of treatment or prognosis.  Clinicians are guided by anchor points at every 10th degree on the scale; these offer an indication of the behavioural functioning displayed by consumers at that level, but the specific behaviours are not required for a particular rating.  The scale and its descriptors are outlined below:

100-91      Superior functioning in all areas (at home, at school, and with peers), involved in a range of activities and has many interests (e.g., has hobbies or participates in extracurricular activities or belongs to an organised group such as Scouts, etc.) Likeable, confident, 'everyday' worries never get out of hand.  Doing well in school.  No symptoms.

90-81       Good functioning in all areas.  Secure in family, school, and with peers.  There may be transient difficulties and 'everyday' worries that occasionally get out of hand (e.g., mild anxiety associated with an important exam, occasional 'blow-ups' with siblings, parents, or peers).

80-71       No more than slight impairment in functioning at home, at school, or with peers. Some disturbance of behaviour or emotional distress may be present in response to life stresses (e.g., parental separations, deaths, birth of a sibling) but these are brief and interference with functioning is transient.  Such children are only minimally disturbing to others and are not considered deviant by those who know them.

| 70-61 | Some difficulty in a single area, but generally functioning pretty well (e.g., sporadic or isolated antisocial acts, such as occasionally playing hooky or petty theft; consistent minor difficulties with schoolwork; mood changes of brief duration; fears and anxieties which do not lead to gross avoidance behaviour; self doubts). Has some meaningful interpersonal relationships. Most people who do not know the child well would not consider him/her deviant but those who do know him/her well might express concern. |
|---|---|
| 60-51 | Variable functioning with sporadic difficulties or symptoms in several but not all social areas. Disturbance would be apparent to those who encounter the child in a dysfunctional setting or time but not to those who see the child in other settings. |
| 50-41 | Moderate degree of interference in functioning in most social areas or severe impairment of functioning in one area, such as might result from, for example, suicidal preoccupations and ruminations, school refusal and other forms of anxiety, obsessive rituals, major conversion symptoms, frequent anxiety attacks, frequent episodes of aggressive or other antisocial behaviour with some preservation of meaningful social relationships. |
| 40-31 | Major impairment in functioning in several areas and unable to function in one of these areas, that is, disturbed at home, at school, with peers, or in the society at large (e.g., persistent aggression without clear instigation; markedly withdrawn and isolated behaviour due to either mood or thought disturbance, suicidal attempt with clear lethal intent). Such children are likely to require special schooling and/or hospitalisation or withdrawal from school (but this is not a sufficient criterion for inclusion in this category). |
| 30-21 | Unable to function in almost all areas, for example, stays at home, in ward or in bed all day without taking part in social activities OR severe impairment in reality testing OR serious impairment in communication (e.g., sometimes incoherent or inappropriate). |
| 20-11 | Needs considerable supervision to prevent hurting others or self, for example, frequent violent, repeated suicide attempts OR to maintain personal hygiene OR gross impairment in all forms of communication, for example, severe abnormalities in verbal and gestural communication, marked social aloofness, stupor, etc. |
| 10-1 | Needs constant supervision (24-hour care) due to severely aggressive or self-destructive behaviour or gross impairment in reality testing, communication, cognition, affect, or personal hygiene. |

## Versions

The CGAS was designed for use by individuals with clinical expertise, training in the use of psychometric measures generally, and the ability to use clinical judgement.[143, 148] In light of the limitations this poses Bird et al[149] devised a non-clinical CGAS with simplified descriptors. Further discussion of the non-clinical version is beyond the scope of the current chapter.

## Psychometric properties

The CGAS has undergone various tests of its psychometric properties.

### *Content validity*

Published commentaries have not expressed any significant concerns about the content of the CGAS. Extensive piloting occurred during its development, with clinical teams using the revised GAS (with its amended, child appropriate wording and behavioural descriptors) to rate consumers with varying levels of functional impairment. The final version followed the resolution of ambiguous wording and inappropriate descriptor examples.[143]

*Construct validity*

There is a dearth of published data on the construct validity of the CGAS.

*Concurrent validity*

Several studies have considered the relationship between CGAS scores and scores on other instruments that examine at least some components of psychosocial functioning, in an effort to assess its construct validity. In general, these studies have found that CGAS scores demonstrate high correlations with independent measures of competence, intellectual and social functioning, and problem solving, and only moderate correlations with measures of symptomatology. Collectively, these findings have been interpreted as evidence of the instrument's construct validity, since the former constructs are closely related to what it purports to measure.[143, 147, 148, 150-152]

Further support for the contention that the CGAS has good construct validity comes from studies that have examined the instrument's ability to distinguish between particular groups of consumers. Lower average scores have been demonstrated for consumer groups who would be expected to have poorer levels of functioning than their peers – inpatients (compared with outpatients),[143] referred consumers (compared with non-referred consumers)[148] cases meeting diagnostic criteria (compared with non-cases),[148] psychiatric services users (compared with non-users).[145]

*Predictive validity*

Little has been reported on the predictive validity of the CGAS. The exception is the work by Sourander and colleagues,[153, 154] who found that CGAS ratings at admission were predictive of functioning and residential status at follow-up.

*Test-retest reliability*

Shaffer et al[143] examined the test-retest reliability of the CGAS by presenting case vignettes to participating clinicians at an initial session and again six months later. Within-rater consistency was excellent for cases representing individuals with a range of diagnoses. Only the vignettes depicting individuals with isolated symptom disorders showed discrepancies between the two rating points.[143]

*Inter-rater reliability*

The design of the above study by Schaffer et al[143] also permitted an examination of the inter-rater reliability of the CGAS, since they could assess the agreement between raters at both time one and time two. They found excellent agreement between raters, a finding that has generally been supported by subsequent studies. Dyborg et al,[144] Bird et al[148] and Green et al,[150] for example, also found good inter-rater agreement, particularly among experienced raters. Weissman[147] found good agreement between ratings made by mothers, children and psychiatrists, but Sourander et al[151] reported poor agreement between ratings provided by parents and teachers.

*Sensitivity to change*

Few studies have considered the ability of the CGAS to detect change. An exception is a study by Weissman et al[147] who observed patterns of CGAS scores for a group of children who had no current or previous psychiatric disorder at initial assessment, some of whom progressed to a first onset within two years. The average difference between CGAS ratings at initial assessment and follow-up was significantly greater for this subgroup than for those who remained disorder-free. These results are indicative of the CGAS's sensitivity to change in clinical status.

***Feasibility and utility***

The CGAS is generally regarded as a useful measure of child and adolescent functioning, providing more detailed information for guiding treatment decisions than diagnosis- or symptom-based measures alone.[144, 146, 147] Published commentaries do, however, put forward a number of criticisms about the measure. Firstly, concerns have been expressed about its vulnerability to rater manipulation, in that raters can assign scores below or above a particular cut-off point to suit their needs. Secondly, the global nature of the scoring has been criticised for failing to consider different domains of functioning in any organised manner. Finally, the instrument's accuracy has been questioned, given its dependence upon the clinician's observations of the consumer at the time of assessment and the availability of relevant background information.[155]

## Summary

The CGAS is a clinician-administered instrument that provides an assessment of levels of functioning in children and adolescents (aged 4 to 16). It provides a single global rating, ranging from 1 (severe dysfunction) to 100 (superior functioning).

In the main, tests of the CGAS's psychometric properties have concentrated on concurrent validity and inter-rater reliability, where it has performed well. There is also evidence, albeit from a smaller number of studies, that the instrument has adequate content validity, predictive validity, test-retest reliability and sensitivity to change. Published commentaries tors have generally regarded the CGAS as having adequate feasibility and utility, although concerns have been expressed about its vulnerability to rater manipulation, its lack of any overarching organising principles, and its accuracy.

# Chapter 9:  Mental Health Inventory (MHI)

## Background

The MHI was developed by Veit and Ware[156] as part of the RAND Health Insurance Experiment in the United States, a study designed to examine the influence of health care financing arrangements on health care utilisation.  Adapted from another mental health instrument, the General Wellbeing Schedule (GWB),[157] the MHI was designed to assess general psychological distress and wellbeing in a non-patient population.

The MHI has been employed in a wide range of studies. For example, it has been used to assess the mental health of consumers with cancer[158-164] and human immunodeficiency virus (HIV),[165-167] as well as that of transplant recipients.[168, 169]  It has also been used as a screening tool for depression in older people[170, 171] and as a general measure of mental illness for primary care consumers.[172, 173]

## Purpose

As noted above, the MHI was designed to measure psychological distress and psychological wellbeing in the general population.

## Availability

The MHI can be downloaded free of charge from the RAND Corporation website.[174]

## Description

The MHI is a consumer-rated measure comprising 38 items:
- Item 1: How happy, satisfied, or pleased have you been with your personal life during the past month?
- Item 2: How much of the time have you felt lonely during the past month?
- Item 3: How often did you feel nervous or jumpy when faced with excitement or unexpected situations in the past month?
- Item 4: During the past month, how much of the time have you felt that the future looks hopeful and promising?
- Item 5: How much of the time, during the past month, has your daily life been full of things that were interesting to you?
- Item 6: How much of the time, during the past month, did you feel relaxed and free from tension?
- Item 7: During the past month, how much of the time have you generally enjoyed the things you do?
- Item 8: During the past month, have you had any reason to wonder if you were losing your mind, or losing control over the way you act, talk, think, feel, or of your memory?
- Item 9: Did you feel depressed during the past month?
- Item 10: During the past month, how much of the time have you felt loved and wanted?
- Item 11: How much of the time, during the past month, have you been a very nervous person?
- Item 12: When have you got up in the morning, during the past month, how often did you expect to have an interesting day?
- Item 13: During the past month, how much of the time have you felt tense or 'high-strung'?
- Item 14: During the past month, have you been in firm control of your behaviour, thoughts, emotions or feelings?
- Item 15: During the past month, how often did your hands shake when you tried to do something?

- Item 16: During the past month, how often did you feel that you had nothing to look forward to?
- Item 17: How much of the time, during the past month, have you felt calm and peaceful?
- Item 18: How much of the time, during the past month, have you felt emotionally stable?
- Item 19: How much of the time, during the past month, have you felt downhearted and blue?
- Item 20: How often have you felt like crying, during the past month?
- Item 21: During the past month, how often have you felt that others would be better off if you were dead?
- Item 22: How much of the time, during the past month, were you able to relax without difficulty?
- Item 23: How much of the time, in the past month, did you feel that your love relations, loving and being loved, were full and complete?
- Item 24: How often, during the past month, did you feel that nothing turned out for you the way you wanted it to?
- Item 25: How much have you been bothered by nervousness, or your 'nerves,' during the past month?
- Item 26: During the past month, how much of the time has living been a wonderful adventure for you?
- Item 27: How often, during the past month, have you felt so down in the dumps that nothing could cheer you up?
- Item 28: During the past month, did you think about taking your own life?
- Item 29: During the past month, how much of the time have you felt restless, fidgety, of impatient?
- Item 30: During the past month, how much of the time have you been moody or brooded about things?
- Item 31: How much of the time, during the past month, have you felt cheerful, light-hearted?
- Item 32: During the past month, how often did you get rattled, upset or flustered?
- Item 33: During the past month, have you been anxious or worried?
- Item 34: During the past month, how much of the time were you a happy person?
- Item 35: How often during the past month did you find yourself having difficulty trying to calm down?
- Item 36: During the past month, how much of the time have you been in low or very low spirits?
- Item 37: How often during the past month, have you been waking feeling fresh and rested?
- Item 38: During the past month, have you ever been under or felt you were under any strain, stress or pressure?[18]

These items aggregate into six subscales (or lower order factors):
- Anxiety: Items 3, 11, 13, 15, 25, 29, 32, 33, 35
- Depression: Items 9, 19, 30, 36
- Loss of behavioural or emotional control: Items 8, 14, 16, 18, 24, 27, 28
- General positive affect: Items 4-7, 12, 17, 26, 31, 34, 37
- Emotional ties: Items 10, 23
- Life satisfaction: Item 1.[175]

It should be noted that Veit and Ware[156] originally proposed five subscales, with the single item in the Life satisfaction subscale subsumed into the General positive affect subscale. The six subscale version was proposed by Davies et al, and is being used in the NOCC data collection protocol.[175]

In turn, the subscales aggregate into two global scales (or higher order factors):
- Psychological distress: Subscales Anxiety, Depression, and Loss of behavioural or emotional control

- Psychological wellbeing: Subscales General positive affect and Emotional ties.[175]

Each item includes a description of a particular state of mind which is scored on a six-point Likert scale (range 1-6). The exceptions are Items 9 and 28, which are each rated on a 5-point scale (range 1-5). The respondent is required to indicate the frequency or intensity to which they have experienced this state during the past month. Item scores are summed to give six subscale scores, global scale scores and a total score. Prior to summing, certain items need to be reverse scored so that higher scores indicate a higher level of the construct named by the item, scale or index (Davies, 1998).

The MHI can be completed either as a self-report measure or as part of a structured-interview; either way it takes approximately 10-15 minutes to complete.[18, 176]

## Versions

Several alternative versions of the MHI have since been developed, including the briefer MHI-18 and MHI-5[172] (the latter of which forms the general mental health scale included the MOS-Short Form 20 (SF-20)[177] and MOS-Short Form 36 (SF-36)[178]). In addition, the MHI has been translated into several other languages, including Hebrew.[179]

The remainder of this chapter focuses on the psychometric properties of the MHI and, to a lesser extent, the MHI-5.

## Psychometric properties

The MHI has undergone various examinations concerning its psychometric properties.

### *Content validity*

Several attempts have been made to explore the content validity of the MHI, and certain items have been found to be problematic with particular sub-populations. Gowans et al,[180] for example, found a multicultural sample had difficulty in understanding the nuances that distinguish some of the items on the Depression subscale (e.g., Item 9 – 'Did you feel depressed during the past month?', Item 19 – 'How much of the time, during the past month, have you felt downhearted and blue?' and Item 36 – 'During the past month, how much of the time have you been in low or very low spirits?'), and consequently could not rate them independently. Huebeck and Neill[181] deleted certain items (e.g., Item 21 – 'During the past month, how often have you felt that others would be better off if you were dead?' and Item 28 – 'During the past month, did you think about taking your own life?') in a study of adolescents, due to ethical concerns raised by the severity of their language. Likewise, Ell et al[162] eliminated items deemed inappropriate for newly diagnosed cancer patients.

### *Construct validity*

Numerous studies have examined the internal consistency of the full-length MHI, as measured by Cronbach's alpha. In these studies, Cronbach's alpha has ranged from 0.63 to 0.93 for the subscales, 0.90-0.97 for the global scales and 0.93 to 0.97 for the total score.[3, 156, 159, 179, 182-194] Together, these studies indicate that the MHI has a high internal consistency.

A number of studies have examined the structure of the MHI, considering how well the observed data fit the original model proposed by Veit and Ware.[156] Most have found support for the original structure.[179, 195, 196] However, some have only found support for the two- factor model (i.e., the global scores)[181, 185] Others have found a poor fit across all models and/or suggested alternative solutions.[160, 197, 198]

*Concurrent validity*

Numerous studies have considered the concurrent validity of the MHI, examining the relationship between its subscales and global scales and comparable constructs on other standardised measures. The following instruments have shown to correlate well with the full-length MHI in the predicted direction: the Brief Pain Inventory (BPI),[199] the Medical Outcome Study Social Support Survey (MOS-SSS),[158] the Life Experience Survey (LES),[158] the Systems of Belief Inventory (SBI),[158] the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire C30 (EORTC),[200] the Positive and Negative Affect Schedule (PANAS),[160] the Dyadic Adjustment Scale (DAS),[160] the Purpose in Life Test (PIL),[201] the Life Regard Index (LRI),[201] the Sense of Coherence Scale (SOC),[201] the MOS 36-item Short-Form Health Survey (SF-36)[3] and the Behaviour and Symptom Identification Scale (BASIS-32®).[3] However, poor correlations have been found between the MHI and the Life Skills Profile (LFS),[3] Role Functioning Scale (RFS),[3] and Health of the Nation outcome scales (HoNOS)[3] and the Global Assessment Scale (GAS).[198]

Several studies have also considered the concurrent validity of the shorter MHI-5, also assessing its performance against other, more established measures. The MHI-5 has been shown to correlate well with the General Health Questionnaire (GHQ-12),[193] the Rhode Island Stress and Coping Inventory (RISCI),[202] the General Health Rating Index (GHRI),[203] the Quality of Well Being Scale (QWB)[203] and the Sickness Impact Profile (SIP).[203] The MHI-5 was also found to be highly correlated with all dimensions of the SF-36, in particular Role limitations due to emotional problems (RE), Social Functioning (SF), Vitality (VT) and General Health Perceptions.[191, 193]

Other studies have assessed the concurrent validity of the full-length MHI to discriminate between subgroups of consumers. Analyses in these studies have found that the MHI can discriminate between those who have experienced stressful life events and those who have not,[158, 159, 193] those with low level of social support and those with good social networks,[158] and those with poor physical health and those with no medical problems.[158, 204, 205] Furthermore, MHI scores have been shown to relate to mental health service use[193] and to discriminate between clinical and non-clinical samples, at least for certain diagnostic groups.[165, 206, 207]

*Predictive validity*

Several studies have explored the predictive validity of the MHI. In particular, these studies have focused on the ability of the instrument to predict future service use. Results from the Health Insurance Study indicated that persons with low total scores on the MHI (i.e., high psychological distress and low psychological wellbeing) were more likely to receive mental health care than those with higher scores.[205, 208]

*Test-retest reliability*

A number of studies have examined the test-retest reliability of the MHI with retest periods of up to one year.[3, 156, 181] Generally the reliabilities reported have been adequately high for subscale, global scale and total scores on the MHI. These studies have reported correlations ranging from 0.56 to 0.97, indicating that the MHI is relatively stable over time.

*Inter-rater reliability*

Only a few studies have examined inter-rater reliability of the MHI.[170, 197] Overall, these studies have found significant correlations between the self-report and interviewer-administered versions of the MHI. However, there are suggestions that some items (e.g., those to do with suicidal ideation) may have lower inter-rater reliability than the majority.

*Sensitivity to change*

Stedman et al[3] investigated the sensitivity to change of the MHI by examining whether consumer's global reports of mental health improvement, stability or deterioration correlated with

changes in MHI scores.  They found that those who rated themselves as having improved showed corresponding improvements on the majority of MHI subscales, those who rated themselves as stable showed no change on the MHI, and those who rated themselves as having deteriorated showed a decline on the MHI.

***Feasibility and utility***

Stedman et al[129] asked consumers to rate the MHI's utility, using a quantitative scale that defined utility in terms of perceived relevance, effectiveness and usefulness.  The MHI scored higher than either the BASIS-32 or the SF-36.  Stedman and colleagues have also qualitatively explored the opinions of consumers and other stakeholders regarding the feasibility and utility of the MHI and have found them to report that it is comprehensive, easy to understand, user-friendly, acceptable and appropriate.  Having said this, some have criticised the wording of specific items, its assessment period, its response options, its relevance to particular subgroups, and its coverage.[3, 129]

## Summary

The MHI is a self-report instrument developed for use in the RAND Health Insurance Experiment to assess psychological distress and wellbeing of people in the general population.  Its full-length version comprises 38 items.  Most items are rated on a scale from 1 to 6 based on frequency or intensity where higher scores reflect more frequent occurrence of favourable mental health symptoms.

In tests of its psychometric properties, the MHI has been shown to have adequate to good content, construct, concurrent, predictive validity, test-retest and inter-rater reliability. The MHI has demonstrated sensitivity to change as well as feasibility and utility.

# Chapter 10: Behaviour and Symptom Identification Scale 32 (BASIS-32®)

## Background

The BASIS-32® was developed by Eisen and colleagues at McLean Hospital in the United States as a consumer-oriented measure of symptoms and behavioural distress.[209, 210] The instrument was originally developed and validated among inpatients, but subsequent studies have supported its use in outpatient and residential settings.[211-217] As a measure of mental health outcomes, the BASIS-32® has been widely used for research and quality-improvement purposes.[218-222]

## Purpose

The purpose of the BASIS-32® is to measure changes in self-reported symptom and problem difficulty over the course of treatment for people with a mental illness.[223]

## Availability

The BASIS-32® is a commercial instrument and is not available in the public domain. Copyright is held by the McLean Hospital, and there is an annual fee and site licence.[18]

## Description

The BASIS-32® is a consumer-rated instrument comprising 32 items:
- Item 1: Managing day-to-day life (e.g., getting places on time, handling money, making everyday decisions)
- Item 2: Household responsibilities (e.g., shopping, cooking, laundry, cleaning, other chores)
- Item 3: Work (e.g., completing tasks, performance level, finding/keeping a job)
- Item 4: School (e.g., academic performance, completing assignments, attendance)
- Item 5: Leisure time or recreational activities
- Item 6: Adjusting to major life stressor
- Item 7: Relationships with family
- Item 8: Getting along with people outside the family
- Item 9: Isolation or feelings of loneliness
- Item 10: Being able to feel close to others
- Item 11: Being realistic about yourself and others
- Item 12: Recognising and expressing emotions appropriately
- Item 13: Developing independence, autonomy
- Item 14: Goals or direction in life
- Item 15: Lack of self confidence, feeling bad about yourself
- Item 16: Apathy, lack of interest in things
- Item 17: Depression, hopelessness
- Item 18: Suicidal feeling or behaviour
- Item 19: Physical symptoms
- Item 20: Fear, anxiety or panic
- Item 21: Confusion, concentration, memory
- Item 22: Disturbing or unreal thoughts or beliefs
- Item 23: Hearing voices, seeing things
- Item 24: Manic, bizarre behaviour
- Item 25: Mood swings, unstable moods
- Item 26: Uncontrollable, compulsive behaviour
- Item 27: Sexual activity or preoccupation
- Item 28: Drinking alcoholic beverages

- Item 29: Taking illegal drugs, misusing drugs
- Item 30: Controlling temper, outbursts of anger, violence
- Item 31: Impulsive, illegal or reckless behaviour
- Item 32: Feeling satisfaction with your life

Collectively, the items constituting the BASIS-32® comprise five subscales:
- Relation to self and others (Items 7, 8, 10, 11, 12, 14, 15);
- Depression and anxiety (Items 6, 9, 17, 18, 19, 20);
- Daily living and role functioning (Items 1, 2, 3, 4, 5, 13, 16, 21, 32);
- Impulsive and addictive behaviour (Items 25, 26, 28, 29, 30, 31); and
- Psychosis (Items 22, 23, 24, 27).[224]

Each item is rated using a five-point scale (0 = no difficulty; 1 = a little difficulty; 2 = moderate difficulty; 3 = quite a bit of difficulty; 4; extreme difficulty) which are used to calculate subscales and total scores by adding the ratings for each item and dividing by the number of non-omitted items. The exception to this rule is the computation of the Daily living and role functioning subscale, which is determined by taking the highest difficulty rating of items 2, 3 and 4 to create a single 'role functioning' rating, and averaging this value in with the remaining six items.[224]

Ratings are based on the difficulty experienced during the preceding period (one week in the original instrument; two weeks in the NOCC collection), ascertained by structured interview (either with a rater present or by telephone) or self-report format (either on-site or through mail-out).[214] The structured interview is generally administered if the consumer is not capable of self-report (e.g. due to illiteracy or an excess of symptoms) in which the rating scale choices are presented to the consumer on large index cards. The interviewer reads each item to the consumer who is then required to indicate their response on the cards.[225]

In its interviewer-administered form, the BASIS-32® may be used by both professional and non-professional staff due to the absence of training requirements. An instruction manual is available which contains a survey form, a scoring algorithm, a reference list and articles relating to methodology, reliability and validity.[223]

The BASIS-32® takes on average 5-10 minutes to administer in the self-report format, while the structured interview takes approximately 15-10 minutes to complete.[226]

## Versions

There are several versions of the BASIS. As noted above, the administration of the original BASIS-32® involves a rating period of one week. A modified version of the original instrument is being used under the NOCC protocol; this Australian version involves a rating period of two weeks.

The original instrument has recently undergone further development resulting in the newly released 24-item version, the BASIS-24® (which involves a one week rating period, in line with the Australian version). Many of the features of the original instrument were discarded in the development process, and, in fact, only three of the original items remain in the new version.[227, 228]

In addition, the BASIS-32® has been translated into many languages including Spanish, French, Japanese, Chinese, Korean, Cambodian, Vietnamese and Tagalog.[215]

The psychometric properties of the BASIS are considered below. The majority of relevant studies in this area have focused on the original BASIS-32®.

## Psychometric properties

Compared with some other instruments, the BASIS-32® has undergone extensive and comprehensive psychometric testing.

### Content validity

Several attempts have been made to assess the content validity of the BASIS-32®. Eisen and colleagues,[217, 229] Cameron et al,[230] Graham et al[231] and Stedman et al[3] assessed consumer acceptability of the BASIS-32® by soliciting comments following the administration of the instrument. Eisen et al's[217, 229] participants reported that the instrument was comprehensive, but participants in the other reports had mixed responses. Concerns identified in the work by Cameron et al,[230] Graham et al[231] and Stedman et al[3] related to ambiguous and complex language, an exclusive focus on difficulties and issues with content areas. Recommendations from consumers included the addition of items to cover the outcome domains of greater relevance to them.[230, 231]

In addition to these general criticisms, respondents in these studies expressed specific concerns over the use of certain items. This led both Cameron et al[230] and Graham et al's[231] to oppose the use of items with multiple content, including Items 17 (Depression, hopelessness), 19 (Physical symptoms) and 21 (Confusion, concentration, memory), and to regard Items 4 (School), 22 (Disturbing or unreal thoughts or beliefs), 27 (Sexual activity or preoccupation) and 29 (Taking illegal drugs, misusing drugs) as invasive or insensitive. In addition, Cameron et al[230] believed Items 11 (Being realistic about yourself and others) and 12 (Recognising and expressing emotions appropriately) contained concepts that were too complicated, and Graham et al[231] expressed concern about the appropriateness of the content of Item 7 (Relationships with family) due to possible variation in its interpretation.

Furthermore, the BASIS-32® has been criticized for requiring a reading level too high for individuals with limited literacy skills, and for being difficult to complete for individuals who are acutely psychotic or intoxicated, or have dementia.[211, 213, 222, 227] Indeed, one of the reasons for the development of the BASIS-24® was to overcome these concerns.[227]

### Construct validity

Numerous studies have examined the internal consistency of the BASIS-32®, as measured by Cronbach's alpha. In these studies, Cronbach's alpha of subscales has ranged from 0.6 to 0.9, indicating the that the BASIS-32® has a high level of internal consistency across settings (inpatients and outpatients) and types of administration (structured-interview and self-administration).[3, 210-214, 217, 221, 229, 232, 233] In these studies, the Impulsive and addictive behaviour subscale and the Psychosis subscale have consistently achieved the lowest Cronbach's alpha (i.e., the poorest internal consistency). Their failure to reach recommended levels was one of the impetuses for the development of the BASIS-24®, according to Eisen et al.[227]

Studies conducted by Eisen and colleagues,[217, 229] and by Hoffman et al,[211] Russo et al,[213] and Chow et al[232] examined the subscale structure of the BASIS-32® and how well the observed data fit the five-factor model derived from the original sample of inpatients.[210] The original five-factor model was confirmed by an analysis of inpatients.[211, 213] It was found to fit adequately to samples of community-based consumers,[217, 229, 232] although items on the Impulsive and addictive behaviour subscale (Item 26 – Uncontrollable, compulsive behaviour; Item 28 - Drinking alcoholic beverages; and Item 29 - Taking illegal drugs, misusing drugs)[217, 232] and Psychosis subscale (Item 23 - Hearing voices, seeing things; and Item 27 - Sexual activity or preoccupation)[217] had low item-scale correlations. Eisen et al[227] suggested that the factor structure may not generalise well to ambulatory care settings, again citing this as one of the reasons for the development of the BASIS-24®.

Several studies have also analysed the intercorrelations between the subscales. These studies have shown that three of the five subscales (Relation to self and others; Depression and anxiety; and Daily living and role functioning) are highly intercorrelated, indicating this instrument may be measuring a single dimension rather than distinct aspects of psychological functioning.[210, 213, 229]

### *Concurrent validity*

Numerous studies have considered the concurrent validity of the BASIS, with the majority comparing individual item scores, subscale scores and total scores on the BASIS-32® with equivalent scores on other standardised measures. The following instruments have been shown to correlate well with the BASIS-32® in the predicted direction: the Client Assessment of Strengths Interests and Goals – Self Report (CASIG-SR),[234] the Client Assessment of Strengths Interests and Goals (CASIG),[235] the Outcome Questionnaire (OQ-45),[220] the Short Form Health Status Profile (SF-36),[3, 217, 229] the Symptom Checklist (SCL-90),[209, 211] the Child and Adolescent Functional Assessment Scale (CAFAS),[211] the Hopkins Symptom Checklist (MSCL-43),[212] the Brief Psychiatric Rating Scale (BPRS),[212] the Mental Health Inventory (MHI),[3] the Camberwell Assessment of Need (CAN),[236] the Psychiatric Symptom Assessment Scale (PSAS),[213] the Outcome Assessment Program (OAP) Questionnaire,[237] the Health of the Nation Outcome Scales (HoNOS),[3] and Lehman's Quality of Life Interview (QOLI).[213] By contrast, performance on the BASIS-32® has been found to be unrelated to performance on the Social and Occupational Functioning Assessment Scale (SOFAS-Revised GAF),[213] the Mini Mental Status Examination (MMSE),[213] the Levels of Recovery from Psychotic Disorders Scale (LORS),[238] the Global Assessment of Relational Functioning Scale (GARF),[239] the Life Skills Profile (LSP)[3] and the Role Functioning Scale (RFS).[3]

A second group of studies has assessed the concurrent validity of the BASIS-32® by examining its ability to discriminate between consumers. The BASIS-32® has been shown to correlate well with objective indicators of functioning, effectively discriminating between consumers who were rehospitalised,[209-211] and those who were currently employed or undertaking studies,[209, 210] and between inpatients and outpatients.[217] It has also been shown to discriminate between consumers from particular diagnostic groups. In general, for example, consumers with depression, psychotic disorders and substance abuse problems have been shown to score highly on the Depression and anxiety, Psychosis and Impulsive and addictive behaviours subscales, respectively.[209-211, 213, 217, 220, 240] Likewise, consumers with comorbid mental health and substance abuse problems have been shown to score significantly higher on all five subscales than their counterparts with less complex conditions.[233, 241, 242] There are some exceptions to this rule, however, with several studies being unable to establish a connection between the BASIS-32® score and diagnosis.[212, 220]

### *Predictive validity*

Few studies have examined the predictive utility of the BASIS-32®, but those that have have found the instrument to be able to predict future service use. These studies have observed that, when assessed at discharge, consumers who subsequently require rehospitalisation score higher than their counterparts who maintain community tenure.[209, 211, 229]

### *Test-retest reliability*

Only a few studies have examined the test-retest reliability of the BASIS-32®, but their findings are uniformly positive, suggesting that the instrument produces consistent results when rated by the same rater at different points in time.[3, 209, 210, 212]

### *Inter-rater reliability*

Only a limited number of studies have considered the inter-rater reliability of the BASIS-32®. These have either assessed the self-report version of the instrument against the interview version, or have compared self-report with reports of a close informant. These studies have consistently found good overall inter-rater reliability.[209, 212]

*Sensitivity to change*

Several studies have assessed the sensitivity of the BASIS-32® to change by following consumers over time and comparing pre- and post-treatment ratings with independent pre- and post- measures of improvement, deterioration or stability. In general, these studies have found the BASIS-32® to be highly sensitive to change following treatment in both inpatient and outpatient samples.[3, 209-211, 213, 217, 229, 236, 243, 244] The Impulsive and addictive behaviours subscale and the Psychosis subscale showed poorer sensitivity to change than the other three subscales.[213, 236]

*Feasibility and utility*

Empirical studies and published commentaries suggest that the BASIS-32® demonstrates adequate feasibility and utility. Stedman et al, for example, found that it was rated favourably by consumers in terms of its utility (defined in terms of its perceived relevance, effectiveness and usefulness), although it was not ranked as highly as the MHI. Others have found it to be applicable to a wide range of people receiving mental health treatment, not limited by diagnoses, symptom patterns or treatment setting, user-friendly, adaptable due to its alternative administration modes. It has also been found to place minimal burden on staff due to its brevity, the simplicity of its design and its absence of training requirements.[212, 216, 220, 223, 229]

Having said this, there have been criticisms of the feasibility and utility of the BASIS-32®. Firstly, the developers themselves have noted that there is 'unnecessary redundancy in the instrument', and cite this as one of the reasons for the development of the shorter BASIS-24®.[227] Secondly, Higgins and Purvis[245] studied the usefulness of the instrument of evaluating program-level outcomes in Californian mental health services. They found that although it had clinical relevance at an individual level, it did not work well when data were aggregated. Consumers consistently and dramatically under-reported their symptoms, and improvements over time were negligible (although statistically significant), rendering it of limited utility at a system level. In addition, the costs and licencing demands meant that other tools were preferable.

## Summary

The BASIS-32® is a consumer-rated measure that was originally developed to assess outcomes among inpatients with mental health problems. Subsequent studies have confirmed its utility as an outcome measure for use across a range of mental health settings. The instrument comprises 32 items which collectively measure symptoms and behavioural distress in people with a mental illness. Each item is rated from 0 (no difficulty) to 4 (extreme difficulty), resulting in 32 individual scores, five subscale scores and a single total score.

The BASIS-32® has been shown to have adequate validity and reliability, and to be sensitive to change during treatment (although the Impulsive and addictive behaviours subscale and the Psychosis subscale perform less well in these areas than do the other three subscales). Arguably, the instrument is also regarded as demonstrating adequate feasibility and utility. Having said this, it should be noted that the developers of the instrument have since prepared a revised version, the BASIS-24®, which only retains three of the original items.

# Chapter 11:  Kessler-10 Plus (K-10+)

## Background

The Kessler-10 (K-10) was developed by Kessler and colleagues, for use as a measure of non-specific psychological distress to be incorporated into the United States National Health Interview Survey (NHIS).[246] Its development began with the selection of an initial pool of 612 items, chosen from existing screening scales.  These items were whittled down to 45 by a process that involved refining the domains of interest and discarding redundant and unclear items.  Further refinements occurred on the basis of analysis of data from a mail pilot survey and a telephone pilot survey, resulting in the 10-item instrument.  The K-10+ includes an additional four items over and above those of the K-10 which relate to the functional impairment associated with identified distress (see below).[247]

## Purpose

As noted above, the K-10 is designed as a screener for non-specific psychological distress.[246] The tool was originally designed for use in epidemiological surveys, but was deliberately constructed in a manner that would allow it to have utility in clinical settings as well.[176] The 'plus' questions in the K-10+ go beyond the concept of psychological distress captured by the original 10 questions, and assess distress-specific global impairment, distress-specific service use and self-reported physical health contribution to distress.

## Availability

The K-10 is readily available and can be downloaded free-of-charge.[248]

## Description

The K-10 is a self-report measure of psychological distress, in which the consumer completes questions about symptoms of depression and anxiety in the past four weeks.  It is designed to span the range from few or minimal symptoms through to extreme levels of distress, and consequently contains both low-threshold items which many people may endorse and high-threshold items which very few will endorse.  For each item, the consumer indicates the amount of time during the four-week period that he or she experienced the particular problem.  There is a five level response scale that ranges from none of the time (1) to all of the time (5).  The maximum score is 50, indicating severe distress, and the minimum score is 10, indicating no distress.[176, 247] The specific items are as follows:

- Item 1: In the last four weeks, about how often did you feel tired out for no good reason?
- Item 2: In the last four weeks, about how often did you feel nervous?
- Item 3: In the last four weeks, about how often did you feel so nervous that nothing could calm you down?
- Item 4: In the last four weeks, about how often did you feel hopeless?
- Item 5: In the last four weeks, about how often did you feel restless or fidgety?
- Item 6: In the last four weeks, about how often did you feel so restless you could not sit still?
- Item 7: In the last four weeks, about how often did you feel depressed?
- Item 8: In the last four weeks, about how often did you feel that everything was an effort?
- Item 9: In the last four weeks, about how often did you feel so sad that nothing could cheer you up?
- Item 10: In the last four weeks, about how often did you feel worthless?[247]

As mentioned above, the K-10+ includes four additional items (Items 11-14) that constitute follow-up questions.[b] These items aim to quantify the level of disruption and disability resulting from the problems reported, in terms of the degree of limitation of normal activity, and/or seeking help for the problems. Items 11 and 12 require a response in terms of number of days; Item 13 requires a response in terms of number of consultations; and Item 14 follows the same response format as Items 1-10.

- Item 11: In the last four weeks, how many days were you totally unable to work, study or manage your day to day activities because of these feelings?
- Item 12: [Aside from those days], in the last four weeks, how many days were you able to work or study or manage your day to day activities, but had to cut down on what you did because of these feelings?
- Item 13: In the last four weeks, how many times have you seen a doctor or any other health professional about these feelings?
- Item 14: In the last four weeks, how often have physical health problems been the main cause of these feelings?[247]

These questions are consistent with the notion that symptoms are necessary but not sufficient for a person to be considered as having a disorder; impairment must also be present.[249] They are also consistent with the intent of similar questions used in numerous population health surveys to elicit information on the functional impact of distress and service utilisation related to distress (e.g., the National Survey of Mental Health and Wellbeing,[250] which used similar questions in addition to the core K-10). The specific four questions that constitute the 'plus' part of the K-10+ were first used in Australia in the 1996 New South Wales Health Survey.

## Versions

Several versions of the K-10 exist, and two are of particular relevance here. As noted above, the K-10+ includes an additional four questions that capture functional impairment associated with psychological distress. The K-10-L3D differs from the K-10 by rating period, and refers to the last three days, rather than the last four weeks. The NOCC protocol employs the K-10+ with its four week rating period at all collection occasions except discharge from an inpatient setting. At the latter collection occasion, the K-10-L3D (including the additional four questions from the K-10+) is employed, in recognition of the brevity of episodes in inpatient settings.

In addition to the above versions of the K-10, there is a six item instrument (the K-6), which was developed at the same time as the K-10, in the event that space requirements in the NHIS demanded greater brevity than that afforded by the K-10.[246] The K-10 has also been translated into many languages, including Arabic, Bosnian, Chinese, Croatian, Farsi, Greek, Hindi, Italian, Korean, Macedonian, Serbian, Spanish, Tagalog, Turkish and Vietnamese.[247]

The current chapter focuses primarily on the original K-10, since this version of the instrument is the one that receives the greatest attention in the psychometric literature.

## Psychometric properties

Various tests of the psychometric properties of the K-10 have been undertaken.

### Content validity

Ensuring that the K-10 had strong content validity was a priority during its original development. The process of selecting and refining the potential items on the instrument was assisted by an expert advisory panel of survey researchers, which rated each potential item for clarity and

---

[b] Items 11-14 are not answered if the consumer responded 'None of the time' to all of the questions covered by Items 1-10.

wording.  Only those items that were consistently rated as clear were included in the pool of items from which the final 10 were chosen.[246]

Having said this, little follow-up work has been done since the original instrument was released in terms of seeking assurance from respondents that the K-10 measures what it purports to measure.  There has been some published commentary and debate in this regard, although it has focused more on the appropriate clinical cut-off points to use when scoring the instrument, and less on the content of individual items.[176]

### Construct validity

Early work on the K-10 indicated that the instrument has good precision in the 90th-99th percentile of the population distribution (i.e., standard errors of standardized scores in the range 0.20-0.25), as well as good internal consistency (i.e., a Cronbach's alpha of 0.93).[246]  Later work has provided support for the construct validity of the instrument.  For example, an examination of data from the Australian National Survey of Mental Health and Wellbeing (NSMHWB) yielded a Cronbach's alpha of 0.92.[246]

### Concurrent validity

K-10 scores have been shown to be significantly correlated with other instruments that measure symptomatology and/or disability, including the General Health Questionnaire (GHQ),[176] the Short Form 12 (SF-12),[176] Comprehensive International Diagnostic Interview – Short Form (CIDI-SF),[251] and the World Health Organization Disability Assessment Schedule (WHO-DAS).[251]

The K-10 has also been shown to discriminate between cases and non-cases of particular DSM-IV/SCID disorders (particularly anxiety disorders and mood disorders, but also other mental disorders) in the general population, and to do so as or more effectively than the GHQ, the CIDI-SF and the WHO-DAS.[176, 246, 248, 251, 252]  For example, when data from the Australian National Survey of Mental Health and Wellbeing[250] were fitted to a standard Receiver Operating Characteristic (ROC) curve to assess the extent to which K10 distress scores predicted the presence or absence of an anxiety or affective disorder (as assessed by the Comprehensive International Diagnostic Interview (CIDI)), the Area Under the Curve (AUC) was 0.89, which is indicative of very good performance (a perfect measure scoring 1.0).[248]

In addition, a positive relationship has been observed between current K-10 scores and the number of consultations for mental health problems in the previous 12 months,[176] and between K-6 scores and levels of work, life and other stressors.[253-255]

### Predictive validity

No published literature exists on the predictive validity of the K-10.

### Test-retest reliability

In the peer-reviewed literature, there are no published studies of the reliability of the K-10.  Reference is made to the instrument's reliability in a brief report from the South Australian Department of Human Services, which says, 'Reliability tests were done on the K-10 … the values of the kappa and weighted kappa scores ranged from 0.42 to 0.74 which indicates that the K-10 is a moderately reliable instrument'.  However, no further detail is provided about the methodology used.[256]

### Inter-rater reliability

The concept of inter-rater reliability is not relevant for the K-10, since it is designed to be self-administered so there is only one rater.  In this respect, it differs from the other adult self-report measures in the NOCC suite – the MHI and the BASIS-32® –  both of which can be self-administered or administered by structured interview.

*Sensitivity to change*

There are no published data on the ability of the K-10 to detect change in psychological distress over time.

*Feasibility and utility*

There are no published studies that have specifically examined the feasibility of the K-10 as a routine outcome measure. Likewise, there are no published studies that have considered the acceptability and utility of the instrument from the perspective of consumers, carers, clinicians and/or managers, although its brevity augurs well in this regard. Having said this, it must be noted that the K-10 is widely used in a range of settings, suggesting that many relevant stakeholders favour its use. In Australia, it has been incorporated into population surveys that have been conducted nationally[250, 257] and at a state/territory level[258-261] (although only the New South Wales Health Survey[262, 263] has incorporated the K10+). In addition, it is being widely used in primary care settings in Australia, where, as one of the key outcome measures of choice in the Better Outcomes in Mental Health Care initiative, it has received a positive response from GPs.[264, 265]

## Summary

The K-10+ is a version of the K-10, which was developed for use as a measure of non-specific psychological distress. It was originally designed for use in the United States National Health Interview Survey, but was deliberately constructed in a manner that would allow it to have utility in clinical settings as well.

The K-10 is a 10-item self-report measure which asks the consumer about symptoms of depression and anxiety in the past four weeks. The K-10+ includes an additional four items that quantify the level of disruption and disability resulting from the problems identified in the first 10 items.

The K-10 is extremely widely used, both as a measure of mental health status in general population surveys and as an outcome measure in primary care settings, suggesting that it is well-regarded by the mental health field. The published studies on the psychometric properties of the K-10 are not extensive, but the instrument appears to have adequate to good content, construct and concurrent validity, and test-retest reliability. Its predictive validity and sensitivity to change require further exploration. So too do its feasibility and utility, although its brevity and widespread use in a range of settings augur well in this regard.

# Chapter 12:  Strengths and Difficulties Questionnaire (SDQ)

## Background

The Strengths and Difficulties Questionnaire (SDQ) was developed by Goodman in the United Kingdom as a brief screening tool that describes children and adolescents' behaviours, emotions and relationships.  It was designed to address some of the short-comings of other similar instruments:  to be concise; to be applicable to both children and adolescents; to have versions that could be completed by both parents and teachers, as well as a self-report version (for older children/adolescents); to represent both negative and positive attributes; and to equally cover five dimensions (namely conduct problems, emotional symptoms, hyperactivity, peer relationships and prosocial behaviour).[266]

## Purpose

According to Goodman,[267] the SDQ can be used for screening, as part of a clinical assessment, as a treatment-outcome measure, and as a research tool.

## Availability

The SDQ can be downloaded free-of-charge.[268]

## Description

There are a number of versions of the SDQ:  Parent-rated and teacher-rated versions for children aged 4-10 and adolescents aged 11-17; and a self-report version for adolescents aged 11-17.  For most items, the major differences between versions are the way in which the child/adolescent is referred to ('your child' in the parent-rated versions, 'this child' in the teacher-rated versions, and 'you' in the self-report version) and some wording modifications to ensure relevance to the age group in question.  The parent-rated and self-report versions of the instrument are mandatory inclusions in the NOCC dataset; the teacher-rated version is optional.  The SDQ is the only consumer *and carer* instrument in the NOCC suite.

In each version, the parent, teacher or child/adolescent is asked to consider the child/adolescent's behaviour over the past six months (or over the last one month in the case of follow-up administrations), and consider whether the statement is 'not true', 'somewhat true' or 'certainly true'.  A modification was made for the purposes of using the instrument at follow-up, such that any SDQ collected at baseline considers the previous six months and any follow-up SDQ considers the past month only.

The core instrument, which can be completed in about five minutes,[269] contains 25 items that depict either a positive or negative attribute.  The specific items are shown in Table 3:

**Table 3:  SDQ core items by version**

| Informant | Parent-rated or teacher-rated | | Self-report |
|---|---|---|---|
| Age | 4-10 | 11-17 | |
| Instruction | … on the basis of your/this child/student's behaviour over the last six months/one month | | … how things have been for you over the last six months/one month |
| Item 1 | Considerate of other people's feelings | Considerate of other people's feelings | I try to be nice to other people. I care about their feelings |
| Item 2 | Restless, overactive, cannot stay still for long | Restless, overactive, cannot stay still for long | I am restless, I cannot stay still for long |
| Item 3 | Often complains of headaches, stomach aches, or sickness | Often complains of headaches, stomach aches, or sickness | I get a lot of headaches, stomach aches or sickness |

| Informant | Parent-rated or teacher-rated | | Self-report |
|---|---|---|---|
| Age | 4-10 | 11-17 | |
| Item 4 | Shares readily with other children, for example toys, treats, pencils | Shares readily with other children, for example CDs, games, food | I usually share with others, for example CDs, games, food |
| Item 5 | Often loses temper | Often loses temper | I get very angry and often lose my temper |
| Item 6 | Rather solitary, prefers to play alone | Would rather be alone than with other young people | I would rather be alone than with people of my age |
| Item 7 | Generally well behaved, usually does what adults request | Generally well behaved, usually does what adults request | I usually do as I am told |
| Item 8 | Many worries or often seems worried | Many worries or often seems worried | I worry a lot |
| Item 9 | Helpful if someone is hurt, upset or feeling ill | Helpful if someone is hurt, upset or feeling ill | I am helpful if someone is hurt, upset or feeling ill |
| Item 10 | Constantly fidgeting or squirming | Constantly fidgeting or squirming | I am constantly fidgeting or squirming |
| Item 11 | Has at least one good friend | Has at least one good friend | I have one good friend or more |
| Item 12 | Often fights with other children or bullies them | Often fights with other young people or bullies them | I fight a lot. I can make other people do what I want |
| Item 13 | Often unhappy, depressed or tearful | Often unhappy, depressed or tearful | I am often unhappy, depressed or tearful |
| Item 14 | Generally liked by other children | Generally liked by other young people | Other people my age generally like me |
| Item 15 | Easily distracted, concentration wanders | Easily distracted, concentration wanders | I am easily distracted, I find it difficult to concentrate |
| Item 16 | Nervous or clingy in new situations, easily loses confidence | Nervous in new situations, easily loses confidence | I am nervous in new situations. I easily lose confidence |
| Item 17 | Kind to younger children | Kind to younger children | I am kind to younger children |
| Item 18 | Often lies or cheats | Often lies or cheats | I am often accused of lying or cheating |
| Item 19 | Picked on or bullied by other children | Picked on or bullied by other young people | Other children or young people pick on me or bully me |
| Item 20 | Often volunteers to help others (parents, teachers, other children) | Often volunteers to help others (parents, teachers, other children) | I often volunteer to help others (parents, teachers, children) |
| Item 21 | Thinks things out before acting | Thinks things out before acting | I think before I do things |
| Item 22 | Steals from home, school or elsewhere | Steals from home, school or elsewhere | I take things that are not mine from home, school or elsewhere |
| Item 23 | Gets along better with adults than with other children | Gets along better with adults than with other young people | I get along better with adults than with people my own age |
| Item 24 | Many fears, easily scared | Many fears, easily scared | I have many fears, I am easily scared |
| Item 25 | Good attention span, sees chores or homework through to the end | Good attention span, sees chores or homework through to the end | I finish the work I'm doing. My attention is good |

The 25 items 'roll up' into five scales:
- Emotional symptoms (Items 3, 8, 13, 16, 24);
- Conduct problems (Items 5, 7, 12, 18, 22);
- Hyperactivity-inattention (Items 2, 10, 15, 21, 25);

- Peer problems (Items 6, 11, 14, 19, 23); and
- Prosocial behaviour (Items 1, 4, 9, 17, 20).

For the majority of items, a response of 'not true' is scored 0, a response of 'somewhat true' is scored 1 and a response of 'certainly true' is scored 2. The exceptions are Items 7, 11, 14, 21 and 25, where the reverse scoring order applies. This scoring method yields a score on each scale of 0-10 for each scale, and a total difficulties score of 0-40, generated by summing the scores from all of the scales except the Prosocial behaviour scale.[267, 269]

Extended versions of the SDQ which can be used at baseline and follow-up include the 25 core items plus a brief impact supplement that asks whether the respondent thinks that the child or youth has a problem, and, if so, inquires further about overall distress, social impairment, burden and chronicity caused by the problem.[270] At baseline, the extended versions also seek information on the respondent's perceptions of how others view the child/adolescent. At follow-up, they also explore the impact of the given service on the child/adolescent's problems. The extended versions (including the impact supplement and the cross-informant questions) have been adopted under the NOCC protocol. Further detail of these additional items can be found on the SDQ website and in relevant NOCC documentation,[18] but is not included here, since the psychometric literature reviewed in the remainder of this chapter deals largely with the 25 core items.

## Versions

As noted above, several different versions of the SDQ exist, varying by informant. In addition, the SDQ is available in more than 40 languages.[267]

## Psychometric properties

Various studies have considered the psychometric properties of the SDQ.

### *Content validity*

There is a dearth of scientific evidence regarding the content validity of the SDQ. Published anecdotal reports and commentaries about the content of the instrument have generally been positive, although it has received some criticism for sampling a restricted range of behaviour.[271]

### *Construct validity*

A number of studies have examined the construct validity of the SDQ by conducting confirmatory factor analyses of the parent-rated,[267, 272-275] teacher-rated[267] and self-report[267, 273, 276, 277] versions in a variety of languages. In general, these studies found five-factor solutions that corresponded with the original scales proposed by Goodman, and have resulted in only minor suggestions about alternative factor structures[278] The exception was a study by Ronning et al, which found 'a variable and questionable fit' when data from the Norwegian self-report version of the SDQ were compared with the Goodman model.[277]

A body of studies has also considered the internal consistency of the SDQ, again considering the parent-rated,[267, 273, 274, 278-281] teacher-rated[267, 278, 79, 281] and self-report[267, 273, 276-279, 281-283] versions in a range of languages. Taken together, these studies suggest that the instrument has very good overall internal consistency, with Cronbach's alphas in the range of 0.70 for total difficulties. The individual scales (including the Impact scale) also generally appear to have satisfactory to good internal consistency, although some – notably Conduct problems and Peer problems – have comparatively poorer internal consistency. The parent-rated and teacher-rated scales tend to have better internal consistency than the self-report scales.

Several studies have also looked at correlations between the scales comprising the SDQ. Again, these studies have typically found support for the construct validity of the instrument.[267, 273]

Typically, they have found low to moderate cross-scale correlations, indicating that the scales measure comparatively independent constructs in terms of difficulties.

### Concurrent validity

A number of studies have pitted the SDQ against other more established measures that tap similar domains. Typically, these studies involve informants of a particular type (i.e., parent, teacher or youth) rating a given child/adolescent using the SDQ and the other measure at the same point in time. Collectively, these studies have shown that:
- the parent-rated SDQ performs well against the parent-rated Rutter questionnaires,[266] the Child Behavior Check List (CBLC),[273, 279, 284, 285] the Children's Depression Inventory (CDI),[273] the Revised Children's Manifest Anxiety Scale (RCMAS),[273] the Attention Deficit Hyperactivity Disorder Questionnaire (ADHDQ),[273] and an abbreviated version of the Child and Adolescent Burden Assessment (CABA),[270];
- the teacher-rated SDQ performs well against the teacher-rated Rutter questionnaires[266]; and
- the self-report version of the SDQ performs well against the self-rated Youth Self Report,[273, 278, 279, 284] self-report versions of the Children's Depression Inventory (CDI),[273] the Revised Children's Manifest Anxiety Scale (RCMAS),[273] and the Attention Deficit Hyperactivity Disorder Questionnaire (ADHDQ).[273]

Additional evidence for the concurrent validity of the SDQ comes from studies that have examined its ability to discriminate between particular groups of children/adolescents, on the basis of independent indicators of mental health problems. These demonstrated that the SDQ could distinguish between clinic-based and community-based samples of children/adolescents (or those receiving treatment and those not receiving treatment), and could do so at least as well as other, more established instruments like the Rutter questionnaires, the CBCL, and the YSR.[266, 270, 275, 280, 282-284, 286]

Within clinic-based samples, the SDQ has further been shown to be able to discriminate between children/adolescents with particular diagnoses or problematic behaviour, and again to do so as well as or better than other instruments like the CBCL and the YSR[275, 279, 284-289] In general, multi-informant SDQs have been shown to have greater sensitivity in identifying particular diagnoses than single informant SDQs.

### Predictive validity

To date, no studies have adequately examined the predictive validity of the SDQ.

### Test-retest reliability

Several studies have purported to assess the test-retest reliability of the SDQ, although in all cases the period between the first and second administration of the instrument was arguably too long to expect the rating to remain the same (ranging from 3-4 weeks to 12 months).[267, 270, 273, 275, 281] Nonetheless, all these studies reported good stability across time, even for younger children. In terms of raters, teacher ratings were the most stable, and youth self-report ratings the least. In terms of scales, total difficulty scores and hyperactivity-inattention scores were the most stable; the impact score was the least reliable.

### Inter-rater reliability

Various tests of the inter-rater reliability of the SDQ have been conducted, although it must be said that most of these have considered the correspondence between different types of raters using the different versions of the instrument, rather than the correspondence between the same types of rater using a single version of the instrument. The majority of studies that have been conducted have found positive correlations between different raters on the individual scale scores, the total difficulties score and the impact score,[266, 267, 273, 278, 279, 281-283] but there are some

exceptions.[272, 290, 291]  Although only moderate in magnitude, these correlations compare favourably with the average inter-rater agreement between different versions of similar instruments, identified in a meta-analysis by Archenbach and colleagues.[292]  In general, correlations between the parent and teacher versions of the SDQ have been found to be higher than correlations between either of these versions and the self-report version.  There is also variability by scale in terms of inter-rater agreement, although this is not consistent across studies.

### *Sensitivity to change*

Mathai, Anderson and Bourne[293] considered the sensitivity to change of the parent-rated, teacher-rated and self-report versions of the SDQ, by following 130 consecutive new attendees at a Victorian Child and Adolescent Mental Health Service for six months.  Changes in SDQ scores from baseline to follow-up were compared with changes in clinician-rated HoNOSCA scores over the same period.  There was a significant improvement over time in the total difficulties score on all versions of the SDQ, which corresponded to improvement as measured by the HoNOSCA.  There were also lowered levels of perceived difficulties and burden on the SDQ impact supplement, and a decrease in the overall impact.  Having said this, it should be noted that the study suffered from substantial loss to follow-up, and it was unclear whether the reference period covered by the second SDQ was the previous six months or the previous month.

### *Feasibility and utility*

Given the relatively extensive psychometric testing of the SDQ, it is surprising to find that very few studies have asked respondents using the different versions of the instrument to comment on its feasibility and utility.  The only study that provides any information in this regard is that of Goodman and Scott,[285] which found that  mothers who had used the SDQ and the CBCL were twice as likely to prefer the former.

## Summary

The SDQ is a brief screening tool that describes children/adolescents' behaviours, emotions and relationships.  Parent-rated, teacher-rated and self-report versions of the SDQ are available.  The core instrument comprises 25 items that depict a positive or negative attribute and 'roll up' into five scales: Emotional symptoms; Conduct problems; Hyperactivity-inattention; Peer problems; and Prosocial behaviours.  Extended versions of the SDQ include these items plus an impact supplement.

Extensive testing of the construct and concurrent validity and test-retest and inter-rater reliability of the SDQ has been undertaken, and the instrument is strong in terms of these psychometric properties.  Less work has been done in the areas of content validity, predictive validity, sensitivity to change and feasibility and utility, but those studies that do provide information in this regard indicate that the instrument performs well on these dimensions as well.

# Chapter 13:  Discussion

## Summary of findings

### *Descriptive characteristics of NOCC measures*

Tables 4 and 5 summarise the key descriptive characteristics of the NOCC measures, identified during the course of the review.  Table 4 shows that, collectively, the measures incorporate clinician- and consumer- (and parent-) perspectives on a range of mental health-related constructs relevant to adults, children/adolescents and older people.  Most of the instruments are short, with the longest (the MHI) comprising 38 items.

Table 4 also shows that some were purpose-designed as outcome measures (e.g., the HoNOS family), whereas others were developed as one-off measures of health status (e.g., as screening tools, like the K-10+) or as casemix measures (e.g., the RUG-ADL).  As Table 5 shows, the adaptation of the latter instruments for use in outcome measurement has sometimes meant that the rating period covered has had to be altered to cater for the brevity of episodes of care in particular settings.  In some cases (e.g., the SDQ), the developer of the instrument has made this change to increase its utility; in other cases (e.g., the HoNOS family), the change has been made specifically for the purposes of the NOCC collection.

**Table 4: Descriptive characteristics of NOCC measures (1): Rater, target age group, overarching constructs, number of items and reason for development**

|  | Rater | Target age group | Overarching constructs | No. of items | Reason for development |
|---|---|---|---|---|---|
| **HoNOS** | Clinician | Adult | Mental health and social functioning | 12 | Outcome measure |
| **HoNOSCA** | Clinician | Children and adolescents | Range of behavioural, symptomatic, social and impairment domains | 15 | Outcome measure |
| **HoNOS65+** | Clinician | Older persons | Mental health and social functioning | 12 | Outcome measure |
| **LSP-16** | Clinician | Adults and older persons | Disability | 16 | Status measure |
| **RUG-ADL** | Clinician | Older persons | Dependency or functional status | 4 | Casemix measure |
| **CGAS** | Clinician | Children and adolescents | Dysfunction | 1 | Status measure |
| **MHI** | Consumer | Adults and older persons | Psychological distress and wellbeing | 38 | Status measure |
| **BASIS-32®** | Consumer | Adults and older persons | Symptom and problem difficulty | 32 | Outcome measure |
| **K-10+** | Consumer | Adults and older persons | Non-specific psychological distress | 10 core; 4 additional | Status measure |
| **SDQ** | Consumer and parent | Children and adolescents | Behaviours, emotions and relationships | 25 core; varying additional depending on version | Status measure |

**Table 5: Descriptive characteristics of NOCC measures (2): Rating periods covered**

| | Inpatient | | | Community residential | | | Ambulatory | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Admission** | **Review** | **Discharge** | **Admission** | **Review** | **Discharge** | **Admission** | **Review** | **Discharge** |
| **HoNOS** | 2 weeks | 2 weeks | 3 days[a] | 2 weeks | 2 weeks | 2 weeks | 2 weeks | 2 weeks | 2 weeks |
| **HoNOSCA** | 2 weeks | 2 weeks | 3 days[a] | 2 weeks | 2 weeks | 2 weeks | 2 weeks | 2 weeks | 2 weeks |
| **HoNOS65+** | 2 weeks | 2 weeks | 3 days[a] | 2 weeks | 2 weeks | 2 weeks | 2 weeks | 2 weeks | 2 weeks |
| **LSP-16** | 3 months | 3 months | 3 months | 3 months | 3 months | 3 months | 3 months | 3 months | 3 months |
| **RUG-ADL** | Unspecified | Unspecified | Unspecified | Unspecified | Unspecified | Unspecified | Unspecified | Unspecified | Unspecified |
| **CGAS** | 2 weeks | 2 weeks | 2 weeks | 2 weeks | 2 weeks | 2 weeks | 2 weeks | 2 weeks | 2 weeks |
| **MHI** | 1 month | 1 month | 1 month | 1 month | 1 month | 1 month | 1 month | 1 month | 1 month |
| **BASIS-32®** | 2 weeks | 2 weeks | 2 weeks | 2 weeks | 2 weeks | 2 weeks | 2 weeks | 2 weeks | 2 weeks |
| **K-10+** | 4 weeks[c] | 4 weeks[c] | 3 days[d] | 4 weeks[c] | 4 weeks[c] | 4 weeks[c] | 4 weeks[c] | 4 weeks[c] | 4 weeks[c] |
| **SDQ** | 6 months | 1 month[e] | 1 month[e] | 6 months | 1 month[e] | 1 month[e] | 6 months | 1 month[e] | 1 month[e] |

\*      Shaded areas represent collection occasions at which collection of the given measure is mandated by the NOCC protocol.

a.    The standard rating period for the HoNOS, HoNOSCA and HoNOS65+ is 'the preceding two weeks', but the NOCC protocol specifies a rating period of 'the preceding three days' at discharge from an inpatient setting, in recognition of the brevity of admissions to such settings.

b.    The standard rating period for the BASIS-32® is 'the last week', but the jurisdictions that have incorporated the instrument into their routine collections have amended the rating period to 'the past two weeks', primarily to align the measure with the majority of clinician-rated measures.

c.    The standard rating period for the K-10+ is 'the last 30 days', but in Australian use the rating period has become 'the last four weeks'.

d.    As noted, the standard rating period for the K-10+ in Australia is 'the last four weeks'.  However, the jurisdictions that have incorporated the instrument into their routine collections are using the K-10L3D at discharge from an inpatient setting.  The rating period for the latter instrument is 'the last three days', which is chosen in recognition of the brevity of admissions to such settings.

e.    The standard rating periods for the SDQ are different at baseline and follow-up.  At baseline, the rating period is six months, and at follow-up it is one month.  These different rating periods are consistent with the original instrument, and not NOCC-specific  modifications.

*Psychometric properties of NOCC measures*

The main focus of the review was a critical appraisal of the psychometric properties of each of the NOCC measures. This task was limited to some extent by the fact that a number of the instruments being used in the NOCC collection constitute alternative versions of an original instrument, and the majority of studies concentrated on the original instrument. Version modifications include reductions in the number of items (e.g., the LSP-16, derived from the LSP-39), increases in the number of items (e.g., the K-10+, which includes additions to the K-10), and differences in the rating period covered by the instrument (e.g., the HoNOS family, described above). It should also be noted that, with the exception of the RUG-ADL, the instruments that are being administered with older persons (i.e., the LSP-16, the MHI, the BASIS-32® and the K-10+) have largely been tested only with adults in the middle age range. While it is likely that the findings of these studies can be generalised to the alternative versions of the instruments (and to different population groups), some caution should be exercised in doing so.

This caveat should be borne in mind in interpreting Table 6, which provides a broad summary of the psychometric properties of the NOCC measures. Taken together, the measures mostly have adequate or good validity, reliability, sensitivity to change, and feasibility and utility. However, some instruments show better overall psychometric performance than others, and for a number of instruments, particular psychometric characteristics are under-investigated.

**Table 6: Psychometric properties of NOCC measures**

| | Validity | | | | Reliability | | Sensitivity to change | Feasibility and utility |
|---|---|---|---|---|---|---|---|---|
| | **Content** | **Construct** | **Concurrent** | **Predictive** | **Test-retest** | **Inter-rater** | | |
| **HoNOS** | Good | Good | Good | Good | Adequate | Adequate | Adequate | Adequate |
| **HoNOSCA** | Unknown | Adequate | Good | Adequate | Adequate | Adequate | Good | Adequate |
| **HoNOS65+** | Unknown | Unknown | Good | Unknown | Unknown | Good | Adequate | Adequate |
| **LSP-16** | Adequate | Adequate | Adequate | Adequate | Good | Good | Good | Adequate |
| **RUG-ADL** | Poor | Unknown | Unknown | Good | Unknown | Adequate | Unknown | Adequate |
| **CGAS** | Adequate | Good | Adequate | Adequate | Adequate | Good | Adequate | Adequate |
| **MHI** | Adequate | Adequate | Good | Good | Good | Adequate | Adequate | Adequate |
| **BASIS-32®** | Adequate | Good | Good | Good | Good | Good | Adequate | Adequate |
| **K-10+** | Adequate | Adequate | Good | Unknown | Adequate | N/A | Unknown | Adequate |
| **SDQ** | Adequate | Good | Good | Unknown | Adequate | Adequate | Adequate | Adequate |

## Future directions

Having said this, there are still a number of unknowns. Individually, some instruments require further testing to establish their value as outcome measures. In particular, consideration should be given to instruments that were not originally designed as outcome measures, and/or have diverged from their original format or design. In addition, some psychometric properties might warrant special attention, given the NOCC context. For example, inter-rater reliability is particularly important, given that a number of raters may be involved in administering measures for the same consumer in the same episode of care. Likewise, sensitivity to change is crucial, in the context of outcome measurement (as opposed to one-off health status measurement). As a by-product of routine outcome measurement in Australia, there is the potential for national-level analyses that replicate existing findings in an Australian context, and contribute to new international knowledge regarding the psychometric properties of the relevant instruments. AMHOCN has a role to play here.

The current review was limited to a consideration of the published literature. This meant that only the standardised measures in the NOCC suite could be examined, and not the additional measures (i.e., factors influencing health status, focus of care, mental health legal status, and principal and additional diagnoses). Indeed, it was sometimes even beyond the scope of the review to consider the standardised instruments in the form in which they are being used in NOCC, particularly in instances where additional items have been added (e.g., there are no published data on the cross-informant items on the SDQ, nor on the additional items that differentiate the K-10+ from the K-10). Again, there is the potential for national-level analyses of NOCC data to contribute to scientific knowledge in this regard, and AMHOCN is well placed to provide input.

## Conclusions

The current review provides evidence that the NOCC suite constitutes a group of measures that can collectively assess outcomes for different groups, from different perspectives, on a range of mental health-related constructs. Where they have been tested, they appear to perform adequately or better in terms of validity, reliability, sensitivity to change, and feasibility and utility. To this extent, they can be regarded as appropriate for the purposes of monitoring outcomes for consumers, with a view to improving the quality and effectiveness of treatment.

# References

1. Australian Health Ministers. *National Mental Health Plan.* Canberra: Australian Government Publishing Service; 1992.
2. Andrews G, Peters L, Teesson M. *Measurement of Consumer Outcome in Mental Health: A Report to the National Mental Health Information Strategy Committee.* Sydney: Clinical Research Unit for Anxiety Disorders; 1994.
3. Stedman T, Yellowlees P, Mellsop G, Clarke R, Drake S. *Measuring Consumer Outcomes In Mental Health: Field Testing of Selected Measures of Consumer Outcome in Mental Health.* Canberra: Department of Health and Family Services; October, 1997.
4. Buckingham W, Burgess P, Solomon S, Pirkis J, Eagar K. *Developing a Casemix Classification for Mental Health Services. Volume 1: Main Report.* Canberra: Commonwealth Department of Health and Family Services; 1998.
5. Burgess P, Pirkis J, Buckingham W, Eagar K, Solomon S. Developing a casemix classification for specialist mental health services. *Casemix.* 1999;1(4):4-20.
6. Australian Health Ministers. *Second National Mental Health Plan.* Canberra: Mental Health Branch, Commonwealth Department of Health and Family Services; 1998.
7. National Mental Health Working Group. *Mental Health Information Development: National Information Priorities and Strategies under the Second National Mental Health Plan 1998-2003.* Canberra: Commonwealth Department of Health and Aged Care; 1999.
8. Australian Health Ministers. *National Mental Health Plan (2003-2008).* Canberra: Australian Government; 2003.
9. National Mental Health Working Group. *Mental Health National Outcomes and Casemix Collection: Technical Specification of State and  Territory Reporting Requirements for the Outcomes and Casemix Components of 'Agreed Data', Version 1.50.* Canberra: Commonwealth Department of Health and Ageing; 2003.
10. Greenhalgh J, Long A, Brettle A, Grant M. Reviewing and selecting outcome measures for use in routine practice. *Journal of Evaluation in Clinical Practice.* 1998;4(4):339-350.
11. Green RS, Gracely EJ. Selecting a rating scale for evaluating services to the chronically mentally ill. *Community Mental Health Journal.* 1987;23:91-102.
12. McDowell I, Newell C. *Measuring Health: A Guide to Rating Scales and Questionnaires.* Oxford: Oxford University Press; 1996.
13. Cohen JA. A coefficent of agreement of nominal scale. *Education and Psychological Measurement.* 1960;20:37-46.
14. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159-174.
15. Bebbington P, Brugha T, Hill T, Marsden L, Window S. Validation of the Health of the Nation Outcome Scales. *British Journal of Psychiatry.* May 1999;174:389-394.
16. Wing JK, Beevor AS, Curtis RH, Park SB, Hadden S, Burns A. Health of the Nation Outcome Scales (HoNOS). Research and development. *British Journal of Psychiatry.* Jan 1998;172:11-18.
17. Wing JK, Lelliott P, Beevor AS. Progress on HoNOS. *British Journal of Psychiatry.* 2000;176:392-393.
18. Department of Health and Ageing. *Mental Health National Outcomes and Casemix Collection: Overview of Clinician Rated and Consumer Self-report Measures, Version 1.50.* Canberra: Department of Health and Ageing; 2003.
19. Wing J, Curtis RH, Beevor A. Health of the Nation Outcome Scales (HoNOS). Glossary for HoNOS score sheet. *British Journal of Psychiatry.* May 1999;174:432-434.
20. Trauer T, Buckingham B. *National Outcomes and Casemix Collection Discussion Paper - The Health of the Nation Outcome Scales (HoNOS): Towards an Agenda for Future Development.* Canberra: Adult Mental Health Outcomes Expert Group; 2004.
21. Lambert G, Caputi P, Deane FP. Sources of information when rating the Health of the Nation Outcome Scales. *International Journal of Mental Health Nursing.* Jun 2002;11(2):135-138.
22. Royal College of Psychiatrists. http://www.rcpsych.ac.uk/cru/honoscales/faqs.htm.

23. Goldney RD, Fisher LJ, Walmsley SH. A pilot study of the Health of the Nation Outcome Scales as a measurement of outcome in a private psychiatric hospital. *Australasian Psychiatry.* 1996;4(6):319-321.

24. Morris-Yates A, Barber H, Harris M, Zapart S. *Trainer's Manual for the Health of the Nation Outcome Scales, Version 5, Adult.* Canberra: Commonwealth Department of Health and Aged Care; 1999.

25. Bech P, Bille J, Schutze T, Sondergaard S, Wiese M, Waarst S. Health of the Nation Outcome Scales (HoNOS): Implementability, subscale structure and responsiveness in the daily psychiatric hospital routine over the first 18 months. *Nordic Journal of Psychiatry.* 2003;57(4):285-290.

26. Bonsack C, Borgeat F, Lesage A. Measuring patients' problems severity and outcomes in a psychiatric sector: A field study with the French version of the Health of Nation Outcome Scales (HoNOS-F)/Mesurer la severite des problemes des patients et leur evolution dans un secteur psychiatrique: Une etude sur le terrain du Health of Nation Outcome Scales en francais: (HoNOS-F). *Annales Medico-Psychologiques.* Sep 2002;160(7):483-488.

27. Lauzon S, Corbiere M, Bonin J-P, Bonsack C, Lesage A, Ricard N. Validation of the French version of the Health of the Nation Outcome Scale (HoNOS). *Canadian Journal of Psychiatry.* 2001;46:841-846.

28. Morosini P, Gigantesco A, Mazzarda A, Gibaldi L. HoNOS-Rome: An expanded, customized, and longitudinally oriented version of the HoNOS. *Epidemiologia e Psichiatria Sociale.* 2003;12:53-62.

29. Lora A, Bai G, Bianchi S, et al. The Italian version of HoNOS (Health of the Nation Outcome Scales), a scale for evaluating the outcome and the severity in mental health services. *Epidemiologia e Psichiatria Sociale.* 2001;10:198-204.

30. Andreas S, Harfst T, Dirmaier J, Koch U, Schultz H. *Research on HoNOS in Germany.* London 2003.

31. Ruud T. *Experiences with HoNOS in Norway.* London 2003.

32. Shergill SS, Shankar KK, Seneviratna K, Orrell MW. The validity and reliability of the Health of the Nation Outcome Scales (HoNOS) in the elderly. *Journal of Mental Health (UK).* Oct 1999;8(5):511-521.

33. Orrell M, Yard P, Handysides J, Schapira R. Validity and reliability of the Health of the Nation Outcome Scales in psychiatric patients in the community. *British Journal of Psychiatry.* May 1999;174:409-412.

34. McClelland R, Trimble P, Fox ML, Stevenson MR, Bell B. Validation of an outcome scale for use in adult psychiatric practice. *Quality in Health Care.* Jun 2000;9(2):98-105.

35. Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to their Development and Use.* Oxford: Oxford University Press; 1995.

36. Trauer T. The subscale structure of the Health of the Nation Outcome Scales (HoNOS). *Journal of Mental Health (UK).* Oct 1999;8(5):499-509.

37. Page AC, Hooke GR, Rutherford EM. Measuring mental health outcomes in a private psychiatric clinic: Health of the Nation Outcome Scales and Medical Outcomes Short Form SF-36. *Australian and New Zealand Journal Psychiatry.* Jun 2001;35(3):377-381.

38. Trauer T. Comment. *Australian and New Zealand Journal of Psychiatry.* Jun 2000;34(3):520-521.

39. Preston NJ. The Health of the Nation Outcome Scales: Validating factorial structure and invariance across two health services. *Australian and New Zealand Journal Psychiatry.* Jun 2000;34(3):512-519.

40. Parker G, O'Donnell M, Hadzi-Pavlovic D, Proberts M. Assessing outcome in community mental health patients: A comparative analysis of measures. *International Journal of Social Psychiatry.* Mar 2002;48(1):11-19.

41. Browne S, Doran M, McGauran S. Health of the Nation Outcome Scales (HoNOS): Use in an Irish psychiatric outpatient population. *Irish Journal of Psychological Medicine.* Mar 2000;17(1):17-19.

42. Amin S, Singh SP, Croudace T, Jones P, Medley I, Harrison G. Evaluating the Health of the Nation Outcome Scales. Reliability and validity in a three-year follow-up of first-onset psychosis. *British Journal of Psychiatry.* May 1999;174:399-403.

43. Rees A, Richards A, Shapiro DA. Utility of the HoNOS in measuring change in a community mental health care population. *Journal of Mental Health.* Jun 2004;13(3):295-304.

44. Hope JD, Trauer T, Keks NA. Reliability, validity and utility of the Health of the Nation Outcomes Scale (HoNOS) in Australian adult psychiatric services. *Schizophrenia Research.* 1998;29(1,2): 9-10.

45. Adams M, Palmer A, O'Brien JT, Crook W. Health of the Nation Outcome Scales for psychiatry: Are they valid? *Journal of Mental Health.* Apr 2000;9(2):193-198.

46. Brooks R. The reliability and validity of the Health of the Nation Outcome Scales: Validation in relation to patient derived measures. *Australian and New Zealand Journal Psychiatry.* Jun 2000;34(3):504-511.

47. Issakidis C, Teesson M. Measurement of need for care: A trial of the Camberwell Assessment of Need and the Health of the Nation Outcome Scales. *Australian and New Zealand Journal Psychiatry.* Oct 1999;33(5):754-759.

48. Hunter R, McLean J, Peck D, et al. The Scottish 700 Outcomes Study: A comparative evaluation of the Health of the Nation Outcome Scale (HoNOS), the Avon Mental Health Measure (AVON), and an Idiographic Scale (OPUS) in adult mental health. *Journal of Mental Health.* Feb 2004;13(1):93-105.

49. Trauer T, Callaly T. Concordance between mentally ill patients and their case managers using the Health of the Nation Outcome Scales (HoNOS). *Australasian Psychiatry.* Mar 2002;10(1):24-28.

50. Slade M, Beck A, Bindman J, Thornicroft G, Wright S. Routine clinical outcome measures for patients with severe mental illness: CANSAS and HoNOS. *British Journal of Psychiatry.* May 1999;174:404-408.

51. Gallagher J, Teesson M. Measuring disability, need and outcome in Australian community mental health services. *Australian and New Zealand Journal Psychiatry.* Oct 2000;34(5):850-855.

52. Allan S, McGonagle I. A comparison of HoNOS with the Social Behaviour Schedule in three settings. *Journal of Mental Health.* Apr 1997;6(2):117-124.

53. Broadbent M. Reconciling the information needs of clinicians, managers and commissioners: A pilot project. *Psychiatric Bulletin.* Nov 2001;25(11):423-425.

54. Schneider J, Wooff D, Carpenter J, Brandon T, McNiven F. Service organisation, service use and costs of community mental health care. *Journal of Mental Health Policy and Economics.* Jun 2002;5(2):79-87.

55. Ashaye K, Seneviratna K, Shergill S, Orrell M. Do the Health of the Nation Outcome Scales predict outcome in the elderly mentally ill? A 1-year follow-up study. *Journal of Mental Health (UK).* Dec 1999;8(6):615-620.

56. Goldney RD, Fisher LJ, Walmsley SH. The Health of the Nation Outcome Scales in psychiatric hospitalisation: A multicentre study examining outcome and prediction of length of stay. *Australian and New Zealand Journal Psychiatry.* Apr 1998;32(2):199-205.

57. Boot B, Hall W, Andrews G. Disability, outcome and case-mix in acute psychiatric in-patient units. *British Journal of Psychiatry.* Sep 1997;171:242-246.

58. Trauer T, Callaly T, Hantz P, Little J, Shields R, Smith J. Health of the Nation Outcome Scales. Results of the Victorian field trial. *British Journal of Psychiatry.* May 1999;174:380-388.

59. Audin K, Margison FR, Clark JM, Barkham M. Value of HoNOS in assessing patient change in NHS psychotherapy and psychological treatment services. *British Journal of Psychiatry.* Jun 2001;178:561-566.

60. Taylor JR, Wilkinson G. HoNOS v. GP opinion in a shifted out-patient clinic. *Psychiatric Bulletin.* Aug 1997;21(8):483-485.

61. Sharma VK, Wilkinson G, Fear S. Health of the Nation Outcome Scales: A case study in general psychiatry. *British Journal of Psychiatry.* May 1999;174:395-398.

62. Ashaye O, Mathew G, Dhadphale M. A comparison of older longstay psychiatric and learning disability inpatients using the Health of the Nation Outcome Scales. *International Journal of Geriatric Psychiatry.* May 1997;12(5):548-552.

63. Stafrace S. Doubts about HoNOS. *Australian and New Zealand Journal Psychiatry.* Apr 2002;36(2):270.

64. Stein GS. Usefulness of the Health of the Nation Outcome Scales. *British Journal of Psychiatry.* May 1999 1999;174:375-377.
65. Gilbody SM, House AO, Sheldon TA. Psychiatrists in the UK do not use outcomes measures: National survey. *British Journal of Psychiatry.* Feb 2002;180(2):101-103.
66. Glover GR, Sinclair Smith H. Computerised information systems in English mental health care providers in 1998. *Social Psychiatry and Psychiatric Epidemiology.* Nov 2000;35(11):518-522.
67. James M, Kehoe R. Using the Health of the Nation Outcome Scales in clinical practice. *Psychiatric Bulletin.* Sep 1999;23(9):536-538.
68. Milne D, Reichelt K, Wood EI. Implementing HoNOS: An eight stage approach. *Clinical Psychology and Psychotherapy.* 2001;8(2):106-116.
69. Trauer T. The Health of the Nation Outcome Scales in outcome measurement: A critical review. *Australasian Psychiatry.* 1998;6(1):11-14.
70. Hunter J, Higginson I, Garralda E. Sytematic literature review: Outcome measures for child and adolescent mental health services. *Journal of Public Health Medicine.* 1996;18(2):197-206.
71. Jensen PS, Hoagwood K, Petti T. Outcomes of mental health care for children and adolescents: II. Literature review and application of a comprehensive model. *Journal of the American Academy of Child and Adolescent Psychiatry.* 1996;35:1064-1077.
72. Gowers SG, Harrington RC, Whitton A, et al. Brief scale for measuring the outcomes of emotional and behavioural disorders in children. Health of the Nation Outcome Scales for children and Adolescents (HoNOSCA). *British Journal of Psychiatry.* May 1999;174:413-416.
73. Gowers S, Bailey-Rogers SJ, Shore A, Levine W. The Health of the Nation Outcome Scales for Child and Adolescent Mental Health (HoNOSCA). *Child Psychology and Psychiatry Review.* 2000;5(2):50-56.
74. Manderson J, McCune N. The use of HoNOSCA in a child and adolescent mental health service. *Irish Journal of Psychological Medicine.* Jun 2003;20(2):52-55.
75. Gowers SG, Harrington RC, Whitton A, et al. *Health of the Nation Outcome Scales: Trainer's Guide.* Liverpool: The University of Liverpool; 1997.
76. Gowers SG, Harrington RC, Whitton A, et al. Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA). Glossary for HoNOSCA score sheet. *British Journal of Psychiatry.* May 1999;174:428-431.
77. Yates P, Garralda ME, Higginson I. Paddington Complexity Scale and Health of the Nation Outcome Scales for Children and Adolescents. *British Journal of Psychiatry.* May 1999;174:417-423.
78. Garralda ME, Yates P, Higginson I. Child and adolescent mental health service use. HoNOSCA as an outcome measure. *British Journal of Psychiatry.* Jul 2000;177:52-58.
79. Brann P. *Routine Outcome Measurement in Child/Adolescent Mental Health: HoNOSCA - Valid Enough? Feasible Enough?* Melbourne, Monash University; Forthcoming.
80. Brann P, Coleman G, Luk E. Routine outcome measurement in a child and adolescent mental health service: An evaluation of HoNOSCA. *Australian and New Zealand Journal of Psychiatry.* Jun 2001;35(3):370-376.
81. Bilenberg N. Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA): Results of a Danish field trial. *European Child and Adolescent Psychiatry.* Dec 2003;12(6):298-302.
82. Gowers S, Levine W, Bailey-Rogers S, Shore A, Burhouse E. Use of a routine, self-report outcome measure (HoNOSCA-SR) in two adolescent mental services. *British Journal of Psychiatry.* Mar 2002;180(3):266-269.
83. Garralda E, Yates P. HoNOSCA: Uses and limitations. *Child Psychology and Psychiatry Review.* 2000;5(3):131-132.
84. Burns A, Beevor A, Lelliott P, et al. Health of the Nation Outcome Scales for elderly people (HoNOS 65+). Glossary for HoNOS 65+ score sheet. *British Journal of Psychiatry.* May 1999;174:435-438.
85. Burns A, Beevor A, Lelliott P, et al. Health of the Nation Outcome Scales for elderly people (HoNOS 65+). *British Journal of Psychiatry.* May 1999;174:424-427.
86. Macdonald AJ. HoNOS 65+ glossary. *British Journal of Psychiatry.* Aug 1999;175:192.

87. College Research Unit. *HoNOS 65+: A Tabulated Glossary for Use with HoNOS65+ (Version 3).* London: Royal College of Psychiatrists; 2002.

88. Spear J, Chawla S, O'Reilly M, Rock D. Does the HoNOS 65+ meet the criteria for a clinical outcome indicator for mental health services for older people? *International Journal of Geriatric Psychiatry.* Mar 2002;17(3):226-230.

89. Allen L, Bala S, Carthew R, et al. Experience and application of HoNOS65+. *Psychiatric Bulletin.* Apr 1999;23(4):203-206.

90. Mozley CG, Huxley P, Sutcliffe C, et al. 'Not knowing where I am doesn't mean I don't know what I like': Cognitive impairment and quality of life responses in elderly people. *International Journal of Geriatric Psychiatry.* Sep 1999;14(9):776-783.

91. Bagley H, Cordingley L, Burns A, et al. Recognition of depression by staff in nursing and residential homes. *Journal of Clinical Nursing.* May 2000;9(3):445-450.

92. Reilly S, Challis D, Burns A, Hughes J. The use of assessment scales in Old Age Psychiatry Services in England and Northern Ireland. *Aging Mental Health.* May 2004;8(3):249-255.

93. Macdonald AJD. The usefulness of aggregate routine clinical outcomes data: The example of HoNOS65+. *Journal of Mental Health (UK).* Dec 2002;11(6):645-656.

94. Rosen A, Hadzi-Pavlovic D, Parker G. The Life Skills Profile: A measure assessing function and disability in schizophrenia. *Schizophrenia Bulletin.* 1989;15(2):325-337.

95. Parker G, Rosen A, Emdur N, Hadzi-Pavlovic D. The Life Skills Profile: Psychometric properties of a measure assessing function and disability in schizophrenia. *Acta Psychiatrica Scandinavica.* Feb 1991;83(2):145-152.

96. Buckingham W, Burgess P, Solomon S, Pirkis J, Eagar K. *Developing a Casemix Classification for Mental Health Services. Volume 2: Resource Materials.* Canberra: Commonwealth Department of Health and Family Services; 1998.

97. Trauer T, Duckmanton RA, Chiu E. The Life Skills Profile: A study of its psychometric properties. *Australian and New Zealand Journal of Psychiatry.* Sep 1995;29(3):492-499.

98. Bream V, Buchanan A. Distress among children whose separated or divorced parents cannot agree arrangements for them. *British Journal of Social Work.* Mar 2003;33(2):227-238.

99. Rosen A, Trauer T, Hadzi Pavlovic D, Parker G. Development of a brief form of the Life Skills Profile: The LSP-20. *Australian and New Zealand Journal of Psychiatry.* 2001;35(5):677-683.

100. Zizolfi S. The Italian version of Life Skills Profile (LSP), an instrument for evaluating performance and disabilities of schizophrenic patients/La versione italiana del Life Skills Profile (LSP), uno strumento per la valutazione del funzionamento e delle disabilita dei pazienti schizofrenici. *Epidemiologia e Psichiatria Sociale.* Sep-Dec 1997;6(3):196-211.

101. Fernandez de Larrinoa P, Bulbena A, Dominguez AI. Estudio de fiabilidad, validez y consistencia interna de la escala LSP (Life Skills Profile) perfil de habilidades de la vida cotidiana. *Actas Luso-Eso. Neurol. Psiquiatr.* 1992;20:71-75.

102. Bulbena A, Fernandez de Larrinoa P, Dominguez AI. Adaptacion castellas de la escala LSP (Life Skills Profile) perfil de habilidades de la vida cotidiana. *Actas Luso-Eso. Neurol. Psiquiatr.* 1992;20:51-60.

103. Fokianos C, Mohr S, Favrod J, Ferrero F, Simon AE. Validation of the French version of the Life Skills Profile in patients with schizophrenia. *Schizophrenia Research.* 2001;49:259.

104. Eu PW, Lee C, Parker G, Loh J. The disability profile of patients with schizophrenia in psychiatric hospital and community settings in Singapore. *Singapore Medical Journal.* 2001;42(12):559-562.

105. Hadzi-Pavlovic D, Rosen A, Parker G. The relevance and use of life skills assessments. In: Kavanagh DJ, ed. *Schizophrenia: An Overview and Practical Handbook.* London: Chapman and Hall; 1992:206-220.

106. Dickinson D, Coursey RD. Independence and overlap among neurocognitive correlates of community functioning in schizophrenia. *Schizophrenia Research.* Jul 2002;56(1-2):161-170.

107. Andrews G, Teesson M, Stewart G, Hoult J. Follow-up of community placement in the chronic mentally ill in New South Wales. *Hospital and Community Psychiatry.* 1990;41:184-188.

108. Wooff D, Schneider J, Carpenter J, Brandon T. Correlates of stress in carers. *Journal of Mental Health.* Feb 2003;12(1):29-40.
109. Simon AE, Giaocomini V, Ferrero F, Mohr S. Dysexecutive syndrome and social adjustment in schizophrenia. *Australian and New Zealand Journal of Psychiatry.* Jun 2003;37(3):340-346.
110. Norman RMG, Malla AK, McLean T, et al. The relationship of symptoms and level of functioning in schizophrenia to general wellbeing and the Quality of Life Scale. *Acta Psychiatrica Scandinavica.* Oct 2000;102(4):303-309.
111. Trauer T, Duckmanton RA, Chiu E. Estimation of costs of public psychiatric treatment. *Psychiatric Services.* Apr 1998;49(4):440-442.
112. Trauer T. *Analysis of Outcome Measurement Data from the Four Victorian 'Round One' Agencies*: Mental Health Branch, Department of Human Services, Victoria; 2003.
113. Trauer T, Duckmanton RA, Chiu E. The assessment of clinically significant change using the Life Skills Profile. *Australian and New Zealand Journal of Psychiatry.* 1997;31(2):257-263.
114. Keller S, Hayes R. The relationship between the Allen Cognitive Level Test and the Life Skills Profile. *American Journal of Occupational Therapy.* Nov-Dec 1998;52(10):851-856.
115. Kirkby K, Daniels B, Jones I, McInnes M. A survey of social outcome in schizophrenia in Tasmania. *Australian and New Zealand Journal of Psychiatry.* 1997;31(3):405-410.
116. Browne G, Courtney M. Measuring the impact of housing on people with schizophrenia. *Nursing and Health Sciences.* 2004;6(1):37-44.
117. Parker G, Hadzi-Pavlovic D. The capacity of a measure of disability (the LSP) to predict hospital readmission in those with schizophrenia. *Psychological Medicine.* Jan 1995;25(1):157-163.
118. Preston NJ. Predicting community survival in early psychosis and schizophrenia populations after receiving intensive case management. *Australian and New Zealand Journal of Psychiatry.* 2000;34(1):122-128.
119. Ballesteros RJ, Martinez SM, Martin CM, Ibarra GN, Bulbena VA. Assessment of the Life Skills Profile and the Brief Psychiatric Rating Scale as predictors of the length of psychiatric hospitalization. *Actas Esplanolas de Psiquiatria.* 2002;30:225-232.
120. Kisely S, Preston N, Rooney M. Pathways and outcomes of psychiatric care: Does it depend on who you are, or what you've got? *Australian and New Zealand Journal of Psychiatry.* 2000;34(6):1009-1014.
121. Sanderson K, Issakidis C, Johnston S, Teesson M, Salkeld G, Buhrich N. *Cost-effectiveness of Intensive Case Management for People with Serious Mental Illness.* Sydney: South Eastern Sydney Area Mental Health Service; 1996.
122. Johnston S, Salkeld G, Sanderson K, Issakidis C, Teesson M, Buhrich N. Intensive case management: A cost-effectiveness analysis. *Australian and New Zealand Journal of Psychiatry.* 1998;32(4):551-559.
123. Craig T, Doherty I, Jamieson-Craig R, Boocock A, Attafua G. The consumer-employee as a member of a Mental Health Assertive Outreach Team: (1) Clinical and social outcomes. *Journal of Mental Health.* Feb 2004;13(1):59-69.
124. Rosen A, Teesson M. Does case management work? The evidence and the abuse of evidence-based medicine. *Australian and New Zealand Journal of Psychiatry.* Dec 2001;35(6):731-746.
125. Hamernik E, Pakenham KI. Assertive community treatment for persons with severe mental disorders: A controlled treatment outcome study. *Behaviour Change.* 1999;16(4):259-268.
126. Hambridge J, Rosen A. Assertive community treatment for the seriously mentally ill in suburban Sydney: A programme description and evaluation. *Australian and New Zealand Journal of Psychiatry.* 1994;28:438-445.
127. Ford R, Ryan P, Beadsmoore A, Craig T, et al. Intensive case management for people with serious mental illness - Site 2: Clinical and social outcome. *Journal of Mental Health.* Apr 1997;6(2):181-190.
128. Ford R, Barnes A, Davies R, Chalmers C, Hardy P, Muijen M. Maintaining contact with people with severe mental illness: 5-year follow-up of assertive outreach. *Social Psychiatry and Psychiatric Epidemiology.* 2001;36(9):444-447.

129. Stedman T, Yellowlees P, Drake S, Chant D, Clarke R, Chapple B. The perceived utility of six selected measures of consumer outcomes proposed for routine use in Australian mental health services. *Australian and New Zealand Journal of Psychiatry.* 2000;34(5):842-849.

130. Fries BE, Schneider DP, Foley WJ, Gavazzi M, et al. Refining a case-mix measure for nursing homes: Resource utilization groups (RUG-III). *Medical Care.* Jul 1994;32(7):668-685.

131. Aronson MK, Cox D, Guastadisegni P, et al. Dementia and the nursing home: association with care needs. *Journal of the American Geriatrics Society.* 1992;40(1):27-33.

132. Ljunggren G. Case-mix analyses in long-term care institutions in Stockholm. *Scandinavian Journal of Primary Health Care.* 1992;10(2):151-156.

133. Uehara ES, Smukler M, Newman FL. Linking resource use to consumer level of need: Field test of the level of need-care assessment (LONCA) method. *Journal of Consulting and Clinical Psychology.* Aug 1994;62(4):695-709.

134. Shugarman LR, Fries BE, James M. A comparison of home care clients and nursing home residents: can community based care keep the elderly and disabled at home? *Home Health Care Services Quarterly.* 1999;18(1):25-45.

135. Carpenter GI, Main A, Turner GF. Casemix for the elderly inpatient: Resource Utilization Groups (RUGs) validation project. *Age and Ageing.* 1995;24(1):5-13.

136. Ljunggren G, Brandt L. Predicting nursing home length of stay and outcome with a resource-based classification system. *International Journal of Technology Assessment in Health Care.* 1996;12(1):72-79.

137. Lombardo NBE, Fogel BS, Robinson GK, Weiss HP. Achieving mental health of nursing home residents: Overcoming barriers to mental health care. *Journal of Mental Health and Aging.* Win 1995;1(3):165-211.

138. Arling G, Zimmerman D, Updike L. Nursing home case mix in Wisconsin. Findings and policy implications. *Medical Care.* 1989;27(2):164-181.

139. Hirdes JP, Smith TF, Rabinowitz T, et al. The Resident Assessment Instrument-Mental Health (RAI-MH): Inter-rater reliability and convergent validity. *Journal of Behavioral Health Services and Research.* Nov 2002;29(4):419-432.

140. Lee L, Kennedy C, Aitken J. Does the NAIP classification predict length of stay in rehabilitation, geriatrics and palliative care? *Australian Health Review.* 1996;19(2):56-74.

141. Mitty EL. Resource utilization groups. DRGs move to long-term care. *Nursing Clinics of North America.* 1988;23(3):539-557.

142. Endicott J, Spitzer RL, Fleiss J, Cohen J. The Global Assessment Scale: A procedure for measuring overall severity of psychiatric disturbance. *Archives of General Psychiatry.* 1976;33.

143. Schaffer D, Gould, M.S., Brasic, J., et al. A children's global assessment scale (CGAS). *Archives of General Psychiatry.* 1983;40:1228-1231.

144. Dyrborg J, Larsen F, Nielsen S, Byman J, Buhl Nielsen B, Gautre-Delay F. The Children's Global Assessment Scale (CGAS) and Global Assessment of Psychosocial Disability (GAPD) in clinical practice: Substance and reliability as judged by intraclass correlations. *European Child and Adolescent Psychiatry.* Sep 2000;9(3):195-201.

145. Steinhausen H-C. Global assessment of child psychopathology. *Journal of the American Academy of Child and Adolescent Psychiatry.* Mar 1987;26(2):203-206.

146. Bird HR, Yager TJ, Staghezza B, Gould MS, et al. Impairment in the epidemiological measurement of childhood psychopathology in the community. *Journal of the American Academy of Child and Adolescent Psychiatry.* Sep 1990;29(5):796-803.

147. Weissman MM, Warner V, Fendrich M. Applying impairment criteria to children's psychiatric diagnosis. *Journal of the American Academy of Child and Adolescent Psychiatry.* Sep 1990;29(5):789-795.

148. Bird HR, Canino GJ, Rubio-Stipec M, Ribera JC. Further measures of the psychometric properties of the Children's Global Assessment Scale. *Archives of General Psychiatry.* Sep 1987;44(9):821-824.

149. Bird HR, Andrews H, Schwab-Stone M, et al. Global measures of impairment for epidemiologic and clinical use with children and adolescents. *International Journal of Methods in Psychiatric Research.* Dec 1996;6(4):295-307.

150. Green B, Shirk S, Hanze D, Wanstrath J. The Children's Global Assessment Scale in clinical practice: An empirical evaluation. *Journal of the American Academy of Child and Adolescent Psychiatry.* Oct 1994;33(8):1158-1164.
151. Sourander A, Piha J. Parent, teacher and clinical ratings on admission to child psychiatric inpatient treatment: A study of cross-informant correlations using the CBCL, Rutter scales and the CGAS. *Nordic Journal of Psychiatry.* 1997;51(5):365-370.
152. Steinhausen H-C, Metzke CW. Global measures of impairment in children and adolescents: Results from a Swiss community survey. *Australian and New Zealand Journal of Psychiatry.* Jun 2001;35(3):282-286.
153. Sourander A, Leijala H, Lehtila A, Kanerva A, Helenius H, Piha J. Short-term child psychiatric inpatient treatment. Place of residence as one-year outcome measure. *European Child and Adolescent Psychiatry.* 1996;5(1):38-43.
154. Sourander A, Helenius H, Piha J. Child psychiatric short-term inpatient treatment: CGAS as follow-up measure. *Child Psychiatry and Human Development.* Win 1996;27(2):93-104.
155. Hodges K, Gust J. Measures of impairment for children and adolescents. *Journal of Mental Health Administration.* Fal 1995;22(4):403-413.
156. Veit CT, Ware JE. The structure of psychological distress and well-being in general populations. *Journal of Consulting and Clinical Psychology.* Oct 1983;51(5):730-742.
157. Dupuy HJ. *The psychological section of the current health and nutritional examination survey.* Washington, DC: National Conference on Health Statistics; 1972.
158. Kornblith AB, Herndon JE, 2nd, Zuckerman E, et al. Social support as a buffer to the psychological impact of stressful life events in women with breast cancer. *Cancer.* 2001;91(2):443-454.
159. Cohen L, Hack TF, de Moor C, Katz J, Goss PE. The effects of type of surgery and time on psychological adjustment in women after breast cancer treatment. *Annals of Surgical Oncology.* 2000;7(6):427-434.
160. Manne S, Schnoll R. Measuring cancer patients' psychological distress and well-being: A factor analytic assessment of the Mental Health Inventory. *Psychological Assessment.* 2001;13(1):99-109.
161. Cull A, Gould A, House A, et al. Validating automated screening for psychological distress by means of computer touchscreens for use in routine oncology practice. *British Journal of Cancer.* 2001;85(12):1842-1849.
162. Ell KO, Nishimoto RH, Mantell JE, Hamovitch MB. Psychological adaptation to cancer: A comparison among patients, spouses, and nonspouses. *Family Systems Medicine.* Fal 1988;6(3):335-348.
163. Ell KO, Mantell JE, Hamovitch MB, Nishimoto RH. Social support, sense of control, and coping among patients with breast, lung, or colorectal cancer. *Journal of Psychosocial Oncology.* 1989;7(3):63-89.
164. Alter CL, Fleishman SB, Kornblith AB, Holland JC, et al. Supportive telephone intervention for patients receiving chemotherapy. *Psychosomatics: Journal of Consultation Liaison Psychiatry.* Sep-Oct 1996;37(5):425-431.
165. Smith MY, Egert J, Winkel G, Jacobson J. The impact of PTSD on pain experience in persons with HIV/AIDS. *Pain.* 2002;98(1-2):9-17.
166. Siegel K, Gluhoski VL, Karus D. Coping and mood in HIV-positive women. *Psychological Reports.* Oct 1997;81(2):435-442.
167. Grummon K, Rigby ED, Orr D, Procidano M, et al. Psychosocial variables that affect the psychological adjustment of IVDU patients with AIDS. *Journal of Clinical Psychology.* Jul 1994;50(4):488-502.
168. Littlefield C, Abbey S, Fiducia D, Cardella C. Quality of life following transplantation of the heart, liver, and lungs. *General Hospital Psychiatry.* Nov 1996;18(6 Suppl):36S-47S.
169. Petrie K. Psychological well-being and psychiatric disturbance in dialysis and renal transplant patients. *British Journal of Medical Psychology.* Mar 1989;62(1):91-96.
170. Pomeroy IM, Clark CR, Philp I. The effectiveness of very short scales for depression screening in elderly medical patients. *International Journal of Geriatric Psychiatry.* Mar 2001;16(3):321-326.
171. Dorfman RA, Lubben JE, Mayer-Oakes A, Atchison K, et al. Screening for depression among a well elderly population. *Social Work.* May 1995;40(3):295-304.

172. Berwick DM, Murphy JM, Goldman PA, Ware JE, et al. Performance of a five-item mental health screening test. *Medical Care.* Feb 1991;29(2):169-176.
173. Weinstein MC, Berwick DM, Goldman PA, Murphy JM, et al. A comparison of three psychiatric screening tests using receiver operating characteristic (ROC) analysis. *Medical Care.* Jun 1989;27(6):593-607.
174. RAND Corporation. http://www.rand.org/health/surveys/section5.pdf.
175. Davies AR, Sherbourne CD, Peterson JR, Ware JE. *Scoring Manual: Adult Health Status and Patient Satisfaction Measures used in RAND's Health Insurance Experiment.* Santa Monica: RAND Corporation; 1988.
176. Andrews G, Slade T. Interpreting scores on the Kessler Psychological Distress Scale (K10). *Australian and New Zealand Journal of Public Health.* 2001;25(6):494-497.
177. Stewart AL, Hays RD, Ware JEJ. The MOS-Short Form. *Medical Care.* 1988;26:724.
178. Ware JE, Sherbourne CD. The MOS 36-item Short Form Health Status Survey (SF-36): (1) Conceptual framework and item selection. *Medical Care.* 1992;30:473-483.
179. Florian V, Drory Y. Mental Health Inventory (MHI): Psychometric properties and normative data in the Israeli population. *Psychologia: Israel Journal of Psychology.* 1990;2:26-35.
180. Gowans SE, DeHueck A, Abbey SE. Measuring exercise-induced mood changes in fibromyalgia: a comparison of several measures. *Arthritis and Rheumatism.* 2002;47(6):603-609.
181. Huebeck BG, Neill JT. Confirmatory factor analysis and reliability of the Mental Health Inventory for Australian adolescents. *Psychological Reports.* 2000;87(2):431-440.
182. Downe-Wamboldt BL, Melanson PM. Emotions, coping, and psychological well-being in elderly people with arthritis. *Western Journal of Nursing Research.* Jun 1995;17(3):250-265.
183. Weiss AS. Psychological distress and well-being in Hare Krishnas. *Psychological Reports.* Aug 1987;61(1):23-35.
184. Brunier G, Graydon J, Rothman B, Sherman C, Liadsky R. The psychological well-being of renal peer support volunteers. *Journal of Advanced Nursing.* 2002;38(1):40-49.
185. Ostroff JS, Woolverton KS, Berry C, Lesko LM. Use of the Mental Health Inventory with adolescents: A secondary analysis of the Rand Health Insurance Study. *Psychological Assessment.* Mar 1996;8(1):105-107.
186. Birnbaum GE, Orr I, Mikulincer M, Florian V. When marriage breaks up: Does attachment style contribute to coping and mental health? *Journal of Social and Personal Relationships.* Oct 1997;14(5):643-654.
187. Heubeck BG, Neill JT. Confirmatory factor analysis and reliability of the Mental Health Inventory for Australian adolescents. *Psychological Reports.* Oct 2000;87(2):431-440.
188. Mikulincer M, Horesh N, Levy-Shiff R, Manovich R, Shalev J. The contribution of adult attachment style to the adjustment to infertility. *British Journal of Medical Psychology.* Sep 1998;71(3):265-280.
189. Kravetz S, Faust M, Shalit M. The mental health of men with a psychiatric disorder or traumatic brain injury: Wives' labeling and husbands' perceived control. *Rehabilitation Psychology.* Nov 2001;46(4):450-465.
190. Rumpf HJ, Meyer C, Hapke U, John U. Screening for mental health: Validity of the MHI-5 using DSM-IV Axis I psychiatric disorders as gold standard. *Psychiatry Research.* 2001;105(3):243-253.
191. McHorney C, Ware JE, Rachel Lu JF, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): (3) Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Medical Care.* 1994;32(1):40-66.
192. McHorney CA, Ware JE. Construction and validation of an alternate form general mental health scale for the Medical Outcomes Study Short-Form 36-Item Health Survey. *Medical Care.* Jan 1995;33(1):15-28.
193. McCabe CJ, Thomas KJ, Brazier JE, Coleman P. Measuring the mental health status of a population: a comparison of the GHQ-12 and the SF-36 (MHI-5). *British Journal of Psychiatry.* 1996;169(4):516-521.
194. Jette A, Davis A. The Functional Status Questionnaire: Reliability and validity when used in primary care. *Journal of General Internal Medicine.* 1986;1:143-149.

195. Tanaka JS, Huba GJ. Structures of psychological distress: Testing confirmatory hierarchical models. *Journal of Consulting and Clinical Psychology.* Aug 1984;52(4):719-721.

196. Liang J, Wu SC, Krause NM, Chiang TL, Wu HY. The structure of the Mental Health Inventory among Chinese in Taiwan. *Medical Care.* 1992;30(8):659-676.

197. Zautra AJ, Guarnaccia CA, Reich JW. Factor structure of mental health measures for older adults. *Journal of Consulting and Clinical Psychology.* Aug 1988;56(4):514-519.

198. Wichstrom L, Anderson AMC, Holte A, Wynne LC. Disqualifying family communication and childhood social competence as predictors of offspring's mental health and hospitalization: A 10- to 14-year longitudinal study of children at risk of psychopathology. *Journal of Nervous and Mental Disease.* Oct 1996;184(10):581-588.

199. Kornblith AB, Herndon JE, 2nd, Zuckerman E, Godley PA, Savarese D, Vogelzang NJ. The impact of docetaxel, estramustine, and low dose hydrocortisone on the quality of life of men with hormone refractory prostate cancer and their partners: A feasibility study. *Annals of Oncology.* 2001;12(5):633-641.

200. Kornblith AB, Herndon JE, 2nd, Silverman LR, et al. Impact of azacytidine on the quality of life of patients with myelodysplastic syndrome treated in a randomized phase III trial: A Cancer and Leukemia Group B study. *Journal of Clinical Oncology.* 2002;20(10):2441-2452.

201. Zika S, Chamberlain K. On the relation between meaning in life and psychological well-being. *British Journal of Psychology.* Feb 1992;83(1):133-145.

202. Fava JL, Ruggiero L, Grimley DM. The development and structural confirmation of the Rhode Island Stress and Coping Inventory. *Journal of Behavioral Medicine.* Dec 1998;21(6):601-611.

203. Read JL, Quinn RJ, Hoefer MA. Measuring overall health: An evaluation of three important approaches. *Journal of Chronic Diseases.* 1987;40(Suppl 1):S7-S26.

204. Wyshak G. Women's college physical activity and self-reports of physician-diagnosed depression and of current symptoms of psychiatric distress. *Journal of Women's Health and Gender-Based Medicine.* May 2001;10(4):363-370.

205. Ware JE, Manning W, Duan W, Wells K, Newhouse J. Health status and use of outpatient mental health services. *American Psychologist.* 1984;39:1090-1100.

206. Rosenthal TL, Downs JM, Arheart KL, Deal N, et al. Similarities and differences on five inventories among mood and anxiety disorder patients. *Behaviour Research and Therapy.* 1991;29(3):239-247.

207. Cassileth B, Lusk E, Strouse T, et al. Psychosocial status in chronic illness. *New England Journal of Medicine.* 1984;311:506-511.

208. Ware JE, Davies-Avery A, Brook RH. *Conceptualization and Measurement of Health for Adults in the Health Insurance Study: (4) Analysis of Relationships Among Health Status Measures.* Santa Monica: RAND Corporation; 1980.

209. Eisen SV, Grob MC, Klein AA. BASIS: The development of a self-report measure for psychiatric inpatient evaluation. *Psychiatric Hospital.* 1986;17(4):165-171.

210. Eisen SV, Dill DL, Grob MC. Reliability and validity of a brief patient-report instrument for psychiatric outcome evaluation. *Hospital and Community Psychiatry.* 1994;45(3):242-247.

211. Hoffmann FL, Capelli K, Mastrianni X. Measuring treatment outcome for adults and adolescents: reliability and validity of BASIS-32. *Journal of Mental Health Administration.* 1997;24(3):316-331.

212. Klinkenberg WD, Cho DW, Vieweg B. Reliability and validity of the interview and self-report versions of the BASIS-32. *Psychiatric Services.* 1998;49(9):1229-1231.

213. Russo J, Roy-Byrne P, Jaffe C, et al. The relationship of patient-administered outcome assessments to quality of life and physician ratings: Validity of the BASIS-32. *Journal of Mental Health Administration.* Spr 1997;24(2):200-214.

214. Eisen SV. Behavior and Symptom Identification Scale (BASIS-32). In: Sederer LI, Dickey B, eds. *Outcomes Assessment in Clinical Practice.* Baltimore: Williams and Wilkins; 1996.

215. Eisen SV, Culhane MA. Behavior and Symptom Identification Scale (BASIS-32). In: Maruish ME, ed. *The Use of Psychological Testing for Treatment Planning and Outcomes Assessment (2nd ed.).* Mahwah, NJ: Lawrence Erlbaum Associates; 1999:759-790.

216. Eisen SV, Leff HS, Schaefer E. Implementing outcome systems: lessons from a test of the BASIS-32 and the SF-36. *Journal of Behavioral Health Services and Research.* 1999;26(1):18-27.
217. Eisen SV, Wilcox M, Leff HS, Schaefer E, Culhane MA. Assessing behavioral health outcomes in outpatient programs: reliability and validity of the BASIS-32. *Journal of Behavioral Health Services and Research.* 1999;26(1):5-17.
218. Joint Commission on Accreditation of Healthcare Organizations. *ORYX: The Next Evolution in Accreditation.* Oakbrook Terrace, IL: Joint Commission on Accreditation of Healthcare Organizations; 1997.
219. Dickey B, Normand S-LT, Hermann RC, et al. Guideline recommendations for treatment of schizophrenia. *Archives of General Psychiatry.* Apr 2003;60(4):340-348.
220. Doerfler LA, Addis ME, Moran PW. Evaluating mental health outcomes in an inpatient setting: Convergent and divergent validity of the OQ-45 and BASIS-32. *Journal of Behavioral Health Services and Research.* 2002;29(4):394-403.
221. Uttaro T, Gonzalez A. Psychometric properties of the Behavior and Symptom Identification Scale administered in a crisis residential mental health treatment setting. *Psychological Reports.* Oct 2002;91(2):439-443.
222. Hawthorne WB, Green EE, Lohr JB, Hough R, Smith PG. Comparison of outcomes of acute care in short-term residential treatment and psychiatric hospital settings. *Psychiatric Services.* 1999;50(3):401-406.
223. McLean Hospital. http://basissurvey.org/basis32/.
224. Department of Human Services. *Measuring Consumer Outcomes in Clinical Mental Health Services. A Training Manual for Services in Victoria (2nd edition)*: Victorian Government; 2003.
225. Eisen SV. Assessment of subjective distress by patients self-report versus structured interview. *Psychological Reports.* 1995;76:35-39.
226. Eisen SV, Dickey B. Mental health outcome assessment: The new agenda. *Psychotherapy.* 1996;33:181-189.
227. Eisen SV, Normand SL, Belanger AJ, Spiro A, Esch D. The revised Behavior and Symptom Identification Scale (BASIS-R). *Medical Care.* 2004;42:1230-1241.
228. McLean Hospital. http://www.basissurvey.org/basis24/.
229. Eisen SV, Wilcox M, Schaefer E. *Use of BASIS-32 for Outcome Assessment of Recipients of Outpatient Mental Health Services.* Belmont, MA: McLean Hospital; 1997. 1094-3412.
230. Cameron IM, Eagles JM, Howie FL, et al. Preliminary validation of a UK-modified version of the BASIS-32. *International Journal of Psychiatry in Clinical Practice.* Mar 2001;5(1):41-48.
231. Graham C, Coombs T, Buckingham W, Eagar K, Trauer T, Callaly T. *Victorian Mental Health Outcomes Measurement Strategy: Consumer Perspectives on Future Directions for Outcome Self-Assessment: Report of the Consumer Consultation Project*: Department of Human Services, Victoria; 2001.
232. Chow JC-C, Snowden LR, McConnell W. A confirmatory factor analysis of the BASIS-32 in racial and ethnic samples. *Journal of Behavioral Health Services and Research.* Nov 2001;28(4):400-411.
233. Pollack LE, Cramer RD. Perceptions of problems in people hospitalized for bipolar disorder: implications for patient education. *Issues in Mental Health Nursing.* 2000;21(8):765-778.
234. Lecomte T, Wallace CL, Caron J, Perreault M, Lecomte J. Further validation of the Client Assessment of Strengths Interests and Goals. *Schizophrenia Research.* Jan 2004;66(1):59-70.
235. Wallace CJ, Lecomte T, Wilde J, Liberman RP. CASIG: A consumer-centered assessment for planning individualized treatment and evaluating program outcomes. *Schizophrenia Research.* 2001;50(1-2):105-119.
236. Trauer T, Tobias G. The Camberwell Assessment of Need and Behaviour and Symptom Identification Scale as routine outcome measures in a psychiatric disability rehabilitation and support service. *Community Mental Health Journal.* Jun 2004;40(3):211-221.

237. Dornelas EA, Botticello AL, Goethe JW, Fischer EH. Validity of a brief measure of post-hospital adjustment for psychiatric patients. *Comprehensive Psychiatry.* Sep-Oct 2001;42(5):410-415.
238. Sousa SA, Frazier R. A nursing tool for adherence and recovery in psychosis: A pilot study. *Journal of Psychosocial Nursing and Mental Health Services.* 2004;42(3):28-36.
239. Wilkins LP, White MB. Interrater reliability and concurrent validity of the Global Assessment of Relational Functioning (GARF) Scale using a card sort method: A pilot study. *Family Therapy.* 2001;28(3):157-170.
240. Eisen SV. Alcohol, drugs and psychiatric disorders: A current view of hospitalised adolescents. *Journal of Adolescent Research.* 1992(7):250-275.
241. Pollack LE, Cramer RD, Varner RV. Psychosocial functioning of people with substance abuse and bipolar disorders. *Substance Abuse.* Sep 2000;21(3):193-203.
242. Johnson ME, Brems C, Burke S. Recognizing comorbidity among drug users in treatment. *American Journal of Drug and Alcohol Abuse.* 2002;28(2):243-261.
243. Eisen SV, Grob MC, Dill DL. Substance abuse in an inpatient population. *McLean Hospital Journal.* 1989;14:1-22.
244. Sederer LI. Case-based reimbursement for psychiatric hospital care. *Hospital and Community Psychiatry.* 1992;43(11):1120-1126.
245. Higgins J, Purvis K. Usefulness of the BASIS-32 in evaluating program level outcomes. *Journal of Psychiatric Practice.* 2002;8(2):125-130.
246. Kessler RC, Andrews G, Colpe LJ, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine.* Aug 2002;32(6):959-976.
247. Department of Health and Ageing. *National Outcomes and Casemix Collection: Overview of Clinical Measures and Data Items. Version 1.02.* Canberra: Commonwealth Department of Health and Ageing; July 2002.
248. Clinical Research Unit for Anxiety and Depression. http://www.crufad.com/cru_index.htm.
249. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders - DSM-IV-TR (Text Revision).* Arlington, VA.: American Psychiatric Association; 2000.
250. Australian Bureau of Statistics. *Mental Health and Wellbeing: Profile of Adults, Australia.* Canberra: Australian Bureau of Statistics; 1998.
251. Kessler RC, Barker PR, Colpe LJ, et al. Screening for serious mental illness in the general population. *Archives of General Psychiatry.* Feb 2003;60(2):184-189.
252. Furukawa TA, Kessler RC, Slade T, Andrews G. The performance of the K6 and K10 screening scales for psychological distress in the Australian National Survey of Mental Health and Well-Being. *Psychological Medicine.* 2003;33(2):357-362.
253. Cole DC, Ibrahim S, Shannon HS, Scott FE, Eyles J. Work and life stressors and psychological distress in the Canadian working population: A structural equation modelling approach to analysis of the 1994 National Population Health Survey. *Chronic Diseases in Canada.* 2002;23(3):91-99.
254. Stephens T, Dulberg C, Joubert N. Mental health of the Canadian population: A comprehensive analysis. *Chronic Diseases in Canada.* 1999;20(3):118-126.
255. Wilkins K, Beaudet MP. Work stress and health. *Health Reports.* 1998;10(3):47-62.
256. Centre for Population Studies in Epidemiology. *The Kessler Psychological Distress Scale (K-10).* Adelaide: Department of Human Services; 2002.
257. Australian Bureau of Statistics. *National Health Survey: Mental Health.* Canberra: Australian Bureau of Statistics; 2003.
258. Dal Grande E, Taylor A, Wilson D. *South Australian Health and Wellbeing Survey.* Adelaide: South Australian Department of Human Services; 2002.
259. d'Espaignet E, Measey M, Dal Grande E. *Northern Territory Health and Wellbeing Survey: Non-Indigenous Population.* Darwin: Department of Health and Community Services; 2002.
260. Victorian Government Department of Human Services. *Victorian Population Health Survey 2001.* Melbourne: Victorian Government Department of Human Services; 2002.
261. Saunders D, Daly A. *2000 Collaborative Health and Wellbeing Survey: Psychological Distress in the Western Australian Population.* Perth: Health Department of Western Australia; 2001.

262. New South Wales Department of Health. New South Wales Adult Health Survey, 2002. *NSW Public Health Bulletin Supplement.* 2003;14(S4).
263. New South Wales Department of Health. New South Wales Adult Health Survey, 2003. *NSW Public Health Bulletin Supplement.* 2004;15(S4).
264. Pirkis J, Kohn F, Morley B, Burgess P, Blashki G. Better outcomes in mental healthcare? *Primary Care Mental Health.* In press.
265. Australian Divisions of General Practice. *Familiarisation Training GP and Practice Manual, Second Edition.* Canberra: Australian Divisions of General Practice; 2003.
266. Goodman R. The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry and Allied Disciplines.* Jul 1997;38(5):581-586.
267. Goodman R. Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry.* Nov 2001;40(11):1337-1345.
268. youthinmind. http://www.sdqinfo.com/.
269. Goodman R. The Strengths and Difficulties Questionnaire (SDQ). In: VandeCreek L, ed. *Innovations in Clinical Practice: Focus on Children and Adolescents.* Sarasota, FL: Ellis Human Development Institute; 2003:109-111.
270. Goodman R. The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. *Journal of Child Psychology and Psychiatry and Allied Disciplines.* Jul 1999;40(5):791-799.
271. Clarbour J, Roger D. The construction and validation of a new scale for measuring emotional response style in adolescents. *Journal of Child Psychology and Psychiatry.* Mar 2004;45(3):496-509.
272. Thabet AA, Stretch D, Vostanis P. Child mental health problems in Arab children: Application of the Strengths and Difficulties Questionnaire. *International Journal of Social Psychiatry.* Win 2000;46(4):266-280.
273. Muris P, Meesters C, van den Berg F. The Strengths and Difficulties Questionnaire (SDQ): Further evidence for its reliability and validity in a community sample of Dutch children and adolescents. *European Child and Adolescent Psychiatry.* Mar 2003;12(1):1-8.
274. Smedje H, Broman JE, Hetta J, von Knorring AL. Psychometric properties of a Swedish version of the 'Strengths and Difficulties Questionnaire'. *European Child and Adolescent Psychiatry.* Jun 1999;8(2):63-70.
275. Hawes DJ, Dadds MR. Australian data and psychometric properties of the Strengths and Difficulties Questionnaire. *Australian New Zealand Journal of Psychiatry.* Aug 2004;38(8):644-651.
276. Koskelainen M, Sourander A, Vauras M. Self-reported strengths and difficulties in a community sample of Finnish adolescents. *European Child and Adolescent Psychiatry.* 2001;10(3):180-185.
277. Ronning JA, Handegaard BH, Sourander A, Morch W-T. The Strengths and Difficulties Self-Report Questionnaire as a screening instrument in Norwegian community samples. *European Child Adolescent Psychiatry.* Apr 2004;13(2):73-82.
278. van Widenfelt BM, Goedhart AW, Treffers PDA, Goodman R. Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *European Child and Adolescent Psychiatry.* Dec 2003;12(6):281-289.
279. Koskelainen M, Sourander A, Kalijonen A. The Strengths and Difficulties Questionnaire among Finnish school-aged children and adolescents. *European Child and Adolescent Psychiatry.* 2000;9(4):277-284.
280. Malmberg M, Rydell AM, Smedje H. Validity of the Swedish version of the Strengths and Difficulties Questionnaire (SDQ-Swe). *Nordic Journal of Psychiatry.* 2003;57(5):357-363.
281. Mellor D. Furthering the use of the Strengths and Difficulties Questionnaire: Reliability with younger child respondents. *Psychological Assessment.* 2004;16(4):396-401.
282. Goodman R, Meltzer H, Bailey V. The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. *European Child and Adolescent Psychiatry.* Sep 1998;7(3):125-130.
283. Goodman R, Meltzer H, Bailey V. The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. *International Review of Psychiatry.* Feb 2003;15(1-2):173-177.

284.    Klasen H, Woerner W, Wolke D, et al. Comparing the German versions of the Strengths and Difficulties Questionnaire (SDQ-Deu) and the Child Behavior Checklist. *European Child and Adolescent Psychiatry.* 2000;9(4):271-276.
285.    Goodman R, Scott S. Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: Is small beautiful? *Journal of Abnormal Child Psychology.* Feb 1999;27(1):17-24.
286.    Mullick MSI, Goodman R. Questionnaire screening for mental health problems in Bangladeshi children: A preliminary study. *Social Psychiatry and Psychiatric Epidemiology.* Feb 2001;36(2):94-99.
287.    Goodman R, Ford T, Simmons H, Gatward R, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *British Journal of Psychiatry.* Dec 2000;177:534-539.
288.    Goodman R, Renfrew D, Mullick M. Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *European Child and Adolescent Psychiatry.* Jun 2000;9(2):129-134.
289.    Mathai J, Anderson P, Bourne A. Comparing psychiatric diagnoses generated by the Strengths and Difficulties Questionnaire with diagnoses made by clinicians. *Australian New Zealand Journal of Psychiatry.* Aug 2004;38(8):639-643.
290.    Cury CR, Golfeto JH. Strengths and difficulties questionnaire (SDQ): A study of school children in Ribeirao Preto. *Revista brasileira de psiquiatria Sao Paulo, Brazil 1999.* 2003;25(3):139-145.
291.    Lindsey C, Frosh S, Loewenthal K, Spitzer E. Prevalence of emotional and behavioural disorders among strictly orthodox Jewish pre-school children in London. *Clinical Child Psychology and Psychiatry.* Oct 2003;8(4):459-472.
292.    Achenbach TM, McConaughy SH, Howell CT. Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin.* 1987;101:213-232.
293.    Mathai J, Anderson P, Bourne A. Use of the Strengths and Difficulties Questionnaire as an outcome measure in a child and adolescent mental health service. *Australasian Psychiatry.* 2003;11:334-337.