

Review of diagnostic screening instruments for alcohol and other drug use and other psychiatric disorders

2ND EDITION

by

Sharon Dawe¹

Natalie J Loxton¹

Leanne Hides¹

David J Kavanagh²

Richard P Mattick³

¹School of Applied Psychology, Griffith University
Mt Gravatt Campus, Brisbane, Queensland 4111

²Department of Psychiatry, University of Queensland,
Royal Brisbane Hospital, Herston, Brisbane, Queensland 4006

³National Drug and Alcohol Research Centre,
University of New South Wales, Sydney, NSW 2052

August 2002

© Commonwealth of Australia 2002

ISBN 0 642 82124 0

This work is copyright. Apart from any use as permitted under the *Copyright Act 1968*, no part may be reproduced by any process without prior written permission from the Commonwealth available from the Department of Communications, Information Technology and the Arts. Requests and inquiries concerning reproduction and rights should be addressed to the Manager, Copyright Services, Info Access, GPO Box 1920, Canberra ACT 2601.

Publication approval number: 3146

Acknowledgments

We would like to thank the clinicians and researchers who reviewed an earlier version of this report. Their comments and suggestions have improved the quality of the report and we appreciate the time given to this task.

We would also like to thank the developers of, and those in the process of further validating the reviewed instruments, who have provided us with updated information. We would like to also thank those who have allowed us to publish their material.

In particular we wish to acknowledge Dr. Jim Gerostamoulos from the Victorian Institute of Forensic Medicine for providing us with information regarding hair analysis for detecting illicit drug use and those clinicians who spoke with us about specific issues relating to the measurement of substance problems and psychiatric disorders in Indigenous Australians.

Finally, we would like to thank Sally Frye for reviewing the literature in regards to Indigenous Australians and Sarah Rothe for her initially proofing of this document.

Contents

PART I. General issues	1
Background and context of the review	3
Aims and limitations of the review	4
Issues specific to the development of screening and diagnostic instruments	4
Reliability	4
Validity	5
Sensitivity and specificity	7
Specific issues for reliability and validity in substance using populations	7
Screening vs severity of dependence measures	8
Methodology	8
PART II. The concept of diagnosis	9
The role of diagnosis	11
Structured Diagnostic Interviews	12
The Composite International Diagnostic Interview (CIDI)	12
PART III. Screening and diagnosis of substance misuse	15
ALCOHOL	17
Overview	17
Assessment of quantity and frequency of alcohol use	18
Khavari Alcohol Test	20
Calendar Methods	23
Timeline Followback Method	23
Prospective records	26
Measures used to screen for alcohol problems	27
Alcohol Screening Instruments – Which measure to use when?	29
Alcohol Use Disorders Identification Test (AUDIT)	30
Michigan Alcoholism Screening Test (MAST)	36
CAGE	40
T-ACE and the TWEAK	43
Use of screening instruments with Indigenous Australians	47
Use of screening instruments with older people	48
Use of screening instruments with Adolescents	50
Alcohol disorder screening in serious mental disorder	51
Measures to assess severity of alcohol dependence	52
Self-report measures of severity of alcohol dependence	53
Severity of Alcohol Dependence Questionnaire (SADQ-C)	56
Short Alcohol Dependence Data Questionnaire (SADD)	58
Alcohol Dependence Scale (ADS)	61
Structured clinical interviews for alcohol abuse and dependence	63
Addiction Severity Index (ASI)	63
Comprehensive Drinker Profile	66
Biochemical measures used in the assessment of alcohol use	67
Blood Alcohol Concentration (BAC)	67
Liver Function Tests (LFT)	68

Nicotine	70
Measures used to determine nicotine use	70
Quantity and Frequency Methods	70
Assessment of severity of nicotine dependence	71
Revised Fagerström Tolerance Questionnaire (RTQ) & Fagerström Test for Nicotine Dependence (FTND)	71
Biochemical measures of nicotine use	75
Expired Air Carbon Monoxide (CO)	75
Nicotine and Cotinine	75
Nicotine use in severe and persistent psychiatric populations	76
Drug use other than alcohol and tobacco	77
Overview	77
Screening for other drug use with Indigenous Australians	78
Measures used to screen for other drug use	78
Drug Abuse Screening Test (DAST)	81
Assessment of quantity and frequency of drug use	82
Opiate Treatment Index (OTI)	82
Assessment of severity of dependence on drugs	85
Severity of Opiate Dependence Scale (SODQ)	85
Benzodiazepine Dependence Questionnaire	88
Leeds Dependence Questionnaire (LDQ)	92
Severity of Dependence Scale (SDS)	94
Structured interview for other drug use	97
Substance Dependence Severity Scale	97
Biochemical measures of drug use	99
Urine analysis	99
Hair analysis	100
PART IV. Screening and assessment of psychiatric problems	103
Measures used to assess general psychological state	105
Symptom Checklist-90-Revised (SCL-90-R)	105
General Health Questionnaire (GHQ)	107
Brief Psychiatric Rating Scale (BPRS)	111
Royal Park Multidiagnostic Instrument for Psychosis	113
Screening and assessment of psychiatric problems in Indigenous Australians	115
Measures used to assess specific disorders	116
Beck Depression Inventory (BDI)	116
Beck Hopelessness Scale (BHS)	119
Beck Scale for Suicidal Ideation (BSS)	121
Spielberger State Trait Anxiety Scale (STAI)	123
Beck Anxiety Inventory (BAI)	125
Impact of Events Scale (IES)	128
Eating Attitudes Test (EAT)	131
Eating Disorders Inventory II (EDI)	133
Bulimic Investigatory Test - Edinburgh (BITE)	135
PART V. Recommendations for future research	139
General Comments	141
Recommendation for future research	141
References	143

Part I

General issues

Introduction

Background and context of the review

In recent years there has been a growing recognition that many people with drug or alcohol problems are also experiencing a range of other psychiatric and psychological problems. The presence of concurrent psychiatric or psychological problems is likely to impact on the success of treatment services. These problems vary greatly, from undetected major psychiatric illnesses that meet internationally accepted diagnostic criteria such as those outlined in the Diagnostic and Statistical Manual (DSM-IV) of the American Psychiatric Association (1994), to less defined feelings of low mood and anxiety that do not meet diagnostic criteria but nevertheless impact on an individual's sense of wellbeing and affect their quality of life.

Similarly, the presence of a substance misuse problem among those suffering from a major psychiatric illness, often goes undetected. For example, the use of illicit drugs such as cannabis and amphetamine is higher among those individuals suffering from schizophrenia (Hall, 1992) and the misuse of alcohol in people suffering from schizophrenia is well documented (e.g., Gorelick et al., 1990; Searles et al., 1990; Soyka et al., 1993). High rates of alcohol misuse have also been reported in a number of groups including women presenting for treatment with a primary eating disorder (Holderness, Brooks Gunn, & Warren, 1994), individuals suffering from post-traumatic stress disorder (Seidel, Gusman and Aubueg, 1994), and those suffering from anxiety and depression.

Despite considerable evidence of high levels of co-morbidity, drug and alcohol treatment agencies and mainstream psychiatric services often fail to identify and respond to concurrent psychiatric or drug and alcohol problems, respectively. The original review was conducted as a first step in providing clinicians with information on screening and diagnostic instruments that may be used to assess previously unidentified co-morbidity. The current revision was conducted to extend the original review by updating psychometric findings on measures in the original review, and incorporating other frequently used measures that were not previously included. The current revision has included information regarding special populations, specifically Indigenous Australians, older persons and adolescents.

The objectives were to:

- update the original review of AOD and psychiatric screening/diagnostic instruments,
- recommend when these instruments should be used, by whom and how they should be interpreted,
- identify limitations and provide recommendations for further research,
- refer the reader to pertinent Internet sites for further information and/or purchasing of assessment instruments.

Aims and limitations of the review

It is hoped that the revised review of screening and diagnostic instruments and procedures contained within serves as a practical resource for clinicians working within mental health settings, hospitals, and general practice. It is not intended to be a comprehensive review of all screening and diagnostic instruments, nor does it provide an exhaustive review of the research findings regarding particular instruments. Due to the nature and scope of this project the authors have been highly selective and those instruments reviewed are widely used, have been demonstrated to be reliable and valid measures of the construct in question and are brief and easy to administer. It should be noted for those familiar with the original review that a small number of newly developed instruments previously reviewed have not been included in this edition, primarily due to a lack of further validation. Furthermore, in this edition while related Internet pages have been provided, it should be noted that due to the fast growing nature of the Internet some pages might become outdated.

Issues specific to the development of screening and diagnostic instruments

One of the central issues discussed regarding the instruments presented in this review is the instrument's reliability and validity (i.e. its psychometric properties). Broadly speaking, the concept of reliability refers to the instrument's ability to measure a construct consistently, while the term validity describes how accurately an instrument measures what it purports to measure (see Anastasi & Urbina, 1997, for further discussion). Reliability is generally more easily established than validity and therefore an instrument's psychometric properties are usually described with reference to reliability first and then validity.

Reliability

Correlation Coefficient

An instrument's reliability is best understood in terms of correlation coefficients, which denote an association between one variable (or score) and another. A correlation coefficient consists of two elements: direction and magnitude. A correlation may be positive or negative (direction) and may range from -1 to 1 (magnitude). A positive correlation between two scores indicates that as one score increases so does the other. A negative correlation between two scores indicates that as one score increases the other score decreases. In terms of magnitude, the closer a coefficient is to 1 the more strongly they are related. A coefficient of 1 indicates a perfect one-to-one association between the scores and a coefficient of 0 indicates that there is no systematic relationship between the two. It should be emphasised that a negative correlation does not denote a weak relationship. Therefore, a score of -1 indicates a perfect relationship whereby as one score increases the other decreases. In psychology, perfect correlations do not exist due to innate imperfection in human nature and inherent error in the measurement of psychological constructs. To make decisions regarding how strong a relationship needs to be between one variable and another, rules of thumb for acceptable strengths of association are used. Generally, for research purposes, a correlation of $.70$ or above is acceptable. However, for clinical work in which treatment choices may depend, in part, on an assessment score, acceptable correlations for instruments should be higher.

An instrument's reliability is determined by the stability of the measurement across time, that is, test-retest reliability; and by internal consistency, or the extent to which items on an instrument measure the same construct.

Test-Retest Reliability

Test-retest reliability is determined by administering the same instrument on two occasions and assessing how similar the scores are; that is, the strength of the correlation. The more stable the measure, the stronger the correlation. The time frame between testing can vary from several days to several years, depending on the construct being measured. For instance, scores on an intelligence test taken in adulthood should remain stable over the years, whereas scores on a reading or vocabulary test will change rapidly during early childhood.

Internal Consistency

In addition to stability over time, an instrument needs to be internally consistent. The internal consistency of an instrument refers to the homogeneity of the item, i.e., items measuring one construct, or a single domain within a construct, should correlate strongly with each other. There are a number of ways to assess internal consistency and the reader is referred to Kaplan and Saccuzzo (2001) for further discussion. One frequently used method is referred to as split-half reliability. An instrument is administered and divided into halves that are scored separately. The results of one half of the test can then be correlated with the results of the other. An alternative method used for determining an instrument's internal consistency is the use of a statistic known as Cronbach's alpha (α), which is based on the average correlation of all the possible split-halves within a test. The recommended minimal criterion for an alpha co-efficient is .7 (Nunnally, 1994). A Kuder-Richardson coefficient (KR20) is also frequently reported and is essentially the same as an alpha. It is calculated for instruments that are dichotomously scored, i.e., scored 0 or 1; true/false, whereas the alpha is calculated for instruments with likert-type scales.

Inter-rater reliability

A final index of reliability is inter-rater reliability, a statistic relevant to assessments in which two or more trained observers arrive at some conclusion, such as diagnosis. Although there are a variety of coefficients assessing inter-rater reliability, the most commonly used coefficient is the Kappa statistic, which measures the level of agreement between two or more observers on categorical outcomes (e.g., illness/no illness). Kappas range from 0-1, with coefficients closer to 1 indicating that two observers agree more often than chance. Generally a Kappa of .70 or greater is considered acceptable. Intra-class correlations (ICC) are also frequently used for agreement on measures that derive a score rather than a categorical decision (Bartko, 1976). For an in depth discussion of reliability see Kaplan and Saccuzzo (2001).

Validity

The validity of a test refers to "what the test measures and how well it does so" (Anastasi & Urbina, 1997, p. 113). Broadly speaking, validity is ascertained by reference to independent, external criteria, and may be roughly divided into content validity, construct validity and criterion validity.

Content Validity

An instrument is considered to have content validity if it measures all aspects of the particular condition. For example, an instrument designed to assess the severity of alcohol withdrawal symptoms would need to include items addressing the broad range of changes that are known to occur on cessation of alcohol use.

Construct Validity

Construct validity requires that the instrument measure only those characteristics of the condition or construct it was designed to measure. Construct validity is one of the most difficult types of validity to determine and is usually determined by the use of a variety of procedures. Statistical methods which function to test construct validity include factor analysis, which is used to ascertain whether the instrument measures the same number of domains of the construct that are theoretically thought to exist. Factor analysis is a statistical procedure that analyses the inter-relationships in a set of data. Questionnaire items that measure a single domain should correlate highly together and those that measure different domains should correlate minimally. Consequently, highly correlated data will “clump” together and load on one factor and other data will “clump” with other like data and load on additional factors. The number of factors “extracted” from a set of data should correspond to the number of hypothesised domains. To illustrate, an instrument developed to measure a construct of “addiction” theorised to have three domains of cravings, compulsive use, and physical dependence should extract three factors – one factor for each domain. Measures of single, unitary constructs should result in a single factor solution. Further, individual factors should have high internal consistency as calculated by Cronbach’s alpha. If different domains relate to a single overarching construct, there will be some correlation between factors (for a detailed discussion of factor analysis see Tabachnick & Fidell; 2001).

Two other related forms of construct validity are **convergent** and **discriminant validity**. Convergent validity is established when the measure correlates highly with other tests or methods of measuring the same construct. For example, a self-report measure of depression should correlate highly with a clinician’s assessment of depression and/or with another established measure of depression, such as the Beck Depression Inventory. Alternatively, discriminant validity is established in the presence of low correlations between the measure and measures of other unrelated constructs. For instance, a measure of anxiety should correlate poorly with a measure of intelligence. Most importantly though, for measures of drug and alcohol use, construct validity implies that there are theory consistent group differences and theory consistent intervention effects. In other words, scores on a test of substance dependence should differentiate between individuals who are dependent on a substance and those who are not. Further, scores on a measure of dependence should decrease following treatment for the dependence.

Criterion Validity

An instrument’s criterion validity refers to the extent to which it corresponds to another accurate measure of the construct. In the context of screening and diagnostic instruments, **concurrent validity** is one aspect of criterion validity that is particularly important. Studies on concurrent validity look at the relationship between an instrument and the criterion, e.g., a diagnosis. Essentially, concurrent validity refers to the relationship between scores on the instrument of interest and another measure of the same construct or a diagnosis at the same time. To illustrate: if researchers develop a brief instrument to measure the severity of

alcohol dependence, they may compare the scores on the new instrument with other standardised measures of alcohol dependence or against a major diagnostic system such as the DSM-IV. **Predictive validity**, on the other hand, is the extent to which an individual's scores on a measure will accurately predict that individual's future scores or behaviour. For instance, a measure of hazardous alcohol use, in which high scores in adolescence consistently predict alcohol use problems in adulthood, would be considered to have predictive validity. A more detailed discussion of general issues in validity can be found in Anastasi and Urbina (1997). Closely related to the concept of validity are issues of sensitivity and specificity.

Sensitivity and specificity

In addition to knowing whether an instrument compares favourably with a previously validated instrument, it is important for clinicians to be able to confidently interpret the scores. Using a statistical procedure (receiver operating characteristics; ROCs) it is possible to determine the score obtained that produces maximum **sensitivity** (i.e., the instrument correctly identifies subjects with a current diagnosis) and **specificity** (i.e., it correctly identifies those who do not meet diagnostic criteria). A ROC curve is obtained by plotting sensitivity against false positive rate for all possible cut-off points of the instrument. Rey, Morris-Yates and Staanslaw (1992), Hanley and McNeil (1982) and Murphy, Berwick, Weinstein, et al. (1987) provide a more detailed review of the use of ROC curves. Whenever possible, information on cutoff scores are included in this review, to guide clinicians in their interpretation of scores.

Specific issues for reliability and validity in substance using populations

The section above briefly reviews reliability and validity issues common to psychological assessment instruments. However, there are some issues that are especially relevant to measures of drug and alcohol use. Notably, there is often some question regarding the validity of self-report measures to accurately assess alcohol and drug use, since drug using individuals may under-report their usage due to social desirability or potential legal ramifications. Generally, though, research in this area has found that self-reported measures of substance use generally provide valid and reliable information regarding client's actual substance use (Darke, 1998; Sobell, Breslin, & Sobell, 1997). There is also some evidence that individuals are actually more open regarding their drug and alcohol use when completing computerised or pen and paper self-report questionnaires, than in face-to-face interviews (e.g., Sobell, Brown, Leo, & Sobell, 1996). Self-report accuracy can be enhanced by ensuring respondents are not intoxicated when assessed, that confidentiality is assured, and that the wording of instruments is clear and understood by the respondent (Sobell & Sobell, 1995). Given the above and taking into consideration the efficiency of using brief, inexpensive, widely available and easily administered questionnaires, this review primarily includes self-report questionnaires, both copyright instruments and those in the public domain, and where possible, computerised versions of such.

Structured clinical interviews, however, do play an important role in assessing substance use, especially in clinical populations, where detailed knowledge of a client's drug taking history and current behaviour are essential to diagnosing and facilitating appropriate care

and treatment. For this reason, we look at a number of frequently used and well validated clinical interviews, such as the Composite International Diagnostic Interview (CIDI; both pen and paper and computerised versions) and the Addiction Severity Index (ASI). The decision regarding which, or how many, sources of information regarding substance use is required depends on a variety of issues, including the evaluation's purpose and the cost-benefit trade-off. It should also be noted that potential consequences of admitting substance use and misuse (both positive and negative) are likely to affect any method of self-report. Accordingly, the context of the assessment should be taken into consideration when interpreting self-reported assessments (Finch & Strang, 1998).

Screening vs severity of dependence measures

Generally, this review organises selected measures in each section in terms of: 1) those instruments that *screen* for the presence of drug and alcohol abuse; 2) those that obtain a descriptive account of frequency and quantity of use; 3) those that assess the *severity* of dependence; and 4) biochemical measures. The decision regarding which type of instrument to use depends on the assessment's purpose. Most screening instruments are sensitive to low-level misuse of a substance, but are less sensitive to determining a range of use and dependence (i.e., they have a ceiling effect). Therefore screening instruments are most suitable for research in the general population and for detecting the presence of potential abuse and dependence. Conversely, measures of severity of dependence are often insensitive to low-level use and are most appropriately used with clients with established dependence, and to monitor treatment outcomes. Frequency/quantity measures are most useful clinically for diagnostic purposes and determining treatment goals. Biochemical measures are both expensive and often insensitive. However, there may be a role for their use if external validation of recent use is necessary. Quantity/frequency and biochemical measures provide measures of drug and alcohol use but fail to provide information on negative psychological, occupational, social and physical consequences – information that is particularly important to treatment planning.

Methodology

The first edition of this review began by identifying screening and diagnostic instruments used in psychiatric and alcohol and other drug using populations. The revised edition has further updated psychometric data on these measures, specifically covering the years 1996-2001. Literature searches were conducted using the MedLine and PsychINFO databases, in addition to searching abstracts of recent relevant conference proceedings, and other sources such as the bibliographies of recently published books. Articles that provided information on instrument's reliability, validity, sensitivity and specificity in detecting disorders, administration time or other relevant features formed the basis of the review. To provide the latest information regarding these measures, primary authors and researchers investigating the psychometric properties of these measures, were approached.

Part II

The concept of diagnosis

The role of diagnosis

The purpose of diagnosis is to provide clear descriptive categories to enable clinicians to identify presenting problems and communicate succinctly about them. The importance of this communication and the determination of related prognosis is generally recognised in health care. Diagnostic categories allow the development of knowledge about the causation and natural course of the problems and information on how they might best be treated.

A diagnosis though does not provide information about the *unique* features of an individual's problems. However, this does not argue against the use of diagnosis – only against its improper use. It is essential to supplement a diagnosis with additional information about factors that affect an individual's presentation and that may exacerbate or affect the problems he/she has. Diagnosis should become part of good clinical management, rather than a mechanistic and simplistic tool for clinical management.

The predominant current diagnostic systems for mental disorders are the Diagnostic and Statistical Manual of Mental Disorders – 4th Edition, known as the DSM-IV and the International Classification of Disease (ICD-10; World Health Organization, 1992), which sets out the classification of mental and behaviour disorders. Over recent years, the two taxonomies have come closer together in terms of their criteria and there is greater agreement about the disorders and the criteria defining them. ICD is used throughout Australia's hospital system and the DSM is also used extensively, particularly in mental health research.

In reaching a decision regarding whether a diagnosis is present or absent, an important distinction is drawn between the presence of symptoms and the making of a diagnosis. Symptoms do not generate a diagnosis unless their nature, number and duration meet the criteria for the disorder. An important and simple example of this distinction is in the case of depression: many individuals complain of unhappiness, sadness or dysphoria but do not have sufficient symptoms or functional impairment for the dysphoria to reach clinical significance or require treatment. This distinction between symptoms and diagnoses underscores the importance of using systematic criteria to make formal diagnosis. However, in many contexts the skills and knowledge required to make a diagnosis may not be present, and because of this there is a need to have methods of screening to detect the *likelihood* that a diagnosis might be present, so that further investigation may occur. Such screening is common in many areas of medicine and well accepted by the general public and health care professionals. For example, foetuses are screened in the early stages of pregnancy for neural tube defect and Down's Syndrome using alpha-foetoprotein levels in the maternal blood. While not a foolproof method, the levels of alpha-foetoprotein are used to detect the likelihood of the presence or absence of a problem, and to suggest the presence or absence of a diagnosis of neural tube defect or Down's Syndrome. Similarly, screening instruments can be used in the context of drug or alcohol problems to indicate the likelihood of a diagnosis or presence of either alcohol abuse or alcohol dependence (mild, moderate or severe). Other screening instruments are useful and frequently employed in detecting psychological disorders such as depressive disorders, anxiety disorders and other mental health problems.

The following two sections of this report provide information about screening, diagnosing and assessment of substance misuse disorders and psychiatric or psychological problems. The role of screening instruments is to allow for a simple, often self-report, approach to gathering information about the presence or absence of certain behaviours or emotional

states, and thereafter to enable assessment and/or consideration of the need for further investigation.

Structured Diagnostic Interviews

The Composite International Diagnostic Interview (CIDI)

Key Reference: Wittchen, H.U. (1994). Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI): A critical review. *Journal of Psychiatric Research*, **28**, 57-84.

Associated Web Page: (URL: <http://www.crufad.com>)

Summary

The Composite International Diagnostic Interview is a comprehensive and fully standardised interview schedule designed for the assessment of psychological disorders to provide ICD and DSM diagnoses. It may be administered by a trained interviewer or administered via computer. Specific diagnoses are generated, with time of onset and duration of each disorder determined.

Description and Development of the CIDI

The Composite International Diagnostic Interview was developed as part of a process conducted by the World Health Organization (WHO, 1993; see Wittchen, 1994 for a review) and the United States' Alcohol, Drug Abuse and Mental Health Administration commencing in the early 1980s. Through the process, a number of instruments were developed including the predecessor of the CIDI – the Diagnostic Interview Schedule (DIS), and the Schedules for Clinical Assessment for Neuropsychiatry (SCAN; WHO, 1992). These instruments were specifically designed to allow trained interviewers to conduct standardised interviews to detect the presence or absence of psychological disorder under existing psychological taxonomies (ICD and DSM). The DIS is used extensively in population epidemiological research. The CIDI was also designed for population surveys and can be used by trained interviewers who have been familiarised with the interview's structure. The CIDI covers eating disorders, organic mental disorders, substance use disorders, schizophrenic disorders, paranoid disorders, affective disorders, anxiety disorders, post-traumatic stress disorders, somatoform disorders, dissociative disorders, and psycho-sexual disorders. Field trials are continuing and other disorders will be added, as data on reliability and validity are gathered.

Reliability and validity of the CIDI

The CIDI has been examined in a large number of studies reviewed by Wittchen (1994) and shown to have acceptable levels of test-retest and inter-rater reliability for depressive disorders, anxiety disorders, organic brain syndrome, schizophrenic disorders and eating disorders. Validity studies have also been conducted to assess the CIDI generated diagnosis against other methods of reaching a diagnosis. Wittchen's (1994) summary of the information suggests that the overall diagnostic concordance between the CIDI and clinical checklists are more than adequate for depressive disorders, anxiety and phobic disorders, and substance abuse disorders. There is also good concordance between ICD-10 diagnoses

and CIDI diagnoses. It appears that the CIDI does provide information likely to approximate a clinically generated diagnosis, and is frequently used in research settings to assign diagnostic group membership. In terms of validating drug and alcohol use disorders, the CIDI demonstrates good re-test reliability across substances, with stronger reliability for dependence and lower reliability for less severe harmful use and abuse (Uestuen et al., 1997). Moderate correlations with a clinical interview for dependence (SCAN) have been found, although there is some variability across substances with higher agreement for alcohol and opiate dependence and lower agreement for amphetamine dependence. Again, the CIDI was more sensitive to dependence, as opposed to abuse, diagnoses (Cottler et al., 1997).

There is some question of the validity of the psychosis module of the CIDI, with recent research suggesting relatively poor sensitivity and specificity. Typically the module has been found to underestimate schizophrenic diagnoses and the presence of DSM and ICD criteria (e.g. the presence or absence of hallucinations, delusions) when compared with clinical diagnoses (Cooper, Peters, & Andrews, 1998). Accordingly, other sources of information should be added to CIDI data to confirm diagnosis.

Suitability for Special Populations

The CIDI has been used extensively in a number of countries and found to perform more than adequately across different cultures and settings. The countries examined included Australia, The Netherlands, Greece, India, China, United Kingdom, Sweden, Eastern Germany, Luxembourg, West Germany, Italy, United States, Portugal, Norway, France, Porta Rica and Brazil. However, its suitability for Indigenous Australians is less clear. Although there have been no specific field trials with this group, discussions with clinicians who have administered the computerised version of the CIDI indicate that there are considerable problems with length, detail and relevance. Further, Indigenous people viewed it with some suspicion and hostility. Presently, clinicians recommend that those sections covering personality disorders and low prevalence disorders be omitted (Earnest Hunter, personal communication, 1997), a strategy that may also be helpful with other population groups.

With populations that have marginal literacy levels in English, the interviewer-administered form of the instrument should be used. Having patients self-administer the computerised version may prove too difficult for some sub-groups that may be seen in drug and alcohol or psychiatric settings, especially those from non-English speaking backgrounds or with limited educational attainment.

Administration and scoring

The CIDI comes in two forms: a paper form, which is completed by an interviewer, or a computerised form, which can be either administered by an interviewer or completed by an informant. As mentioned previously, the CIDI generates specific diagnoses with time of onset and the most recent occurrence. The latest version (v2.1) was released in 1997 and comes with two forms – lifetime and 12 months. Using the computerised form, no scoring is required by the interviewer. However, when using the pen and paper form, a scoring protocol needs to be applied.

Because of the extensive training required for the paper and pen form, its relatively lengthy administration time (an average of 75 minutes), and the clerical time required for data entry and scoring, its use in routine clinical practice may be limited (Peters, Morris- Yates, and

Andrews, 1994). Accordingly, a short screening version (CIDI-SF), which takes less than 10 minutes to complete, has been developed and is currently undergoing validation. Reports at this stage suggest good concordance with the full version of the CIDI (URL: <http://www.unsw.edu.au/crufad/journal/acidi-sf.htm>).

Professor Gavin Andrews and colleagues have developed the computerised version of the CIDI, entitled CIDI-AUTO. Run on desktop IBM compatible computers, it takes between 20 minutes to an hour and a half to complete depending on the specific diagnoses selected. The interviewer selects specific diagnoses that may be of interest. The program's latest version (v2.1) can be self- or interviewer-administered, and has been translated into Dutch, French, German, Portuguese and Spanish. The reliability and validity of this version is currently under assessment, but earlier versions have been found to have acceptable to excellent inter-rater and test-retest reliability and acceptable validity, although the CIDI-Auto (like the CIDI) appears to be less sensitive than diagnoses made by clinicians ($\kappa = .40$). The authors report that clients tend to feel more comfortable completing the CIDI-AUTO than the interview version of the CIDI, and that it allows them the opportunity to disclose information that has not been addressed in previous interviews (Andrews & Peters, 2000, p.5).

The CIDI may be used by any suitably qualified mental health worker who has been oriented to its use; however, actual administration and scoring should be supervised by a fully trained mental health professional who has conducted recognised training in the CIDI. In NSW this training may be obtained from the Clinical Research Unit for Anxiety Disorders, St. Vincents Hospital in Sydney. It is important that those administering the CIDI ensure that clients understand the terms used, especially when using the self-administered, computerised format.

The CIDI is a particularly useful instrument for drug and alcohol settings where the assessment process allows for the making of a diagnosis on site, or for those settings where referral to a clinical psychologist or a psychiatrist is made so difficult that the facility to make a diagnosis is important.

Availability

The CIDI is available from the St. Vincents Hospital in Sydney and training costs, documentation and floppy disk for the instrument are approximately AU\$700 per licence. A licence allows for automatic update of the CIDI software with new versions. Details should be confirmed by contacting the WHO CIDI Training and Reference Centre, 299 Forbes St, Sydney on +61 (0)2 9332-4316, or at URL: <http://www.crufad.com/cidi/supportindex.htm>

Part III

Screening and diagnosis of substance misuse

ALCOHOL

Overview

Increasingly, the importance of detecting harmful and hazardous drinking in all health care settings has been recognised and incorporated into the National Drug Strategy's strategic plan. Screening instruments need to be short, easily understood by the client, simply scored by the clinician and provide reliable information to enable the clinician to decide whether further assessment and intervention is required. Ideally, screening for alcohol problems should be incorporated into routine practices, particularly, medical practices, in general hospitals, the workplace and welfare and general counselling services (Mattick and Jarvis, 1993).

Determining the quantity and frequency of alcohol use is an essential part of an assessment when an alcohol problem is suspected. However, it may be time consuming. The Khavari Alcohol Test offers a quick method of estimating the overall quantity of alcohol consumed over a specific time period, including periods of very heavy drinking. However, it does not determine the *pattern* of drinking. On the other hand, the Timeline Followback method, although time consuming, assesses both recent alcohol consumption and obtains detailed information on amount consumed, the time period in which it was consumed, and derives a pattern of drinking, which is particularly useful for treatment planning. Quantity and frequency measures however do not provide information regarding the impact of alcohol misuse on people's social, occupational, and psychological functioning.

There are a number of self-report measures designed to screen for the presence of alcohol-related problems; in this report we have reviewed the Alcohol Use Disorders Identification Test (AUDIT), the Michigan Alcoholism Screening Test (MAST) and the CAGE. Also included are the T-ACE and the TWEAK, which were developed specifically for the purpose of screening for hazardous drinking in pregnant women. The TWEAK, though, is now being validated for use with men and women in the general population. The MAST and the CAGE were both developed in North America. They have been widely used and there is considerable information on their psychometric properties. The AUDIT is a comparatively new instrument; however, it has the advantage of good documentation on reliabilities and validity from a range of cultural groups. While each instrument has its strengths and weaknesses, we recommend that the AUDIT be used when an instrument is required to screen for harmful or hazardous alcohol use.

Determining the severity of alcohol dependence is important to assist in developing an appropriate treatment response. The three reviewed questionnaires measuring the severity of alcohol dependence (the Severity of Alcohol Dependence Questionnaire; SADQ; the Short Alcohol Dependence Data Questionnaire; SADD; and the Alcohol Dependence Scale; ADS) were all developed based upon Edwards and Gross's (1976) formulation of the "alcohol dependence syndrome". The following are the key elements: narrowing of drinking repertoire (i.e., a lifestyle where drinking is a major focus), salience of drink-seeking behaviour, increased tolerance to alcohol, repeated withdrawal symptoms, relief or avoidance of withdrawal symptoms by further drinking, subjective awareness of a compulsion to drink, and reinstatement after abstinence. More recently, concern has been expressed that the ADS fails to incorporate the subjective sense of loss of control over

alcohol and the inability to abstain from drinking (e.g., Laranjeira, 1995). The SADQ goes some way to countering this criticism by the addition of a companion scale, the Impaired Control Scale (ICQ), in the revised SADQ-C (Stockwell et al., 1994).

For more detailed information on the drinking behaviour of clients identified with established alcohol problems, the use of structured interviews may be useful in treatment planning. This report reviews the Comprehensive Drinker Profile and the more recently developed Addiction Severity Index (ASI). The ASI is a widely used structured interview that provides detailed information on alcohol and illicit drug use, and related psychosocial problems.

Finally, the use of biochemical measures as either screening or diagnostic measures have been the focus of considerable attention in the alcohol field. However, they pose relatively limited value in detecting an alcohol use disorder. Traditionally used measures of liver function that are routinely used as indicators of excessive drinking, generally lack sensitivity in detecting hazardous alcohol use. More specific measures of alcohol use such as carbohydrate-deficient transferrin (CDT), while more specific to detecting alcohol misuse than more liver function tests, still tend to be less sensitive to alcohol misuse than self-report questionnaires.

Although there are many techniques and instruments for assessing and measuring alcohol misuse, ultimately, as Sobell and Sobell (1995, p. 55) note, “the utility of a drinking measure for research and/or clinical purposes will rest on its intended use”. They suggest that when choosing an assessment instrument consideration is given to the following:

- the assessment’s purpose (e.g., screening, effect of intervention, severity of dependence)
- the assessment’s clinical usefulness (e.g., will it assist in treatment planning?)
- ease of use and relevance for respondents
- length of time to administer and score
- availability and costs involved (Sobell & Sobell, 1998).

Assessment of quantity and frequency of alcohol use

Despite some reservations about the accuracy of self-report there is much to commend this method of obtaining information about the quantity and frequency of alcohol use. Ensuring that both the interviewer and client are working from a shared knowledge base is, however, an essential first step in increasing the accuracy and reliability of the information obtained. Introducing the concept of a standard drink in which one standard drink is equal to 10 gm of ethanol, provides a common unit of measurement across a range of alcoholic drinks. If the client is not already familiar with the concept, this familiarisation also provides an important educative function. It is also important to impart accurate information about low risk, harmful and hazardous drinking. The following are the current guidelines from the National Health and Medical Research Council (NH&MRC, 2000), which can be obtained from: URL: <http://www.nhmrc.gov.au/advice/alcomp.htm>.

Generally, the Australian National Health and Medical Research Council’s recommended guidelines for minimising health risks and gaining long term benefits are 14 standard drinks

for women and 28 standard drinks for men per week, with at least two alcohol free days (NHMRC, 2000). The guidelines also recommend that within this weekly limit that men should not drink any more than 6 standard drinks on any one day and that women should not drink any more than 4 standard drinks on any one day. People involved in high risk, or highly skilled activities should avoid alcohol consumption before or during such activities. Additionally, the guidelines make recommendations for low risk drinking across a range of specific populations, including adolescents, the elderly, young adults, pregnant women, and those with pre-existing social or health problems that may be exacerbated by alcohol use.

What is a “standard drink”?

In Australia a standard drink typically contains 10 grams of alcohol. Beer, wine and spirits vary in the amount of alcohol they contain. The easiest way to estimate how many standard drinks you consume is to look at the drink container, most of which state their alcoholic content. Generally, a “stubby” or can of “full strength” beer contains 1.5 standard drinks; the same volume of “mid-strength” beer contains 1 standard drink. A 120ml glass of table wine or a 30ml “shot” of spirits (e.g., vodka, rum, gin, bourbon) also equal 1 standard drink. A brief guide to alcohol content is provided below (Table 1). A detailed guide is available from the Queensland Health Alcohol, Tobacco and Other Drug Services website:

URL: <http://www.health.qld.gov.au/atods/resources/standard2.htm>

Table 1.

Approximate number of “standard drinks” for commonly consumed alcoholic beverages

Beverage type	Quantity	Approximate number of standard drinks
Regular strength beer (4-5% alcohol)	285ml glass (pot, middy, ten, seven)	1
	375ml can or “stubby”	1.5
	425ml glass (schooner, handle, pint)	1.7
	750ml bottle (Tall-ee)	3
Light strength beer (2-3%)	285ml glass (pot, middy, ten, seven)	0.5
	375ml can or “stubby”	0.7
	750ml bottle (Tall-ee)	1.6
Table Wine (10-13%)	110ml glass	1
	750 bottle	7-8
	2litre cask	16-20
	4litre cask	32-40
Fortified wines (sherry, port, vermouth – 18-20%)	60ml glass	1
	750ml bottle	12
Spirits (rum, whisky, vodka, tequila – 40%)	30ml nip/shot	1
	375ml half-bottle	11
	700ml bottle	22

Quantity-frequency methods

Alcohol intake can be measured directly by asking a client how much they drink (quantity) and how often they drink (frequency). There are many ways in which quantity and frequency information has been obtained (see Room, 1990, for a review). The most common method used in surveys of drinking patterns was developed in the late 1960's by Cahalan and colleagues (Cahalan, Cisin and Crossley, 1967). For each type of alcoholic beverage, i.e., wine, beer, spirits, respondents answer the following questions:

Think of all the times you have had recently.

When you drink....., how often do you have five or six?

When you drink....., how often do you have three or four?

When you drink....., how often do you have one or two?

The frequency scale for each alcoholic drink is as follows:

Three or more times a day

Two times a day

Once a day

Nearly every day

Three or four times a week

Once or twice a week

Two or three times a month

About once a month

Less than once a month but at least once a year

Less than once a year

I have never had....

More recently the KAT (see next section) has been developed as a measure of quantity and frequency of both regular drinking behaviour and periods of binge-drinking. This instrument calculates an estimate of annual alcohol consumption.

Khavari Alcohol Test

Key reference: Khavari, K.A., & Farber, P.D. (1978). A profile instrument for the quantification and assessment of alcohol consumption. *Journal of Studies on Alcohol*, 39, 1525-1539.

Summary

The KAT provides a method for calculating an estimate of yearly alcohol consumption, taking into account episodes of heavy drinking. A simple mathematical formula allows researchers and clinicians to calculate volume of beer, wine and spirits consumed over the year in term of millilitres of alcohol, as well as calculate alcohol consumption in terms of pure ethanol ingested. The KAT has good validity and reliability and has been used in alcohol misusing, geriatric and college populations. It does fail, however, to assess any psychological and social repercussions of drinking and does not access information as to whether the respondent perceives his/her drinking to be problematic.

Description of the KAT

The Khavari Alcohol Test (KAT) consists of 12 items addressing the frequency and quantity of alcohol typically consumed, and instances of heavy consumption. Four items refer to the consumption of beer, four to the consumption of wine and four to the consumption of spirits. For each class of alcohol, the client is asked the frequency of drinking this type of alcohol, the number of drinks usually consumed, the maximum number of drinks consumed and the frequency at which the maximum number of drinks is consumed. Based on the responses, annual absolute alcohol intake can be estimated via a mathematical calculation. This process is thought to capture both regular and binge drinking consumption.

The KAT has been found to have good temporal stability with 2-week test-retest reliability coefficients ranging from .78 to .98. The authors found annual consumption, as calculated by the KAT, to differentiate groups of participants with alcohol problems from participants without problems. Furthermore, annual intake was predictive of Short Michigan Alcoholism Screening Test (SMAST) scores. While the KAT appears to be a valid and reliable measure of alcohol consumption, it does not provide information regarding any problems related to alcohol intake. In cases where more information other than simple consumption is required, the AUDIT would be a preferable alternative as the AUDIT also assesses quantity and frequency of regular and binge drinking episodes and has been found to correlate highly with the KAT ($r = .79$; Loxton, 1998). The KAT has been used in a variety of populations, such as geriatric patients and college students, to assess annual absolute alcohol intake.

As the KAT, along with other quantity-frequency measures, provides an overall estimate of total consumption, it lacks sensitivity in detecting changes in drinking patterns. As such, the KAT is not particularly useful for assessing treatment outcome. Diary-type measures, such as the Timeline Followback and prospective self-monitoring, are better measures of change in drinking over treatment.

Administration and scoring

Based on table 2, frequency of alcohol consumption is given a loading ranging from 0-365.0 for each class of alcohol. Quantity of alcohol is measured in millilitres (e.g., pot of beer = 285mls; glass of wine = 115mls; nip of spirits = 30mls). The following calculation is then used to estimate annual alcohol intake for each type of alcohol:

$$\text{Volume} = \{[\text{usual frequency-maximum frequency}] \times \text{usual volume}\} + \text{maximum frequency} \times \text{maximum volume}.$$

Add all 3 subtotals to calculate annual amount of alcohol in terms of millilitres.

To calculate annual amount of alcohol consumed in terms of milligrams of pure ethanol, multiply each type of alcohol by its ethanol content (i.e., mls of beer x .05; mls of wine x .11; mls of spirits x .37).

Table 2.**Loadings for calculation of estimates for use with the Khavari Alcohol Test**

Ordinal value	Category	Loadings
0	Never had (beer/wine/spirits)	0
1	Tried, but not currently drinking	0.1
2	Once a year	1.0
3	Twice a year	2.0
4	Three or four times a year	3.5
5	Once a month	12.0
6	Twice a month	24.0
7	Three of four times a month	42.0
8	Once a week	52.0
9	Twice a week	104.0
10	Three or four times a week	182.0
11	Daily	365.0

Availability and cost

A copy of the KAT and scoring guide can be ordered for a nominal fee, from Rutgers University Center of Alcohol Studies, 607 Allison Road, Piscataway, NJ 08854-8001, 732-445-2190; (fax) 732-445-3500 or @ **URL: <http://www.rci.rutgers.edu/~cas2/>**.

Reliability of self-reported alcohol consumption

The reliability of problem drinkers' self-report has often been questioned. A commonly held view is that problem drinkers either deliberately under report their alcohol use due to an unconscious "defence mechanism" or consciously, to avoid the repercussions of admitting to an alcohol problem. In fact, there is very little empirical evidence that suggests problem drinkers deliberately lie or minimise their alcohol use when responding to self-report questionnaires (Sobell et al., 1997). There is a body of literature that reports a high level of agreement between problem drinkers self-reported use of alcohol and other informational sources, such as family members or spouses (e.g., Miller, Crawford and Taylor, 1979). In a review of the literature, Midanik (1982) concluded that research studies investigating the validity of self-reported alcohol use found a high degree of agreement between problem drinkers and collaterals with no consistent detection of error.

There are, however, factors that may affect the reliability and accuracy of self-reported alcohol use that a clinician should always be sensitive to. Obtaining corroborative information in such circumstances is prudent and may be done through interview of family members or spouses. However, collateral information is subject to availability of specific information and reporting biases. Even in severe and persistent mental disorder, collateral data rarely provides any more information than is given by the participants themselves (Carey & Simons, 2000). Fals Stewart et al. (2000) suggest that accuracy of self-report can be enhanced when clinicians and researchers tell respondents that their reports will be validated against another measure, such as urine analysis and/or collateral information. They also suggest using interviewers who have already established rapport with the respondent and to use clinically trained interviewers where possible. Furthermore, Miller (1983) provides evidence that an interviewer who is confrontational, authoritative and challenging is less likely to obtain accurate information regarding alcohol use than an interviewer who is empathic and non-threatening.

Limitations

While quick and easy to administer, Quantity-Frequency methods generally do not provide information on atypical drinking periods but assume that drinkers' patterns are stable. They are likely to be less accurate when substantial fluctuations in use are present. Therefore, simple Quantity-Frequency methods lack the precision and information that can be obtained by more time-consuming calendar methods such as the Timeline Followback (Fals Stewart et al., 2000). Further, by focusing only on consumption, there is no indication regarding whether the individual has a concern about their alcohol consumption. Administration of a standardised instrument such as the AUDIT is preferable, as an initial screening tool in a setting where the primary purpose is to detect potentially hazardous or harmful alcohol use. However it should be recognised that the Quantity-Frequency questions on the AUDIT are subject to the same limitations as other applications of these questions, and a more detailed assessment of the alcohol consumption and its determinants is required before planning treatment.

Availability and cost

Nil

Calendar Methods

Timeline Followback Method

Key references: Sobell L.C. and Sobell, M.B. (1992). Timeline Followback: A technique for assessing self-reported alcohol consumption. In R. Litten and J. Allen (eds) *Measuring Alcohol Consumption*. Humana Press Inc.

Sobell, L.C. and Sobell, M.B. (1995). *Alcohol Timeline Followback Users' Manual*. Addiction Research Foundation: Toronto: Canada.

Summary

The Timeline Followback Method (TLFB) obtains precise information on the amount of alcohol consumed and duration of each drinking session over a specified period of time, usually 3 months. The key feature of this method is a blank calendar that the clinician and client complete together to obtain a detailed description of alcohol consumption. This procedure produces a detailed pattern of consumption. The time taken to complete the TLFB depends upon the time period covered and an individual drinker's pattern of consumption. It is not a useful method to use in a primary care setting. However, in a treatment or research setting the TLFB has been demonstrated to have high test-retest reliability, and the quantity measures obtained correlate well with other indices of alcohol problems, such as severity of dependence and biochemical measures of liver function. A computerised version and a shortened version suitable for telephone interviews have been developed and have shown comparable validity and reliability to the original measure. The shortened version, though, has some limitations, which may affect its utility in clinical and research programs.

Description of the TLFB

The Timeline Followback Method (TLFB) is a technique developed to enable an accurate retrospective account of alcohol consumption to be made over a specified time period. The key element of this approach is the use of a calendar on which the client provides an estimate of the amount of alcohol drunk on each drinking occasion during the time period. To assist in memory recall and to provide a framework for the client to work within, the first task is to note all events that may assist with recall. Such events may include national holidays, newsworthy events and significant personal events. It is then possible to build a picture of alcohol consumption around these significant dates. Any personal diary that might assist in recall is also included in the procedure. The client is then able to proceed with filling in the drinking days, noting amount consumed and, if required, the number of hours taken to consume alcohol. As the TLFB is sensitive to changes in drinking behaviour, this measure is particularly useful for assessing treatment outcome.

Reliability and validity of the TLFB

The TLFB has been used extensively in the research literature for over ten years and has been found to have high test-retest reliability, with co-efficients ranging from .79 to .96 over 30- to 90 day follow-up periods across a range of drinking populations (Sobell & Sobell, 2000). There is also a high degree of agreement between client self-report and official records such as days in jail or treatment facilities. The more recent computerised version has been found to have good temporal stability, with most studies finding test-retest correlations exceeding .85 (Fals Stewart et al., 2000). Alternative forms of the TLFB (pen and paper, modified for telephone, computerised) have provided similar information, with high agreement between forms on number of drinks consumed and pattern of consumption (Sobell et al., 1996).

Recently, Sobell & Sobell (2000) demonstrated overall consumption, number of heavy drinking days, and number of mean drinks per drinking day, to be positively correlated with a standardised severity of dependence scale (ADS) and scores on the short MAST indicating that the level of alcohol problems or dependence was directly related to drinking behaviour as determined by the TLFB method. There was a similarly high level of agreement between those drinking variables derived from the TLFB method and biochemical indices of alcohol-related liver dysfunction (Sobell & Sobell, 2000).

Studies comparing daily alcohol consumption (either clients while in treatment or self-monitoring non-alcohol abusing community members) with retrospective TLFB data over 30 days, have found that information gathered by TLFB tends to underestimate actual alcohol consumption (e.g., Carney, Tennen, Affleck, del-Boca, & Kranzler, 1998). However, the disparity between TLFB and actual consumption is minimal and appears to be influenced by individual differences in reporting.

Special Populations

There are no specific issues regarding the use of the TLFB method in special populations, with psychometric properties maintaining robustness across a variety of populations, including those with severe mental illnesses, those in alcohol treatment facilities, and college students. Further, the TLFB has recently been validated for collecting information on other drug use, in addition to alcohol consumption (Carey & Correia, 1998; Fals Stewart et al., 2000). The TLFB has been validated across a number of countries including Australia,

Canada, Mexico, Poland, and Sweden (Annis et al., 1996; Sobell et al., 2001). There are no reports of the use of the TLFB with Indigenous Australians.

Administration and scoring

The TLFB can be administered by the interviewer as a pen and paper instrument or via a client administered computer program (Sobell & Sobell, 1996). A modified version for telephone interviews has also been developed (Sobell et al., 1996). The interviewer administered, pen and paper version is a time consuming method of obtaining information on alcohol consumption. Although it has been suggested that information for a 3-month period should be obtained in 10-15 minutes, it is the first authors' experience that at least 30 minutes should be allowed. To assist clinicians in the procedure, a training video has been developed and is available at the address below. It is also recommended that clinicians read the key papers.

The modified TLFB for telephone interviews asks clients about their daily consumption of alcohol, in four categories: 0 drinks, 1-4 drinks, 5-9 drinks, and ≥ 10 drinks. Specifically, interviewers state a specific time period (i.e., from 1st April to 30 April) and ask how many days the respondent did not drink anything. For the remaining number of days, the interviewer asks how many days they drank 1-4 drinks and so on, until all the days of interest (typically 30 days) have been accounted for by the 4 categories. This modified version, though, has some limitations. First, due to the respondents giving a range of drinks rather than a specific number, a total score for number of drinks consumed cannot be calculated. Second, reliability has only been assessed on a 30-day period and therefore this form awaits further validation.

The computerised version of the TLFB provides clients with detailed instructions for self-administration and country-specific information regarding standard drinks, and allows them to incorporate their own personal events or holidays into the provided calendar. The computerised version allows measurement of time intervals up to 12 months and takes the same amount of time to administer as the pen and paper version (L.Sobell, personal communication, May 2001). This version has been well received by clinicians, who noted that clients enjoyed completing computerised assessments and perceived the results to be more credible than pen and paper assessments (Sobell et al., 1996).

Availability and cost

The TLFB is under copyright but the pen and paper version may be obtained free of charge from Linda C. Sobell, Center for Psychological Studies, Nova Southwestern University, 3301 College Avenue, Fort Lauderdale, FL 33314, (phone 954-262-5811; fax 954-262-3895) or via e-mail: sobell@cps.nova.edu. A copy of the TLFB is also available on CD-ROM included with the following publication: American Psychiatric Association (Eds.). (2000). *Handbook of psychiatric measures*. Washington, DC: Author.

Cost information for the users' guide, computerised version and training video is available from Marketing and Sales Services, Centre for Addiction and Mental Health, 33 Russell Street, Toronto, Ontario, Canada, M5S 2S1 (phone 416-595-6059), or @

URL:http://www.camh.net/resources/clinical_tools.html#TimelineFollowBack

Measures used to screen for alcohol problems

In the following section we have listed five of the most widely used and validated self-report screening measures for identifying individuals engaging in hazardous or harmful alcohol misuse and/or who may meet diagnostic criteria for alcohol abuse or dependence. These measures include the Alcohol Use Disorders Identification Test (AUDIT), the Michigan Alcoholism Screening Test (MAST), the CAGE, the T-ACE and the TWEAK. As noted previously, the choice of screening measures to use depends on the assessment's purpose, the cost and availability of the instrument, time to administer and, most importantly for screening devices, the sensitivity of the instrument to detecting alcohol misuse.

Furthermore, the utility of these devices has been found to vary across populations of interest and the type of problematic drinking under investigation. Therefore, to assist the reader in determining the most appropriate measure to use in a particular situation, a flow chart has been developed to aid decision-making. It should be noted that the flow chart is merely a guide.

Generally, the AUDIT appears to be the most sensitive instrument for current alcohol use disorders across a range of countries and populations, and is the best instrument for identifying low-level hazardous drinking behaviour in adults and adolescents. There has been some suggestion, although with little empirical investigation, that cut-off points may need to be lowered whenever used with women or adolescents (e.g., Chung et al., 2000; see p.50 for a review of alcohol misuse assessment in adolescent populations). Furthermore, although the utility of the AUDIT varies across sub-populations of drinkers (e.g., Cherpitel, 1995), compared to both the MAST and the CAGE, the sensitivity of the AUDIT remains consistently high (e.g., Cherpitel, 1998). One salient exception is in the case of older people, where the few studies investigating alcohol-screening measures have found the AUDIT less sensitive to the presence of alcohol problems.

The CAGE on the other hand, while less sensitive to alcohol misuse in young people, appears to be the best measure for alcohol problems in the elderly. Although the CAGE tends to lack sensitivity to alcohol problems in the general public, when accompanied by a brief assessment of quantity-frequency, its brevity (4 items) and capacity for verbal administration, it becomes useful for screening purposes within a routine medical or psychological interview. The TLFB can be further added to the assessment if warranted by responses to the CAGE and Q-F questions.

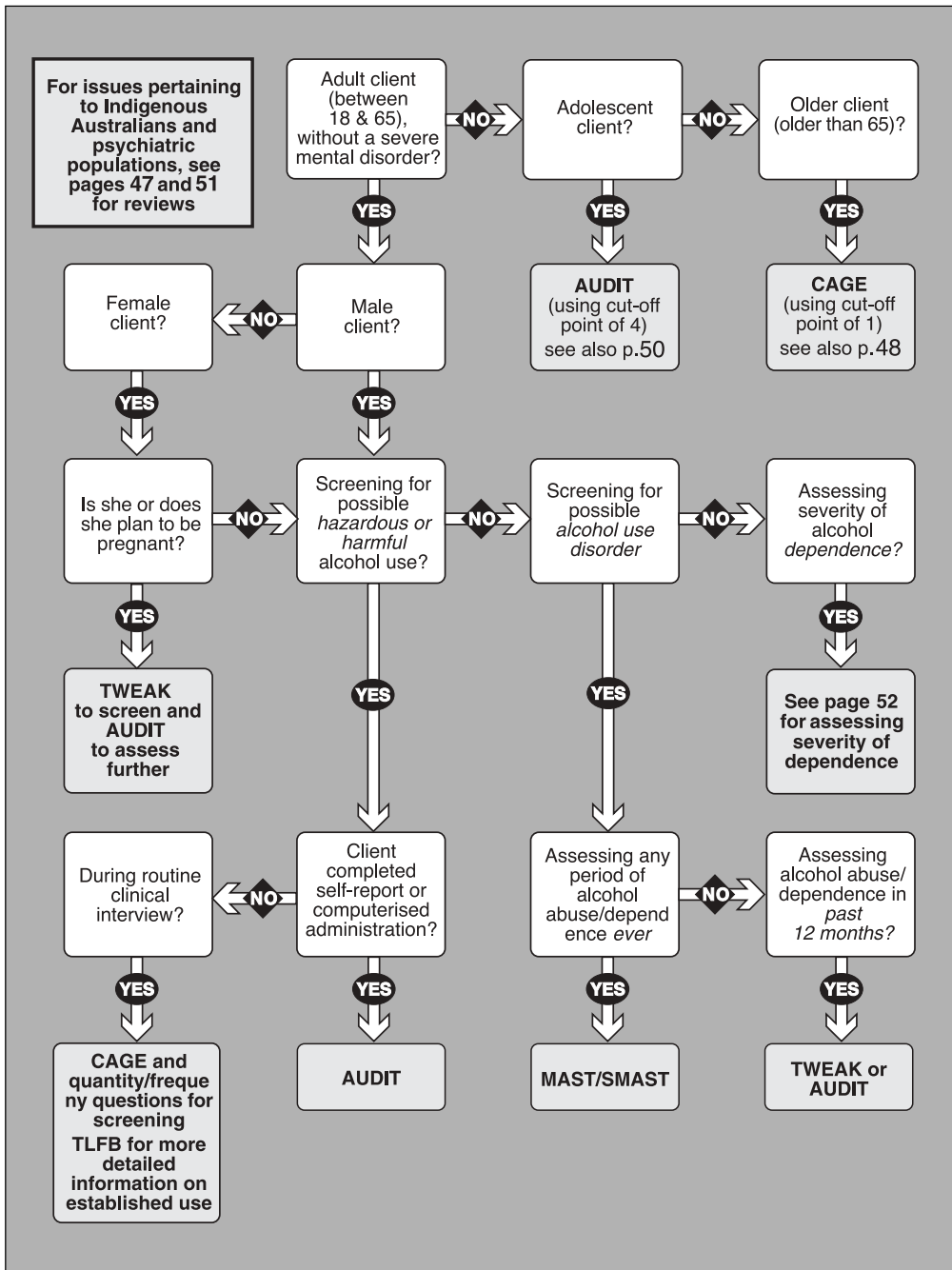
Overall the MAST, while useful for detecting a past history of alcohol abuse/dependence, tends to be less sensitive to current alcohol problems in the general population. Furthermore, its utility is hampered by its length (24-items vs 10-items for the AUDIT) and lack of items assessing consumption and episodes of binge-drinking, which are typical of younger, non-dependent problem drinkers. Indeed, studies have consistently found instruments such as the AUDIT and the TWEAK to outperform the MAST and the CAGE (e.g., Bradley, Bush, McDonell, Malone and Fihn, 1998; Cherpitel, 1995).

The T-ACE and the TWEAK are two new measures that incorporate items from the CAGE and the MAST, in addition to a question on tolerance. Originally developed to detect very low-level hazardous drinking in pregnant women, the TWEAK in particular has been found to be a highly sensitive and quickly administered measure of risky drinking in this population. Despite its development for pregnant women, there is an increasing body of evidence showing the TWEAK to be a useful screening device for the presence of alcohol use disorders in men and non-pregnant women. Indeed the TWEAK has been found to be

more sensitive in men than in women (e.g., Cherpitel, 1999). The utility of the TWEAK and the AUDIT in assessing current alcohol use disorders appear comparable, although in a review of the evidence, Bradley, Boyd Wickizer, Powell, & Burman (1998) propose that the TWEAK may outperform the AUDIT in Caucasian women. Further, they recommend the use of the “hold” version of the TWEAK with a cut-off point of 2 for screening adult women for heavy drinking or an alcohol use disorder. Overall, it is suggested that the brevity of the TWEAK and its psychometric robustness across men and women make it useful for incorporation into a standard interview, while administration of the AUDIT will provide a clearer picture of overall drinking consumption, evidence of dependence and problems related to drinking.

For specific information on screening for alcohol problems in Indigenous Australians, elderly clients, adolescents and psychiatric patients, see the relevant sections following the reviews of screening measures.

Alcohol Screening Instruments – Which measure to use when?



Alcohol Use Disorders Identification Test (AUDIT)

Key reference: Saunders, J.B., Aasland, O.G., Babor, T.F., de le Fuente, J.R. and Grant, M. (1993). Development of the alcohol use disorders identification test (AUDIT). WHO collaborative project on early detection of persons with harmful alcohol consumption – II: *Addiction*, 88, 791-804

Relevant webpage: URL: http://www.who.int/substance_abuse/topic_assessment.htm

Summary

The AUDIT is a 10-item screening instrument developed by a WHO collaborative study conducted in six countries: Australia, Kenya, Bulgaria, Norway, Mexico and the USA. It is designed to screen for a range of drinking problems and in particular for hazardous and harmful consumption. It is particularly suitable for primary health care settings and has been used in a number of different countries and with diverse cultural groups. The psychometric properties of the AUDIT have been assessed across a range of populations, including university students, women, psychiatric patients, geriatric populations and the unemployed. A score of 8 is associated with harmful or hazardous drinking. As a general guide, a score of 13 or more is likely to indicate alcohol dependence. Overall, the AUDIT is a comprehensive brief screening device, providing information on hazardous, harmful use, abuse and dependence. Whilst a relatively new measure, it has already gained substantial credibility.

Description and development of the AUDIT

The AUDIT was developed as an instrument that (i) would identify individuals who were drinking alcohol at harmful or hazardous levels before they sustained alcohol-related harm or developed physical dependence and (ii) had cross-cultural applicability. In the initial development phase, an international analysis of the prevalence of hazardous and harmful drinking was conducted in clinic attendees from six culturally diverse countries. Information on the drinking patterns, medical history and present condition, alcohol-related problems, psychological reactions to alcohol and family history (among others) was obtained on 1888 individuals. Based on this study, the researchers concluded that there was sufficient uniformity in patterns of alcohol consumption among culturally diverse groups to warrant the development of a single standardised instrument (Saunders, Aasland, Amundsen, & Grant, 1993).

The final 10 items from the AUDIT were selected from a 150-item interview schedule. The basis of selection was determined both thorough statistical analysis and face validity. The questions were selected from four conceptual domains: alcohol consumption (items 1 – 3), drinking behaviour related to dependence (items 4 – 6), adverse psychological reactions (items 7 – 8) and alcohol-related problems (items 9 – 10). Later, *adverse psychological reactions* and *related-problems* were combined to form a single *alcohol related problems* domain (Karno, Granholm, & Lin, 2000). Accordingly, the AUDIT is thought to measure three aspects of alcohol misuse: consumption, dependence, and related-problems.

An optional supplement to the AUDIT is the Clinical Screening Instrument (CSI), which assesses physiological consequences of harmful alcohol use. The CSI consists of three

sections, trauma history, a clinical exam and blood test to assess levels of γ -GT (a biochemical marker of liver damage). While the CSI contains only 8 items, it is invasive and must be conducted by a trained health worker, thereby limiting its administration by non-medical personnel. Furthermore, it shows lower sensitivity in detecting hazardous alcohol use, alcohol abuse and dependence than the “core” AUDIT in most countries (although there is substantial variability in sensitivity across cultures; Babor, De La Fuente, Saunders, & Grant, 1992). Notwithstanding the above, the CSI is recommended over, or at least in conjunction with, the “core” AUDIT in settings in which a lack of confidentiality may compromise valid self-reporting (McPherson & Hersch, 2000). For the sake of clarity and brevity though, the remainder of this review refers to the 10-item “core” AUDIT only. Further information on the CSI can be found in the AUDIT manual.

Reliability and validity of the AUDIT

The psychometric properties of the AUDIT have been explored in a number of populations, including geriatric inpatients, rural and urban primary care and emergency room patients, the unemployed and college students (see Allen, Litten, Fertig, & Babor, 1997, for a comprehensive review). With the exception of geriatric populations, the AUDIT has been found to have good internal reliability across these populations, with Cronbach alphas ranging from .80 to .94. Surprisingly, little research has examined the temporal stability of the AUDIT. In one of the few studies conducted to date, Daepfen, Yersin, Landry, Pecoud and Decrey (2000), found the AUDIT, embedded within a larger general health questionnaire, to have good test-retest reliability ($r = .88$) over a 6-week period. More recently, Reid Hester and colleagues found the AUDIT to have excellent test-retest reliability over a one-week period in a pilot study for a computerised motivational interviewing program for problematic adult drinkers ($r = .96$; personal communication, May, 2001; also see URL: <http://www.behaviortherapy.com>) As the AUDIT assesses alcohol misuse over a 12 month period, further research is required to test temporal stability across longer time frames.

In the initial development study, a comparison of AUDIT scores and diagnoses based on a comprehensive structured interview, physical examination and laboratory findings, two cut-off points of 8 and 10 produced maximal sensitivity and specificity (Saunders, Aasland, Babor, De La Fuente, & Grant, 1993). Of those individuals who scored 8 or more on the AUDIT, 95-100% were classified in the hazardous alcohol consumption group; 93-100% were classified as having abnormal drinking behaviour; 100% were alcohol dependent. MacKenzie, Langa, & Brown (1996) recommend using cut-offs of ≥ 8 , and ≥ 10 for detecting hazardous and harmful alcohol consumption, respectively, and a cut-off of ≥ 19 for diagnosing abuse and dependence.

However, cut-off scores may need to be modified depending upon the characteristics of the client group (Cherpitel, 1995). For instance, the cut-off points for potentially hazardous consumption in the AUDIT do not differentiate between males and females. As with all of the alcohol screening instruments included in this report, the quantity questions on the AUDIT do not take into account that women sustain alcohol related damage at lower levels of consumption than do men (Smith, 1986). Therefore, it is possible to argue that a score of 8 for a woman would be associated with a greater risk of alcohol related physical harm. Indeed, Bradley, Boyd Wickizer, et al. (1998) suggest that lowering the recommended cut-off point from 8 to 4 may enhance the sensitivity of the AUDIT in adult women. However, no research appears to have validated this use of a lower cut-off.

Validating the use of various cut-off points for “hazardous” and “harmful” drinking is somewhat problematic, as operational definitions for these two concepts vary from study to study. Generally, researchers have typically defined this type of drinking in terms of medical consequences, psychological harm, social problems, work problems or experienced trauma, among others.

Although the AUDIT was developed to measure 3 domains – alcohol consumption, alcohol dependence, and alcohol-related consequences, and a single global score is typically considered representative of an individual’s overall drinking behaviour – three recent studies have revealed two distinct factors (Karno et al., 2000; Maisto, Conigliaro, McNeil, Kraemer, & Kelley, 2000; O’Hare & Sherrer, 1999). Using a large sample of primary care attendees, Maisto et al. (2000) found that AUDIT items loaded on 2 factors – “Alcohol Consumption” (items 1-3) and “Dependence/Consequences” (items 4-9), which accounted for 43% and 13% of the variance explained, respectively. A subsequent confirmatory factor analysis verified the 2-factor solution as having the best fit. O’Hare and Sherrer (1999) found almost identical results in a sample of college students. Likewise, using exploratory factor analysis and a smaller sample of psychiatric out-patients, Karno et al. (2000) found two similar factors described by Maisto et al. – “Consumption” (items 1-3) and “signs of problem drinking” (items 1-8 and item 10) - which accounted for 24% and 31.7% of the variance explained, respectively. Based on their findings, Maisto et al. suggest that the goals of the assessment should dictate whether to use a single global AUDIT score or to examine scores based on the multifactorial structure (e.g., to differentiate level of alcohol consumption from problems arising from such consumption). In a related vein, Karno et al. propose that clients with high alcohol consumption but with few alcohol -related problems might require different interventions from those clients who engage in high consumption and also report considerable problems related to that consumption. The AUDIT may be a useful instrument for identifying these two groups of clients and assist in treatment planning.

Predictive validity has been demonstrated in three studies in which AUDIT scores were used to predict alcohol-related physical disorders and social problems (Conigrave, Hall, & Saunders, 1995; Conigrave, Saunders, & Reznik, 1995), and the likelihood of remaining unemployed after a two-year period (Claussen & Aasland, 1993). Conigrave, Saunders, et al. for instance, found AUDIT scores predicted 41% of people who would experience trauma and 43% of those who would develop hypertension over a three-year period using the standard cut-off. The AUDIT score was a better predictor of subsequent alcohol-related medical and social problems than standard biochemical markers.

Studies examining the construct validity of the AUDIT have primarily focussed on the correlation of AUDIT scores to other measures of alcohol misuse. On the whole, the AUDIT has displayed convergent validity with moderate to high correlations observed with other self-report measures such as the MAST and the CAGE. Lower correlations are typically found with biochemical measures (Allen et al., 1997), however, as noted in a later section, biochemical markers are notably less reliable measures of low-level hazardous alcohol use. Indeed, Aertgeerts, Buntinx, Ansoms, & Fevery (2001) found the AUDIT to be vastly superior for detecting alcohol use and dependence than biochemical markers in a large sample of men and women attending a general practice. Generally, the AUDIT has been consistently found to be more sensitive to less severe cases of alcohol misuse than other measures, including the MAST and CAGE, thereby demonstrating its superior sensitivity to non-dependent problem drinking (e.g., Barry & Fleming, 1993; MacKenzie et al., 1996).

Suitability for special populations

The items for the AUDIT were derived from a cross-national data set and only those items that could be translated literally and idiomatically were included (Saunders, Aasland, Amundsen et al., 1993). The cultural diversity represented by the original six countries participating in the study, suggests that the AUDIT may be used with a range of cultures. However, further research trials may be warranted: for instance, the AUDIT performed rather poorly when used with a Nigerian population and DSM III-R diagnosis as the criterion standard, with a reported sensitivity of only .32 (Gureje et al., 1992). Although cultural differences may have influenced the sensitivity, the comparatively good sensitivity of the AUDIT with a Kenyan population argues against this as a major contributing factor (Saunders, Aasland, Babor et al., 1993). Indeed, Cherpitel (1998) found the AUDIT maintains its psychometric robustness across a variety of ethnic groups, primarily Caucasian, Hispanic, and African-American men and women. Overall, it is fair to conclude that research findings to date indicate the AUDIT is a useful screening instrument that accurately detects hazardous drinkers from a range of cultural groups, at least within the United States.

Recently, a modified version of the AUDIT has been developed for use with the Australian population. The AusAUDIT (Conigrave & Elvy, 1998; Degenhardt, Conigrave, Wutzke, & Saunders, 2001) retains the 10 items of the original AUDIT but makes some changes to detect lower, but still harmful drinking. Response options for question 2 (quantity of drinking), for instance, were changed from: (0) 1 or 2, (1) 3 or 4, (2) 5 or 6, (3) 7 to 9, (4) 10 or more, to: (0) 1, (1) 2, (2) 3 or 4, (3) 5 or 6, (4) 7 or more. Additionally, lower recommended cut-offs were suggested and some terms and phrases were altered slightly, such as the term “drinks” replaced with the term “standard drinks”. By making these changes the developers of the AusAUDIT hoped to enhance the detection of those drinking over recommended *safe* levels of drinking, but who did not evidence any harm from this use. A full version of the AusAUDIT can be found in Degenhardt et al. Initial validation of the AusAUDIT has been promising, although this version was found to show good sensitivity but somewhat limited specificity (Degenhardt et al., 2001). Using AusAUDIT recommended cut-offs of 6 for women and 7 for men, Degenhardt et al found the AusAUDIT to have a sensitivity of 93.6% and specificity of 57.6% for women, and a sensitivity of 100% and specificity of 28.6% for hazardous drinking (drinking above NH&MRC recommended limits). To improve specificity, Degenhardt et al recommend increasing the cut-offs to 7 for women (sens = .85; spec = .70) and to 10 for men (sens = .85; spec = .70). Using a cut-point of 7 for both men women, the AusAUDIT was somewhat less sensitive to detecting those participants who met ICD-10 diagnoses of dependence and/or harmful use, with sensitivities ranging from 85.7% to 87.2%. Again, specificities were low (ranging from 33.7% to 41.2%). The authors concluded that while the AusAUDIT appears to be a suitable screening measure for use in the Australian community, identified cases should be followed up with clinical assessments.

Despite the development of the AusAUDIT, very little research has been conducted regarding Indigenous Australians. The Addiction Research Institute of Victoria have developed a computerised multimedia presentation of the AUDIT within an Aboriginal cultural context. Initial feedback from community members has been positive (Greg Powell, personal communication) although the cultural relevance of several items still needs to be determined. See the section on screening instruments for Indigenous Australians below, for further details on the use of the AUDIT in this population.

As noted earlier, while the AUDIT appears equally appropriate for both men and women, a lower cut-off point may improve sensitivity in adult women. Compared with other screening instruments, it appears the AUDIT may be somewhat less sensitive to female alcohol abuse and dependence than the TWEAK (Bradley, Boyd Wickizer et al., 1998). However, the TWEAK does not provide a picture of the client's pattern of consumption. Therefore, a positive identification by the TWEAK may be supplemented by the AUDIT.

The AUDIT has been used with university student populations across the United States, Belgium and Australia (Aertgeerts, Buntinx, Bande Knops et al., 2000; Fleming, Barry, & MacDonald, 1991; Roche & Watt, 1999). In a sample of college students, O'Hare and Sherrer (1999) found the AUDIT to have good psychometric properties and to be a valid screening instrument for hazardous and harmful alcohol use in this population. Aertgeerts et al. (2000), however, found the AUDIT to have low sensitivity (48.1%) to any alcohol use disorder at a cut-off point of ≥ 9 in a sample of Belgian university students. A cut-off point of ≥ 6 improved the sensitivity of the instrument in this population (80.2%). Nevertheless, the sensitivity at both these cut-off points is far lower than that found in Fleming et al.'s (1991) sample of American college students (.94% at cut-off point of ≥ 8 & ≥ 7). There does not appear to have been any research into the optimal cut-off point for Australian students, although the only study to have used the AUDIT with Australian students (Roche & Watt, 1999) used a cut-off point of ≥ 7 to categorise hazardous drinking for women and the standard cut-off point of ≥ 8 for men.

While a useful instrument for younger persons, the utility of the AUDIT as a screening device for alcohol misuse in elderly populations in Australia and the UK is questionable, with the few studies to date finding the AUDIT to have unacceptably low sensitivity. See the section below on specific issues related to screening alcohol misuse in the elderly.

Administration and scoring

The AUDIT is designed as a self-report measure. It is scored by adding each of the 10 items. Items 1 to 8 are scored on a 0 – 4 scale, and items 9 and 10 are scored 0, 2, 4. A score of 8 or above has been frequently used to indicate the presence of alcohol problems. While studies have found this cut-off point to have adequate sensitivity and specificity for adult men, a lower cut-off point of 4 may be more useful for women and adolescents.

In addition to the pen and paper version, a computerised version has been developed and compared with the pen and paper version in a British alcohol day-treatment centre (Chan Pensley, 1999). This study found the computerised version to derive similar results to the paper equivalent and to be an acceptable alternative, even for those clients anxious about using computers. A noted advantage was the immediacy of feedback to the respondent regarding their level of drinking.

The AUDIT is easy to read and has been shown to be understood by individuals with a minimum reading level of seventh grade (Hays, Merz and Nicholas, 1995). Therefore, it would be suitable for people for whom English was a second language who were able to read a broadsheet newspaper (reading age of 9 years required). The pen and paper version takes between 2-5 minutes to complete (Claussen and Aasland, 1993). The AUDIT may be used by any health worker who requires a reliable and brief screening instrument to identify an individual with alcohol problems.

Availability and cost

The AUDIT is in the public domain and is reproduced below. It may be used without cost but with due acknowledgment of the source. A copy of the AUDIT and manual is available free of charge from http://www.who.int/substance_abuse/docs/audit2.pdf.

ALCOHOL USE DISORDERS IDENTIFICATION TEST SCREENING INSTRUMENT

Please circle the answer that is correct for you:

1. How often do you have a drink containing alcohol?

NEVER	MONTHLY OR LESS	2-4 TIMES A MONTH	2-3 TIMES A WEEK	4 OR MORE TIMES A WEEK
-------	--------------------	----------------------	---------------------	---------------------------

2. How many drinks containing alcohol do you have on a typical day when you are drinking?

1 OR 2	3 OR 4	5 OR 6	7 TO 9	10 OR MORE
--------	--------	--------	--------	------------

3. How often do you have six or more drinks on one occasion?

NEVER	LESS THAN MONTHLY	MONTHLY	WEEKLY	DAILY OR ALMOST DAILY
-------	-------------------	---------	--------	-----------------------

4. How often during the last year have you found that you were not able to stop drinking once you had started?

NEVER	LESS THAN MONTHLY	MONTHLY	WEEKLY	DAILY OR ALMOST DAILY
-------	-------------------	---------	--------	-----------------------

5. How often during the last year have you failed to do what was normally expected from you because of drinking?

NEVER	LESS THAN MONTHLY	MONTHLY	WEEKLY	DAILY OR ALMOST DAILY
-------	-------------------	---------	--------	-----------------------

6. How often during the last year have you needed a first drink in the morning to get yourself going after a heavy drinking session?

NEVER	LESS THAN MONTHLY	MONTHLY	WEEKLY	DAILY OR ALMOST DAILY
-------	-------------------	---------	--------	-----------------------

7. How often during the last year have you had a feeling of guilt or remorse after drinking?

NEVER	LESS THAN MONTHLY	MONTHLY	WEEKLY	DAILY OR ALMOST DAILY
-------	-------------------	---------	--------	-----------------------

8. How often during the last year have you been unable to remember what happened the night before because you had been drinking?

NEVER	LESS THAN MONTHLY	MONTHLY	WEEKLY	DAILY OR ALMOST DAILY
-------	-------------------	---------	--------	-----------------------

9. Have you or someone else been injured as a result of your drinking?

NO	YES, BUT NOT IN THE LAST YEAR	YES, DURING THE LAST YEAR
----	-------------------------------	---------------------------

10. Has a relative or friend or a doctor or other health worker, been concerned about your drinking or suggested you cut down?

NO	YES, BUT NOT IN THE LAST YEAR	YES, DURING THE LAST YEAR
----	-------------------------------	---------------------------

Michigan Alcoholism Screening Test (MAST)

Key reference: Selzer, M.L. (1971). The Michigan Alcoholism Screening Test: The quest for a new diagnostic instrument. *American Journal of Psychiatry*, 127, 1653-1658.

Summary

The MAST is a 24-item screening instrument designed to identify and assess alcohol abuse and dependence. The MAST has been widely used and early studies report that it was reliable and valid. Later studies question the extent to which it measures a single core concept. Despite this reservation, the MAST has been demonstrated to have adequate sensitivity and specificity with a cut-off score of 13 in identifying individuals meeting diagnostic criteria for alcohol abuse and dependence. Shortened 13-item and 10-items versions of the MAST, (SMAST and BMAST respectively), can reliably be used as self-administered screening instruments.

Description and development of the MAST

The Michigan Alcoholism Screening Test (MAST) was originally developed as a structured instrument able to detect alcoholism that could be administered by a range of clinicians (Selzer, 1971). The original 25-item MAST was developed by Selzer using a population of hospitalised alcoholics, persons convicted of drunk and disorderly behaviour or drink driving offences, and a control group; all subjects were male and interviews were conducted in Michigan. The original items included in the MAST were derived from empirical investigations and surveys of alcoholism (Selzer, 1971). In subsequent development, item 7 was deleted and the MAST was reduced to 24 items (Selzer, Vinokur, and van Rooijen, 1975). Despite the apparent lack of systematic selection of individual items, the MAST has been widely used and there is considerable information regarding its reliability and validity. Two shortened versions, the 13-item Short MAST (SMAST; Selzer et al., 1975) and the 10-item Brief MAST (BMAST; Pokorny, Miller, & Kaplan, 1972) can also be reliably used.

Reliability and validity

The first study to investigate the MAST's reliability was conducted by Selzer and colleagues (1975). The MAST appears to have high internal consistency with an alpha coefficient of .95 reported in the original validation study (Selzer et al., 1975). However, more recent reports cast doubt on these conclusions. Crook, Oei and Young (1994) found that a number of items were not highly correlated with the total MAST score. For example, item 17 ("Have you ever been told you have liver trouble? Cirrhosis?") was endorsed by only 36% of the sample.

Some factor analytic studies of the MAST suggest that it is measuring one single factor that may be labelled "alcoholism" (Zung, 1982). However, Crook et al. (1994) identified a three-factor solution labelled as "alcohol related disabilities", "help-seeking" and "recognition of problem". Based on these findings, Crook et al. argued that important information about drinking patterns and alcohol symptomatology is lost if only the total score is used. Instead, calculation of three subscales would provide additional information about the three domains identified by the factor analysis. Other studies suggest the MAST to measure both a general over-arching higher-order "alcohol dependence" factor, incorporating several lower-order symptoms, including denial, antisocial drinking, discord,

and vocational impairment (Parsons, WallBrown, & Myers, 1994; Saltstone, Halliwell, & Hayslip, 1994).

Although a slightly outdated view, some critics of self-report instruments suggest that their reliability is reduced by the reluctance of those with problem drinking to accurately report the extent and nature of their alcohol use. In the initial study, the association between social desirability, as measured by the Deny-Bad scale of the Crowne-Marlow Social Desirability Scale, and scores on the MAST indicated that although the two measures were correlated, the relationship was weak. Further analysis controlling for individual Deny-Bad scores found little change in MAST scores. The authors concluded that any tendency to deny undesirable characteristics does not materially affect the validity of the MAST as a screening instrument (Selzer et al., 1975).

ROC analyses indicate that the optimal cut-off point for the MAST is 13 for detecting the presence of DSM-III alcohol abuse and dependence. At this score the MAST has a sensitivity of .91 and a specificity of .76 for a DSM III diagnosis of alcohol dependence (Ross, Gavin and Skinner, 1990). Selzer (1971) suggested a lower cut-off point of 5 to identify harmful or hazardous drinking. However, the MAST appears to be best used as a measure of lifetime alcohol abuse and dependence in clinical populations, rather than as a screening devise of lower levels of hazardous and harmful use in the general population. When assessment of hazardous drinking is desired, the AUDIT should be used instead.

One major criticism of the MAST is that it does not discriminate between past and present drinking. As such, the MAST can be somewhat problematic in over-detecting individuals with a past alcohol problem but who no longer engage in problematic alcohol use. Generally, the MAST tends to be a better measure of detecting any alcohol misuse across a person's lifetime. If the assessment's purpose is to establish current drinking problems, the AUDIT or the TWEAK should be used. Further, as with the CAGE, the MAST does not assess an individual's current frequency and quantity of alcohol consumption. Again, if quantity and frequency information is required, the AUDIT assesses this information in a brief form.

Suitability for special populations

The MAST is an American instrument and includes many phrases and words that are not part of Australian vernacular. However, the findings reported by Crook and colleagues (1994) indicate that Australian problem drinkers understand the MAST. The items from the MAST do however reflect the experience and worldview of the North American male problem drinker and this may reduce sensitivity in detecting problem drinking in Australian women. The term "partner" or "spouse" should be used instead of "girlfriend" in Item 12. Other items pose greater problems. Several of the behavioural consequences listed in the MAST are more applicable to men and probably occur infrequently with women problem drinkers. For example, item 9 refers to physical fights and item 24 refers to drunken behaviour in a public place; items 13 and 14 presuppose paid employment. Conversely, there are no questions relating to children although item 15 does include "obligations" and "family". Cherpitel (1998) found the BMAST, along with other alcohol screening instruments, to be relatively less sensitive to detecting alcohol problems in female drinkers compared with men. Using a cut-off point of 6 the BMAST had a sensitivity of 65% and specificity of 94% for alcohol dependence and a sensitivity of 42% and specificity of 97% for any type of alcohol misuse, in a sample of female drinkers. The low sensitivity percentages suggest that the BMAST fails to detect a large proportion of women with

drinking problems. Other measures, such as the AUDIT and the TWEAK appear to be more sensitive screening instruments in the female population. Likewise, the validity of the BMAST varies across ethnic groups, being less sensitive to non-Caucasian American populations (Cherpitel, 1998). No studies to date, though, have examined the validity of the MAST or BMAST with Indigenous Australians.

The MAST has been found to lack sensitivity in older populations. As such, a version specifically for assessing older individuals with alcohol problems, the 24-item MAST-G, was developed (Blow et al., 1992; see section below on screening instruments for elderly populations).

Administration and scoring

The MAST is a self-report measure that takes about 10 minutes to complete. A “YES” answer on items 3, 5, 9 and 16 are scored as 1; a “YES” answer on items 1,2,4,6,7,10 – 15, 17,18, 21-24 are scored as a 2; items 8,19 and 20 are scored as 5. The total score is 53.

The MAST may be used by any health worker who requires a reliable and brief screening instrument to identify individuals with alcohol dependence. However, for screening of lower levels of current alcohol misuse, the AUDIT is probably the instrument of first choice as it is shorter, takes less time to administer and is more sensitive to non-dependent, hazardous drinking.

Availability and cost

The MAST is in the public domain and is reproduced below. It may be used without cost but with due acknowledgment of the source.

MICHIGAN ALCOHOL SCREENING TEST

1. Do you feel you are a normal drinker? (By normal we mean you drink less than or as much as most other people)	TRUE	FALSE
2. Have you ever wakened the morning after drinking the night before and found that you could not remember a part of the evening?	TRUE	FALSE
3. Does your wife, husband, a parent, or other near relative ever worry or complain about your drinking?	TRUE	FALSE
4. Can you stop drinking without a struggle after one or two drinks?	TRUE	FALSE
5. Do you ever feel guilty about your drinking?	TRUE	FALSE
6. Do friends or relatives think you are a normal drinker?	TRUE	FALSE
7. Are you able to stop drinking when you want to?	TRUE	FALSE
8. Have you ever attended a meeting of Alcoholics Anonymous?	TRUE	FALSE
9. Have you ever gotten into physical fights when drinking?	TRUE	FALSE
10. Has drinking ever created a problem between you and your wife, husband, a parent, or other near relative?	TRUE	FALSE
11. Has your wife, husband, a parent, or other near relative ever gone to anyone for help about your drinking?	TRUE	FALSE
12. Have you ever lost friends or a girlfriend because of your drinking?	TRUE	FALSE
13. Have you ever gotten into trouble at work because of your drinking?	TRUE	FALSE
14. Have you ever lost a job because of drinking?	TRUE	FALSE
15. Have you ever neglected your obligations, your family, or your work for two or more days in a row because you were drinking?	TRUE	FALSE
16. Do you drink before noon fairly often?	TRUE	FALSE
17. Have you ever been told you have liver trouble? Cirrhosis?	TRUE	FALSE
18. After heavy drinking have you ever had delirium tremens (DTs) or severe shaking, or heard voices or seen things that weren't really	TRUE	FALSE
19. Have you ever gone to anyone for help about your drinking?	TRUE	FALSE
20. Have you ever been in a hospital because of your drinking?	TRUE	FALSE
21. Have you ever been a patient in a psychiatric hospital or on a psychiatric ward of a general hospital where drinking was part of the problem that resulted in hospitalisation?	TRUE	FALSE
22. Have you ever been seen at a psychiatric or mental health clinic or gone to any doctor, social worker, or clergyman for help with any emotional problem, where drinking was part of the problem?	TRUE	FALSE
23. Have you ever been arrested for drunken driving, driving while intoxicated, or driving under the influence of alcoholic beverages?	TRUE	FALSE
24. Have you ever been arrested, even for a few hours, because of other drunken behaviour?	TRUE	FALSE

CAGE

Key reference: Ewing, J.A. (1984). Detecting alcoholism: The CAGE questionnaire. *JAMA*, 252, 1905-1907.

Summary

The CAGE is a 4-item screening instrument designed to identify and assess potential alcohol abuse and dependence. It is not a diagnostic instrument; an affirmative answer to two or more questions indicates that further assessment of potential alcohol abuse is warranted. The CAGE is extremely short and easily administered taking less than a minute to complete. The CAGE is particularly useful for detecting alcohol problems in geriatric populations and can be easily incorporated into routine psychological or primary care assessments.

Description and development of the CAGE

The CAGE is a four-item screening questionnaire designed to identify problem drinking. Each letter reflects the core concept of each of the items: **C**utdown; **A**nnoyed; **G**uilty; **E**ye-Opener. The CAGE consists of the following four items:

1. Have you ever felt you ought to **cut down** on your drinking?
2. Have people **annoyed** you by criticising your drinking?
3. Have you ever felt bad or **guilty** about your drinking?
4. Have you ever had a drink first thing in the morning to steady your nerves or get rid of a hangover?

The CAGE questions were developed from a clinical study conducted in the late 1960's in North Carolina, USA. One hundred and thirty medical and surgical patients answered questions that had previously been demonstrated to be sensitive to detecting alcoholism. Subjects were classified into alcoholics and non-alcoholics and the minimum number of questions that would usefully divide the respondents into two groups was ascertained (see Ewing, 1984, for further information). The CAGE can be administered as self-report or may be used as part of a clinical or medical interview. The mode of administration appears to have no influence on the accuracy of the client's report (Aertgeerts, Buntinx, Fevery, & Ansoms, 2000).

Reliability and validity of the CAGE

The items from the CAGE have good internal reliability, with all four items correlating with each other indicating that the CAGE is measuring a single homogenous construct (Mischke & Venneri, 1987).

However, responses to the CAGE may not be stable over time. In the one longitudinal survey conducted, individuals did not answer the individual CAGE questions consistently over a period of seven years (Green and Whichelow, 1994). As items 1, 3 & 4 ask "Have you ever..." the authors argue that these responses should be stable over time. The data obtained suggest that these items are being interpreted to mean "recently". In line with this

finding Watson et al. (1995) found the CAGE to be a better identifier of alcohol dependence in the previous 12 months, while the MAST was the better identifier of lifetime alcohol problems. Therefore, even though the CAGE is considered a measure of lifetime alcohol abuse/dependence, the CAGE tends to more accurately screen for relatively recent problems.

It has been suggested that the increased awareness of the dangers associated with alcohol consumption will produce an increase in the number of non-problem drinkers answering YES to items 1 and 3. Waterson and Murray-Lyon (1988) have reported some data that supports this. They found that in a sample of individuals who would have had a raised awareness of alcohol consumption (i.e., expectant fathers), there was a higher positive response rate to these questions compared with a general population sample. This is an important issue in Australia as there is an active and growing public health campaign aimed at informing the general public about low risk drinking.

King (1986) reported that the CAGE performed well at detecting current at-risk drinking (defined as 8 or more standard drinks a day) in a UK sample, with a sensitivity of 84%, specificity of 95% and positive predictive value of 45% using a cut-off point of 2 or more affirmative responses. MacKenzie et al. (1996) examined the ROCs of the CAGE across varying cut-off points. They recommend using a cut-off point of 1 for detecting hazardous/harmful use and a cut-off point of 3 for identifying those likely to meet a DSM diagnosis. However, the difficulty with using a cut-off point of 1 for harmful use is that although sensitivity is increased, specificity is compromised, resulting in a marked increase in false positives and thereby rendering “the instrument impractical” (Watkins, Eisele, & Matthews, 2000, p. 597). Other researchers (e.g., Ewing, Bradley, & Burman, 1998) have suggested that one solution is to use the CAGE as part of any medical history taking if the respondent answers positively to “Do you ever drink alcohol?”. However, if the assessment’s purpose is screening for low-level hazardous/harmful drinking, and a face-to-face interview is not plausible, the AUDIT should be used.

Suitability for special populations

The CAGE has been widely used across the world and in a range of cultures. As with any questionnaire, careful translation and back translation is required to ensure that words such as “annoyed”, “bad” and “guilty”, for example, are accurately translated (see Indran, 1992 for a discussion regarding Malay, Chinese and Tamil versions of the CAGE). For instance, in a study of alcohol consumption among Indigenous Australians in the Kimberley region, researchers consulted with community members and changed the wording for each of the four items (Hunter, Hall and Spargo; 1991). The following questions were used in this study.

1. Do you sometimes think you shouldn't drink, or maybe drink less?
2. Do you feel angry or upset when other people get on your back about drinking, or tell you to cut down?
3. Do you ever feel shame or guilty about drinking?
4. Do you sometimes take a drink early in the morning for headache or because you feel no good, a reviver?

The CAGE has been used with Indigenous Australians in two separate studies. In the first, 106 homeless men in the Port Hedland area completed the CAGE (Skowrow and Smith, 1986). Those with a high score on the CAGE consumed significantly more alcohol both on the day before interview, and on a typical drinking day, drank more often. In a later study, scores on the CAGE were related to both quantity and frequency of alcohol intake (Hunter, et al., 1991).

The CAGE has been used extensively in hospital settings (e.g., Niles and McCrady, 1991), in population surveys (e.g., Smart, Adlaf & Knoke, 1991), and in primary health care settings (e.g., Nilssen and Cone, 1994; Chan, Pristach and Welte, 1994) in North America, England and the West Indies. There is evidence though, that the CAGE lacks the ability to discriminate between heavy drinkers and non-heavy drinkers in the general population (Bisson, Nadeau, & Demers, 1999) and is therefore recommended for use in clinical populations of already identified alcohol misusers rather than for screening individuals in the general population.

The CAGE has been found to perform poorly in younger populations (O’Hare & Tran, 1997). In one study using university students, for instance, the CAGE failed to identify over half the respondents who engaged in alcohol abuse or were dependent, as assessed by the CIDI, even at the lowest cut-off point of 1 (Aertgeerts, Buntinx, Bande Knops et al., 2000). One possible reason for this lack of sensitivity may be that the CAGE assesses aspects of problematic drinking that are irrelevant to young people, such as morning drinking, which is indicative of long-term alcohol dependence, rather than more recent problematic drinking (O’Hare & Tran, 1997). To increase the sensitivity of the CAGE in this population, Aertgeerts et al. suggested replacing the “Annoyed” item with “*Have you ever been under the influence of alcohol in a situation where it increased your chances of getting hurt, for example, when riding a bicycle or driving a car*”, referred to as the “Under the influence” item. The authors refer to this version of the CAGE as the CUGE. This instrument, though, requires cross validation before it can be recommended as the measure of choice in young people.

The utility of the CAGE for detecting alcohol problems in person over 65 is discussed in the section below regarding the use of screening instruments with elderly populations. Generally speaking, the CAGE is the best measure of problematic drinking behaviour in those over 65 years, when using a cut-off point of 1.

In terms of its utility in assessing women, like the MAST, the CAGE lacks sensitivity, especially with the usual cut-off point of 2 (Bradley, Boyd Wickizer et al., 1998; O’Hare & Tran, 1997). In particular, the CAGE shows less sensitivity for assessing dependence or harmful drinking in Caucasian women as compared with men and non-Caucasian women (Cherpitel, 1998). Due to this apparent insensitivity of the CAGE and the MAST with women, the T-ACE and the TWEAK were developed.

Administration and scoring

The CAGE is easily administered and takes less than one minute to complete. It is scored by adding the number of YES answers. A score of 2 or more should be taken as an indication that the client may be drinking at harmful or hazardous levels and that further assessment or referral is warranted.

The CAGE may be used by any health worker who requires a reliable and brief screening instrument to identify individuals with an alcohol problem. However, given the relative

insensitivity of the CAGE to lower levels of problematic drinking in the general population and that only an additional few minutes are required to complete the AUDIT, this is recommended as the screening instrument of first choice. The CAGE, though, is a very useful adjunct to an already planned clinical interview.

Availability and cost

The CAGE is in the public domain and is reproduced above. It may be used without cost but with due acknowledgment of the source.

T-ACE and the TWEAK

Key references: T-ACE: Sokol, R.J., Martier, S.S., & Ager, J.W. (1989). The T-ACE questions: Practical prenatal detection of risk-drinking. *American Journal of Obstetrics and Gynecology*, 160, 863-868.

TWEAK: Russell, M. (1994). New assessment tools for risk drinking during pregnancy: T-ACE, TWEAK, and others. *Alcohol Health and Research World*, 18, 55-61.

Summary

The T-ACE and the TWEAK are two instruments that were initially developed to specifically identify at-risk drinking in pregnant women. However, further research has suggested the utility of the T-ACE and the TWEAK for assessing alcohol problems in men and non-pregnant women. Both measures are easily administered and scored. Studies have consistently found them to be superior instruments to either the CAGE or the MAST in samples attending prenatal clinics, although the TWEAK appears to be the more frequently used and validated instrument of the two. A cut-off score of 2 or more on either measure is used to indicate that at-risk drinking may be present in pregnant and non-pregnant women. Recent studies have found the TWEAK to perform well for screening for alcohol problems in both men (using a cut-off point of 3) and women in the general population (using a cut-off point of 2).

Description and development of the T-ACE

It has been consistently argued that the MAST may be less sensitive to the presence of alcohol problems in female drinkers, given its focus on problem areas associated with male activities and lifestyle. Similarly, although the CAGE does not appear to have questions that would make it less sensitive with women, Waterson and Murray-Lyon (1988) found that it was less sensitive in detecting problem drinking in prenatal clinics than in psychiatric settings. Chang (1997) argues that limitations inherent in the MAST and CAGE such as the lack of items assessing quantity and frequency of consumption, episodes of binge drinking and the focus on symptoms of dependence rather than symptoms of hazardous drinking, may limit the usefulness of these measures for women. Concern was also expressed that women generally, and pregnant women specifically may have greater pressure to minimise their alcohol use. As even a low level of alcohol consumption may be harmful to unborn foetuses, the detection of non-abusive but still risky alcohol consumption in pregnant women is of the utmost importance. To address the apparent lack of an appropriate screening device for pregnant women, the T-ACE and the TWEAK were developed.

Subsequent research into the utility of these two instruments, especially the TWEAK, has found them to be equally useful with non-pregnant women and with men.

The initial study on which the T-ACE is based was conducted with pregnant women in a prenatal clinic in Detroit. In addition to the administration of the MAST and the CAGE a “tolerance” question was included: “*How many drinks does it take to make you feel **high**?*” (Sokol et al., 1989). The T-ACE consists of the Tolerance question and three CAGE questions:

- T** How many drinks does it take to make you feel high?(**Tolerance**)
- A** Have people **annoyed** you by criticising your drinking?
- C** Have you ever felt you ought to **cut** down on your drinking?
- E** Have you ever had a drink first thing in the morning to steady your nerves or get rid of a hangover?

Reliability and validity of the T-ACE

Recent comparisons have revealed that the T-ACE shows consistently higher sensitivity and specificity than the MAST or CAGE in pregnant women (Chang et al., 1998; Russell et al., 1996). For instance, using a cut-off point of ≥ 1 . The T-ACE had a sensitivity of 76% in predicting risky drinking during pregnancy compared with 59% and 76% for the CAGE and the MAST, respectively. Specificities for the T-ACE, CAGE and MAST were 79%, 82% and 76% (Sokol et al., 1989). More recent studies using the T-ACE have found a cut-off point of ≥ 2 to maximally discriminate non-risky drinkers from risk drinkers in a sample of pregnant African-American women (sensitivity = 88%; specificity = 79%).

The Tolerance question was later changed to “*How many drinks can you **hold**?*” This question is scored positive (i.e., receives 2 points) if women report being able to consume more than five drinks before falling asleep or passing out. This increased the sensitivity and specificity of the T-ACE to 91% and 81% (Russell, Martier, Sokol & Mudar, 1994).

Compared with the AUDIT, the T-ACE has been found to be slightly less sensitive to prenatal alcohol consumption (AUC = .70 and .65, respectively) in a sample of pregnant women. Sensitivity of the T-ACE, though, was significantly enhanced by the addition of clinical predictors, such as being >30 years of age, craving alcohol during the previous week, routine care and early recognition of pregnancy. This additional information did not enhance the sensitivity of the AUDIT (Chang, Goetz, Wilkins Haug, & Berman, 1999).

Description and development of the TWEAK

An alternative self-administered form of the MAST was developed for use in a female population (Russell and Bigler, 1979). Questions were eliminated if they pertained to behaviour that was more typical of men e.g., fights, or identified individuals whose problems with alcohol had already been recognised. Russell and Skinner (1988) found three questions that identified 70% of women reporting two or more indications of problem drinking. The three questions covered blackouts, feeling the need to cut down on drinking and having close friends or relatives worry or complain about drinking in the last year. Essentially the TWEAK contains 2 questions from the CAGE (Cutdown and Eyeopener), 2 from the MAST (Amnnesia and Worried) and a Tolerance item. As with the T-ACE, the Tolerance question appears to be a sensitive aspect indicative of female problem drinking.

Reliability and validity of the TWEAK

The TWEAK has been compared with the T-ACE, CAGE, MAST and AUDIT. When the “How many drinks does it take to get you **high?**” version of the Tolerance question was included, sensitivity scores for the TWEAK and T-ACE were 79% and 70% and specificity scores were 83 and 85 percent (Russell, 1994). By way of comparison sensitivity scores for the CAGE and the MAST were only 49% and specificity scores were 93% and 95%, respectively. Like the T-ACE, using the “hold” version of the Tolerance question, the TWEAK sensitivity increased from 79% to 91% using a cut-off point of 2 (Russell, 1994). Based on this study, Russell concluded that the TWEAK appeared to be somewhat more sensitive and less specific than the T-ACE but with both clearly outperforming the MAST and the CAGE test in screening for risk drinking during pregnancy. The superiority of the TWEAK for screening risky drinking in pregnant women has been found across a range of socioeconomically and ethnically diverse populations in the United States (Chang, Wilkins Haug, Berman, & Goetz, 1999; Russell et al., 1996). Chang et al. (1999), for instance, assessing the validity of the TWEAK in a socioeconomically diverse sample of pregnant women and using the TWEAK as a stand-alone instrument, found high sensitivity for detecting lifetime alcohol use disorders, risky drinking and current alcohol consumption (ranging from 84.1% - 92.3%), but markedly lower specificity (ranging from 25.0% - 28.9%) using a cut-off point of ≥ 2 . The decision regarding which cut-off point to use depends on the purpose for the assessment. However, since there is no completely “safe” level of pre-natal drinking, it seems appropriate to use a cut-off point with high sensitivity, even if this results in a high proportion of false negatives.

Although the TWEAK was developed with pregnant women, questions have been raised regarding its validity for use with men and non-pregnant women. Some studies of emergency room and primary care patients indicate the TWEAK performs well for screening alcohol abuse and dependence in women in the general population, outperforming the CAGE and the MAST (Bradley, Boyd Wickizer et al., 1998; Cherpitel, 1997). Other studies, though (e.g., Cherpitel, 1999), have found the CAGE to be more sensitive to alcohol use disorders in Caucasian and Hispanic, but not African American women, than the TWEAK. There is some evidence that the TWEAK outperforms the AUDIT in detecting female alcohol abuse and dependence (Bradley, Boyd Wickizer et al., 1998). Interestingly, even though the TWEAK was developed to detect high-risk drinking in pregnant women, it has been found to be more sensitive to detecting current alcohol dependence in men in the general public than the CAGE (Cherpitel, 1999). Overall, it appears that the TWEAK is a sensitive instrument for detecting alcohol problems in both pregnant women and in the general population (Cherpitel, 1998). However, like the CAGE and the MAST, the TWEAK does not provide a picture of the client’s pattern of consumption. Therefore, a positive identification by the TWEAK may be supplemented by the AUDIT.

Chan et al. (1993) have found a cut-off point of three to show optimal sensitivity and specificity for assessing heavy drinking and alcohol dependence in a combined sample of men and women. Cherpitel (1998; 1999) found this cut-off point to be a reasonable cut-off for men (sensitivity ranging from 76%-91% and specificity ranging from 74%-86%), but to result in markedly lower sensitivity for women (sensitivity ranging from 57%-80%) in clinical and general populations. Consequently, a lower cut-off point of two has been suggested for women. Indeed, this cut-off point has been found to have optimal sensitivity and specificity for detecting alcohol problems in women with sensitivities ranging from 89% - 91% and specificity ranging from 77%- 87% (Cherpitel, 1995; Russell et al., 1996).

Suitability for special populations

These questionnaires still need to be evaluated in Australia among a range of cultural groups to ensure that initial US findings are generalisable across cultures and class in this country.

Administration and scoring

The T-ACE is easily administered, takes less than one minute to complete and is scored by adding the number of YES answers. A score of 2 or more should be taken as an indication that the client may be drinking at harmful or hazardous levels and that further assessment or referral is warranted.

The TWEAK is scored slightly differently and care should be taken in scoring this instrument. Item 1 scores 2 points if a woman reports she can **hold** more than five drinks, a YES response to the Worry question scores 2 points and a YES response to the last three questions scores 1 point each. Therefore the maximum score obtained is 7. A total score of ≥ 3 for men and ≥ 2 for women is suggestive of alcohol problems.

Availability and cost

The T-ACE is a copyrighted instrument and permission to publish it can be acquired by contacting: S. Martier, Ob/Gyn, 4707 Saint Antoine, Detroit, MI, 48201, USA or Robert Sokol MC, Director, Mott Center, 275 East Hancock, Detroit, MI 48201. The TWEAK is in the public domain and reproduced below. It may be used without cost but with due acknowledgment of the source. Copies of the TWEAK and scoring instructions are available from: Marcia Russell, Ph.D., Research Institute on Addictions, 1021 Main Street, Buffalo, NY 14203, USA. 716-887-2507.

TWEAK Test

- T** How many drinks can you hold?
- W** Have close friends or relatives worried or complained about your drinking in the last year?
- E** Do you sometimes take a drink in the morning when you first get up?
- A** Has a friend or family member ever told you about things you said or did while you were drinking that you could not remember?
- K©** Do you sometimes feel the need to cut down on your drinking?

Key. **T** : Tolerance; **W**: Worried; **E**: Eye opener; **A**: Amnesia; **K©**: Cut down

Use of screening instruments with Indigenous Australians

Indigenous patterns of alcohol and drug use continue to be an area of concern for both indigenous and non-indigenous health care providers. Although there is a growing awareness of the need to use culturally appropriate measures within alcohol and drug services, there continues to be a reliance on traditional diagnostic and screening instruments with indigenous populations. The appropriateness of this use, however, is questionable, as there have been few investigations of the reliability and validity of screening instruments with this population. There appears to be a general acceptance in the field that “a brief albeit imperfect screen is better than nothing” (Commonwealth Department of Health and Aged Care, 1999, p. 25).

It is important to acknowledge that Australian indigenous people are a culturally diverse group. In remote areas some indigenous people have been able to maintain traditionally oriented lifestyles, while others living in towns or cities participate in more Westernised lifestyles. Such diversity has a bearing on the types of assessment strategies used within the different regions of Australia and the notion of developing uniform standardised assessment measures for indigenous populations throughout Australia appears implausible. Assessment instruments need to reflect the region’s specific needs and clinicians should seek to modify or change procedures as appropriate to the specific needs of their clients (personal communication, Corralie Ober, May, 2001).

Measures used to screen for alcohol problems: The Alcohol Use Disorder Identification Test (AUDIT).

Of the screening instruments used with indigenous clients, widespread use of the AUDIT has been reported. As the items for the AUDIT were derived from a cross national data set (Saunders et al., 1993), the cultural diversity represented by participating countries in the study suggest that the AUDIT is appropriate for a range of cultural groups. However, specific field trials with Australian indigenous populations have not been conducted and although use of the AUDIT is often regarded as ‘best practice’, the absence of controlled research signals the need for caution in both the delivery of the instrument and the interpretation of findings.

Feedback from the field indicates that practitioners often modify aspects of the standardised procedures with indigenous clients. Delivery of the AUDIT within an interview format is particularly common when there are concerns regarding client literacy levels and comprehension skills (personal communication, J. Parr, April, 2001). Establishing good rapport within a non-threatening, non-confrontational format is regarded as essential to this style of delivery.

In Victoria, the AUDIT has also delivered, using a computerised touch sensitive program that is incorporated into a general health screening assessment. The program features Koori faces, Koori voices, art and locally relevant material. In some cases the kiosk automatically prints out a ticket with the AUDIT scores and a contact phone number for further counselling if the score is high (Commonwealth Department of Health and Aged Care, 1999). Different regions in Australia have adapted the pictorial aides relating to standardised drinks to reflect those used within local indigenous communities (see example provided in Commonwealth Department of Health and Aged Care, 1999, p.90). The concept

of calculating individual alcohol consumption rates is considered problematic, particularly where alcohol consumption often takes place within a communal setting. In response, workers in the field have introduced creative ways of measuring individual consumption rates such as counting ‘number of swigs from the flagon’. As the cask has now almost replaced the flagon in the parks, another way is to measure the number of times the stubby or can has been refilled from the cask bladder (personal communication, Graham Lena, February, 2001).

Feedback from an exploratory study, which gathered information on indigenous interpretations of AUDIT questions, suggests the possibility of semantic differences in question interpretation. In particular Question 4 of the AUDIT was found to elicit particular confusion; “What does it mean? People stop drinking when they run out”. It is suggested that continuation of drinking is often dependent on supply rather than levels of dependency. Likewise question 7, which aims to determine if the individual feels sorry or guilty about drinking, may reflect general attitudes held by Indigenous people about any level of drinking *per se* rather than individual attitudes (Carol Atkinson, personal communication, February, 2001).

Practitioners working with indigenous clients with alcohol related issues in primary care settings are referred to the document National Recommendations for the Clinical Management of Alcohol Related Problems in Indigenous Primary Care Settings (1999) published by the Commonwealth Department of Health and Aged Care. This document aims to assist primary care practitioners with decisions about managing indigenous clients who present with alcohol related problems.

Use of screening instruments with older people

The prevalence of alcohol and other drug misuse is greatly underestimated in older persons, despite the more serious physical problems created by even a moderate amount of alcohol consumption in this age group. The underestimation of alcohol problems in this population may be due to both the perpetuation in the health care system of the “myth” that few people over the age of 45 abuse alcohol, and that symptoms of alcohol abuse often mimic or are masked by other age-specific conditions, such as dementia or delirium (DeHart & Hoffman, 1997). The early detection of alcohol problems is extremely important in this age group, since there are as many hospitalisations for alcohol-related conditions as there are for myocardial infarcts, in those over 65 years of age. Further, older alcohol misusers tend to respond very successfully to interventions (DeHart & Hoffman). Accordingly, brief screening for alcohol problems in older people should be incorporated within routine medical checkups and psychological intake interviews.

Despite the necessity of screening for potential substance problems in those over 65, researchers in the area have noted that diagnostic criteria for alcohol abuse and dependence, as defined by the DSM-IV and the ICD-10, are not representative of the signs and symptoms of disordered alcohol use in this age group. For instance, many symptoms of alcohol use disorders in the DSM-IV refer to vocational and interpersonal problems related with problematic alcohol use. Older people, though, are generally retired from employment and may have less contact with family and peers. Therefore, alcohol abuse may not affect these areas of psychosocial functioning and therefore remain hidden (DeHart & Hoffman, 1997). Alcohol abuse and dependence in older persons tends to be reflected in other more age-

specific consequences, such as increased accidents, inadequate nutrition, increasing social isolation, depression, liver disease, dementia, incontinence, seizures, hypertension, and interaction effects with other medication (Conigliaro, Kraemer, & McNeil, 2000). Based on these age-related differences, Gomberg (1990) proposes a different set of symptoms of alcohol abuse in older persons based on the above age-specific consequences of abuse.

In accord with these differences, alcohol-screening devices that are designed for and validated with much younger age groups, tend to lack sensitivity to alcohol misuse in the elderly. Indeed, reviews of the most commonly used measures (i.e., AUDIT and MAST), generally find lower sensitivity to alcohol abuse and dependence in older populations. The AUDIT in particular appears to be quite insensitive to alcohol misuse in older persons. Powell and McInness (1994), for instance, found the AUDIT had low sensitivity (57%) to alcohol abuse in a large sample of hospitalised Australian inpatients over 65 years of age. Similarly, Morton et al. (1996) found the AUDIT to have a sensitivity of 33% in assessing the presence of alcohol use disorder in American male war veterans over 65 years. The poor performance of the AUDIT in the elderly may be due to the emphasis on the actual consumption of alcohol, which appears to be less relevant to alcohol misuse and related problems in this age group (Conigliaro et al., 2000). Likewise, although not empirically tested, both the T-ACE and the TWEAK may not be useful with older persons, as Tolerance (a significant component of these instruments) tends to be a poor indicator of alcohol misuse in this age group (DeHart & Hoffman, 1997).

To create a measure specifically for screening older persons, Blow et al. (1992) developed the MAST-G. The MAST-G is a 24 item, dichotomously scored (yes/no) self-report questionnaire developed to take into account age-related differences in alcohol-related symptoms. Using a cut-off of ≥ 5 the MAST-G has been found to have good sensitivity (ranging from .70 - .95) and specificity (ranging from .65-.84) in Americans over 65 years (Fingerhood, 2000). However, the MAST-G had poor sensitivity to excessive alcohol intake (.54) and alcohol dependence (.50) in British emergency room attendees over 65 years, leading the authors to suggest cultural differences. Also, despite the positive findings of the utility of the MAST-G, at least within American populations, the MAST-G is somewhat limited by its length and its necessity to be administered in a pen and paper format. A shortened 10-item version, the S-MAST-G, has yet to be validated (Conigliaro et al., 2000).

The CAGE, on the other hand, is brief and can be easily accommodated within a medical or psychological consultation. Generally, it appears to be as sensitive as the MAST-G in the United States, with sensitivity for alcohol problems and dependence ranging from .77 to .94 at a cut-off point of ≥ 1 for persons over 60 years (Conigliaro et al., 2000). Further, Buchsbaum, Buchanan, Welsh, Centor and Schnoll (1992) found the CAGE effectively discriminated between primary care patients over 60 with a history of alcohol misuse from those who did not. However, like the MAST-G, the CAGE showed extremely low sensitivity in British emergency room patients over 65 years (.15 and .13 for excessive alcohol intake and dependence, respectively), suggesting that these American instruments may need to be further tested in non-American populations.

Notwithstanding these limitations, the CAGE and the MAST-G are currently the best instruments for screening alcohol misuse in elderly populations. Conigliaro et al. (2000) recommend the CAGE over the MAST-G, due to its brevity and ease of use during routine consultations, in addition to differentiating those with and without a history of problem drinking. Research using the CAGE with older persons suggests the use of a cut-off point of

≥1 (e.g., Buchsbaum et al., 1992). Additionally, Conigliaro et al. promote Gombert's (1990) position in indicating the need to add extra questions that tap episodes of binge-drinking and age-specific consequences to enhance the utility of the CAGE. The disparity in the psychometrics of the CAGE between the USA and the UK also suggests its utility within older Australian populations should be further investigated.

Use of screening instruments with adolescents

Alcohol misuse by Australian adolescents is a major public health concern, with approximately 40% of 14-19 year old Australian boys and approximately 30% of Australian adolescent girls engaging in hazardous levels of alcohol consumption on a regular basis (Australian Institute of Health and Welfare; AIHW, 2000; Loxton & Dawe, 2001). Furthermore, while adolescent drinking is widespread in the community, with 70.5% of 14-19 year olds drinking occasionally, and 30% drinking at least weekly (AIHW, 1999), those individuals who begin drinking at a very early age (particularly between 11 and 14 years) are at an extremely high risk of heavy drinking in adolescence and developing alcohol use disorders in adulthood (DeWit, Adlaf, Offord, & Ogbourne, 2000; Grant & Dawson, 1997). Fergusson, Horwood, & Lynskey (1995) for instance, found that the younger participants were when they first used alcohol the more likely they were to be heavy alcohol users at age 14, which in turn, was predictive of heavy, frequent and/or problematic alcohol use at age 16. Generally then, it appears that early adolescent alcohol use and misuse is both common and may lead to longer-term alcohol problems. Reliable screening instruments to detect potentially harmful use of alcohol in this age group are essential for detecting those at high risk for later alcohol problems.

Signs and symptoms of alcohol abuse vary across the life span. For instance, while adolescents typically drink less than adults, when they do drink, they tend to engage in episodes of binge-drinking and consequently, to suffer from more acute effects, such as hangovers and blackouts, rather than more chronic effects, such as liver disease, delirium, tremors and withdrawal. A number of studies have found alcohol-misusing adolescents typically drink over a longer period than intended, reduce non-alcohol related activities, express concerns over use and cravings, and experience social and academic problems due to alcohol misuse. By contrast, physiological dependence symptoms, such as those typically assessed by measures of alcohol dependence (e.g., withdrawal, alcohol-related medical problems) are rare, even in clinical samples (see Martin & Winters, 1998 for a review; White & Labouvie, 1989). It should be noted that increased tolerance has been demonstrated in alcohol abusing adolescents. However, Martin and Winters argue that tolerance generally does not distinguish between differing levels of alcohol-related problems in adolescents and as such may reflect a normal developmental path, thereby reducing specificity in identifying problematic adolescent drinkers. Accordingly, the best indices of adolescent alcohol misuse would appear to be repeated episodes of binge-drinking, loss of control, psychological craving and academic and interpersonal problems resulting from alcohol use. Generally, adolescents tend not to meet diagnostic criteria for abuse or dependence, but rather engage in problematic use that increases their risk of developing an alcohol problem. Therefore, the best measure for this age group is one that assesses low-level hazardous and harmful alcohol use.

The AUDIT, which assesses adolescent-specific aspects, such as episodes of binge drinking, level of alcohol consumption, and loss of control over drinking, appears to be a suitable

measure of hazardous alcohol use in this age group. Chung et al. (2000) recently assessed the utility of the AUDIT, TWEAK and CAGE in detecting DSM-IV diagnoses of alcohol abuse and dependence in a sample of 13-19 year-old American emergency room attendees. After slightly modifying some of the items to enhance participant understanding, they found the AUDIT to be the superior measure for this population (with AUCs for the AUDIT, TWEAK and CAGE at be .93, .86, and .75, respectively). The AUDIT was the more sensitive measure even when the lowest cut-off point (i.e., ≥ 1) for the TWEAK and the CAGE were used. Importantly, the authors noted that sensitivity of the AUDIT was greatly enhanced when the cut-off point was reduced from ≥ 8 (.55) to ≥ 4 (.94). Specificity was somewhat compromised by this reduction (.96 vs .80); however, the marked improvement in sensitivity justifies this loss, a reasonable trade-off in screening instruments. Although less sensitive than the AUDIT, the TWEAK showed adequate sensitivity (84%) and specificity (80%) for this population using a cut-off point of 1. As such, the TWEAK may be useful to incorporate within a general health consultation when the use of a pen and paper instrument may not be appropriate.

Two additional findings emerged from Chung et al. (2000). First, the CAGE was found to be of little use with young alcohol misusers, containing items that are generally irrelevant to teenage drinkers. Second, specific symptoms of problematic alcohol use in this age group were found to centre on alcohol-related amnesia (as tapped by both the TWEAK and the AUDIT) and episodes of high alcohol consumption (as addressed by the consumption items of the AUDIT). In light of this, instruments that tap psychological symptomatology of alcohol misuse, such as cravings and loss of control (i.e., the AUDIT & TWEAK) are more appropriate to the assessment of adolescents than those that index physiological dependence and alcohol-related problems (i.e., the MAST & CAGE).

Alcohol disorder screening in serious mental disorder

Alcohol abuse often occurs but is frequently under diagnosed in psychiatric populations (Ananth et al., 1989). This is unfortunate as problems with alcohol use can emerge at much lower levels of intake in people with serious mental disorder than in the general community due to poor or baseline functioning, susceptibility to symptomatic effects from alcohol, and limited resources (Drake, Osher, & Wallach, 1989). Screening instruments therefore need to be sensitive to problems that may emerge at quite low levels of intake and physical dependence. Therefore, a brief screening instrument that is sensitive to alcohol misuse but discriminates alcohol induced symptoms from non-alcohol induced symptoms (such as tactile delusions and disorientation) is vital.

A review of the literature has found the MAST to have greater sensitivity than comparative biochemical measures in correctly identifying alcohol-dependent patients with schizophrenia (classified according to DSM III criteria) but to have low specificity (many false positives; Toland & Moss, 1989). The MAST has been found to differentiate between non-alcoholic and alcoholic patients with schizophrenia with an overall detection rate of 80% (compared with 56% on the MacAndrew Alcoholism Scale; Searles, Alterman and Purtill, 1990). It has also been used successfully on an outpatient basis with women who were psychiatric patients (Swett, Cohen, Surrey & Compaine, 1991) and with clients undergoing methadone treatment programs (Stastny and Potter, 1991). A less positive

finding was reported in a recent study of inpatients with severe psychiatric disorders, with a sensitivity of 63% and specificity of 68% (Wolford et al., 1999). However, in a recent meta-analysis the MAST has shown a strong degree of sensitivity (average sensitivity = 87.7%) and a moderate degree of specificity (average of 68.1%) across nine MAST validity studies in psychiatric populations (Teitelbaum & Mullen, 2000). As in the general population, the MAST shows lower sensitivity to subclinical levels of alcohol misuse. Importantly, the meta-analysis found the psychometric properties of the MAST were unaffected by respondents' psychiatric condition. In other words, the MAST is as valid for clients with psychotic diagnoses as those with mood or anxiety disorders.

The CAGE is less effective than the MAST at detecting lifetime dependence in psychiatric populations (Breakey, Calabrese, Rosenblatt, & Crum, 1998; Watson et al., 1995). Wolford et al. (1999) report the CAGE as having a sensitivity of 61% and specificity of 69% in detecting alcohol-related disorders, while the TWEAK and T-ACE respectively had sensitivity of 58% and 47%, and a specificity of 85% and 87% in inpatients with serious mental disorders.

Seven items from several existing alcohol screening instruments were combined from the Dartmouth Assessment of Lifestyle Inventory (DALI; Rosenberg et al., 1998). The DALI was found to have a sensitivity of 80% and specificity of 85% in identifying alcohol use disorders. Further replication of these results is required before the scale is recommended for routine use.

Preliminary research in this area has shown the AUDIT to be an appropriate and valuable screening instrument in serious mental disorder (Hulse, Saunders, Roydhouse, Stockwell, & Basso, 2000). Using the standard cut-off of 8, Dawe, Seinen & Kavanagh (2000) found the AUDIT had a sensitivity of 87% and a specificity of 90% in detecting past 12 month CIDI-diagnosed alcohol disorders in a sample of Australian male psychiatric inpatients with a primary diagnosis of schizophrenia. A further study with young inpatients at the first to third psychotic episode found that the standard cut-off on the AUDIT had a sensitivity of 100% and specificity of 77% in detecting current alcohol disorder (Kavanagh et al., 1999). A US study on outpatients with severe and persistent psychiatric disorder found that the same cut-off produced a sensitivity of 90% and specificity of 70% (Maisto et al., 2000).

The AUDIT also maintained its psychometric robustness in a drug-dependent sample, with comparable sensitivity and specificity to the MAST in detecting alcohol use disorders and superior sensitivity in detecting hazardous alcohol use. Using the cut-off point of 8, the AUDIT had a sensitivity of 97% and a specificity of 69% for hazardous drinking, and a sensitivity and specificity for a diagnosable alcohol use disorder of 91% and 84%, respectively (Skipsey, Burlison, & Kranzler, 1997). Generally, the AUDIT appears to be the screening instrument of choice in psychiatric and drug dependent populations.

Measures to assess severity of alcohol dependence

While measures such as the AUDIT and the MAST are useful measures to *screen* individuals for possible alcohol use disorders, measures to assess the *severity* of dependence are valuable for treatment planning and assessing treatment outcome in those already identified as having an alcohol problem. While the MAST may be used for this purpose, the

lifetime timeframe may be insensitive to treatment-related changes. Self-report measures specifically designed to measure severity of dependence, such as the ADS, the SADD and the SADQ-C, are useful for classifying low and high dependence in planning treatment. For example, a brief outpatient treatment program may be best suited to an individual with low dependence, whereas an intensive, inpatient program may be best for a highly dependent person. Such measures are also useful for assessing potential severity of withdrawal symptoms and for determining treatment goals, such as controlled drinking or abstinence. Additional information for planning treatment can be provided by a structured interview, such as the ASI. For further information on the use of outcome measures in alcohol and drug treatment see Teesson, Clement, Copeland, Conroy, & Reid (2000).

Self-report measures of severity of alcohol dependence

Severity of Alcohol Dependence Questionnaire (SADQ-C)

Key reference: Stockwell, T., Sitharthan, T., McGrath, D. and Lang, E. (1994). The measurement of alcohol dependence and impaired control in community samples. *Addiction*, **89**, 167-174.

Summary

The Severity of Alcohol Dependence Questionnaire (SADQ-C) is a 20-item questionnaire designed to measure the severity of dependence on alcohol. It is divided into five subscales: physical withdrawal symptoms, affective withdrawal symptoms, craving and withdrawal relief drinking, consumption and reinstatement. It is a widely used measure of severity of dependence, particularly in Britain and Australia, and has demonstrated reliability and validity. It is relatively quick to complete (approximately 5 minutes) and is easy to score. It is probably most useful as an assessment tool for use with problem drinkers rather than a screening instrument. However, a shortened form has been used successfully with a non-clinic population of drinkers.

Description and development of the SADQ

The SADQ is a 20 item questionnaire based upon the premise formulated by Edwards and Gross (1976) that alcohol dependence comprises a cluster of symptoms derived from a single syndrome centred around a “drive” to consume alcohol. This “drive” is focused upon the need to drink to either avoid or to alleviate alcohol withdrawal symptoms (Stockwell et al., 1979). The original SADQ is divided into five sections corresponding to (i) physical withdrawal symptoms, (ii) affective symptoms of withdrawal, (iii) craving and withdrawal-relief drinking, (iv) typical daily consumption and (v) reinstatement of withdrawal symptoms after a period of abstinence. The more recent 16-item version, the SADQ-C, has an additional companion scale, the Impaired Control Scale (ICQ) complimenting the SADQ questions that focus on the physical and affective aspects of alcohol dependence. The ICQ assesses the extent to which clients perceive themselves to be out of control regarding their alcohol use.

The SADQ was developed at the Maudsley Hospital in London in the late 1970's. The original published study was based on a small sample of alcoholics who were admitted to the Maudsley Hospital for treatment (80 males, 24 females). In the later version, the SADQ-C, the sample size was extended to include 944 subjects from the general population and 197 subjects attending a clinic for controlled drinking.

The SADQ has been validated on in-patient, out-patient and community-based treatment samples and is useful for predicting withdrawal severity and the likelihood of achieving a moderate drinking outcome. The SADQ is also frequently used as an outcome measure for research purposes.

Reliability and validity

The SADQ is a widely used measure of the severity of alcohol dependence and has the most evidence of reliability and validity of all the major self-report questionnaires (Davidson, 1986). The original SADQ is considered a valid and reliable instrument that has shown high test-retest reliability when administered within an interval of two weeks on 45 inpatients in an Alcohol Treatment Unit (Stockwell, Murphy and Hodgson, 1983). Reid Hester and associates found the SADQ-C and the ICQ subsection to have very good test-retest reliability ($r = .89$ for both) in a sample of adult drinkers (personal communication, May, 2001; also see URL: <http://www.behaviortherapy.com>).

In the original published study (Stockwell et al., 1979) there were significant correlations between the sections measuring physical withdrawal symptoms, affective withdrawal symptoms, withdrawal relief drinking and craving, typical daily consumption and reinstatement. The SADQ was also sensitive to the degree of alcohol dependence assessed by retrospective analysis of case notes, with 82% concordance between the clinician's ratings of alcohol dependence and the scores of the SADQ. These findings were largely replicated by an independent team of researchers using a sample of Irish drinkers (Meehan, Webb and Unwin, 1985). Stockwell et al. (1983) suggest that a score of 30 should be taken to indicate severe dependence, a recommendation with which Meehan et al. (1985) concur. In a further study of construct validity, Stockwell et al. reported significant correlations between (i) scores on the SADQ and observed withdrawal severity and (ii) SADQ scores and narrowing of drinking repertoire assessed by the Drinking Pattern Interview.

Suitability for special populations

The SADQ has been used with a range of cultural groups (e.g., Ee Heok Kua et al., 1990), although to date, the SADQ has always been administered in English. The questionnaire requires a reasonable understanding of English, and if used with clients for whom English is a second language, close attention should be paid to ensure that they understand the questions.

It has been suggested that 30 is used as the cut-off point for severe dependence. However, it is arguable that given the contribution of the consumption questions to the total score, a lower cut-off score may be more appropriate for females. As women typically represent a small proportion of samples examined in research, further research regarding this point should be conducted.

A shortened version of the SADQ-C has been used in a general population sample of 1272 subjects (Stockwell et al., 1994). The modified and shortened SADQ-C appeared to be both

reliable and valid when compared with the original SADQ. Stockwell et al. (1994) suggest that this is an appropriate instrument to use when a measure of alcohol dependence is required in a general community sample.

Administration and scoring

The SADQ-C does not require specialised training and takes between 5 – 10 minutes to complete. Items 1, 3 and 4 of the ICQ are scored on a 4-point scale ranging from 0 (never or almost never) to 3 (nearly always). Items 2 and 5 are scored in reverse with a score of 0 (nearly always) to a score of 3 (never or almost never). The twenty items of the SADQ are all scored as follows: 0 = never or almost never, 1 = sometimes, 2 = often, 3 = nearly always.

Availability and cost

The SADQ-C is in the public domain and is reproduced below. It may be used without cost but with due acknowledgment of the source.

Self-report measures of severity of alcohol dependence

Severity of alcohol dependence questionnaire Form-C (SADQ-C)

SADQ-C

NAME: SEX: M / F DATE OF BIRTH: .../.../... AGE: ...

Have you drunk any alcohol in the past six months? YES / NO

If YES, please answer all the following questions by circling the most appropriate response.

(Section A – ICQ) DURING THE PAST SIX MONTHS:

1. After having just one or two drinks, I felt like having a few more.

NEVER or ALMOST NEVER	SOMETIMES	OFTEN	NEARLY ALWAYS
-----------------------	-----------	-------	---------------

2. After having two or three drinks, I could stop drinking if I had other things to do.

NEVER or ALMOST NEVER	SOMETIMES	OFTEN	NEARLY ALWAYS
-----------------------	-----------	-------	---------------

3. When I started drinking alcohol, I found it hard to stop until I was fairly drunk.

NEVER or ALMOST NEVER	SOMETIMES	OFTEN	NEARLY ALWAYS
-----------------------	-----------	-------	---------------

4. When I went drinking, I planned to have at least six drinks.

NEVER or ALMOST NEVER	SOMETIMES	OFTEN	NEARLY ALWAYS
-----------------------	-----------	-------	---------------

5. When I went drinking, I planned to have no more than two or three drinks.

NEVER or ALMOST NEVER	SOMETIMES	OFTEN	NEARLY ALWAYS
-----------------------	-----------	-------	---------------

(Section B – SADQ, Form-C) DURING THE PAST SIX MONTHS:

1. The day after drinking alcohol, I woke up feeling sweaty.

NEVER or ALMOST NEVER	SOMETIMES	OFTEN	NEARLY ALWAYS
-----------------------	-----------	-------	---------------

2. The day after drinking alcohol, my hands shook first thing in the morning.

NEVER or ALMOST NEVER	SOMETIMES	OFTEN	NEARLY ALWAYS
-----------------------	-----------	-------	---------------

3. The day after drinking alcohol, I woke up absolutely drenched in sweat.

NEVER or ALMOST NEVER	SOMETIMES	OFTEN	NEARLY ALWAYS
-----------------------	-----------	-------	---------------

4. The day after drinking alcohol, my whole body shook violently first thing in the morning if I don't have a drink.

NEVER or ALMOST NEVER	SOMETIMES	OFTEN	NEARLY ALWAYS
-----------------------	-----------	-------	---------------

5. The day after drinking alcohol, I dread waking up in the morning.

NEVER or ALMOST NEVER	SOMETIMES	OFTEN	NEARLY ALWAYS
-----------------------	-----------	-------	---------------

6. The day after drinking alcohol, I was frightened of meeting people first thing in the morning.

NEVER or ALMOST NEVER	SOMETIMES	OFTEN	NEARLY ALWAYS
-----------------------	-----------	-------	---------------

- | | | | | |
|---|-----------------------|-----------|-------|---------------|
| 7. The day after drinking alcohol, I felt at the edge of despair when I awoke. | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |
| 8. The day after drinking alcohol, I felt very frightened when I awoke. | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |
| 9. The day after drinking alcohol, I liked to have a morning drink. | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |
| 10. The day after drinking alcohol, in the morning I always gulped my first few alcoholic drinks down as quickly as possible. | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |
| 11. The day after drinking alcohol, I drank more alcohol in the morning to get rid of the shakes. | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |
| 12. The day after drinking alcohol, I had a very strong craving for an alcoholic drink when I awoke. | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |
| 13. I drank more than a quarter of a bottle of spirits in a day (or 1 bottle of wine or 7 middies of beer). | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |
| 14. I drank more than half a bottle of spirits in a day (or 2 bottles of wine or 15 middies of beer). | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |
| 15. I drank more than one bottle of spirits per day (or 4 bottles of wine or 30 middies of beer). | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |
| 16. I drank more than two bottles of spirits per day (or 8 bottles of wine or 30 middies of beer). | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |

(SECTION C – SADQ, Form-C) IMAGINE THE FOLLOWING SITUATION:

You have HARDLY DRUNK ANY ALCOHOL FOR A FEW WEEKS.

You then drink VERY HEAVILY for TWO DAYS.

How would you feel THE MORNING AFTER THOSE TWO DAYS OF HEAVY DRINKING?

- | | | | | |
|-------------------------------------|------------|----------|------------|-------------|
| 17. I would start to sweat. | NOT AT ALL | SLIGHTLY | MODERATELY | QUITE A LOT |
| 18. My hands would shake. | NOT AT ALL | SLIGHTLY | MODERATELY | QUITE A LOT |
| 19. My body would shake. | NOT AT ALL | SLIGHTLY | MODERATELY | QUITE A LOT |
| 20. I would be craving for a drink. | NOT AT ALL | SLIGHTLY | MODERATELY | QUITE A LOT |

Short Alcohol Dependence Data Questionnaire (SADD)

Key reference: Raistrick, D., Dunbar, G and Davidson, R. (1983). Development of a questionnaire to measure alcohol dependence, *British Journal of Addiction*, **78**, 89-95

Summary

The SADD is used to measure the severity of alcohol dependence. It has many similarities with the SADQ although is less focused on the experience of withdrawal symptoms and includes behavioural and subjective aspects of alcohol dependence. It has good test-retest reliability and construct validity, and correlates highly with the SADQ. The authors argue that it is relatively independent of socio-cultural influences and there is some independent evidence to support this.

Description and development of the SADD

Like the SADQ, the SADD is based upon the Edwards and Gross formulation of the Alcohol Dependence Syndrome (Davidson and Raistrick, 1986). It is a 15-item self-report questionnaire designed to provide a measure of the severity of dependence on alcohol, based upon a continuum of mild problem drinking to severe alcohol dependence. The major difference from the SADQ is the inclusion of items reflecting behavioural and subjective changes associated with problem drinking. Davidson and Raistrick (1986) suggest that the SADD is more sensitive to drinkers in the mild to moderate problem range than the SADQ because it includes cognitive and behavioural indices of problem drinking. Therefore, it has greater sensitivity in identifying those drinkers not yet experiencing alcohol withdrawal phenomena.

The present 15 item SADD was derived from a 39 item Alcohol Dependence Data (ADD) questionnaire given to three groups: regular drinkers (41), psychiatric patients (30) and patients admitted to an alcohol treatment unit (174). Items were included from the original 39-item version if: (i) most respondents endorsed “never” and fewest responded “always”; (ii) those items correlated significantly with overall score.

Reliability and validity

In the first report, Raistrick et al. (1983) showed acceptable internal consistency using split-half reliability (a correlation between total score on odd and even numbered questions). The SADD also showed good test-retest reliability ($r = .87$) over a 19-40 day time period in a sample of young British male offenders (McMurran & Hollins, 1989).

Davidson, Bunting and Raistrick (1989) demonstrated that the SADD measured the alcohol dependence syndrome. Their statistical analysis (confirmatory factor analysis) was consistent with the proposition that the SADD was measuring a single concept.

The construct validity of the SADD has been investigated in several studies, in which the SADD has been compared with a variety of measures related to the alcohol dependence syndrome. Davidson and Raistrick (1986) reported the results of three separate studies conducted using patients of the Leeds Addiction Unit. SADD scores were significantly correlated with (i) alcohol intake (most recent heavy drinking period and a problem checklist); (ii) SADQ scores and (iii) an interview-based assessment of alcohol dependence.

There was a high correlation between SADQ scores and SADD scores in a sample of 160 Irish problem drinkers, providing further evidence that the two questionnaires are measuring the same theoretical construct (Doherty and Webb, 1989).

Suitability for special populations

Raistrick et al. (1983) considered the SADD to be relatively free of socio-cultural influences, although it has not been widely used with other cultural groups. The applicability of the SADD was investigated in a Brazilian study (Jorge and Masur, 1985). Phase one involved administering the English version and the translated Portuguese version, two weeks apart, to bilingual university students with no history of problem drinking. Scores from the Brazilian translation of the SADD were highly correlated with the original English form. To investigate the usefulness of the SADD with an illiterate population, the authors administered the SADD in two ways: first in using the self-completion version on two occasions and second by using self-completion and an interview one week apart. The test-retest reliability was extremely high, as was the correlation between the self-completion and interview administration. Based on this finding, and bearing in mind the difficulties involved in the accurate translation of questionnaires, we suggest that the SADD may be used with a range of ethnic groups and cultures within Australia. Whether it is appropriately used with Indigenous Australians, requires further research.

Both SADQ and SADD are routinely administered to adolescent clients of the NSW Juvenile Justice System. Although the use of instruments has not been subjected to empirical investigation, the clinicians who administer them report that the SADD is more easily understood and completed by their client group than the SADQ (Jennifer Barton, personal communication, 1996). Likewise, McMurran & Hollins (1989) found changing difficult-to-read words for young offenders resulted in comparable reliability as the original wording. For example, item 6 was changed from “Do you drink as much as you want irrespective of what you are doing the next day?” to “Do you drink as much as you want without considering what you’ve got to do the next day?”. A full version of the modified SADD is available in McMurran & Hollins (1989). No specific studies appear to have tested the validity and reliability of the SADD with women.

Administration and scoring

The SADD takes less than 5 minutes to administer. Each item is scored as follows: never = 0; sometimes = 1; often = 3; nearly always = 4. A total score is obtained by adding the score from each of the items.

A score of 1 – 9 indicates low dependence, 10 – 19 medium dependence and a score of 20 or more high dependence.

Availability and cost

The SADD is in the public domain and is reproduced below. It may be used without cost but with due acknowledgment of the source.

Self-report measures of severity of alcohol dependence

SADD: The following questions cover a wide range of topics to do with drinking. Please read each question carefully but do not think too much about its exact meaning. Think about your **MOST RECENT** drinking habits and answer each question by placing a tick (✓) under the **MOST APPROPRIATE** heading. If you have any difficulties **ASK FOR HELP**.

	NEVER	SOME-TIMES	OFTEN	NEARLY ALWAYS
1. Do you find difficulty in getting the thought of drink out of your mind?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Is getting drunk more important than your next meal?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Do you plan your day around when and where you can drink?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Do you drink in the morning, afternoon and evening?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Do you drink for the effect of alcohol without caring what the drink is?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Do you drink as much as you want irrespective of what you are doing the next day?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Given that many problems might be caused by alcohol do you still drink too much?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Do you know that you won't be able to stop drinking once you start?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. Do you try to control your drinking by giving it up completely for days or weeks at a time?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. The morning after a heavy drinking session do you need your first drink to get yourself going?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. The morning after a heavy drinking session do you wake up with a definite shakiness of your hands?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. After a heavy drinking session do you wake up and retch or vomit?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. The morning after a heavy drinking session do you go out of your way to avoid people?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14. After a heavy drinking session do you see frightening things that later you realise were imaginary?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15. Do you go drinking and the next day find you have forgotten what happened the night before?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Alcohol Dependence Scale (ADS)

Key Reference: Skinner, H.A. and Horn, J.L. (1984). *Alcohol Dependence Scale (ADS): Users Guide*. Toronto: Addiction Research Foundation.

Summary

The Alcohol Dependence Scale (ADS) is designed to identify and assess alcohol abuse and dependence. Like the SADQ-C and SADD, it is based upon Edwards and Gross (1976) conceptualisation of the alcohol dependence syndrome. It is widely used and has demonstrated reliability and validity. It has been used with specific cultural groups, and across a range of samples, including homeless and incarcerated women and Australian university students.

Description and development of the Alcohol Dependence Scale

The ADS is a 25 item self-report questionnaire that is used to measure severity of alcohol dependence by asking about alcohol use in the past 12 months. It reflects four key aspects of the alcohol dependence syndrome: loss of behavioural control, psycho-perceptual withdrawal symptoms, psychophysical withdrawal symptoms and obsessive-compulsive drinking style. Like the SADQ and SADD reviewed previously, it is based on the concept of the alcohol dependence syndrome as described by Edwards and Gross (1976). The ADS is used both as a clinical and research instrument, in clinical and community samples.

The ADS was derived from the Alcohol Use Inventory (Wanberg, Horn, & Foster, 1977). In the original validation study, 225 subjects who sought treatment for alcohol problems from the Addiction Research Foundation, Toronto, Canada, were administered the 174-item Alcohol Use Inventory, the ADS and the MAST (among other instruments).

Reliability and validity of the ADS

Items forming the ADS were found to have good internal consistency ($\alpha = .92$). The scale consists of three factors: the first major factor accounted for items reflecting withdrawal symptoms, the second and third smaller factors were made up of items reflecting obsessive-compulsive drinking patterns and loss of behavioural control Skinner and Horn (1984). These findings were replicated by Kivlahan, Sher and Donovan (1989).

The ADS has good concurrent validity. Skinner and Horn (1984) reported that the ADS score was correlated with both daily consumption of alcohol and lifetime use of alcohol, social consequences from drinking, prior treatment for alcohol abuse, use of alcohol to change mood and feelings of guilt over drinking. The ADS was also significantly correlated with the MAST (Skinner and Horn, 1984; Ross et al., 1990). The ADS has been shown to have strong diagnostic validity regarding DSM diagnoses. In a study investigating the diagnostic validity of the MAST and ADS, Ross et al. (1990) reported that a cut-off score of 9 correctly identified 91% of patients (primarily male) with a current alcohol abuse or dependence disorder (sensitivity), it correctly identified 82% of non abusers (specificity). The overall accuracy of the ADS at this cut-off point was 89%. A more recent study investigating the validity of the ADS with a sample of homeless women found a cut-off point of 8 to be optimal for mild – moderate dependence and from 8-15 for severe dependence (Chantarujikapong, Smith, & Fox, 1997).

Suitability for special populations

The concept of the alcohol dependence syndrome and the use of the ADS was investigated in a sample of American and Russian inpatients and outpatients in alcohol treatment programs (Allen et al., 1994). The ADS had a similar factor structure for both samples leading the authors to suggest that the ADS may be a useful instrument in future cross cultural research. The ADS has also been successfully adapted for use with a sample of Tamil alcoholics in India. The translated version of the ADS was found to have high internal reliability (Rajendran and Cheridan, 1990). However, there have not been any reports of its use with Indigenous Australians.

The ADS has been validated with incarcerated male offenders, with ADS scores correlated with alcohol –related problems and post-release drinking goals (Hodgins & Lightfoot, 1989). Likewise, Peters et al., (2000) recently found the ADS, in combination with the Drug Use section of the ASI, to be one of the most effective screening instruments for detecting substance use disorders in a large sample of male prison inmates. Using a cut-off point of ≥ 14 on the ADS and a cut-off point of $\geq .11$ on the ASI – Drug section, the combined measure had a sensitivity of 74% and a specificity of 92% in detecting alcohol or drug dependence. Overall, 83% of the sample was correctly assigned.

The ADS appears to be an equally valid measure of alcohol dependence in women (Drake, McHugo, & Biesanz, 1995). Internal consistency is also high in this population (a ranging from .85 [dependent participants] to .99 [total sample]; Chantarujikapong et al., 1997). Several studies investigating alcohol dependence in homeless and incarcerated women, and studies of alcohol misuse in female university students have successfully used the ADS (e.g., Biron, Brochu, & Desjardins, 1995; Chantarujikapong et al., 1997; Williams, Connor, & Ricciardelli, 1998).

Administration and scoring

Scoring instructions are given in the manual. No special training or expertise is required to interpret the ADS. A computerised version is available.

Availability and cost

The ADS is copyrighted. It can be purchased from Marketing and Sales Services, Centre for Addiction and Mental Health, 33 Russell Street, Toronto, Ontario, Canada, M5S 2S1 (phone 416-595-6059), or @

URL: [http://www.camh.net/resources/clinical_tools.html#Alcohol Dependence Scale](http://www.camh.net/resources/clinical_tools.html#Alcohol%20Dependence%20Scale).

Structured clinical interviews for alcohol abuse and dependence

Addiction Severity Index (ASI)

Key References: McLennan, A.T., Luborsky, L., Woody, G.E., & O'Brien, C.P. (1980). An improved diagnostic instrument for substance abuse patients: The Addiction Severity Index. *Journal of Nervous and Mental Disease*, 168, 26-33.

McLellan, A.T., Kushner, H., Metzger, D., Peters, R., Smith, I., Grissom, G., Pettinati, H., & Argeriou, M. (1992). The fifth edition of the Addiction Severity Index: Historical critique and normative data. *Journal of Substance Abuse Treatment*, 9, 199-213.

Summary

The Addiction Severity Index is a 45min, 155-item multidimensional structured interview for assessing alcohol and drug dependence. The ASI assesses frequency of drug and alcohol use as well as other psychosocial areas affected by substance use. Although there have been mixed findings, the ASI has been found to have generally good reliability and validity as a measure of treatment outcome. An adolescent version (Teen-ASI) and a female-specific version (Psychosocial History) have been developed for these special populations. Use of the ASI with drug dependent clients with severe and persistent psychiatric disorders is not recommended.

Description and development of the Addiction Severity Index

The ASI is a 155-item multi-dimensional structured interview used to assess severity of dependence in both drug and alcohol dependent people. The ASI consists of 7 sub-scales assessing past 30 day and lifetime alcohol use, drug use, medical problems, psychiatric problems, family/social problems, employment and legal problems. The scoring of the ASI takes into account both subjective ratings of problems by clients, and objective tests of use (such as laboratory tests) across each scale to provide an overall severity rating. The ASI is a good measure for assessing the severity of consequences caused by drug and alcohol use and was developed for use with individuals seeking treatment, rather than for screening individuals in the general community.

The ASI was initially developed as an outcome measure to evaluate treatment across a 6-program treatment network (McLellan, Luborsky, O'Brien, & Woody, 1980). It was designed to be suitable for administration by non-health professionals, to be sensitive to treatment-related change, to be used in follow-up studies, and able to be generalised to suit a variety of treatment programs. Now in its fifth revision (McLellan et al., 1992), the ASI is recognised as the most widely used structured interview for alcohol and other drug abuse in the United States. The fifth edition included several new items to address changes in frequently used substances and accommodate advances in knowledge regarding factors influencing drug use, such as a family history of alcohol and drug use, and factors influencing poor treatment prognosis, such as anti-social behaviour.

The ASI differs from other measures of alcohol and drug use by assessing *frequency* of use without addressing *quantity* of use. The authors argue that an estimate of quantity is of little

utility, as a) quantity correlates with frequency, b) frequency is easier to recall than quantity, c) there is no standard measure such as “standard drinks” for other substances, and d) what respondents *believe* they have consumed, may not in fact be what they *actually* consumed (McLellan et al., 1992).

Reliability and validity of the ASI

Generally, the ASI has been found to have good test-retest and inter-rater reliability. Drake et al. (1995) found scale scores to have test-retest reliability coefficients of $\geq .60$ across a 7-day period in a sample of homeless people. Reliability increased when respondents were younger, female, had fewer psychiatric problems and items referred to more recent time frames. Likewise, Stoffelmayr, Mavis, & Kasim (1994) found the ASI to have high inter-rater reliability, as measured by intra-class correlation coefficients for composite scores (ranging from .83 – 1.00 across domains), but considerably lower reliability for severity scores (ranging from .40-.87 across domains), which remained stable over a two year period in a sample of clients seeking treatment for substance abuse. Internal consistency varied across scales, ranging from .65 for the employment problems scale to .89 for medical problems (Leonhard, Mulvey, Gastfriend, & Shwartz, 2000).

While there have been mixed findings across populations, the ASI has been shown to generally have good content, construct and criterion validity for most substance abusing populations. The items of the ASI appear to assess relevant aspects of alcohol and drug use and associated problems, indicating good content validity across populations, although several researchers have noted that some items may be irrelevant for specific populations, such as homeless people and psychiatric patients (e.g., Carey, Cocco, & Correia, 1997; Corse, Hirschinger, & Zanis, 1995). In particular, the ASI has consistently been found to perform poorly in psychiatric populations. Studies validating the ASI with drug dependent clients with severe and persistent psychiatric disorders have consistently found weak reliability and validity (e.g., Carey et al., 1997; Corse et al., 1995; Zanis, McLellan, & Corse, 1997). Accordingly, the use of the ASI in this population is not recommended.

As the ASI is used for assessing outcome rather than as a screening instrument or a diagnostic tool, the degree of sensitivity and specificity is difficult to ascertain. Nevertheless, Svikis et al. (1996), using a definition of heavy drinking as ≥ 3 drinks per occasion on ≥ 3 days/week over at least the previous 12 months on the ASI as the predictor, found the measure to be highly sensitive (.96) and specific (.94) to DSM-III-R diagnoses of alcohol abuse or dependence, among a sample of drug-abusing women.

The ASI's authors warn against its use as a self-administered instrument due to the high level of literacy required. Rosen, Henson, Finney and Moos (2000), however, recently developed a self-administered pen and paper version that appears to be a viable alternative to the interviewer-administered version. Using this version in a predominantly male in-patient substance abuse treatment centre, they found good internal consistency ($\alpha > .7$) for the alcohol, drug, psychiatric and medical problems scales, with somewhat lower consistency in the remaining scales. Composite scores on the alcohol and drug use scales were similar across the two formats ($r = .87$ and $.73$, respectively). In line with previous findings, respondents reported significantly higher drug and alcohol use on the self-report format as compared with the interviewer-based format. Similar patterns were found on the psychosocial scales. Further validation of this questionnaire is required in other populations.

Suitability for special populations

The ASI has been validated and is frequently used across a variety of substance abusing populations, including psychiatric patients, homeless people, pregnant women and incarcerated prisoners, and has been used to assess treatment outcome across a range of substances, including opiates, cocaine and alcohol (McLellan et al., 1992).

The ASI has been translated into nine languages and validated across a range of countries, notably the USA, France, Costa Rico, Germany and Holland. However, there have been no validation studies in Australia, nor has there been any reports of its use with Indigenous Australians. Nevertheless, Teesson, et al., (2000) argues that the ASI does meet minimum requirements for use in routine assessment and outcome measurement in Australia. It should be noted that a number of items on the ASI are American-specific, such as requesting information like social security numbers and the use of GED for educational level. Such items may need to be altered to take into account Australian-specific terminology and cultural norms.

While the ASI has been shown to be a reliable and valid measure in female drug abusing populations, Comfort, Zanis, Whiteley, Kelly Tyler and Kaltenbach (1999) note that the ASI fails to assess female-specific aspects of drug and alcohol dependence, such as pregnancy-related medical issues, care-giving responsibilities, family relationships and violence-related issues. Such issues are particularly relevant when making decisions regarding treatment planning and outcome goals. Accordingly, Comfort and Kaltenbach (1996) developed a female-specific version – the Psychosocial History (PSH), a 2-hour, 300-item structured interview that incorporates items to assess such female-specific issues. Preliminary validation research indicates the PSH to have adequate internal consistency, temporal stability, inter-rater agreement and concurrent validity in a small sample of substance abusing adult women (Comfort et al.). Indeed, the PSH appears to outperform the ASI, particularly in those specifically targeted areas, commending it as a measure of utility for substance abusing women. The measure, however, is considerably lengthier and more time-consuming than the ASI.

The ASI is not considered a suitable interview for adolescent clients as it pre-supposes personal independence and neglects adolescent-related issues, such as school and relationships with teenage peers, parents and siblings. An adolescent version, the Teen-ASI is available and initial studies show it to be a promising measure of adolescent drug abuse within hospitalised teenagers (Kaminer, Bukstein, & Tarter, 1991; Kaminer, Wagner, Plummer, & Seifer, 1993).

Administration and scoring

Scoring instructions for the original interviewer-administered format are included in the accompanying manual. Administration of the interview form is somewhat lengthy and takes between 30-60 minutes. Administration of the ASI does not require any specialist educational pre-requisites and may be administered by physicians, drug treatment personnel, research technicians and other interested persons who have been trained in its use.

A computer program was developed to assist with administration and scoring (McLellan et al., 1992). Recently a computerised multimedia CD-ROM has been developed (ASI-MV) and has shown promising psychometric characteristics, leading the authors to suggest that this form may be a viable alternative to potentially unreliable interviewer format (Butler et al., 2001).

Availability and cost

The ASI interview is in the public domain and can be used at no cost. A copy of the ASI can be found in Teesson et al. (2000). Alternatively, it is available for a minimal cost to cover photocopying and postage from DeltaMetrics/TRI ASI Information Line – 1- 800-238-2433 (USA) or via URL: <http://www.DeltaMetrics.com/>. Training videos are also available for approximately US\$500 from DeltaMetrics.

The computerised version is available from Biomedical Computer Research Institute, 9743 Redd Rambler Place, Philadelphia, PA 19115, USA, Tel:1-215-676-9743.

Copies of the self-administered questionnaire are available from Craig S. Rosen, Center for Health Care Evaluation, VA Palo Alto Health Care System (152-MPD), 795 Willow Road, Menlo Park, CA, 94025, USA. Tel: 1-650-493-5000 ext 22846, fax: 1-650-617-2736, or at crosen@stanford.edu.

The Teen-ASI is not under copyright but may not be reproduced without permission from the author, Yifrah Kaminer, Director, Bradley Substance Abuse Intervention Center (BASIC), Bradley Hospital, 1011 Veterans Memorial Parkway, East Providence, RI, 02915.

Comprehensive Drinker Profile

Key reference: Miller, W.R. and Marlatt, A. (1984). *Manual for the Comprehensive Drinker Profile, Brief Drinker Profile and Follow up Drinker Profile*. Psychological Assessment Resources, Inc. PO Box 98, Odessa, Florida, 33556.

Related Webpage: URL: <http://casaa.unm.edu>.

Summary

The Comprehensive Drinker Profile (CDP), Brief Drinker Profile (BDP) and Follow-up Drinker Profile (FUDP) are structured interviews that provide information on alcohol consumption, drinking patterns, alcohol related problems and demographic background. They are easily understood by both clinician and client, and cover areas that are of relevance in the assessment and treatment of alcohol problems. Some comparative information is included, but norms from a diverse sample of drinkers are not provided. These instruments make a good assessment tool for clinicians working with clients who present with alcohol problems. The duration of time required to administer do not make them suitable screening instruments.

The Comprehensive Drinker Profile, (CDP), Brief Drinker Profile, (BDP), and Follow-Up Drinker Profile, (FUDP) are a family of semi-structured interviews developed in the early 1970's by Marlatt and colleagues to provide comprehensive assessment of an individual's drinking history. All three instruments are derived from the CDP and closely follow the structure and format of this instrument. A brief description of the CDP is provided, and the reader is referred to other sources for additional information on the BDP and FUDP.

Description of the CDP

The 88-item CDP provides a systematic and extensive assessment of areas of life functioning. Section A focuses on demographic information, Section B focuses on drinking

history, including an interviewer-administered version of the MAST, and Section C obtains motivational information.

Regarding determining the quantity and frequency of drinking, Section B includes a subsection entitled Present Drinking Pattern. The information obtained is used to classify drinkers as periodic drinkers, steady drinkers or a combination of both. Further detailed information is obtained by charting daily drinking patterns including: type of alcohol consumed, % alcohol content, amount drunk on each occasion, and approximate time span during which it is consumed. A table is provided to enable approximate blood alcohol concentration (mg%) reached after three hours of drinking. Normative data are provided on the quantitative variables based upon 103 outpatients from a clinic for problem drinkers at the Department of Psychology, University of New Mexico. In addition to obtaining quantity and frequency information, a quantitative index of strength of family history of alcoholism may be obtained. The CDP is most useful as a measure of baseline alcohol consumption from which to plan and assess treatment.

Administration and scoring

The CDP takes approximately 2 hours to complete and should be administered by clinicians trained in its administration.

The manual is comprehensive, well written and contains a number of references to treatment outcome studies to support all aspects of the CDP. It is a useful instrument for both researchers and clinicians working with people presenting for treatment for alcohol problems.

Availability and cost

All 3 interviews and the CDP manual can be downloaded from URL:

<http://casaa.unm.edu>.

(Centre on Alcoholism, Substance Abuse and Addiction, University of New Mexico, USA).

Biochemical measures used in the assessment of alcohol use

Blood Alcohol Concentration (BAC)

Blood alcohol concentration refers to the concentration of alcohol in the blood and is measured in milligrams of alcohol per 100 ml of blood (mg%). Blood alcohol concentrations can be reliably measured using breath alcohol testing equipment, a non-invasive procedure in which the concentration of alcohol in end-expiratory breath is measured. This is an accurate reflection of the acute body burden of alcohol and the alcohol concentration of the pulmonary blood circulation (Dubowski, 1991).

The advantages of using BAC include the ease of administration, the immediacy of feedback, affordability and portability. BAC, though, cannot distinguish between acute and chronic alcohol use. As BAC only detects recent alcohol consumption, it is insensitive to binge-drinking patterns and long-term alcohol abuse (Chan, 1993). It has been suggested by The National Council on Alcoholism and Drug Dependence (USA), that the following guidelines may indicate a longer term problem with alcohol: having a BAC above .10 during a routine examination, above .15 in a person not exhibiting signs of intoxication, or above

.30 at any time (cited in Salaspuro, 1994). However, these guidelines have received little empirical support (Chan, 1993).

The ideal instrument for assessing BAC is a breathalyser. There is a range of breathalyser equipment available, and manufacturers provide detailed information on their reliability. Drager Australia Pty Ltd is one company that manufactures a range of breathalyser equipment.

If there is no direct access to breathalyser equipment or an estimate of a previous blood alcohol concentration is required, there are both computer programs and tables that enable estimations to be made. Tables in the Comprehensive Drinkers Profile provide the approximate blood alcohol concentration reached after three, four and five hours of drinking for women and men of different body weight (Miller and Marlatt, 1984). Professor B. Miller has also produced a computer program called BACCUS, which calculates blood alcohol concentration for a specified time period. This is available from the University of New Mexico, Albuquerque, New Mexico 87131-1161, USA and may be downloaded at no cost from URL: <http://casaa-0031.unm.edu>. It should be noted that these tables use imperial units of volume and body weight, and therefore require metric conversion when used with Australian clients.

Liver Function Tests (LFT)

Excessive alcohol intake may produce a number of physical complications (see Saunders, 1993 for a review). Laboratory tests can detect abnormalities in body chemistry that have been caused by heavy drinking. One organ that is particularly susceptible to the effects of alcohol is the liver, and a range of laboratory tests are available that provide information on the overall impact of alcohol on the body. However, generally tests of liver function are neither sensitive nor specific to alcohol abuse (Chan, 1991). More recently introduced tests, such as CDT tests, are more sensitive and specific to long-term alcohol abuse; however, they still tend to be less sensitive when compared with self-report measures, especially in non-dependent populations.

Gamma-Glutamyl Transferase (γ -GT).

Gamma-glutamyl transferase (γ -GT) is an enzyme found in the liver, blood and brain. It is a non-specific indicator of liver disease. It has been reported to be elevated in between 60 – 80% of “alcoholics” (Lancet, 1980). γ -GT tends to be raised before either AST or ALT. It is one of the standard laboratory liver function tests. γ -GT has a half-life of 14-26 days (Allen, Litten, Fertig, & Sillanaukee, 2001).

Aspartate aminotransferase (AST/SGOT) and Alanine aminotransferase (ALT/SGPT)

These enzymes also reflect the overall health of the liver and can be routinely obtained using standard laboratory procedures.

Carbohydrate-deficient (CDT).

Unlike γ -GT, AST and ALT, elevated levels of CDT are related specifically to the metabolism of alcohol and are dependent upon the amount of alcohol consumed. Further, CDT levels return to normal after a period of abstinence (Stibler, 1991). CDT has a half-life of approximately 15 days (Allen et al., 2001).

Reliability and validity of LFT

Many heavy drinkers have normal LFT results and many non-drinkers may have elevated levels on biochemical tests that are routinely used to assess impact of alcohol. With the exception of CDT, the markers available have two drawbacks: either they are indicators of liver disease that may or not be related to alcohol use, or they lack sensitivity in detecting hazardous alcohol use (Stibler, 1991).

CDT is dependent on ethanol for its metabolism, appears in serum after high alcohol intake, and is unrelated to other forms of liver disease. Unlike the biochemical markers discussed above, raised levels of CDT have been shown to have good sensitivity (proportion of excessive drinkers with an abnormal test result) and specificity (proportion of non excessive drinkers with normal results). For example, in studies using quantitative microchromatic methods, sensitivities of 81-94% and specificities of 91-100% have been reported for current alcohol intake at definite risk levels (>60 g/day; Stibler, 1991). The majority of participants in this study, though, were individuals with an established pattern of heavy alcohol intake.

In a recent review of 54 studies comparing the sensitivity of CDT to other biochemical markers, Salaspuro (1999) found CDT to be slightly more sensitive than γ -GT in reflecting changes to a moderate or fixed amount of drinking over a 3-4 week period and to be more sensitive in detecting relapse in male alcohol misusers. CDT was similar to γ -GT in detecting alcohol use in male drinkers, although there was mixed findings in the few studies incorporating heavy drinking women. CDT showed low sensitivity (ranging from 0-61%) to lower levels of hazardous drinking in studies using the general community or samples of young people; findings which are similar to other biological markers. CDT, however, was superior to γ -GT in detecting alcohol abuse in individuals with alcohol-related and non-alcohol-related liver disease. All in all, CDT showed marginal superiority to conventional biological markers, although it has been suggested that a combination of CDT and γ -GT may increase sensitivity without lowering specificity (Babor, Steinberg, Anton, & Del Boca, 2000; Helander & Tabakoff, 1997).

While biochemical indicators offer reasonably objective measurement of heavy alcohol use unaffected by motivational or cognitive factors, they are invasive, do not provide immediate feedback, are relatively insensitive to low levels of alcohol misuse, and may even be less reliable than patient self-report (Salaspuro, 1999). Babor et al. (2000) for instance, found relatively low agreement between client self-report and biochemical markers following treatment and 15 months later (39.7% & 51.6% agreement, respectively) as compared with very high agreement between self-report and collateral report (97.1% & 84.7% agreement, respectively). Contrary to popular belief, the evidence suggests that clients are more likely to over-report than under-report alcohol use. Babor et al. note that while biochemical information may increase confidence in the validity of reports, they add little to information that can be obtained cheaply and more efficiently by self-report. Likewise, Aertgeerts et al. (2001) found biochemical markers of alcohol abuse to be far less sensitive to alcohol abuse and dependence in a large sample of general practice attendees. They found that while sensitivities of self-report questionnaires (AUDIT, CAGE and the “five-shot” questionnaire) ranged from 62%-93% in men and from 37% - 80.4% in women (across a range of cut-off points), biochemical markers (γ -GT, ASAT, ALAT, CDT) ranged in sensitivity from 6.8% (γ -GT) to 18.2% (CDT) in men and from 0% (ALAT) – 15.2% (CDT) in women. Of the self-report questionnaires used in this study, the AUDIT was the most sensitive instrument while the CAGE was the least sensitive in this population.

NICOTINE

Overview

The morbidity and mortality associated with tobacco smoking and the attending economic costs associated with these, has led to a major international public health campaign aimed at reducing the numbers of people who smoke tobacco. In Australia 27% of the population are regular and occasional smokers (National Drug Strategy Household Survey, 1998). Of these, the age group with the highest proportion of recent smokers is the 20-29 year olds (32% of all smokers). One in four teenagers (aged between 14 and 19 years) smoke with approximately one in six being regular smokers (16% of all smokers).

A range of self-report instruments and a number of biochemical measures are available to determine smoking status and nicotine dependence. The reliability of self report measures of smoking have been questioned in recent years and are believed to be susceptible to bias, and in particular underestimation as social pressures not to smoke have increased. With this caveat in mind, it would appear that there are several reasonable instruments that may be used to assess nicotine dependence. The most widely used are the 10-item Revised Fagerström Tolerance Questionnaire (RTQ), the six item Fagerström Test for Nicotine Dependence (FTND) and the Heaviness of Smoking Index (HSI) which uses two items from the FTND.

Biochemical measures are often proposed as a more reliable method of assessment, as such procedures are not influenced by social norms and social desirability. The by-products of smoking detectable in body fluids include cotinine (in plasma and saliva), thiocyanate (in plasma and saliva) and carbon monoxide (CO in expired air). Expired air carbon monoxide monitoring is the most economical method. Both cotinine and thiocyanate measurement require specialist laboratory analysis. Expired air CO is used more routinely to determine smoking status.

Measures used to determine nicotine use

Quantity and Frequency Methods

Summary

Quantity and frequency of cigarette smoking is best ascertained by asking the client if they smoke and to provide an estimate of number of cigarettes smoked per day. This is sometimes referred to as the “aggregate method” and provides a reasonably accurate estimate of cigarettes smoked. More accurate information can be obtained by either asking the client to monitor his/her smoking over a period of a week, or using the timeline follow back method retrospectively (see alcohol section earlier).

Ninety per cent of individuals who smoke cigarettes are nicotine dependent (Gust, Hughes and Pechacek, 1986). Therefore, asking a client if he/she smokes will in almost all cases indicate that they are dependent on nicotine. Obtaining an estimate of the number of cigarettes per day is the most time efficient method used to obtain a quantity and frequency estimate. However, accuracy can be improved by use of the timeline followback method (Brown et al., 1998). This allows for a more detailed assessment of smoking patterns, including weekend fluctuations, over an extended period of time (Gariti et al., 1998).

Reliability and validity of quantity-frequency measures

Reliability can be improved if the client is required to self-monitor their cigarette use. It is more accurate if a client is required to record the time at which each cigarette is smoked (Frederiksen, Epstein and Kosevsky, 1975). However, it is also important to note that self-monitoring alone will likely produce a decrease in consumption, particularly if the client is motivated to cut down or stop. Timeline followback is a more reliable and valid way to determine exact numbers of cigarettes smoked. It is, however, a more consuming method.

Measurement

A weekly diary may be used to monitor the number of cigarettes smoked. However, more detailed information may be considered helpful. Additional information, for example, a record of the time the cigarette was smoked, mood state at the time and need for a cigarette, are often recorded.

Assessment of severity of nicotine dependence

Revised Fagerström Tolerance Questionnaire (RTQ) & Fagerström Test for Nicotine Dependence (FTND)

Key References: Tate, J.C. and Schmidt, J.M. (1993) A proposed revision of the Fagerström Tolerance Questionnaire. *Addictive Behaviors*, 18, 135-143.

Etter, J. F., Vu Duc, T., & Perneger, T. V. (1999). Validity of the Fagerström test for nicotine dependence and of the Heaviness of Smoking Index among relatively light smokers. *Addiction*, 94(2), 269-281.

Summary

The Revised Fagerström Tolerance Questionnaire (RTQ) is a ten-item questionnaire designed to measure the severity of nicotine dependence. It has both high internal validity and test-retest reliability and appears to be a valid measure of severity of nicotine dependence as it correlates significantly with expired air CO. Further studies investigating the relationship with cotinine are underway.

The Fagerström Test for Nicotine Dependence (FTND) consists of six items from the RTQ. The HSI is a two-item measure extracted from the RTQ. On balance, it would appear that despite widespread use, these measures perform no better (and in some cases worse) than asking for the number of cigarettes smoked per day.

Description and development of the RTQ & FTND

There are now a number of modifications of the original 8-item Fagerström Tolerance Questionnaire (FTQ; Fagerström, 1978), including the 6-item Fagerström Test of Nicotine Dependence (FTND; Heatherton, Kozlowski, Frecker and Fagerström, 1991), the 10-item Revised Fagerström Tolerance Questionnaire (RTQ; Tate and Schmidt, 1993), and most recently the modified Fagerström Tolerance Questionnaire (mFTQ; Rojas, Killen, Haydel & Robinson, 1998). A number of reports have appeared in the research literature since the development of the RTQ in which either the FTND or the mFTQ is used. It is therefore important in reading any research report in which smoking measures are evaluated, to pay careful attention to which of the Fagerström instruments were used.

The RTQ is a uni-dimensional measure that assesses the severity of nicotine dependency, tolerance and withdrawal. Items cover number of cigarettes smoked, smoking topography, smoking to relieve nicotine withdrawal and difficulty in refraining from smoking. All items are scored on a 5-point scale from 1 to 5. A global score is the mean rating across the 10 items with higher scores indicating greater nicotine dependence.

The FTND consists of six items contained in the RTQ, excluding items 2, 8, 9, & 10, which assess nicotine yield and inhalation. These four items had previously been found to be unrelated to biological measures and be the least valid items on the original FTQ (Heatherton et al., 1991).

The Heavy Smoking Index (Heatherton et al., 1991) is based on two items from the RTQ: item 1 that measures number of cigarettes a day and item 4, which assesses time when the first cigarette is smoked. Heatherton et al found the HSI to be the better predictor of biochemical measures of heavy smoking, compared to the FTND, and may be a useful alternative to the longer questionnaires.

Reliability and validity of FTQ family

Tate and Schmidt (1993) compared RTQ scores across four groups of smokers: regular smokers, smokers attempting to quit, inpatient substance abusers and outpatient substance abusers and found the RTQ to have adequate internal (Cronbach's $\alpha = .83$) and test-retest reliability ($r = .88$ over 4-6 weeks) although there was higher temporal stability for the total score than for individual items. Factor analysis verifying the uni-dimensional structure of the RTQ and a .49 correlation between expired air CO samples and RTQ scores supported the construct validity of the RTQ. The RTQ was found to have greater internal consistency ($\alpha = .83$) than previously reported for either the original FTQ or the FTND. In addition to a high degree of internal consistency, factor analytic procedures produced a single common factor indicating that the RTQ measures a unidimensional, underlying construct. Furthermore, a strong correlation between RTQ score and expired air CO samples ($r = .49$) demonstrated preliminary construct validity for the RTQ.

The relationship between nicotine dependence and success in smoking cessation treatments is not clearly established. Kozlowski et al. (1994) compared the predictive validity of the FTQ, FTND and the HSI. Each of these three measures was able to predict smoking cessation to only a small degree. Further, neither of these measures was clearly superior, leading the authors to conclude that the shorter two-item HSI is more practical in clinical and research settings.

The reliability and validity of the FTND and the HSI have been compared in a recent study by Etter and colleagues (Etter et al., 1999) using saliva cotinine, and psychological measures related to addiction, including self efficacy measures of quitting and stages of change which have previously been found to be predictive of successful quitting. While both measures had good test-retest reliability, the internal consistency of the FTND and HSI was moderate (.70 and .72 respectively). Unlike previous studies, the FTND contained only one factor that accounted for approximately 41% of the variance.

The more important findings from this study related to the construct validity of the measures. While both measures were strongly associated with number of cigarettes a day, saliva cotinine and measures of self efficacy towards quitting (among others), when entered into a regression equation with absolute number of cigarettes a day as step one, they added little to the prediction of the criterion variables listed above. Therefore, based on this study, it would appear that asking people to simply estimate their typical number of cigarettes a day is a better measure of nicotine use than either the FTND or HSI.

Suitability for special populations

The increasing focus on the use of cigarettes by young people and in particular adolescents (see Killen et al, 2001) has led researchers to question the reliability and validity of the current group of Fagerström instruments for this younger population. While at least one study has used the FTQ to assess smoking in pregnant adolescents (Albrecht et al., 1999), another modified version (mFTQ) has now been used in several studies of adolescent smoking. The mFTQ was created by rescaling the items on the FTQ to provide a greater range of response choices, and to allow separate analysis of daily cigarette consumption (Killen, Fortman, Newman, Varady, 1990). The mFTQ has shown good test-retest reliability and internal validity. Rojas et al. (1998) recently found mFTQ score to be highly correlated with number of cigarettes smoked in the past 30 days, with a measure of nicotine withdrawal symptoms and with saliva cotinine levels. Both the mFTQ and a measure of depression were significant predictors of nicotine withdrawal symptoms accounting for approximately 35% of the variance.

Administration and scoring

All instruments are in the public domain and may be used without cost but with due acknowledgment of the source.

REVISED FAGERSTRÖM TOLERANCE QUESTIONNAIRE

1. How many cigarettes a day do you smoke? (circle one)

10 or less	11-15	16-20	21-25	26 or more
------------	-------	-------	-------	------------

2. How deeply do you inhale? (circle one)

1	2	3	4	5
I do not inhale		Moderately	Very Deeply	

3. How often do you smoke more in the morning than the rest of the day? (circle one)

1	2	3	4	5
Never		About half the time	Always	

4. How often do you smoke your first cigarette of the day within 30 minutes of waking? (circle one)

1	2	3	4	5
Never		About half the time	Always	

5. How difficult would it be for you to give up your usual first cigarette of the day? (circle one)

1	2	3	4	5
Not Difficult		Somewhat difficult	Extremely Difficult	

6. How difficult do you find it to refrain from smoking in places where it is forbidden (e.g., in church, at the library, cinema, etc.)? (circle one)

1	2	3	4	5
Not Difficult		Somewhat difficult	Extremely Difficult	

7. How often do you smoke when you are sick with a cold, the flu, or are so ill that you are in bed most of the day? (circle one)

1	2	3	4	5
Never		About half the time	Always	

8. On average, about how much of each cigarette do you smoke? (circle one)

1	2	3	4	5
1/2 or less	1/2	2/3	3/4	ALL

9. On average, how often do you inhale? (circle one)

1	2	3	4	5
Never		About half the time	Always	

10. On average, how often do you hold cigarette smoke in your lungs for a moment or two before exhaling? (circle one)

1	2	3	4	5
Never		About half the time	Always	

1. Scoring: All items are scored on a 5-point scale. For items 1 and 8, the five anchors are assigned numbers (e.g., 26 or more is assigned a score of 5 and 1/2 or less is assigned a score of 1).

Biochemical measures of nicotine use

Summary

The most convenient and economical measure of nicotine intake is expired air carbon monoxide (CO) monitoring. It is most accurate if a reading is taken later in the day when nicotine levels tend to plateau. A cut-off point of 8 ppm correctly identifies 66 - 97 % of smokers; 87% if atypical smokers are excluded. By products of smoking such as cotinine and thiocyanate are detectable in blood, urine and saliva. Whilst having longer half-lives than CO and therefore providing more accurate information on intermittent smokers, these measures are more invasive and costly.

Expired Air Carbon Monoxide (CO)

CO is a combustion by-product of smoking and is absorbed into the blood stream. It has a relatively short half-life of 4 to 5 hours, and therefore the most accurate readings of CO are taken towards the end of the day. CO can be measured in exhaled air with a carbon monoxide monitor.

Reliability and validity of CO

CO levels are positively correlated with the number of self-reported cigarettes smoked per day. CO levels also increase with the number of cigarettes smoked throughout the day, although a plateau is reached 9 hours after the commencement of smoking.

The sensitivity of CO for identifying active smokers is maximal if a cut-off point of 8 ppm is adopted; this will correctly identify between 66% to 97% of smokers and 96% to 99% of non smokers. If atypical smokers are excluded (i.e., pipe smokers, those who smoke less than 10 cigarettes a day, non-inhalers), then sensitivity increases from 66% to 87%.

Nicotine and Cotinine

Nicotine and its metabolic by-product, cotinine, can be detected in saliva, urine and blood. Nicotine metabolism varies considerably between individuals. Consequently, the most accurate index of nicotine intake involves measuring both nicotine blood concentrations and elimination rate. It is recommended that nicotine plasma levels be measured towards the end of the day, as plasma levels tend to plateau after 6 - 8 hours of smoking.

Cotinine has a longer half-life than nicotine (6-16 hours compared with 1-2 hours) and therefore it is generally a preferred measure of nicotine exposure. It is a highly specific measure of nicotine use; a mean salivary cotinine value of 310 ng/ml is the typical value for smokers compared with a 1.7 ng/ml in non-smokers (Jarvis et al., 1984). Cotinine values also correlate with the number of cigarettes smoked per day (Benowitz et al., 1983).

Salivary cotinine is the most accurate index of smoking, however, it is expensive and laboratory analysis is complicated. If a measure other than self-report is required, expired air carbon monoxide reading is the biochemical measure of choice.

A comparison of CO, thiocyanate and cotinine by Waage, Silsand, Urdal, & Langård (1992) showed that CO had the highest sensitivity (98%) and specificity (100%) to current

smoking, but the relatively short half-life of CO (3-5 hours) compared with cotinine (6-16 hours) and thiocyanate (about 14 days) renders CO less useful in cases where intermittent smoking is suspected. Thiocyanate is relatively insensitive to light smoking (Pre & Vassy, 1992).

Suitability for special populations

Expired air carbon monoxide monitoring is a non-invasive procedure producing an immediate result that can be explained in a straightforward manner by the clinician. It is a particularly suitable measure for a range of cultural groups and has been used successfully in a number of trials involving the delivery of smoking cessation programs to diverse groups.

There are, however, racial differences in the levels of cotinine when compared with the reported number of cigarettes smoked per day. The Coronary Artery Risk Development in (Young) Adults study (CARDIA) found black smokers to have higher serum cotinine levels compared with white smokers, even when estimated nicotine intake is lower than matched white controls (Wagenknecht, Cutter, Hayley et al., 1990). Given that black smokers have higher rates of most smoking related diseases, despite smoking fewer cigarettes a day (Stellman and Garfinkel, 1986), a biological explanation is most likely. Some of the possible explanations proposed include genetic differences in metabolism or clearance of the chemical products of smoking (Henningfield, 1990). Additional studies of race-specific risk factors and tobacco metabolism are required to further the study of racial differences in adverse health outcomes associated with cigarette smoking.

Nicotine use in severe and persistent psychiatric populations

A recent US survey found that one-third of people with a current or past mental disorder were current smokers, compared to 22% of the general population (Lasser et al., 2000). Smoking is most prevalent in people with schizophrenia. A large Australian community survey in 1998 found that 69% of Australians with psychosis had smoked in the previous 12 months (or 2.6 times the population rate; Kavanagh, McGrath, & Jenner; 2000). While the risk of respiratory cancers in smokers with schizophrenia is lower than the rest of the population (Hoffer & Foster, 2000), smoking is second to suicide as a reason for the excess mortality in schizophrenia (Brown, Inskip, & Barraclough, 2000). Smokers with schizophrenia show significantly higher levels of cotinine than other smokers (Olincy, Young, & Freedman, 1997) reflecting both the number of cigarettes they consume and the way they smoke them (e.g., depth of inhalation and amount of each cigarette smoked). The assessment of smoking in mental disorders has not been systematically examined. Currently, researchers use the same assessment procedures as for other smokers (George et al., 2000).

DRUG USE OTHER THAN ALCOHOL AND TOBACCO

Overview

The following section reviews instruments and measures used to screen for, or to diagnose problematic drug use other than alcohol and tobacco. A wide range of substances such as psychomotor stimulants, opiates, benzodiazepines, solvents and cannabis are included in this section.

Unlike alcohol and tobacco, the prevalence of other illicit drug use is relatively low in the general community. Although the use of some illicit drugs, such as amphetamine and heroin are on the increase (AIWH, 1999), the use of illicit drugs is relatively low in the general population (46% of Australians over 14 years report ever using any type of illicit drug and 14% using an illicit drug in the previous 12 months, compared with 90% who have ever drunk alcohol; AIWH, 1999). Given this relatively low incidence of illicit drug use widespread routine screening in the manner suggested for alcohol is not indicated. However, the high mortality and morbidity that is associated with illicit drug use, emphasises the need to effectively identify those who are using these drugs, and to implement harm minimisation practices.

Presently there are relatively few standardised measures of (other) drug use. The Drug Abuse Screening Test (DAST), reviewed in detail in the present report, is a well-validated screening measure. Other measures that have been developed and are well accepted include the Personal Experience Screening Questionnaire (PESQ; Winters, 1994) and the Drug Use Screening Inventory (DUSI; Tarter, 1990). The PESQ is a 40-item questionnaire that provides three subscores: problem severity and frequency, defensiveness and psychosocial indicators, and a summary of drug use. The DUSI is a 149-item instrument covering 10 domains: substance use behaviour, behaviour problems, health status, psychiatric disorder, social skills, family system, school work, peer relationship, leisure and recreation. It takes 20 - 40 minutes to complete and has been used successfully with clients as young as 11 years. As with the DAST, these instruments were developed in North America, particularly for use with youth; however, they are extremely comprehensive and are more appropriate as assessment rather than screening instruments.

Determining the quantity and frequency of illicit drug use poses greater problems than either alcohol or tobacco. It is far easier to arrive at reasonably accurate estimates of total alcohol and tobacco consumption than it is with many other drugs such as heroin, cocaine or amphetamine. This is related to a number of characteristics of illicit drugs, including their purity, route of administration and availability. Darke and colleagues have developed the Opiate Treatment Index (OTI; Darke, Ward, Hall, Heather & Wodak, 1991), a method of assessing drug use without reference to actual quantity consumed, focusing instead on the three most recent occasions of drug use. The Addiction Severity Index (McLellan, et al., 1980) also provides an indication of overall severity of dependence by incorporating into the Index the number of drug use days in the preceding 30 days. More information on the ASI can be found in the alcohol section above.

Following the extension of the dependence syndrome concept to drugs other than alcohol (Edwards, Arif and Hodgson, 1981), several instruments have been developed that measure severity of dependence: the Severity of Opiate Dependence Questionnaire (SODQ; Sutherland et al., 1986), the Severity of Amphetamine Dependence Questionnaire (SAmDQ; Churchill et al., 1993), the Benzodiazepine Dependence Questionnaire (BDEPQ) and the Leeds Dependence Questionnaire (LDQ; Raistrick et al., 1994). The Severity of Dependence Scale (SDS; Gossop et al., 1995) originally included in the SODQ, has been used as a separate instrument and has been proposed as a measure of subjective dependence. All these measures are included in the present review.

Biochemical measures used to detect drug use are urine and blood analysis and, more recently, hair analysis. These measures are reliable, although the detection of substances depends both on the amount of drug administered and on the pharmacokinetics of the individual substance. Hair analysis requires specialist laboratory procedures.

Screening for other drug use with Indigenous Australians

Illicit drug use is a growing concern among indigenous communities across Australia. Use of kava is a particular concern in areas of the Northern Territory (Clough, 2000). Petrol sniffing and other inhalant use are of concern particularly within the central Australian region and “Top End” communities (Brady, 1992). In urban and rural areas use of cannabis, amphetamines and heroin is reported (McKelvie, 2000).

Screening tools to detect and classify illicit drugs consumption are not currently well developed nor has their use been formally evaluated with indigenous populations. Faced with limited options, feedback from the field suggests that using mainstream screening and diagnostic tools generally provides appropriate information for the clinician, although changes in delivery format to an interview style is a common recommended strategy. It is noted that formal studies conducted of illicit drug use among indigenous people have generally not used standardised instruments, instead developing their own interview formats in collaboration with indigenous stakeholders (Larson, Shannon, & Eldridge, 1999).

Measures used to screen for other drug use Drug Abuse Screening Test (DAST)

Key Reference: Gavin, D.R., Ross, H.E. and Skinner, H.A. (1989). Diagnostic validity of the Drug Abuse Screening Test in the assessment of DSM-III drug disorders. *British Journal of Addiction*, **84**, 301-307.

Summary

The DAST is a 20-item screening instrument designed to identify individuals who have had a drug abuse problem (excluding alcohol) in the past 12 months. It includes some features of the dependence syndrome such as inability to abstain, withdrawal symptoms and a range of social and emotional problems associated with drug misuse. It is based on the MAST and may have limited use in an Australian context, particularly with women and Indigenous Australians.

Description and Development of the Drug Abuse Screening Test (DAST)

The DAST was developed to provide a brief instrument for clinical screening and treatment evaluation research. The original 28-item version was modelled on the MAST and the items parallel those contained in that instrument (see Skinner, 1982 for a description). It covers the use of drugs, physical and medical complications, and emotional and personal problems arising from drug use in the preceding 12 months. Respondents are required to answer either YES or NO to each of the items. Two shorter forms of the DAST, 20-item and 10-item versions, have been developed. The shorter forms are now routinely used, as they are highly correlated with the original measure (Skinner, 1982; Cocco & Carey, 1998).

Reliability and validity of the DAST

The DAST has been shown to have a high degree of internal consistency (28-item DAST, $\alpha = .92$; 20-item DAST, $\alpha = .95$) and factor analysis of the 28-item DAST indicated it measures a single dominant dimension of problems associated with drug abuse among 256 alcohol and drug abuse treatment seekers (Skinner, 1982).

To determine whether answers on the DAST were likely to be influenced by factors such as concern about reporting socially unacceptable behaviour, the relationship between DAST scores and measures of social desirability and denial was investigated. There was a small positive correlation between these measures that was not statistically significant (Skinner, 1982).

Scores on the 28-item DAST are highly correlated with the frequency of use for a range of drugs including cannabis, barbiturates, amphetamine and opiates. DAST scores also discriminated accurately between alcohol and drug problems, and the 10-item DAST was not significantly correlated with alcohol use, MSAST or CAGE scores (Skinner & Goldberg, 1986). The DAST was recently found to be correlated highly with the ASI, providing further evidence for its criterion related validity (Appleby, Dyson, Altman & Luchins, 1997).

The sensitivity and specificity of the 28-item DAST was evaluated by Gavin et al. (1989). Using the Diagnostic Interview Schedule, subjects were classified according to the presence or absence of any current DSM-III drug disorder (excluding alcohol and tobacco). The DAST attained 85% overall accuracy in identifying subjects who met DSM-III diagnosis; maximum sensitivity (96%) was obtained with a cut-off score of 6 to 7. Based on Receiver Operating Characteristics analyses, the authors recommend that a score of 5 to 6 be used as the cut-off score. In a similar study, a cut-off score of 3 on the 10-item DAST correctly classified 93% of patients (Bohn, Babor & Kranzler, 1991).

Suitability for special populations

The DAST only appears to have been used in North America; there is no information regarding its applicability to specific cultural groups, of which the authors are aware. The relatively high accuracy in using a cut-off score on the DAST and DSM-III diagnosis of drug abuse/dependence, however, indicates that the DAST may be an appropriate screening instrument in populations in which the DSM-III diagnostic system also produces reliable and valid diagnoses.

A recent body of work has examined the utility of the DAST among psychiatric patients. Cocco and Carey (1998) and Teitelbaum and Carey (2000) examined the psychometric

properties of the DAST (10 and 20 item versions) among psychiatric outpatients. They both displayed acceptable internal consistency ($\alpha > .85$) and temporal stability ($r > .70$). Unlike previous research, factor analysis indicated a multidimensional scale. Notwithstanding, DAST scores correlated well with past and current substance use and DSM-III-R diagnosis of a drug use disorder, the MAST and the ASI. A cut-off score between 2 and 4 on the 20-item DAST and 1 or 2 on the 10-item DAST yielded high specificity and moderate sensitivity. Similarly, Maisto et al. (2000) found that a cut-off score of 2 provided good sensitivity and specificity for identifying a current diagnosis of an alcohol or drug use disorder among outpatients with serious persistent mental illness.

A 27-item adolescent version (DAST-A) of the DAST has recently been developed for use with adolescent psychiatric inpatients (Martino, Grilo & Fehon, 2000). Similar to adult studies, initial validation procedures found good internal consistency ($\alpha = .91$), high test-retest reliability ($r = .89$) and a unidimensional factor structure. A cut-off score of 6 yielded 78.6% sensitivity and 84.5% specificity for identifying DSM-IV diagnoses of drug related disorders.

We suggest that the limitations associated with using the MAST with women may also apply to the DAST. The questions on social and occupational functioning are more relevant to employed people than to people whose primary occupation is home duties. It is likely that few women would answer in the affirmative to items regarding use of violence. However, due to the nature of the lifestyle associated with illicit drug use, the other items relating to arrest and illegal activities may be more relevant to female drug users. As with the MAST, there are no questions that focus specifically on children or home duties.

The term “abuse” is consistently used in the DAST rather than the terms “drug use” or “misuse”. Whether this is likely to influence accurate completion of the DAST is a moot point; however, we draw attention to this issue and, in the absence of empirical information, leave it to individual clinicians to decide on whether the wording should be altered. Item 17 refers to withdrawal symptoms as “sick” in parentheses. It is unclear whether this term would adequately describe cocaine or amphetamine withdrawal.

To the authors’ knowledge, the DAST has not been used with Indigenous Australians and in the absence of consultation with Indigenous peoples, the appropriateness of this instrument remains unknown.

Administration and scoring

The DAST takes less than five minutes to complete and is scored by adding the number of items indicating drug use problems. “No” responses to items 4 and 5 indicate problems with drug use so are scored 1. It can be administered without specific training.

Availability and cost

The DAST is reproduced below, and can be used with due acknowledgment of the authors.

More information on the DAST can be obtained from Centre for Addiction and Mental Health, 33 Russell Street, Toronto, Ontario, Canada, M5S 2S1 (phone 416-595-6059), or online at <http://www.camh.net/resources/#Clinical Tools>.

Drug Abuse Screening Test (DAST)

INSTRUCTIONS: The following questions concern information about your potential involvement with drugs not including alcoholic beverages during the past 12 months. Carefully read each statement and decide if your answer is “Yes” or “No”. Then circle the appropriate response beside the questions.

In the statements “drug abuse” refers to (1) the use of prescribed or over the counter drugs in excess of the directions and (2) any non-medical use of drugs. The various classes of drugs may include: cannabis (e.g., marijuana, hash), solvents, tranquilisers (e.g., Valium) barbiturates, cocaine, stimulants (e.g., speed), hallucinogens (e.g., LSD) or narcotics (e.g., heroin). Remember that the questions do not include alcoholic beverages. Please answer every question. If you have difficulty with a statement, then choose the response that is mostly right.

-
- | | | |
|---|-----|----|
| 1. Have you used drugs other than those required for medical reasons? | yes | no |
| 2. Have you abused prescription drugs? | yes | no |
| 3. Do you abuse more than one drug at a time? | yes | no |
| 4. Can you always get through the week without using drugs? | yes | no |
| 5. Are you always able to stop using drugs when you want to? | yes | no |
| 6. Have you had “blackouts” or “flashbacks” as a result of drug use? | yes | no |
| 7. Do you ever feel bad or guilty about your drug use? | yes | no |
| 8. Does your spouse (or parents) ever complain about your involvement with drugs? | yes | no |
| 9. Has drug abuse created problems between you and your spouse or your parents? | yes | no |
| 10. Have you lost friends because of your use of drugs? | yes | no |
| 11. Have you neglected your family because of your use of drugs? | yes | no |
| 12. Have you been in trouble at work because of drug abuse? | yes | no |
| 13. Have you lost a job because of drug abuse? | yes | no |
| 14. Have you gotten into fights when under the influence on drugs? | yes | no |
| 15. Have you engaged in illegal activities in order to obtain drugs? | yes | no |
| 16. Have you been arrested for possession of illegal drugs? | yes | no |
| 17. Have you ever experienced withdrawal symptoms (felt sick) when you stopped taking drugs? | yes | no |
| 18. Have you had medical problems as a result of your drug use (e.g., memory loss, hepatitis, convulsion, bleeding, etc?) | yes | no |
| 19. Have you gone to anyone for help for a drug problem? | yes | no |
| 20. Have you been involved in a treatment program specifically related to drug use? | yes | no |
-

Assessment of quantity and frequency of drug use

Opiate Treatment Index (OTI)

Key Reference: Darke, S., Ward, J., Hall, W., Heather, N. and Wodak, A. (1991). *The Opiate Treatment Index (OTI) Manual. Technical Report Number 11*. National Drug and Alcohol Research Centre, Sydney, Australia. URL: <http://www.med.unsw.edu.au/ndarc/>.

Summary

The Opiate Treatment Index (OTI) is a structured interview primarily developed to allow comparability between research findings. It consists of six independent outcome domains: drug use, HIV risk-taking behaviour, social functioning, criminality, health status and psychological adjustment as measured by the GHQ-28. It takes between 20–30 minutes to administer. While it was not intended to be an alternative to a clinical assessment, it is a useful clinical tool providing quantitative information from which to evaluate a treatment program.

Description and development of the Opiate Treatment Index (OTI)

The Opiate Treatment Index (OTI) is a structured interview primarily developed to provide a comprehensive measure that can be used to determine the relative effectiveness of treatments or interventions in the substance misuse field. The authors suggest that use of a comprehensive instrument that objectively measures six independent outcome areas will enable comparisons to be made between different research studies. The OTI was developed at the National Drug and Alcohol Research Centre, Sydney:
URL: <http://www.med.unsw.edu.au/ndarc/>.

The six independent outcome domains that comprise the OTI have previously been demonstrated to be variables that change following a treatment intervention. They consist of: Drug Use, HIV Risk-taking Behaviour, Social Functioning, Criminality, Health Status and Psychological Functioning. Psychological Functioning is assessed using the General Health Questionnaire (28 item version; see section IV for a separate discussion on the GHQ). The Drug Use domain differs from many typical drug use scales (e.g., Addiction Severity Index) in that it focuses on drug use on only the three most recent occasions. For example, when answering the OTI regarding heroin use, subjects are asked:

- (1) On what day did you last use heroin?
- (2) How many hits did you have on that day?
- (3) On which day before that did you use heroin?
- (4) How many hits did you have on that day?
- (5) And when was the day before that that you used heroin?

The following example is taken from the Opiate Treatment Index Manual (p. 6).

Last use?	Friday
How much	4 hits (q1)
Time before	Thursday (or 1 day before) (t1 = 1 day)
How much	4 hits (q2)
Time before that?	Wednesday (or 1 day before)(t2 = 1 day)

Estimates of recent consumption are calculated by adding the consumption on the two use days and dividing by the intervals between the use days. An estimate of recent consumption is obtained by the formula:

$$Q = \frac{q1 + q2}{t1 + t2}$$

Using the formula above the Quantity/Frequency estimate is calculated as follows:

$$Q = \frac{4 + 4}{1 + 1} = 4$$

The manual provides a guideline to the interpretation of the Quantity/Frequency estimate obtained.

The HIV Risk-taking Behaviour scale (HRBS) consists of 11 items focusing on injecting practices and sexual behaviour that places individuals at risk of either contracting or spreading Human Immunodeficiency Virus (HIV). Addition of all items scored on a 6-point scale (0 - 5) provides a single overall score; the higher the score the greater the risk of contracting or spreading HIV. The Social Functioning scale consists of items reflecting overall social stability and social support. It also incorporates in the overall score a measure of involvement in the drug subculture, a component that is not normally included in more general measures of social support and social functioning. The Criminality scale assesses involvement in recent criminal activity in four areas: property crime, drug dealing, fraud and crimes involving violence. Each of these categories are mutually exclusive. Finally, Health Status provides an indication of subjects' current state of health, particularly regarding illnesses or medical problems associated with drug users' lifestyle. The interviewer reads out a list of common health problems and the respondent indicates whether they are currently suffering from this problem. The overall health score is obtained by summation of "Yes" responses.

Reliability and validity of the OTI

The OTI was originally validated in an Australian opiate treatment setting (primarily methodone) in 1992 (Darke, Hall, Wodak, Heather & Ward, 1992). The psychometric properties of a slightly modified instrument have also been examined in a UK methodone treatment setting (Adelekan, Green, Dasgupta, Tallack, Stimson & Wells, 1996). The OTI demonstrated high levels of test-retest reliability (.77-.99) on all scales in both settings regardless of whether the same or a different interviewer administered the test (Adelekan, Green, et al., 1996; Darke et al., 1992). It also demonstrated generally high levels of internal consistency ranging from $\alpha = .38-.83$ in the Australian study and $\alpha = .34-.93$ in the UK study.

Factor analysis in both studies extracted two factors - 'drug using lifestyle' and 'health and well being' - which accounted for 60% and 55.4% of the total variance in the Australian and London studies respectively (Adelekan, Green, et al., 1996; Darke et al., 1992).

The convergent validity of the OTI has also been examined. Darke et al. (1992) found the overall correlations between each of the subscales compared well with relevant subscales on the Addiction Severity Index (ASI). All scales were significantly correlated with the ASI counterpart except for Legal (ASI) and OTI Crime scale. The authors suggest this disparity reflects an emphasis on *conviction* for crimes in the ASI compared with crimes *committed* on the OTI (Darke et al., 1992). Scores on individual scales of the OTI were also compared with other relevant measures in both studies. Significant correlations were found for scores on the Health Status scale with independent medical examinations, between drug use scores and urinalysis results and between reported recent behaviours and partner collateral reports (Darke et al., 1992). These findings were replicated in the UK study, providing support for the convergent validity and cross-cultural validity of the measure (Adelekan, Green, et al., 1996).

A number of limitations of the OTI for use in a clinical setting have been suggested. These centre on the clinician's ability to use the instrument in a standardised manner and the validity of information given by informants if they perceive there could be treatment consequences for answers that are not congruent with the agency's treatment philosophy. In addressing these issues, two studies have examined the inter-rater reliability of clinician versus research assistant ratings in New Zealand and UK methadone treatment settings. Both studies found good levels of inter-rater reliability, with few or no significant differences between information given to clinicians compared with research assistants (Adelekan, Metrebian, Tallack, Stimson, & Shanahan., 1996; Deering & Sellman, 1996).

Suitability for special populations

The OTI has been translated into Vietnamese, Chinese and Spanish (Liu et al., 2000; Ruz, Gonzalez & Ruiz, 1998; Swift, Maher & Sunjic, 1999). It has recently been used to evaluate the effectiveness of two treatment programs, the first for people with serious mental illness and substance use problems in inner city Sydney (Teesson & Gallagher, 1999); the second a parenting intervention for families on methadone maintenance (Dawe, Harnett, Rendalls & Staiger, submitted). The OTI is easily understood by individuals proficient in English.

Administration and scoring

The manual is clearly written and the scoring system for each of the scales well described. The OTI can be administered in about 20-30 minutes. While the instrument was designed primarily as a research tool it can be used in a clinical setting. It is not intended to replace a clinical history taking but it is a useful adjunct, providing quantitative information that can be used to evaluate clinical progress. In particular, the HRBS scale and the Health Scale provide a useful structure for obtaining information about risk behaviour and health status.

Assessment of severity of dependence on drugs

Severity of Opiate Dependence Scale (SODQ)

Key Reference: Sutherland, G., Edwards, G., Taylor, C., Phillips, G., Gossop, M. and Brady, R. (1986) the measurement of opiate dependence, *British Journal of Addiction*, **81**, 479-484.

Summary

The SODQ is a five-section questionnaire designed to assess severity of opiate dependence. The SODQ contains items addressing the demographics of drug consumption, as well as items related to four aspects of the dependence syndrome: physical withdrawal, affective withdrawal, withdrawal relief drug-taking and rapidity of reinstatement after abstinence. Single items also reflect narrowing of behavioural repertoire and tolerance to opiates. Items have a multiple-choice response format. Whilst the questionnaire has been validated on samples of British and Australian opiate users, a cut-off score indicative of dependence is yet to be ascertained.

Description and Development of the SODQ

Following the extension of the dependence syndrome concept to drugs other than alcohol (Edwards, Arif and Hodgson, 1981), Sutherland and colleagues (1986) drew attention to the need for a conceptually founded instrument to measure the severity of opiate dependence. This would allow further exploration of the nature of opiate dependence, its natural history, and the matching of treatment to client needs. With such aims in mind, the SODQ was developed through a series of pilot stages as a parallel instrument to the SADQ. The SODQ consists of five main sections: quantity and pattern of opiate use; physical symptoms of withdrawal; affective symptoms of withdrawal including craving; withdrawal relief drug-taking; and rapidity of reinstatement of withdrawal symptoms after a period of abstinence. Single items also relate to the notions of tolerance and narrowing of drug use repertoire.

The SODQ was completed by 100 consecutive outpatients applying for treatment for opiate dependence in New York (Sutherland et al., 1986). Factor analyses were conducted on each of the four main sections of the SODQ. A single factor emerged that accounted for a satisfactory proportion of the variance in responses. Further, Cronbach's α indicated acceptable internal consistency for each section. Factor analyses performed on items from all four main sections of the SODQ found a single factor that accounted for 39% of the variance, indicating that the structure of the questionnaire is dominated by a single underlying construct. Very high correlations ($>.95$) between section scores and factor scores justify the simple addition of section scores to form an overall SODQ score. Some preliminary evidence for construct validity was provided by the significant correlations between SODQ score and number of opiate injections per day, and with subjective feelings of dependence.

Reliability and validity of the SODQ

The original application of the SODQ demonstrated preliminary evidence for the internal consistency and the construct validity of the instrument. The SODQ has been used with

British (Phillips, Gossop, Edwards, Sutherland, Taylor and Strang, 1987), Australian (Burgess, Stipp, Pead and Holman, 1989) and North American opiate users (Sutherland, Edwards, Taylor, Phillips and Gossop, 1988). Structural analyses of the questionnaire in these validation studies found results strikingly similar to those of Sutherland et al. (1986), with satisfactory proportions of variance in responses accounted for by factor analyses, and acceptable levels of internal consistency. Construct validity was again demonstrated with significant correlations between SODQ scores and other measures of opiate use, including subjective sense of dependence and some items of the Psychoactive Substance Dependence and Abuse section of the Structured Clinical Interview for DSM-III-R (Spitzer, Williams and Gibbon, 1986).

Amphetamine version

Recognition among clinicians that amphetamine, a drug historically considered to be relatively safe and non-addictive, could induce dependence, led to the SADQ being adapted to measure severity of amphetamine dependence. The adaptation involved changing all references from opiates to amphetamine, and included additional items assessing amphetamine-specific withdrawal symptoms such as tiredness, depression and lethargy. This adapted questionnaire, the Severity of Amphetamine Dependence Questionnaire (SAMdQ; Churchill et al., 1993) was administered to 101 consecutive patients seeking treatment for amphetamine dependence. Analyses of the questionnaire revealed results that were very similar to those obtained by Sutherland et al. (1986) for the SODQ, indicating the questionnaire to possess acceptable internal consistency and unidimensionality. Further, construct validity was demonstrated through significant correlations between total SAMdQ scores and other measures of amphetamine use, such as Severity of Dependence Scale scores, frequency of administration, etc. The SAMdQ has since been administered to a sample of amphetamine users diagnosed as amphetamine dependent by DSM-III-R criteria, as assessed by the CIDI (Topp and Mattick, 1997). Psychometric analyses indicated SAMdQ score correlated strongly ($r = .46$) with CIDI ascertained severity of dependence. Furthermore, SAMdQ score differentiated between participants diagnosed as mild/moderate and severely amphetamine dependent, further supporting its criterion validity. The similarities between the SODQ and the SAMdQ suggests that the dependence syndromes for opiates and amphetamines may be more alike than was previously thought (Topp, Mattick and Lovibond, 1995).

Administration and Scoring

Both the SODQ and the SAMdQ were designed for self-completion; however, they can also be administered by the researcher. Items are scored on a four-point scale ranging from “never or almost never” (scored 0) through “sometimes” (1), “often” (2), to “always or nearly always” (3). Total scores are calculated by summing together scores from the withdrawal sections. The reinstatement section has not been included in these total scores due to some conceptual and practical difficulties with this section.

Availability and Cost

Both the SODQ and the SAMdQ have been used in research settings only. It is considered that such instruments are more suitable for research than clinical applications. They are both in the public domain and the SODQ is reproduced below. They may be used without cost, but with due acknowledgment of the source.

SODQ

NAME: SEX: M / F DATE OF BIRTH:/..../... AGE:

First of all, we would like you to recall a recent month when you were using opiates heavily in a way which, for you, was fairly typical of a heavy use period. Please fill in the month and the year.

Month:..... Year:

Answer every question by circling one response only

1. On waking, and before my first dose of opiates:

- | | | | | |
|---|-----------------------|-----------|-------|---------------|
| (a) My body aches or feels stiff | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |
| (b) I get stomach cramps | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |
| (c) I feel sick | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |
| (d) I notice my heart pounding | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |
| (e) I have hot and cold flushes | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |
| (f) I feel miserable or depressed | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |
| (g) I feel tense or panicky | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |
| (h) I feel irritable or angry | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |
| (i) I feel restless and unable to relax | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |
| (j) I have a strong craving | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |

2. Please complete all sections (a-f) of this question:

- | | | | | |
|---|-----------------------|-----------|-------|---------------|
| (a) I try to save some opiates to use on waking | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |
| (b) I like to take my first dose of opiates within two hour of waking up | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |
| (c) In the morning, I use opiates to stop myself feeling sick | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |
| (d) The first thing I think of doing when I wake up is to take some opiates | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |
| (e) When I wake up I take opiates to stop myself aching or feeling stiff | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |
| (f) The first thing I do after I wake up is to take some opiates | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |

3. Please think of your opiate use during a typical period of drug taking for these questions:

- | | | | | |
|---|-----------------------|----------------|-----------------|---------------|
| (a) Did you think your opiate use was out of control | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |
| (b) Did the prospect of missing a fix (or dose) make you very anxious or worried? | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |
| (c) Did you worry about your opiate use? | NOT AT ALL | A LITTLE | QUITE A LOT | A GREAT DEAL |
| (d) Did you wish you could stop? | NEVER or ALMOST NEVER | SOMETIMES | OFTEN | NEARLY ALWAYS |
| (e) How difficult would you find it to stop or go without | IMPOSSIBLE | VERY DIFFICULT | QUITE DIFFICULT | NOT DIFFICULT |

Benzodiazepine Dependence Questionnaire

Key Reference: Baillie, A & Mattick, R. (1996). The Benzodiazepine Dependence Questionnaire: Development, reliability and validity. *British Journal of Psychiatry*, 169, 276-281.

Summary

The Benzodiazepine Dependence Questionnaire (BDEPQ) is a 30-item self-report questionnaire for measuring dependence on benzodiazepine (BZD) tranquillisers, sedatives and hypnotics. It was designed as a measure of a wider concept of dependence than withdrawal symptoms. The scale has demonstrated good internal consistency and temporal stability amongst a sample of BZD users. Factor analysis indicated a 3-factor solution comprised of general dependence, pleasant effects and perceived need factors. There is moderate evidence for the convergent validity of the scale with other measures and behavioural indicators of dependence. ROC analysis indicated cut-off scores of 23 and 34 were equally efficient in terms of specificity and sensitivity.

Description and Development of the BDEPQ

The BDEPQ was developed to address the lack of a suitable comprehensive measure to assess severity of benzodiazepine dependence (BZD; Baillie & Mattick, 1996). Despite the World Health Organisation broadening its definition of dependence to include psychological dependence, existing measures of BZD had focused on exclusively measuring tolerance and withdrawal symptoms (Kan, Breteler, Timmermans, van der Ven, & Zitman, 1999). Consequently, based on the WHO concept of dependence and information gathered from a sample of BZD users, an initial pool of 51 items was derived and reviewed by experienced clinicians. This resulted in a 30-item scale, which was then piloted on a group of BZD using patients in treatment for anxiety disorders.

Reliability and validity of the BDEPQ

The BDEPQ's psychometric properties were evaluated using the responses of 263 BZD users who had used BZDs more than 7 times in the previous three weeks. The BDEPQ displayed good internal consistency ($\alpha = .92$) and test-retest reliability ($r = .88$) over a 3 to 4 month period. A principal components analysis with an oblique rotation revealed 3 factors, which accounted for 44.7% of the variance – a general dependence factor, pleasant effects and perceived need. The BDEPQ displayed a moderate relationship with aspects of BZD dependence (e.g., experiences of BZD withdrawal, CIDI diagnosis of dependence) and had small correlations with indices of BZD use including dose, duration and frequency of use. Predictive validity was also examined in a sample of BZD users who had attempted to cease or reduce use. The BDEPQ was moderately correlated with severity of withdrawal and future doses of BZD during a 3 to 4 month follow up period. However, this influence was not independent of other predictor variables such as baseline BDI score, alcohol consumption and BZD dose (Baillie & Mattick, 1996).

A ROC analysis was conducted on a sample of 88 BZD users using telephone interviews. Cut-off scores of 23 and 34 were equally efficient in terms of sensitivity and specificity,

correctly identifying 74% of individuals with CIDI BZD dependence (Baillie & Mattick, 1996).

To date, the BDEPQ has only been validated in Baillie & Mattick's (1996) study. Baillie & Mattick note that their sample are most likely non-representative of BZD users in Australia (they were more socioeconomically homogenous, and had taken BZDs more frequently and over a longer time period than suggested by population statistics), thereby limiting the generalisability of these findings. Furthermore, Baillie and Mattick's preliminary analyses were based on the BDEPQ total score rather than the three subscale scores indicated by its multidimensional factor structure. Generally though, these findings suggest the BDEPQ may be useful as a research tool for use in a general practice setting. Further studies are required to examine the validity of the three-factor structure of the BDEPQ and utility of the BDEPQ in a clinical setting.

The Benzodiazepine Dependence Self-Report Questionnaire (Bendep-SRQ; Kan et al., 1999) was recently developed as an alternative measure of the severity of BZD dependence. Unlike the BDEPQ it incorporates physiological aspects of dependence and psychological and social factors. The authors sought to explore the multidimensional nature of BZD dependence suggested by the factor structure of the BDEPQ. The measure was validated with representative samples of regular BZD users (at least once per week) from general practice, psychiatric outpatient and self-help group settings. Four Rasch-homogeneous scales were extracted from the measure – problematic use, preoccupation, lack of compliance and withdrawal. Preliminary findings indicated the scale had an acceptable level of internal consistency ($KR-20 > .7$) and temporal stability ($r = .63$ to $.33$) across the three groups. Principle components analysis indicated the scale had a four-factor structure. The four-factor structure displayed good concurrent and discriminant validity with SCAN ICD-10 and DSM-III-R BZD dependence, ASI and SCL-90 scales. The Bendep-SRQ provides a multidimensional measure of BZD dependence for use in research and clinical settings. However, the predictive validity of the Bendep-SRQ requires further investigation.

Administration and Scoring

The BDEPQ asks respondents to think of their experiences with BZD use in the past month and rate their responses to items based on a four-point likert scale with a variety of different response options. The authors advise that the BDEPQ is best administered following detoxification from BZDs. A total score and three subscale scores can be calculated. Most items are scored by assigning 0 (never), 1 (sometimes), 2 (often), 3 (always) except for items 2,5,6,9,12,16,17 and 23 which are reversed scored. Two part items are scored 0 if the first part scores 0. Item 14a is not scored and item 11 is scored from 3 to 0 in descending order. Refer to the test manual for determination of the three subscale scores. Higher scores indicate individuals are at greater risk of future withdrawal symptoms and a CIDI diagnosis of BZD dependence. However, given the preliminary nature of the scale, test scores should be interpreted with caution.

Availability and Cost

The BDEPQ is in the public domain and is reproduced below. It may be used without cost, but with due acknowledgment of the source. The test manual is available from the National Drug and Alcohol Research Centre, University of New South Wales, Sydney, NSW, 2052 at a cost of \$12.

Benzodiazepine Dependence Questionnaire (BDEPQ)

Instructions: In the questions that follow you will be asked about your experience using medications known as sleeping pills, sedatives, hypnotics, 'benzos' or minor tranquillisers. These medications are also known by their trade names of 'Valium', 'Serapax', 'Mogadon', 'Normison', and 'Rohypnol' to list a few. All of these will be called **sedatives, tranquillisers or sleeping pills** in the questions. When answering the questions please think about your experiences **over the last month**. Place a mark in the box below the response that best suits your experience in the last month.

1. In the last month, have you taken another sedative or tranquilliser as soon as the effects of the previous one began to wear off?
Never Sometimes Often Always
2. Have you taken sedatives, tranquillisers or sleeping pills in the last month because you like the way they make you feel?
Always Often Sometimes Never
3. In the last month, have you felt you cannot face anything out of the ordinary without a sedative or tranquilliser?
Never Sometimes Often Everyday
4. Do you feel you cannot get through the day without the help of your sedatives or tranquillisers?
Never Sometimes Often Everyday
5. Do you need to carry your sedatives or tranquillisers with you?
Always Often Sometimes Never
6. Have you tried to reduce the number of sedatives, tranquillisers or sleeping pill you take because they interfere with your life?
A great deal Somewhat A little No
7. Have you found that you needed to take more tranquillisers, sedatives or sleeping pills to get the same effect in the last month compared with when you first took them?
No Sometimes Often Always
8. Do you need to take sedatives, tranquillisers or sleeping pills to deal with the problems in your life?
Never Sometimes Often Everyday
9. Do you feel terrible if you do not take a sedative, tranquilliser or sleeping pill?
Everyday Often Sometimes Never
10. a) In the last month, have you been worried that your doctor might not continue to prescribe the sedative, tranquillisers or sleeping pills you are taking?
Never Sometimes Often A Lot
Go to Question 11
1. How strong has this worry been?
Mild Moderate Severe
11. Could you stop taking sedatives, tranquillisers or sleeping pills tomorrow without any difficulties?
No, it would be impossible Perhaps, with a lot of difficulty
With some difficulty Without difficulty
12. Do you count down the time until you can take your next sedative, tranquilliser or sleeping pill?
Always Often Sometimes Always
13. a) Have you experienced relief when you have taken sedative, tranquilliser or sleeping pills in the last month?
Never Sometimes Often Always
Go to Question 14
b) How strong is that relief?
Mild Moderate Intense

14. a) In the last month, have you felt bad or sick as the effects of sedative, tranquilliser or sleeping pills wore off?
 Yes Answer the next question No Skip to question 15
- b) Have you taken another sedative, tranquilliser or sleeping pill to reduce these unpleasant after effects?
 Never Sometimes Often Always
15. In the last month. Have you taken sedative, tranquilliser or sleeping pills against your doctor's advice or more frequently than recommended
 Never Occasionally Sometimes Often
16. Are you concerned about the number of sedative, tranquilliser or sleeping pills you have taken in the last month?
 A Great Deal A Lot A Little Not At All
17. Have you taken more sedative, tranquilliser or sleeping pills in one day or night than you planned?
 Everyday Often Sometimes Never
18. a) Have you found the effects of sedative, tranquilliser or sleeping pills pleasant?
 Never Sometimes Often Always
- Go to Question 19 b) How strong is the pleasant feeling?
 Mild Moderate Intense
19. Have you taken sedative, tranquilliser or sleeping pills for a longer period than you intended to when you started?
 Never Sometimes Often A Lot
20. a) Have you felt tense or anxious as your prescription for sedative, tranquilliser or sleeping pills began to run out?
 Never Sometimes Often Every Time
- Go to Question 21 b) How strong have these feelings been?
 Mild Moderate Severe
21. a) Have you felt an urge or a desire to take sedative, tranquilliser or sleeping pills in the last month?
 Never Sometimes Often Every Day
- Go to Question 22 b) How strong is that urge or desire?
 Mild Moderate Intense
22. Have you taken sedatives, tranquilisers or sleeping pills in the last month when you did not really need them?
 Never Sometimes Often Every Day

Instructions: In the next set of questions please tick the box below the answer that matches what you think

21. I feel powerless to prevent myself taking a sedative or tranquilliser when I am anxious, uptight or unhappy
 Strongly Disagree Somewhat Disagree Somewhat Agree Strongly Agree
22. I would not be able to handle my problems unless I take a sedative or tranquilliser?
 Strongly Agree Somewhat Agree Somewhat Disagree Strongly Disagree
23. I get so upset over small arguments that I need to take a sedative or tranquilliser.
 Strongly Agree Somewhat Agree Somewhat Disagree Strongly Disagree

Leeds Dependence Questionnaire (LDQ)

Key Reference: Raistrick, D., Bradshaw, J., Tober, G., Weiner, J., Allison, J. and Healey, C. (1994). Development of the Leeds Dependence Questionnaire (LDQ): A questionnaire to measure alcohol and opiate dependence in the context of a treatment evaluation package, *Addiction*, **89**, 563-572.

Summary

The Leeds Dependence Questionnaire (LDQ) was developed as part of a treatment evaluation package. It is a 10-item, multiple choice self completion questionnaire designed to measure dependence upon a variety of substances, and has been used with alcohol and opiates. The LDQ was designed to be sensitive to change over time. Factor analyses have indicated the LDQ measures a unidimensional concept of alcohol and opiate dependence. It has also demonstrated good levels of internal consistency and temporal stability as well as concurrent, discriminant and convergent validity for alcohol and opiate dependence. No cut-off score indicative of dependence has been established.

Description and Development of the LDQ

The Leeds Dependence Questionnaire (LDQ) was developed to provide a measure of the severity of dependence that was not derived from measures of consumption and substance specific withdrawal symptoms, but incorporated broader notions of psychological dependence. In the LDQ, items address how drug effects are maximised, the re-administration of a drug when its effects are beginning to wear off, and the importance or primacy of the drug's effect. Raistrick et al. (1994) argue that the questionnaire is most sensitive to individuals who are not physically dependent upon a substance but who experience a range of psychological symptoms such as craving, compulsion to use and narrowing of behavioural repertoire.

The LDQ was developed through a series of eight pilot stages, each involving between five and 50 subjects. At each stage of the pilots, items were checked for their comprehensibility and their emotional neutrality. A 10-item, multiple choice response questionnaire emerged with the following items: preoccupation, salience, compulsion to start use, compulsion to continue use, planning to procure and use the drug, maximisation of drug effect, narrowing of repertoire, the primacy of drug effect, the constancy of drug-induced states and cognitive set. The final version of the questionnaire was administered to three samples of alcohol users and a sample of opiate users (Raistrick et al., 1994). A principal components analysis suggested that the questionnaire taps an unidimensional concept, and Cronbach's α indicated very high (.94) internal consistency.

Reliability and validity of the LDQ

Raistrick et al. (1994) were commendably thorough in their attempts to assess the LDQ's reliability and validity. As mentioned, a principal components analysis was conducted on the items, and a single factor emerged that accounted for 64% of the variance in responses, with all LDQ items correlating positively (greater than .65) with the factor score. Internal consistency was confirmed with a high Cronbach's α (.94). Test-retest reliability on a subsample ($n=33$) of both alcohol and opiate users was high ($r= .95$).

The LDQ was significantly correlated with the SODQ and the SADQ. Concurrent validity was also demonstrated through the significant differences in LDQ scores of three different samples of drinkers (alcoholic inpatients had higher scores than college students, who had higher scores than attendees at a G.P. clinic). Discriminant validity was shown in the lack of differences in LDQ scores based on age or gender. Finally, given the well-established relationship between dependence and psychological morbidity, convergent validity was demonstrated in the significant correlations between LDQ scores and GHQ scores (Raistrick et al., 1994).

While the LDQ has demonstrated good psychometric properties as a measure of alcohol and opiate dependence, its psychometric properties for measuring the severity of dependence on other illicit substances has not been examined.

Special Populations

The utility of the LDQ for measuring drug dependence among tertiary students and juvenile delinquents was recently examined (Lennings, 1999). Results indicated the LDQ had good internal consistency in both samples and measured a unidimensional factor of dependence. Among students the LDQ was predictive of alcohol use even after controlling for other variables.

The LDQ was also recently validated as a measure of cannabis dependence among a sample of people diagnosed with Schizophrenia. The LDQ had a high level of internal consistency ($\alpha = .82$) and was found to be predictive of a CIDI diagnosis of cannabis dependence or abuse (Hides, Dawe, Young & Kavanagh, submitted for publication).

Administration and Scoring

The LDQ is a self-completion questionnaire. Respondents are instructed to think about their substance use in the last week when answering the questions, and to tick the relevant response. Each of the items is scored on a “never” (0), “sometimes” (1), “often” (2) and “nearly always” (3) scale, yielding a maximum score of 30. The authors suggest the LDQ could be used in clinical or research settings, and describe it as “brief and user-friendly” (Raistrick et al., 1994, p.571).

Availability and Cost

The LDQ is in the public domain, and is reproduced below. It may be used without cost, but with due acknowledgment of the source.

The Leeds Dependence Questionnaire – LDQ

In answering this questionnaire:

- think about the last week
- think about your main substance or substance group, please specify
- tick the answer that's most appropriate to you

Never Sometimes Often Nearly Always

1. Do you find yourself thinking about when you will next be able to have another drink or take drugs?
2. Is drinking or taking drugs more important than anything else you might do during the day?
3. Do you feel your need for drink or drugs is too strong to control?
4. Do you plan your days around drinking or taking drugs?
5. Do you drink or take drugs in a particular way in order to increase the effect it gives you?
6. Do you take drink or drugs morning, afternoon and evening?
7. Do you feel you have to carry on drinking or taking drugs once you have started?
8. Is getting the effect you want more important than the particular drink or drug you use?
9. Do you want to take more drink or drugs when the effect starts to wear off?
10. Do you find it difficult to cope with life without drink or drugs?

Severity of Dependence Scale (SDS)

Key Reference: Gossop, M., Darke, S., Griffiths, P., Hando, J., Powis, B., Hall, W. and Strang, J. (1995) The severity of dependence Scale (SDS): Psychometric properties of the SDS in English and Australian samples of heroin, cocaine and amphetamine users. *Addiction*, **90**, 607-614.

Summary

The Severity of Dependence Scale (SDS) is a five-item questionnaire designed to measure the degree of dependence on a variety of drugs. The SDS focuses on the psychological aspects of dependence, including impaired control of drug use, and preoccupation with and anxiety about use in the past 12 months. The SDS was originally included as the final section in the SODQ. The SDS appears to be a reliable measure of the dependence construct. It has demonstrated good psychometric properties with heroin, cocaine, amphetamine, and methadone maintenance patients across five samples in Sydney and London. Its utility as a measure of cannabis and benzodiazepine dependence has recently been examined. Preliminary analyses have indicated cut-off scores of 4, 3 and 6 are indicative of amphetamine, cannabis and benzodiazepine dependence respectively.

Description of the Severity of Dependence Scale (SDS)

The SDS is a brief five-item scale that was developed to measure the degree of dependence on a variety of drugs. Unlike other measures of severity of dependence such as the SODQ and SAMDQ, which include sections on withdrawal and tolerance, the SDS focuses only on

the psychological aspects of dependence such as impaired control over drug use, anxiety about use and difficulty stopping.

The SDS was originally included as the final section in the SODQ. Later validation of the instrument occurred with a South London population of heroin users ($n = 200$) who were attending treatment on an outpatient basis (community drug team) or inpatient treatment. The most recent study involved a collaborative project between the National Addiction Centre (UK) and the National Drug and Alcohol Research Centre (Australia) in which the SDS was administered to a number of samples of drug users, heroin and cocaine users in London, and amphetamine users and methadone maintenance patients in Sydney (Gossop et al., 1995).

Reliability and validity of the SDS

Studies among heroin, amphetamine and cocaine users have shown the SDS to be a reliable measure of the dependence construct. Factor analysis produced a single factor solution and scores on each item of the SDS were almost perfectly correlated with factor scores (Gossop et al., 1995). The SDS has been found to have good internal consistency (α ranging from .8 to .9) across the five samples and good test-retest reliability (.89) over a one day interval in a sample of heroin users (Gossop et al., 1995; Gossop, Best, Marsden & Strang, 1997).

The construct validity of the SDS has been supported by significant correlations with behavioural indices of dependence including dose, frequency and duration of use (Darke, Ross & Hall, 1996). Further, those subjects who had previously received treatment for their heroin problem scored significantly higher on all SDS items, than those subjects who had never been in treatment. Severity of dependence was also influenced by route of drug administration with heroin “chasers” having significantly lower dependence scores than their injecting counterparts. (“Chasing the dragon” refers to a process in which heroin is heated on tin foil and the vaporised heroin inhaled by a tube, usually made from a rolled money note.)

The utility of the SDS for assessing cannabis dependence is less clear. Only one study has examined the psychometric properties of the SDS for cannabis dependence among a sample of long-term cannabis users in Australia (Swift et al., 1998). The SDS consisted of a single factor which accounted for 48.4% of the variance, all items correlating positively (greater than .50) with the factor score. However, it had only a moderate level of internal consistency ($\alpha = .72$) for cannabis dependence compared with alcohol and opiate dependence. Support for the SDS’s predictive and criterion validity for measuring cannabis dependence was also less clear. The SDS identified less than a third as many participants as cannabis dependent compared with the DSM-III-R and ICD-10 criteria, and correlated poorly with a number of behavioural indicators of dependence (i.e., quantity of cannabis used, age at first use, past use of illicit drugs) which were correlated with measures of ICD-10 and DSM-III-R criteria. Furthermore, no combination of variables (i.e., age, number of illicit drugs ever used, typical quantity of cannabis used) that were found to be predictive of ICD and DSM-III-R scores were predictive of SDS score. However, the SDS total score was more strongly related to the belief that cannabis use was a problem than the ICD or DSM-III-R scores. Receiver operating characteristics analysis of the SDS calibrated against the CIDI indicated a cut-off score of 3 as indicative of cannabis dependence with a corresponding sensitivity of 64% and specificity of 82% for cannabis users. This is a more liberal estimate compared with the cut-off score of 4 for identifying amphetamine dependence (Topp & Mattick, 1997).

A recent study examined the utility of the SDS as a screening instrument for benzodiazepine dependence among 100 regular benzodiazepine users attending a mental health service (de las Cuevas et al., 2000). The internal consistency of the SDS was supported in two ways. Reliability analysis using Cronbach's α yielded a value of .81. Item total correlations were statistically significant ($p < .01$) ranging from .45 (question 4) to .69 (question 2). ROC analysis indicated a cut-off score of 6 was an appropriate threshold for problematic benzodiazepine use identified by the CIDI, resulting in a specificity of 94.2% and a sensitivity of 97.9%.

Suitability for special populations

The SDS has been used with a variety of drug users in South London and in Sydney's inner city and Western suburbs. The wording of the SDS is straightforward and the concepts appear to be understood by a variety of drug users.

The SDS has been translated (and back translated) into Vietnamese in a Sydney study of heroin users (Swift et al., 1999). A Portuguese version of the instrument recently demonstrated good reliability and validity among Brazilian drug users for cocaine (snorted), crack cocaine (smoked), cannabis and alcohol (Ferri, Marsden, DeAraujo, Laranjeira & Gossop, 2000).

The utility of the SDS as a measure of cannabis dependence was recently examined among a sample of people diagnosed with Schizophrenia. The SDS had a relatively low level of internal consistency ($\alpha = .52$) but was found to be predictive of a CIDI diagnosis of cannabis dependence or abuse (Hides et al., submitted for publication).

SEVERITY OF DEPENDENCE SCALE

1. Did you ever think your use of (drug) was out of control?

Never or almost never	0	Sometimes	1
Often	2	Always or nearly always	3

2. Did the prospect of missing a shot/snort make you very anxious or worried?

Never or almost never	0	Sometimes	1
Often	2	Always or nearly always	3

3. How much did you worry about your use of (drug)?

Not at all	0	A little	1
Quite a lot	2	A great deal	3

4. Did you wish you could stop?

Never or almost never	0	Sometimes	1
Often	2	Always or nearly always	3

5. How difficult would you find it to stop or go without (drug)?

Not difficult	0	Quite difficult	1
Very difficult	2	Impossible	3

Structured interview for other drug use

Substance Dependence Severity Scale

Key Reference: Miele, G. M., Carpenter, K. M., Cockerham, M.S., Trautman, K.D., Blaine, J. & Hasin, D. S. (2000). Substance dependence severity scale (SDSS): reliability and validity of a clinician-administered interview for DSM-IV substance user disorders, *Drug and Alcohol Dependence*, 59, 63-75.

Summary

The SDSS is a newly developed semi-structured clinical interview designed to obtain a measure of severity of DSM-IV substance user disorders. It provides DSM-IV diagnoses of dependence for alcohol, cocaine, heroin, stimulants, illicit opiates, sedatives, methadone, cannabis, hallucinations and 'other' substances in the past 30 days. Preliminary validation studies supported the internal consistency and test-retest reliability of the SDSS for alcohol, heroin, cocaine and sedative use. The SDSS was found to have concurrent and predictive validity in assessing the severity of DSM-IV alcohol, cocaine and heroin dependence. The SDSS is a useful instrument for assessing treatment outcome.

Description and development of the SDSS

The SDSS was developed to provide a comprehensive measure of DSM-IV severity of dependence on a variety of substances, for the evaluation of substance abuse treatment efficacy. Items were keyed to assess substance-specific DSM-IV criteria for substance dependence and abuse. The SDSS assesses dependence severity for alcohol, cocaine, heroin, stimulants, licit opiates, sedatives, methadone, cannabis, hallucinations and 'other' substances (e.g., inhalants). Each substance is rated for symptom severity (usual severity & worst severity) and frequency (total number of days symptom occurred and most severe) in the last 30 days.

Reliability and validity of the SDSS

The SDSS was validated on a sample of 175 recent admissions to four different treatment settings including an inpatient alcohol rehabilitation program, outpatient drug and alcohol program, inpatient dual diagnosis unit and two methadone maintenance programs (Miele et al., 2000a; 2000b). The internal consistency of the four sub-scales was assessed for alcohol, cocaine, heroin, cannabis and sedative use. Internal consistency estimates on the usual severity scale ($\alpha = .79-.91$), worst severity scale ($\alpha = .67-.89$), total number of days the symptom occurred ($\alpha = .75-.91$) and total number of days symptom at worst severity ($\alpha = .66-.82$) were mostly in the acceptable range for all substances with estimates for cannabis tending to fall in the lower end of the acceptable range (Miele et al., 2000a). Test-retest reliability on the usual severity, worst severity, total number of days the symptom occurred, and total number of days symptom at worst severity were generally good to excellent for alcohol, heroin, cocaine and sedatives; with cannabis falling mostly in the fair range (Miele et al., 2000a).

Concurrent validity of the SDSS was demonstrated with usual and worst severity scale scores for all substances, except sedatives, being significantly correlated with clinical

dependence severity ratings. SDSS severity and frequency subscales were significantly correlated with frequency of alcohol, heroin, cocaine and cannabis use (Miele et al., 2000a). The concurrent and discriminant validity of the SDSS was further examined for alcohol, cocaine, heroin and cannabis use (Miele et al., 2000b). SDSS alcohol scales were significantly related to the ASI alcohol composite score but not to the ASI drug composite score, while the SDSS cocaine and heroin scales were significantly related to the ASI drug composite score but not the ASI alcohol composite score. All SDSS scales were significantly correlated with substance-specific measures of the consequences of substance use. Furthermore, there were significant relationships between decreases in SDSS scores from baseline to follow up and changes in ASI, Drinker Inventory of Consequences (DrInC) and Global Assessment Scale (GAS) over the same time period. Evidence for the predictive validity of the SDSS was supported by findings that SDSS scores were significantly related to time to first post treatment alcohol, cocaine and heroin use.

Administration and Scoring

The SDSS was designed to be administered by clinicians with at least a master's degree and clinical experience with patients with substance abuse or mental disorders. Substance specific screening questions are used to assess frequency, recency and amount of use in the past 30 days. The interview takes approximately 30 to 45 minutes to administer depending on the number of substances used in the past 30 days. A separate score sheet is used for every substance used in the past 30 days. This results in each substance having four sum scores for usual severity, worst severity, total number of days symptom occurred and total number of days symptom at worst severity. The two severity variables are scored on a 6 point scale ranging from 0 (absent) to 5 (extreme), with a score of 2 indicating that diagnostic criteria have been met. Total scores for the 11 dependence items range from 0-43 for usual severity and 0-35 for worst severity. The two frequency variables are scored on an 8-point scale ranging from 0 (symptom did not occur) to 7 (symptom occurred every day of past 30).

Suitability for Special Populations

The SDSS has displayed acceptable levels of reliability and validity among samples of American patients from an inpatient alcohol rehabilitation program, outpatient drug and alcohol program, inpatient dual diagnosis unit and two methadone maintenance programs. Its utility as a measure of the severity of substance dependence, in an Australian context, has not been established. The SDSS' sensitivity to change for assessing treatment outcome could benefit from being assessed over a longer period of time. Furthermore, the SDSS could benefit from comparison to other clinical interviews for drug dependence (e.g., CIDI).

The SDSS cannabis scale did not display high levels of test-retest reliability or internal consistency, and while it was significantly related to frequency of cannabis use and consequences of cannabis use, it was not related to other indicators of dependence (e.g., ASI or GAS). Further research examining the validity of the SDSS cannabis scale among primary cannabis users needs to be conducted. The reliability and validity of the SDSS for assessing sedative and other drug dependence also requires further investigation.

Availability and Cost

A copy of the SDSS may be requested from Gloria Miele via e-mail at gmm23@columbia.edu. The SDSS training manual is available from the Elsevier web site URL: <http://www.elsevier.nl/homepage/sab/drugalcdep/supmat.htm>.

Biochemical measures of drug use

The clear majority of research studies have concluded that people provide valid information about their illicit drug use under favourable conditions (e.g., confidential reports). However, recent studies have suggested that self-report of drug use may be more context dependent than previously identified. For example, Hamid and colleagues found that among drug users, the rate of agreement between self-reported drug use in the past 48 hours and urine tests was 58% if the urine test was performed after the interview, and 93% if performed before the interview (Hamid, Deren, Beardsley & Tortu, 1999). The validity of self-reports also appears to vary according to the substance being assessed. Other studies have indicated that self-reported drug use is more consistent with urinalysis when measured over time (Mieczkowski et al., 1991; Yacoubian, 2000). Therefore, agreement between self-reported drug use and biochemical measures such as urinalysis and hair analysis is dependent upon a number of factors, including the type of population studied (e.g., parolees), stage of treatment (e.g., assessment versus maintenance treatment), interviewing conditions (e.g., research versus treatment setting, confidentiality issues) and types of drugs under investigation. In some circumstances urinalysis and hair analysis play an important role in identifying drug use. Clinical decisions regarding whether to use urinalysis, should weigh up the relative merits of costs versus treatment and assessment context (Kilpatrick, Howlett, Sedgwick & Ghodse, 2000; Wish, Gray, Sushinsky, Yacoubian & Fitzgerald, 2000).

Urine analysis

Urine drug testing aims to detect the presence or absence of specific drugs or drug metabolites in urine. However, urine drug testing cannot be used to determine dosage, time of drug administration or the extent of any drug effects in the subject (Blanke, 1986). To differentiate between recent drug use and continued excretion of the drug from previous (heavy and prolonged) use, it is possible to perform a semi-quantitative analysis in which the concentration of the drug in urine is monitored over time. If the person has ceased to use the drug, then the concentration of drug in urine would be expected to decrease each time a urine sample is assayed. Increases or no change in concentration of drug in urine is consistent with continued use (Manno, 1986).

There are a number of factors that influence whether a urine drug screen is positive or negative. Firstly, the higher the dose, the more likely the drug will be detected. For example, whereas a dose of 30 mg of codeine might be detected for 1 – 6 hours after use by a particular method, a 60 mg dose may be detected for 1 – 10 hours (Manno, 1986). Frequency of use is also an important factor influencing detection. As a general rule, most drugs tend to accumulate in the body with regular use. Therefore, the more frequently a drug is used the more likely it is that it will be detected in a drug screen. Drugs are also metabolised at different rates. Cocaine, for example, is eliminated from the body relatively rapidly and depending on the method used, a single dose may only be detectable for a day or less. Continued use on a daily basis though, may cause the drug to be detected for 2 to 3 days after cessation of use. Cannabis, by contrast, can be detected for up to 3 weeks after cessation of use if it has been used regularly on a daily basis (Manno, 1986). For a detailed description of typical screening and confirmation techniques for a range of drugs see Hawks and Chiang (1986).

Hair analysis

Recent technological advances have led to the development of hair analysis to determine the use of a range of substances. Unlike urine screening, in which most drug use is ascertained within a relatively brief window of 2 – 3 days, hair analysis will detect the presence of drugs for the duration of the growth of the hair. Drugs and their metabolites, present in the blood plasma, become embedded in the hair structure during the process of keratinisation and remain there throughout the life of the hair (Marsh, Evans and Strang, 1995). Hair analyses have even been conducted posthumously; Napoleon's hair was found to contain arsenic while Keat's hair was found to contain laudanum (see Strang, Black and Marsh, 1993).

As human hair located on the posterior vertex region of the scalp grows at an average rate of 1 cm/month, it is possible to estimate the approximate time at which the drug was used. Interestingly, for methadone and its metabolites at least, commercially available hair colorants and peroxide bleach reduce drug levels but do not totally eliminate them (Marsh, Evans and Strang, 1995).

For a detailed description of the procedures used to detect opiates, cocaine, amphetamines and cannabis in hair samples see Sachs and Kintz (1998). Hair analysis is available in Australia at the Victorian Institute of Forensic Medicine. Specific protocol instructions and contact details are provided below.

Procedures for detecting most drugs have now been developed but analytical and interpretative problems still remain. Other problems include being unable to clearly establish if the hair originated from a particular individual and determining the influence of

environmental contamination, differences in gender and ethnic groups and cosmetic treatment on results. Overall, hair analysis presents a powerful tool for the detection of substance use, that should be applied widely in clinical practice and research. However, presently it offers little practical help to clinicians; whether hair analysis will offer any clinical advantages over urine analysis in the future, remains to be seen.

HAIR ANALYSIS FOR DRUGS OF ABUSE

Victorian Institute of Forensic Medicine

Hair is useful to target for exposure to drugs of abuse including amphetamines, cocaine, cannabis and heroin when a longer detection period is desired. Depending on the length, hair can provide evidence of exposure to drugs over a few months. Therefore, this allows a much longer detection window than urine testing which typically only allows detection for about 2 days.

Owing to the mechanism of drug incorporation into hair, the parent (ingested) drug is preferentially present rather than one or more metabolites in urine. Heroin users predominately contain heroin and 6-acetylmorphine (6-AM) in hair with lesser amounts of morphine. Cannabis users contain predominately tetrahydrocannabinol (THC). Cocaine users also tend to contain significant amounts of cocaine, in addition to the hydrolysis product benzoylecgonine.

The journal *Forensic Science International* (1995; 1997) has devoted two volumes to the issue of hair analysis including the Proceedings of the Workshops at Abu Dhabi (Dupont & Baumgartner, 1995). The United Nations (1998) has also prepared guidelines for the applicability of hair analysis.

Typically, whole hair is analysed for drug content. The laboratory provides an approximate concentration of drug in the hair, however the drug level cannot usually be used to infer a dose of drug.

To avoid surface contamination of hair, extensive washing procedures are employed prior to any drug isolation technique. Results are normally expressed as ng drug per mg hair.

Screening tests are based either on the ELISA technique or by direct analysis by gas chromatography-mass spectrometry. Confirmatory analyses are conducted using GC-MS.

Collection Protocol

1. Collect approximate 50 mg of hair from the back of the head (nape). This amount is approximately the thickness of a pencil for average length hair.
2. This should be cut close to the scalp and placed into the plastic bag provided.
3. Record the identity of the specimen, name and date of birth, or another form of identifying information that will enable you to connect the result with a person.
4. Seal the bag with security seal provided making sure seal is placed over the opening to the bag. Sign over the seal and the bag, and date.
5. Complete the chain-of-custody form ensuring full details of the specimen identity is provided.
6. The specimen can be stored at room temperature until transportation.
7. Send the specimen and completed form to the Special Projects & Research section of the Victorian Institute of Forensic Medicine, 57-83 Kavanagh Street, Southbank, 3006,

attention Dr Jim Gerostamoulos. The telephone number is (03) 9684 4342, and fax is (03) 9682 7353. The e-mail address for correspondence is jimg@vifp.monash.edu.au.

Part IV

Screening and assessment of psychiatric problems

Measures used to assess general psychological state

Symptom Checklist-90-Revised (SCL-90-R)

Key reference: Derogatis, L.R. (1994) *Symptom Checklist-90-Revised: Administration, scoring and procedures manual, 3rd edition*. National Computer Systems, Inc., Minneapolis, MN 55440.

Summary

The Symptom Checklist-90-Revised (SCL-90-R) is a 90-item self-report instrument designed to measure current psychological and psychiatric symptoms. It is a widely used instrument that is reliable and valid. It has been used with substance abuse populations and has been found to perform better than other general measures of psychological functioning. A brief 53-item version is currently available (Brief Symptom Inventory, BSI). The SCL-90-R and the BSI can only be purchased by Registered Psychologists with post graduate qualifications in Psychology. Scoring and interpretation must be supervised by a Registered Psychologist.

Description and development of the Symptom Checklist-90-Revised (SCL-90-R)

The SCL-90-R is a revised and updated version of the Hopkins Symptom checklist and the SCL-90. It is a 90-item self-report questionnaire designed to assess psychological problems and symptoms of psychopathology. Nine primary symptom dimensions are measured: Somatisation, Obsessive Compulsive, Interpersonal Sensitivity, Depression, Anxiety, Hostility, Phobic Anxiety, Paranoid Ideation, and Psychoticism. It should be noted that these are symptom dimensions and therefore do not correspond directly with a diagnosis based upon either DSM-IV or ICD-10 nomenclature. However, elevated scores on any subscale are an indication that further assessment of the client's mental state is warranted.

The SCL-90-R also provides three summary scores. The Global Severity Index (GSI) is a composite score obtained by summing the scores on the nine symptom dimensions and dividing this score by the total number of items (usually 90 if there are no missing responses). According to Derogatis (1994) the GSI is the best single indicator of severity of disorder and should be used in most instances where a single summary measure is required. The Positive Symptom Distress Index and Positive Symptom Total reflect the intensity and extensiveness of symptoms respectively.

Reliability and validity of the SCL-90-R

The SCL-90 and the revised SCL-90-R have been extensively used in research and clinical practice. There is considerable information regarding reliability and validity of the instrument as a whole and for each individual subscale. A review of the literature is beyond the scope of the present review and therefore only a brief overview of the information contained in the manual (Derogatis, 1994) is provided. Additionally, selected research reports, in which the SCL-90-R has been used with individuals with substance misuse problems, will be reviewed. The nine symptom dimensions have shown very good internal reliability with *alpha* coefficients for each of the dimensions ranging from .79 for Paranoid

Ideation to .9 for Depression (Horowitz et al., 1988). Furthermore, the SCL-90-R has shown good test-retest reliability (Horowitz, et al., 1988).

In terms of validity, there are a large number of published reports indicating that the SCL-90-R performs better than most instruments in both assessment and in measuring change following treatment (see Derogatis, 1994). However, the nine-factor structure of the SCL-90 and SCL-90-R has been questioned when used with non-normative samples. Some studies with inpatient and outpatient samples have indicated a more variable factor structure (Cyr, McKenna-Folwy & Peacock, 1985). The majority of factor analytic studies, though, have indicated the checklist taps a single predominant factor reflecting general psychological distress (Bonyng, 1993; Rauter, Leonard & Swett, 1996). Carpenter & Hittner, (1995) for instance found a unidimensional factor structure that accounted for a large proportion of the variance in a sample of co-morbid inpatient and outpatient substance users. Similarly, a confirmatory factor analysis of the BSI (a shorter version of the SCL-90-R) with substance users failed to confirm the nine-factor structure, again extracting a unidimensional factor of global psychological distress (Benishek, Hayes, Bieschke & Stoffelmayr, 1998). Overall evidence for the nine-factor structure of the SCL-90, SCL-90-R and BSI is sparse. Rather, the SCL-90-R and the BSI appear to provide a unidimensional measure of general psychological distress.

There are two significant studies investigating sensitivity and specificity of the SCL-90-R. In the first, the sensitivity and specificity of the SCL 90-R was compared with that obtained with the Present State Examination (a comprehensive diagnostic interview) in two samples of patients: one with chronic physical disease (diabetes mellitus) and the other with bulimia nervosa (Peveler and Fairburn, 1990). The majority of the subscales of the SCL-90-R corresponded closely with their PSE counterparts. The optimum Global Severity Index cut-off points were .68 (sensitivity 72%, specificity 87%) in the diabetic sample and 1.00 (sensitivity 77%, specificity 91%) in the bulimic sample. A more recent study, Mattick et al. (submitted for publication) compared the SCL-90-R, the General Health Questionnaire (GHQ) and the Beck Depression Inventory (BDI) with DSM-III diagnostic categories “depression”, “anxiety” and “any diagnostic category” derived from the CIDI in methadone maintained subjects. The SCL-90-R Global Severity Index and the symptom dimensions Depression and Anxiety were superior to the GHQ-28 in detecting the relevant disorder.

The utility of the SCL-90 compared with the Addiction Severity Index was recently examined for the screening and diagnosis of anxiety and mood disorders in substance abuse patients. The SCL-90 emerged as a better predictor of CIDI diagnosed anxiety and mood disorders than the ASI—psychiatric problems scale displaying a moderate degree of specificity and high sensitivity (Franken & Hendriks, 2001). More recently, a Mania scale was developed from existing items of the SCL-90, which correctly identified approximately two-thirds of manic patients (Hunter et al., 2000).

Short forms of the SCL-90-R and the SCL-90 have been developed. The Brief Symptom Inventory (BSI) consists of 53 items and measures the same 9 primary symptom dimensions as the SCL-90-R (Derogatis & Melisaratos, 1983). Correlations of the BSI dimensions with the SCL-90-R range from .92 to .99. The BSI has demonstrated adequate internal consistency ranging from .70 (phobic anxiety) to .89 (depression) and test-retest reliability ranging from .68 (somatisation) to .91 (phobic anxiety) over a two-week period (Boulet & Boss, 1991; Derogatis & Melisaratos).

The Symptom Assessment-45 (SA-45; Davison et al., 1997) consists of nine 5-item scales assessing the same symptom domains as the SCL-90. The majority of scales display adequate internal consistency reliabilities (.7 - .8) across different age and patient status variables. It has also demonstrated discriminant validity between controls and adolescent and adult patients and between patients at intake and follow up.

A recent study developed 6- and 10-item indices of psychological distress based on the SCL-90 (Rosen et al., 2000). Items that were most indicative of psychological distress were identified based on a review of eight factor analytic studies of the SCL-90. The convergent validity of the 2 new indices and the previously developed SCL-10 was examined in a sample of 323 patients with PTSD. The three indices correlated .87 to .97 with the SCL-90 and Brief Symptom Inventory respectively.

Suitability for special populations

The SCL-90-R has been used with diverse populations and has been translated into many languages. It is suitable for both male and female respondents. To our knowledge specific field trials or assessments have not been conducted with Indigenous Australians. The SCL-90's predecessor, the Hopkins Checklist, though has been used successfully with Indigenous people (Hunter, 1993) and it seems likely that the SCL-90-R would also be able to provide helpful diagnostic information.

Administration and scoring

There is both a pen and paper and computerised version of the SCL-90-R. The former takes 12 – 15 minutes to complete, is designed for adolescents over the age of 13 years and for adults. A Year 8 reading age is required.

Availability and cost

The SCL-90-R is a copyrighted instrument and therefore cannot be reproduced. It is published by NCA Assessments and distributed in Australia by Psychological Assessments Australia (PAA). All test purchasers must be Registered Psychologists with post-graduate qualifications in Psychology.

General Health Questionnaire (GHQ)

Key reference: Goldberg, D. and Williams, P. (1988). *A Users's Guide to the General Health Questionnaire*. NFER-NELSON Publishing Co. Ltd. Windsor, Berkshire, UK.

Summary

The General Health Questionnaire (GHQ) is a self-administered screening test sensitive to the presence of psychological symptoms. It was designed as a screening instrument and it is less sensitive to the presence of psychological disorders than the more comprehensive SCL-90-R. It has both adequate reliability and validity and is easily administered and scored. In recent years the 12-item version of the instrument has been extensively used. It has been found to be a good screening instrument, performing as well as the longer GHQ-28. The GHQ may be used by a range of health and mental health professionals and is incorporated into the OTI.

Description and development of the General Health Questionnaire (GHQ)

The General Health Questionnaire (GHQ) was developed at the Maudsley Hospital, South London, England. It was designed as a self-administered screening test sensitive to the presence of psychiatric disorders in individuals presenting in primary care settings and non-psychiatric clinical settings. The GHQ is not designed to detect symptoms that occur with specific psychiatric diagnoses such as psychotic disorders; rather, it provides a measure of overall psychological health or wellness. To assess this, the GHQ focuses on two major classes of phenomena: (i) inability to continue to carry out normal “healthy” functions and (ii) symptoms of a distressing nature (Goldberg and Williams, 1988, p 5).

There are several versions of the GHQ. The original GHQ containing 60 items was derived from factor analysis of a checklist of 140 items. Shorter versions of the GHQ have been developed from the GHQ-60, the most widely used being the GHQ-30, although there is also a GHQ-28. The GHQ-28 provides four specific subscales: somatic symptoms, anxiety and insomnia, social dysfunction, and severe depression. In recent years the 12 item GHQ (GHQ-12) has been used extensively. Factor analytic studies of the GHQ-12 have yielded 2 and 3 factor solutions (Burvill & Knuimans, 1983; Worsley & Gribbin, 1977). It is important to note that the subscales of the GHQ do not necessary correspond to psychiatric diagnoses, nor are the subscales independent of each other (Goldberg and Williams, 1988, p 41).

Reliability and validity of the GHQ

The GHQ is a widely used measure of psychological health and consequently, there is a large literature on its reliability and validity. A comprehensive review of the literature is contained in the manual (Goldberg and Williams, 1988), and a brief summary of this will be reported.

The GHQ has reasonable test-retest reliability although a definitive study is yet to be conducted. The reliability co-efficients are higher in studies in which there is a high prevalence of disorder and in which the GHQ is administered within a relatively short period of time (e.g., 5-7 days) and range from .85 to .90 (De Paulo and Folstein, 1978; see also Goldberg and Williams, 1988). When using a sample drawn from the general population, the reliability coefficients decrease substantially. For example, when the reliability of the GHQ was assessed twelve months apart in a sample of school leavers and men facing redundancy, the test-retest correlations were .58 and .51 respectively (Layton, 1986).

The GHQ has both content validity and construct validity. Studies regarding these issues are reviewed in detail by Goldberg and Williams (1988). Studies in which criterion validation has been reported are also discussed in detail in Chapter 6 (Goldberg and Williams, 1988). There are 22 studies reporting on the correlations between GHQ scores and a standardised psychiatric assessment. The median correlation between GHQ and the criterion interview was .72 for the GHQ-60, .59 for the GHQ-30, .76 for the GHQ-28 and .70 for the GHQ-12. To assess overall sensitivity and specificity, Williams, Goldberg and Mari (cited in Goldberg & Williams) used information from 43 independent studies to obtain an overall measure. The GHQ-12 had a sensitivity of 89% and specificity of 80%; the GHQ-28 had a sensitivity of 84% and specificity of 82%; the GHQ-30 has a sensitivity of 74% and specificity of 82% and the GHQ-60 had a sensitivity of 78% and a specificity of 85%.

While these overall figures are high, two studies conducted with substance misuse populations indicate that the GHQ's sensitivity and specificity is generally lower than those in general medical or community samples. Ross and Glasser (1989) found a sensitivity of 82% and a specificity of 55% for the GHQ-60 using 11 to 12 as the optimal threshold score. Mattick et al. (submitted) reported a sensitivity of 68% and a specificity of 69% for the GHQ-28 using a cut-off score of 6. Despite these findings, the GHQ has sufficient reliability to warrant its use as a screening instrument for clinicians working with a substance misusing population.

In recent years, the GHQ-12 has been extensively used. It has been found to provide a reliable measure of psychological symptoms with Cronbach alphas ranging from .82 to .86 (Sriram, Chandrashekar, Isaac & Shanmugham, 1989; Politi, Piccinelli & Wilkinson, 1994; Winefield, Goldney, Winefield & Tiggermann, 1989). A review of 17 studies conducted across nine countries found high median sensitivity (83.7%) and specificity (79.0%) values with a modal optimal cut-off score of 2-3 (Goldberg et al., 1997).

A number of recent studies have compared the utility of the GHQ-12 and GHQ-28. Werneke, Goldberg, Yalcin & Uestuen (2000) investigated the stability of the factor structure of these two measures. Factor analysis of the GHQ-12 extracted 2 factors comprising depression and social dysfunction. The GHQ-28 identified four factors consisting of the original social dysfunction and depression and two less stable additional factors – somatic symptoms and anxiety factors. Overall, the factor structure of the GHQ-28 item was more stable than the GHQ-12. However, these values varied considerably, depending upon the clinical comparison used. A WHO comparison of the validity of the GHQ-28 and GHQ-12 was conducted using a sample of 5438 patients in general health care settings across 15 sites including North America, Europe and developing countries (Goldberg et al., 1997). The ability of the GHQ-12 and GHQ-28 to correctly identify CIDI diagnoses was examined using ROC analysis. Across the 15 sites the GHQ-12 displayed an overall sensitivity of 83.4% and specificity of 76.3% for detecting CIDI diagnoses. Optimal cut-off scores varied across sites ranging from 1 to 7 but tended to be optimal with a cut-off score of 1-2 in most sites. Similarly, the GHQ-28 exhibited an overall high level of sensitivity (79.7%) and specificity (79.2%). However, these were no higher than the GHQ-12, although there tended to be less variation in optimal cut-off scores across centres. The use of simple versus complex scoring methods, translation of the instruments, use in developing countries, and gender, age and educational variables had no significant effect on the sensitivity or specificity of the instruments. Therefore, the authors concluded the use of the GHQ-12 with simple scoring methods to be equivalent to the lengthier GHQ-28.

Previous studies have suggested the GHQ-60, 30, 28 and 12 item versions as susceptible to retest effects with scores tending to decline over time with multiple applications. However, a recent study examining retest effects of the GHQ-12 over a 7-year period found no retest effects even after controlling for the effects of age and gender (Pevalin, 2000).

Suitability for special populations

The GHQ has been translated into 38 languages and has been used in diverse cultural groups. As it is primarily concerned with detecting "psychological illness", the items appear to have cross-cultural relevance despite cultural variations in the expression of mental illness. Indeed, a recent body of work has focused on the psychometric properties of the GHQ-12 across a range of cultural groups. The validity of the GHQ-12 has been established among samples of German primary care patients (Schmitz & Kruse, 1999) and English

health care services employees (Hardy, Shapiro, Haynes & Rick, 1999). Additionally, it has been found to be predictive of CIDI current and lifetime diagnoses among Mexican adults (Caraveo, Jorge, Martinez, Saldivar, Lopez & Jorge, 1998) and to exhibit high sensitivity and specificity among ethnic Indians living in the UK (Jacob, Bhurgra & Mann, 1997). It has also been validated among Turkish primary care attendees (Kilic et al., 1997) and Chinese primary care patients (Pan, Pey-Chyou & Goldberg, 1990).

Given the above, it is probable to suggest that the items on the GHQ reflect universal aspects of psychological distress that are equally relevant to Indigenous Australians. Furthermore, based on work with the GHQ's predecessor, the Hopkins Symptom Checklist-20, items scored on a Likert scale rather than dichotomous responses tend to be more acceptable to Indigenous people living in remote parts of Australia (Hunter, 1993). Indeed, Hunter's experience suggests that the most successful scoring of a questionnaire involved the development of a response category that relied upon visual discrimination. Further research is required to determine whether this is also the case with the GHQ. There are no issues specific to women or other special interest groups.

Administration and scoring

The great advantage of the GHQ-12 and GHQ-28 is its short administration time (2-10 minutes) and its availability to a range of professionals. Unlike the SCL-90-R, its use is not restricted to Registered Psychologists. Indeed, it was specifically developed for use by a broad range of clinicians working in community and non-psychiatric settings, and is therefore a very appropriate instrument for use by Drug and Alcohol Workers.

Furthermore, a recent comparison of the GHQ-12 and SCL-90-R for identifying SCID diagnoses in primary care patients found no differences between the two measures for detecting cases (Schmitz, Kruse, Heckrath, Alberti & Tress, 1999).

The manual is clearly written and provides a comprehensive review of the literature. The sections describing the statistical procedures used in the development of the GHQ are particularly well written and provide sufficient information for the non-statistician to understand the basic principles and procedures involved. Its use is strongly recommended for clinicians working in the substance misuse field.

Availability and cost

The GHQ-28 is incorporated into the OTI. All other versions are copyright protected.

Brief Psychiatric Rating Scale (BPRS)

Key reference: Overall, J.E. and Gorham, D.R. (1962). The brief psychiatric rating scale. *Psychological Reports*, 10, 799-812.

Summary

The Brief Psychiatric Rating Scale (BPRS) is a clinician-administered scale in which 18 items are rated on a continuum of “not present” to “extremely severe”. When used by clinicians trained in the assessment and diagnosis of psychopathology it is reliable and valid. However, the degree of expertise required for reliable administration is considerable and therefore it is most appropriately administered by clinical mental health professionals with specific training in the administration and scoring of the BPRS.

Description and development of the Brief Psychiatric Rating Scale (BPRS)

The Brief Psychiatric Rating Scale (BPRS) is a clinical rating scale widely used in psychiatric clinical practice. It was developed in the early 1960's (Overall and Gorham, 1962) with an original sample of 6000 psychiatric inpatients. Since then at least four versions of the BPRS have been developed. For the purposes of the present report, only the 18-item version will be reviewed.

The BPRS, unlike the SCL-90-R and GHQ, is a clinician rated instrument. Ratings are made after a brief (15-20 minutes) unstructured interview with the patient. Each item is rated on a 7-point scale ranging from “not present” to “extremely severe”.

Reliability and validity of the BPRS

The BPRS has been used in a large number of research studies and consequently there is a substantial literature in which the reliability and validity of the scale has been investigated. A 1980 review of the psychometric studies on the BPRS concluded that “when the BPRS is properly used, inter-rater reliability is generally satisfactory”. The BPRS is a sensitive and effective measure both of psychopathology and of treatment-related symptom changes (Hedlund and Vieweg, 1980; cited in Hafkenscheid 1991). Early reports of the BPRS found that measures were stable across time and had high inter-rater reliability (e.g., Flemenbauch and Zimmermann, 1973). However, the generality of these findings has been questioned by Hafkenscheid (1991), who found unacceptably low inter-rater reliability. While this does not mean that the BPRS is without clinical utility, it does emphasise the need for users to have extensive knowledge of diagnostic concepts and categories, and be trained in administering the BPRS.

Suitability for special populations

There are no gender or cultural issues specific to the BPRS. However, as it is a clinician rated instrument, the attitudes and beliefs held by the clinician will have some bearing on the rating of particular diagnostic categories. While broadly speaking the validity studies of the BPRS indicate that this is not a major issue, use of the BPRS with people from ethnic or cultural groups should take into account the individual's cultural frame of reference when making a judgement on presence of psychopathology.

To the authors' knowledge the BPRS has not been tested for use with Indigenous Australians. The usefulness of this scale needs to be evaluated for cultural relevance.

The BPRS has been used with individuals with a primary substance misuse problem. Steer and Schut (1979) reported its use with heroin addicts and identified anxiety-depressive symptoms as the most prevalent disorder. None of the substance-abusing participants showed as high levels of thought disturbance as psychiatric controls. A similar finding was reported by Westermeyer, Tucker and Nugent (1995) in a sample of newly abstinent alcoholic and substance abuse patients. The BPRS has also been used in various research studies in which the relationship between major psychiatric illness and substance misuse has been investigated (Caspari, 1999; Dawe et al, submitted; Dixon et al., 1991; Sembhi & Lee, 1999; Warner et al., 1994).

A recent pilot study examined the usefulness of the BPRS as an acute inpatient outcome measurement tool (Varner, Chen, Swann & Moeller, 2000). The BPRS was found to demonstrate significant change during brief stays of 1 week or less. This suggests that the BPRS may be useful for aiding clinical decisions regarding suitability for discharge.

Administration and scoring

The BPRS is a reliable and valid instrument when used by individuals who are trained in the assessment and diagnosis of psychopathology. Generally, this is most likely to refer to clinical psychologists or psychiatrists and obviously limits the usefulness of the BPRS. However, if an agency decided that routine use of the BPRS could be usefully incorporated into a battery of assessment and screening instruments then training could be given to clinicians with a background in psychology or psychiatric nursing. In view of the difficulties that have been reported regarding inter-rater reliability, it is advisable that those clinicians involved in administration of the BPRS, conduct ongoing training and reliability checks.

As this scale is clinician rated, it does not require the client to read or write. This may be helpful when assessing individuals who are either illiterate or who are unable to read English.

Availability and cost

The BPRS is best used by adequately trained mental health professionals only.

Royal Park Multidiagnostic Instrument for Psychosis

Key References: McGorry, P. D., Copolov, D. L. & Singh, B. S. (1990). Royal Park multidagnostic instrument for psychosis: Part 1. Rationale and Review, *Schizophrenia Bulletin*, 16 (3), 502-515..

McGorry, P. D., Singh, B. S., Copolov, D. L., Kaplan, I., Dossetor, C.R. & van Riel, R. J. (1990). Royal Park multidagnostic instrument for psychosis: Part II. Development, reliability and validity, *Schizophrenia Bulletin*, 16 (3), 517-535.

Summary

The Royal Park Multidiagnostic Instrument for Psychosis (RPMIP) was developed to assess psychopathology during an acute psychotic episode. Multidiagnostic criteria, based on multiple information sources, provide a diagnostic profile of patients. Acceptable levels of inter rater reliability of the RPMIP had been found despite the multiple sources of information used. There is satisfactory evidence that the scale has good agreement with clinician derived consensus diagnoses. It only had a modest level of agreement with other methods used to assign DSM-III-R diagnoses of psychotic disorders. The RPMIP requires specialised training ; the level of expertise and training required for its raters and its high level of complexity suggests that use should be restricted to appropriately trained mental health professionals.

Description and Development of the RPMIP

The Royal Park Multidiagnostic Instrument for Psychosis (RPMIP) was developed to assess psychopathy during an acute psychotic episode (McGorry, Copolov, & Singh, 1990). It is based on serial interviews and multiple information sources and produces a diagnostic profile of patients. This profile provides a comprehensive picture of the acute psychotic episode from onset to termination or stabilisation. The RPMIP was developed using a “validity orientated” approach due to dissatisfaction with the unreliability of psychiatric diagnoses, and the lack of comparative validity of different operational criteria of disorders. It seeks to overcome these limitations by focusing primarily on the acute psychotic episode, uses multidagnostic criteria based on a range of operationalised diagnostic concepts and uses multiple information sources. Although it was developed primarily for an acute psychotic episode, the RPMIP can be used to assess a current episode among patients with an established psychotic disorder who relapse.

The RPMIP consists of a number of components. The first component, the interview schedule, provides an overview of the instrument followed by assessment of duration and onset, alcohol/drugs screening, affective disorders, psychotic symptoms, prodromal/residual symptoms, syndromal pattern and observational and miscellaneous ratings. It is administered on two separate occasions. The illness duration interview is the second component of the RPMIP. It is administered to relevant informants of the patient and details the presence, duration and sequence of psychopathology before admission. A discharge score sheet is used to collate data from the multiple information sources. A glossary and guideline document assists the completion of the interview. Diagnostic decision rules based on algorithms are used either manually or in a computerised version to derives a list of diagnoses for the patient. Finally, a summary sheet of diagnoses is completed for each patient.

Reliability and validity of the RPMIP

The inter-rater reliability of the RPMIP was examined in a sample of 50 young adults with recent onset psychosis (McGorry, Singh, et al., 1990). Generally, the reliability of nearly all diagnostic categories, except three, was good to excellent ($\kappa > .60$). The reliability of individual items and subsections of the RPMIP was also examined. The mean (unadjusted) κ for all 260 items was good at .70, with over 90% of items falling above the minimal acceptable κ (.40). Similarly all subsections of the RPMIP (e.g., depression, drug & alcohol use) achieved good to excellent test-retest reliability ranging from .62 to .82 (McGorry, Singh et al., 1990).

The procedural validity of the RPMIP was also evaluated by comparing RPMIP DSM-III diagnoses with clinician diagnostic consensus among 87 young adults with recent onset psychosis. Results indicated that there were satisfactory levels of agreement between the two methods. In a later study the procedural validity of the RPMIP was evaluated by examining its level of concordance with other diagnostic procedures. There was a moderate level of agreement between the RPMIP and other methods for deriving DSM-III-R diagnoses including the DSM-IV field trial instrument ($\kappa = .67$; 76% agreement), DSM-III-R consensus diagnosis ($\kappa = .65$; 74% agreement) and the Munich Diagnostic Checklist ($\kappa = .65$; 74% agreement; McGorry et al., 1995).

An exploratory factor analysis of the RPMIP was conducted to determine the dimensional structure of first episode psychosis in a large representative sample. A four factor solution emerged which accounted for 43.6% of the variance, comprising depression, mania, negative/disorganised/catatonic features and positive psychotic symptoms.

Administration and Scoring

The RPMIP is time consuming, taking approximately 6 to 7 hours (4 hours with computer assistance) to administer and score. The rater must have a high level of expertise to elicit and rate complex psychopathology from multiple sources. The level of expertise required may include psychiatrists, clinical psychologists and experienced psychiatric nurses who have been trained in administering the instrument. The training required is intensive, incorporating videotape review and discussion, interviewing under direct supervision, extensive reading in psychopathology and direct training in the definition of RPMIP items and their ratings.

Use with special populations

Given the specialised and time consuming nature of the RPMIP, its complexity and the level of expertise and training required for its raters, its use is probably restricted to research settings primarily focused on patients with acute psychosis.

Availability

Requests for a copy of the RPMIP can be directed to Dr P. D. McGorry, The National Health and Medical Research Council Schizophrenia Research Unit, Royal Park Hospital, Private Bag 3, Parkville, Melbourne, Victoria, 3052, Australia.

Screening and assessment of psychiatric problems in Indigenous Australians

The publication of the National Consultancy Report on Aboriginal and Torres Strait Islander Mental Health proposes that a holistic approach to Indigenous health be taken, in which mental, physical, cultural and spiritual health are seen as interlinked (Swan, 1999). It is also proposed that Indigenous mental health be approached within a socio-emotional context that encompasses issues such as “oppression, racialism, environmental circumstances, economic factors, stress, trauma, grief, cultural genocide, psychological processes and ill health.” (p.15). Culturally appropriate models and frameworks for intervention with indigenous clients have been proposed (Vicary, 2000; Wright, 2000). Such frameworks of indigenous mental health take into account the unique social and historical contexts contributing to Indigenous health and wellbeing.

Within such culturally sensitive frameworks it is acknowledged that existing diagnostic screening measures can only provide a partial insight into indigenous mental health functioning. Although psychiatric assessment measures are used with indigenous people there has been no formal research validating the reliability of such measures with this population, nor has research defined the nature and extent of mental health issues for indigenous people. Although practitioners in the field reluctantly use traditional diagnostic screening assessments with indigenous clients, the absence of appropriate norms, coupled with concerns regarding the inherent cultural bias of psychological testing (Drew, 2000), indicate that results must be interpreted with caution and care. In Australia, the principal document for mental health practitioners working with Indigenous Australians is the Australian Psychological Association’s “Guidelines for the Provision of Psychological Services for and the Conduct of Psychological Research with Aboriginal and Torres Strait Islander People of Australia” (Australian Psychological Society, 1996). Reference to this document is recommended to familiarise practitioners with relevant issues. Another valuable resource for practitioners working in this area is the book *Working with indigenous Australians: A handbook for psychologists* (Dudgeon, Garvey, & Pickett, 2000), which provides clear direction and practical advice for practitioners working within a cross cultural context.

In response to the increasing prevalence of suicide and self-harming behaviour in Indigenous communities, indigenous psychologist Tracey Westerman of Curtin University, has developed a suicide-risk screening device - The Westerman Aboriginal Symptom Checklist - Youth (WASC-Y). The WASC-Y aims to identify 13-17 year old Indigenous youth at risk of suicide, anxiety and depression. This measure has the potential to enable large scale screening of indigenous youth to identify those at risk and also provides a reliable measure on which to assess level of change. The psychometric properties of the WASC-Y are currently being investigated in Perth and the north-western region of Western Australia. For further information about this measure and details regarding use and accreditation contact t.westerman@psychology.curtin.edu.au.

Measures used to assess specific disorders

Beck Depression Inventory (BDI)

Key references: Beck, A.T., and Steer, R.A., (1978). *Beck Depression Inventory Manual*. The Psychological Corporation, San Antonio.

Beck, A.T., Steer, R.A., and Brown, G.K. (1996). *Beck Depression Inventory II Manual*. The Psychological Corporation, San Antonio.

Summary

The Beck Depression Inventory (BDI) is one of the most widely used self-report measures of depression. It is a reliable and valid measure of depression in a range of cultural groups and has been validated with both psychiatric and non-psychiatric populations. The BDI has been widely used with substance misusers and is therefore recommended as a useful screening instrument.

Description and development of the Beck Depression Inventory (BDI)

The Beck Depression Inventory (BDI) has become one of the most widely used instruments to assess depression. It was initially based upon clinical observations and descriptions of symptoms frequently experienced by depressed patients. The items were chosen to assess the severity of depression and do not reflect any particular theory of depression (Beck & Steer, 1978). A more recent version of the instrument was developed to correspond to DSM-IV criteria for depression (Beck, Steer & Brown, 1996). The BDI and BDI-II, although sensitive to the presence of depressed mood, are not diagnostic instruments, and therefore an elevated score on the scale does not equate with a diagnosis of depression but rather indicates the presence of depressed mood. A diagnosis of depression should only be arrived at after conducting a clinical interview (Beck & Steer, 1978).

The BDI assesses 21 symptoms and attitudes including pessimism, sense of failure, self-dissatisfaction, guilt, self dislike, suicidal ideas, social withdrawal, indecisiveness, body image change, insomnia, fatigability, weight loss, somatic preoccupation, and loss of libido in the week preceding administration (see Andrews, Peters and Teesson, 1994). The BDI-II still measures 21 symptoms but the somatic preoccupation, weight loss, body image change and work difficulty items have been replaced with agitation, concentration difficulties, worthlessness and loss of energy to correspond more closely with DSM-IV criteria. Furthermore, the timeframe of the instrument has been changed from one to two weeks. Items on both scales are rated on a 4-point scale ranging from 0 to 3. Beck and Steer (1978) emphasise the importance of attending to elevated scores (i.e., 2 - 3) on items relating to suicide ideation (item 9) and hopelessness (item 2) as these items have been found to be nearly as predictive of eventual suicide as the 20-item Hopelessness Scale (see also Keller and Wolfersdorf, 1993).

Reliability and validity of the BDI

The psychometric properties of the BDI have been widely studied and have been extensively reviewed (e.g., Steer et al., 1986). *Alpha* reliability coefficients range from .76 to .95 in psychiatric samples and from .73 to .92 in non-psychiatric samples (Beck, Steer and Garbin,

1988) indicating that the BDI has good internal consistency. It also has high test-retest reliability, with correlations ranging from .48 to .86 with psychiatric patients and from .60 to .83 with non psychiatric groups (Beck, Steer and Garbin). Regarding substance misusers, the BDI has been found to be a reliable measure of depression (e.g., Kleinman et al., 1990).

The BDI's validity in measuring the construct of depression has been extensively researched. Beck et al. (1988) report significant correlations between clinical ratings of depression and scores on the BDI. High correlations between the BDI and other rating scales have also been reported, particularly the depression subscale of the SCL-90-R. The BDI has demonstrated discriminant validity between depressed and non-depressed patients although its ability to differentiate anxiety from depression has been criticised (Richter, Werner, Heerlein, Kraus & Sauer, 1998).

Studies of the factor structure of the BDI have indicated that it measures between 1 and 7 factors, depending on the composition of the sample, type of extraction method used and criteria for the estimation of the factor number (Beck, Steer & Garbin, 1988). The majority of studies have supported a unidimensional factor structure assessing a general syndrome of depression, that can be subdivided into three highly interrelated factors - negative attitudes towards self, performance impairment, and somatic disturbance (Beck, Steer & Garbin, 1988; Gotlib & Cane, 1989; Welch, Hall & Walkey, 1990 in Dozois et al, 1998). The BDI appears sensitive to change and is often used to evaluate treatment outcome (e.g., Richter et al., 1998).

A recent meta-analysis of 63 studies that examined the psychometric properties of the BDI concluded that the BDI had a number of limitations and advantages (Richter et al., 1998). Limitations of the BDI included its high item difficulty despite its wide application, lack of content validity for assessing DSM-IV criteria and absence of representative normative data, its variable factor structure, unstable temporal stability and poor discriminant validity with anxiety. The advantages of the BDI include its high internal consistency, discriminant validity between depressed and non-depressed patients and its sensitivity to change.

Regarding sensitivity and specificity in substance misuse populations, Rounsaville et al. (1979) found that the BDI had greater sensitivity and specificity than either the SCL-90 or the Raskin Depression Scale when compared with Research Diagnostic Criteria for depression in opiate addicts. The usefulness of the BDI in screening for depression in cocaine addicts was investigated by Weiss et al. (1989). In this study, the BDI offered the best combination of sensitivity and specificity compared with the Hamilton Rating Scale for Depression and the SCL-90. However, the low specificity of all three scales led the authors to propose that the BDI may be of limited use as an initial screening instrument in cocaine abusers. A recent Australian study (Mattick et al., submitted for publication) found that the BDI performed and the SCL-90-R and better than the GHQ in detecting depression. Maximum sensitivity (73%) and specificity (73%) was obtained with a cut-off score of 18.

Psychometric evaluation of the BDI-II indicated it had a high level of internal consistency with estimates of Cronbach's α ranging from .89 to .93 (Dozois et al., 1998; Steer & Clark, 1997; Whisman, Perez & Ramel, 2000). Beck et al. (1996) reported a 2 factor solution consisting of somatic-affective and cognitive symptoms of depression among psychiatric outpatients and college students. This has been replicated in studies of adolescent psychiatric outpatients, undergraduate college students and primary medical care patients (Arnau, Meagher, Norris & Bramson, 2001; Dozois et al.; Steer & Clark, 1997; Steer, Kumar, Ranieri & Beck, 1998; Whisman et al.). These findings suggest the BDI-II has a

more stable factor structure than the BDI. The BDI-II has displayed construct validity for clinical and self-rated depression correlating highly with the BDI, revised Hamilton Rating Scale for Depression and the depression subscale of the SCL-90-R (Dozois et al.; Riskind, Beck, Brown & Steer, 1987; Steer, Ball, Ranieri & Beck, 1997). It has also displayed discriminant validity between clinician and self-rated anxiety and depression, correlating more positively with scores on the revised Hamilton Psychiatric rating scale for depression and depression subscale of the SCL-90-R than scores on the revised Hamilton Psychiatric rating scale for anxiety and anxiety subscale of the SCL-90-R, respectively (Riskind et al., 1987; Steer et al., 1997). The optimum cut-off scores of the BDI-II differ from the BDI. The optimal cut-off scores for the BDI-II are: 0-12 non-depressed, 13-19 dysphoric; 20-63 dysphoric or depressed. These cut-off scores correctly classified 91% of clinically depressed college yielding a sensitivity of 81% and specificity of 92%.

Suitability for special populations

The BDI has been used in a range of cultural settings. A review of the literature since 1990 found reports attesting to the reliability and validity of the BDI in Spanish (Bonicatto, Dew & Soria, 1998; Torres et al., 1991), Arabic (Hamdi et al., 1988), German (Richter et al., 1991), Bulgarian (Byrne, Baron & Balev, 1996), Swedish (Byrne, Baron, Larsson & Melin, 1996) and Chinese (Xu, 1991). In addition to this, the BDI was found to be reliable in a number of different ethnic groups such as African-American college students (Blanton Lacy, 1997) and Black South Africans (Westaway & Wolmarans, 1992). An examination of the readability of the BDI indicated an eighth or ninth grade education was required for individuals to comprehend the inventory (Beckman & Lueger, 1997). Given that the BDI performs well in these diverse cultural settings, it is probable that the BDI may be appropriately used with Indigenous Australians. However, to the authors' knowledge, the BDI has not been routinely used with this population.

Administration and scoring

The BDI and BDI-II take approximately 5 - 10 minutes to complete when self-administered. Oral administration may take considerably longer depending upon the setting and characteristics of the individual completing it. Each statement is scored on a 4-point scale (0 - 3) and a total score is obtained by summing the ratings for each statement. As a general guideline a score from 0 - 9 is considered to be within the normal range or asymptomatic; a score of 10 - 18 indicates mild-to-moderate depression; a score of 19 - 29 indicates moderate-to-severe depression and a score of 30 or more indicates extremely severe depression (Beck and Steer, 1978, p. 7). Full administration and scoring guidelines are provided in the Manual (Beck and Steer, 1978). Computer-administered versions of the BDI and the BHS have been shown to have good discriminant reliability (Steer et al., 1995).

Availability

The Beck Depression Inventory is copyright protected and may not be reproduced without permission. A copy may be purchased from The Psychological Corporation, PO Box 9959, San Antonio, TX 78204-0959 or at URL: <http://www.psychcorp.com>. Scoring and interpretation must be supervised by a Registered Psychologist.

Beck Hopelessness Scale (BHS)

Key reference: Beck, A.T. and Steer, R.A. (1987). *Beck Hopelessness Scale Manual*. The Psychological Corporation, San Antonio.

Summary

The Beck Hopelessness Scale (BHS) is a well-validated measure of hopelessness and specifically targets negative attitudes about the future. Elevated scores on the BHS have been associated with suicide attempts. The scale has been widely used in a range of clinical groups and is reliable and valid. Scoring and interpretation must be supervised by a Registered Psychologist.

Description and development of the Beck Hopelessness Scale (BHS)

The Beck Hopelessness Scale (BHS) was developed to assess hopelessness and, in particular, negative attitudes about the future. It is based upon a concept of hopelessness in which negative expectations about the future are a common element in a system of cognitive schemas. The BHS comprises 20 statements that are answered true or false. These items were selected from a large pool of statements made by patients when they were depressed and not depressed and had good face validity. It is extremely important to note that elevated scores on the BHS have been associated with suicide attempts. Close scrutiny of items and further assessment of suicidal intent is always recommended if clinically significant scores are obtained.

Reliability and validity of the BHS

The BHS has been extensively used in both clinical and research settings. The manual (Beck & Steer, 1987) provides a comprehensive review of the use of the BHS across diverse samples of both clinical and non-clinical groups. The BHS has been shown to have high internal consistency ($\alpha = .93$) across seven clinical groups (Beck & Steer, 1987). It has also displayed high test-retest reliability ($r = .85$) among a sample of college students over a 3-week period (Holden & Fekken, 1988).

The factor structure of the BHS has been extensively studied. The original validation study reported a 3-factor solution – feelings about the future (41.7% of variance), loss of motivation (6.2%) and future expectations (5.6%). Some studies have replicated this factor structure and others have yielded between 1 and 5 factor solutions (see Glanz, Hass & Sweeney, 1995 for a review). The majority of studies have indicated that the first factor of the BHS assesses hope for the future, which accounts for the greatest proportion of variance (Glanz et al., 1995).

There is substantial evidence for the concurrent, criterion and discriminant validity of the BHS. The BHS has been found to be highly correlated with clinician ratings of hopelessness ($r = .64-.74$) among general outpatients and suicide attempters and was significantly correlated with the pessimism item of the BDI (.64; Beck, Weissman, Lester & Trexler, 1974). The scale's construct validity was supported by findings the BHS was more highly correlated with seriousness of suicide attempt and measures of suicidal intent than depression (Beck, Kovacs & Weissman, 1975; Minkoff, Bergman, Beck & Beck, 1973). Furthermore, suicidal patients were found to score higher on the BHS irrespective of the

severity of their depression (Ellis & Ratliff, 1986). A more recent study compared the usefulness of the BHS, Beck Scale for Suicidal Ideation (BSS), BDI and Beck Anxiety Inventory (BAI) for predicting the decision to admit suicidal patients (Cochrane-Brink, Lofchy & Sakinofsky, 2000). The BHS was the third best predictor after the BSS and the High Risk Construct Scale. A cut-off score of 15 on the BHS yielded 100% sensitivity and 71% specificity for predicting hospital admission among suicidal patients (Cochrane-Brink et al.).

The ability of the BHS for predicting suicide has been extensively investigated. Large scale prospective follow up studies using suicidal inpatients and outpatients found a cut-off score of 9 yielded high sensitivity (94.1%) in the prediction of completed suicide, although moderate to poor specificity (41.0%; e.g., Beck, Brown, Berchick & Stewart, 1990). Psychiatric outpatients with a score of 9 or more were found to be at 11 times the relative risk of committing suicide (Beck et al., 1990). However, the evidence for the predictive validity of BHS scores for predicting eventual suicide among suicide attempters is less convincing (Beck & Steer, 1987; Nimeus, Traskman-Bendz & Alsen, 1997).

Suitability for special populations

The BHS has been used in a range of cultural groups and with a diverse sample of clinical groups including substance misusers. It has been used with adolescents from age 13 years but the manual (Beck & Steer, 1987) recommends its use for individuals aged 17 years or more. Good internal consistency, stability and reliability were found in a Spanish version of the BHS in a sample of Spanish psychiatric patients (Aguilar, Hidalgo, Cano, & Lopez, 1995). A Japanese version displayed moderate internal consistency and was negatively correlated with indicators of psychological wellbeing (Tanaka, Sakamoto, Ono, Fujihara et al., 1996).

To the authors' knowledge, the BHS has not been used with Indigenous Australian groups. The actual items of the BHS refer to emotional states such as feelings of pessimism about future prospects, rather than specific daily living problems or occurrences. It is tentatively suggested that this would make it a slightly less "culture bound" instrument and therefore may be a valid measure of pessimistic views about the future. Whether this measure of hopelessness is, however, related to suicidal intent cannot be assumed in a cultural group who have experienced considerable displacement and other life events that may lead them to hold a generally pessimistic view about the future. Further research is required to evaluate the appropriateness of this instrument.

The BHS has also been used with male prison inmates and found to be predictive of suicidality among this population (Ivanoff et al., 1994). A large study of non-treatment intravenous drug users seeking HIV testing and counselling in New York investigated the relationship between self-reported severity of suicidal ideation, depression and subsequently confirmed HIV seropositivity (Steer, Iguchi & Platt, 1994). The BHS structure was comparable with that described for psychiatric patients.

Administration and scoring

The BHS may be self-completed or administered orally. It takes between 5 - 10 minutes to complete and is easily scored by summing the keyed responses of hopelessness for each of the items. As a general guideline 0 - 3 is within the normal range, 4 - 8 is mild, 9 - 14 is moderate and greater than 14 is severe. Scores range from 0 to 20. Full administration and

scoring guidelines are provided in the manual (Beck & Steer, 1987). However, readers are again cautioned about making clinical decisions about suicide risk based on self-report inventories. The BHS provides a useful adjunct to a comprehensive assessment of suicide risk. However, given the inherent difficulties in predicting suicide, the reader is cautioned against using cut-off scores to make clinical decisions about suicide risk. Such decisions should be based on a comprehensive suicide risk assessment.

The BHS may be administered by a range of mental health workers but the interpretation must be supervised by an appropriately trained clinical psychologist or psychiatrist. As it is a measure that is sensitive to suicidal intention, particular attention should be paid to individual items and individuals with scores in the clinical range should always be assessed further.

Availability

The Beck Hopelessness Scale is copyright protected and may not be reproduced without permission. A copy may be purchased from The Psychological Corporation, PO Box 839954, San Antonio, TX 78283-3954 or at URL: <http://www.psychcorp.com>.

Beck Scale for Suicidal Ideation (BSS)

Key References: Beck, A. T. & Steer, R. A. (1991a). *Beck Scale for Suicidal Ideation: Manual*. The Psychological Corporation, San Antonio

Summary

The Beck Scale for Suicidal Ideation (BSS) is a 19-item self-report scale, assessing a person's thoughts, plans and intent to commit suicide. Two additional items provide information about the number and seriousness of previous suicide attempts. The BSS has demonstrated high levels of internal consistency, temporal stability, moderate concurrent validity and discriminant validity. The BSS has also recently demonstrated predictive validity for the decision to admit suicidal patients to hospital. The earlier clinician-rated version of the BSS, the Scale for Suicidal Ideation (SSI), has recently been demonstrated to predict suicide. However, the validity of the BSS for predicting eventual suicide has yet to be demonstrated.

Description and development of the BSS

The original Scale for Suicide Ideation (SSI) was designed to be rated by a clinician to measure the severity of current suicidal ideation in psychiatric patients (Beck, Kovacs & Weissman, 1979). It was found to have high levels of internal consistency and moderately high correlations with suicide items from the BDI and clinical estimates of suicide risk (Beck, Kovacs et al., 1979). However, the SSI lacked predictive validity as psychiatric inpatients who eventually committed suicide did not score higher than inpatients who did not commit suicide (Beck, Steer, Kovacs & Garrison, 1985). The scale was modified to include standardised administration instructions and prompts, which resulted in the Modified Scale for Suicidal Ideation (MSSI).

The Beck Scale for Suicidal Ideation (BSS) was developed to provide a self-report measure to accompany the clinician rated version (Beck & Steer, 1991a). The content of the BSS

was closely based on the content of the SSI. The BSS consists of 19 self-report items assessing a person's thoughts, plans and intent to commit suicide. Two additional items provide information about the number and seriousness of previous suicide attempts. The BSS was designed to be used with adult psychiatric patients. A small number of adolescents were included in the normative sample. However, the authors note that the reliability and validity of the BSS has not been directly tested with adolescents or populations of normal adults.

Reliability and validity of the BSS

The original validation sample was based on a sample of 50 inpatients and 55 outpatients. The BSS demonstrated a high level of internal consistency with both samples ($\alpha = .87 - .90$). A moderate degree of test retest reliability was found among a subsample of 60 inpatients ($r = .54$).

Factor analytic studies have indicated the BSS had three and four factor solutions (Beck, Kovacs et al., 1979; Wetzel, 1977). A five factor solution was found among inpatient suicide ideators. The BSS has demonstrated concurrent validity with the SSI, BDI and BHS (Beck & Steer, 1991a; Beck, Steer & Ranieri, 1988) and the ability to differentiate mood disorders from other disorders and inpatient from outpatient suicidal ideators (Beck, Steer, & Ranieri, 1988).

The predictive validity of the BSS has recently been examined. The usefulness of the BHS, BSS, BAI and recently developed High-Risk Construct Scale for predicting the decision to admit suicidal patients were compared in one study (Cochrane-Brink et al., 2000). The BSS was the second best predictor of admission to hospital. A cut-off score of 24 was derived yielding 100% sensitivity and 90% specificity for predicting hospital admission among suicidal patients. Compared with other measures (BHS, NEW, BAI, BDI) the BSS had the best specificity and positive predictive value (71%; Cochrane-Brink et al., 2000). Based on these findings the BSS was considered the scale of choice for predicting admission based upon suicidal concerns.

A recent innovative study compared the predictive validity of the BHS and the Scale for Suicidal Ideation current (SSI-C) and worst (SSI-W) point in a patient's life for predicting suicide among psychiatric outpatients (Beck, Brown, Steer, Dahlsgaard & Grisham, 1999). It was hypothesised that a retrospective report of suicide ideation at its worst point would be more predictive of eventual suicide than reports of current suicidal ideation. ROC analysis indicated the optimal cut-off scores for SSI-C was 2 resulting in 58% sensitivity and 83% specificity. The optimal cut-off score for SSI-W was 16 resulting in 80% sensitivity and 78% specificity.

Administration and Scoring

The BSS takes between 5 and 10 minutes to administer. Items are rated on a three-point scale (0-2). Items one to five screen for suicidal ideation. If a respondent circles zero for items 4 and 5 they are instructed to skip items 6 to 19. Respondents only rate item 21 if they have made a previous suicide attempt. The total BSS score is based on items 1 to 19 and ranges from 0 to 38. A computer-based version of the BSS is also available. A preliminary study found the internal consistency and concurrent validity of the computer version was comparable to the paper-and-pencil version (Beck, Steer & Ranieri, 1988).

Suitability for special populations

The SSI exhibited discriminative validity between students with and without a history of suicidality and suicide attempters with African American college students (Blanton Lacy, 1997)

Availability

Paper and pencil and computer based formats are available. The BSS may be used by a range of mental health professionals. However, the administration, scoring and interpretation of the measure should be supervised by a registered psychologist. The BSS is copyright protected and may not be reproduced without permission. A copy may be purchased from The Psychological Corporation, PO Box 839954, San Antonio, TX 78283-3954 or via URL: <http://www.psychcorp.com>.

Spielberger State Trait Anxiety Scale (STAI)

Key reference: Spielberger, C.D., Gorsuch, R.L., Lushene, R., Vagg, P.R. and Jacobs, G.A. (1983). Manual for the State-Trait Anxiety Inventory (Form Y). Palo Alto, Consulting Psychologist Press, Inc.

Summary

The State-Trait Anxiety Inventory (STAI) is one of the most widely used measures of both situational or transitory anxiety (state) and more enduring personality characteristics (trait) associated with anxiety. It is reliable and valid and has been used with clinical and non-clinical populations. Scoring and interpretation must be supervised by a Registered Psychologist.

Description and development of the State-Trait Anxiety Inventory (STAI)

The STAI is a 40-item self-report questionnaire used to measure current anxiety, described as feelings of tension, apprehension, nervousness and worry (20 items) and a more enduring stable personality characteristic referred to as trait anxiety (20 items). The S-Anxiety scale is completed by rating each of 20 items on a 4-point scale. The T-Anxiety scale asks for ratings to be made on a further 20-items with reference to how an individual generally feels using the same scale.

The original STAI was developed in the 1960's. Since then the STAI-Form X and the STAI-Form Y have been developed, the latter being the most recent version. Therefore, the STAI-Form Y builds upon extensive item development occurring in the earlier versions. In the construction and standardisation of Form Y, more than 5,000 subjects were tested. In addition to obtaining reliability and validity data from this sample, the STAI Form Y has been used extensively, resulting in a considerable body of literature attesting to its reliability and validity.

Reliability and validity of the STAI

Based upon the original normative sample, the test-retest reliability of the STAI T-Anxiety scale is relatively high with a median reliability coefficient of .76 for college students and .69 for high school students (Spielberger et al., 1983). Although lower for the STAI S-Anxiety (a median reliability coefficient of .33) this scale is still considered to have good

internal consistency as an *alpha* coefficient of .90 was obtained. This is arguably a better measure of internal consistency than test-retest reliability when measuring current emotional state.

The T-Anxiety and S-Anxiety scales have been demonstrated to distinguish reliability between psychiatric and non-psychiatric patients. Further, scores on the S-Anxiety scale are significantly higher in stressful situations (e.g., immediately before an exam, during military training) compared with non-stressful settings (e.g., after a relaxation class). Although there is a high correlation between scores on the S-Anxiety and T-anxiety, S-Anxiety scores increase under conditions of greater *a priori* stress and decrease under more relaxed conditions. The T-Anxiety correlates with other measures of trait anxiety (Spielberger et al., 1983).

Suitability for special populations

The STAI has been widely used and there are no special issues or concerns regarding its appropriateness as a measure of anxiety in women. It has not been used with Indigenous Australian groups to the authors' knowledge, indicating that further research addressing the suitability of the measure with this group is warranted. A children's version of the scale is available – the State-Trait Anxiety Inventory for Children (STAIC; Spielberger, 1973). The STAI has been validated for use with older adults with and without anxiety disorders (Stanley, Beck & Zebb, 1996).

Administration and scoring

The STAI takes between 5 - 10 minutes to complete. A reading level of 4th or 5th grade is required. Each item is given a weighted score of 1 to 4. A rating of 4 indicates the presence of a high level of anxiety for 10 of the S-Anxiety and T-Anxiety items; a high rating indicated an absence of anxiety for the remaining items. Therefore, when scoring the STAI close attention must be paid to those items in which a reverse score must be applied. A template is provided with the manual to facilitate scoring.

As with the SCL-90-R this test can only be purchased by Registered Psychologists with post graduate training. Full administration and scoring guidelines are provided in the Manual.

Availability

The STAI is copyright protected and may not be reproduced without permission. A copy may be purchased from ACER Press: Australian Council for Educational Research, ACER Customer Service, 347 Camberwell Road (Private Bag 55), Camberwell Victoria Australia 3124, Tel: (03) 9835 7447 Fax: (03) 9835 7499 via e-mail: sales@acer.edu.au or the internet : URL: <http://www.acerpress.com.au>.

Beck Anxiety Inventory (BAI)

Key Reference: Beck, A. T. & Steer, R. A. (1990). *Beck Anxiety Inventory Manual*, San Antonio: Psychological Corporation.

Summary

The BAI is a 21-item self-report measure of the severity of anxiety in adults and adolescents. The BAI has consistently been found to have high levels of internal consistency and concurrent validity with other measures of anxiety amongst clinical and non-clinical samples. It has been criticised for its lack of divergent validity from measures of depression. The latest version of the manual presents a revision of the diagnostic ranges and descriptive labels for interpreting the scale.

Description and development of the BAI

The BAI was developed by Aaron Beck and colleagues at the Centre for Cognitive Therapy, University of Pennsylvania School of Medicine, Department of Psychiatry. The BAI was constructed to assess symptoms of anxiety that were minimally shared with depression. The original scale items were drawn from three earlier measures of anxiety – the Anxiety Checklist, PDR Checklist and the Situational Anxiety Checklist. An item pool of 86 symptoms of anxiety was created from the responses of 810 outpatients with mood and anxiety disorders to these scales. A series of principal components analyses and item analysis resulted in the 21-item scale. The 21-items of the scale consist of descriptive statements of anxiety symptoms (e.g., “terrified”; “fear of dying”) that are rated on a 4-point scale.

Reliability and validity of the BAI

The psychometric properties of the 21-item version of the BAI were first examined among a sample of 160 outpatients with mood and anxiety disorders (DSM-III & DSM-III-R; Beck, Epstein, Brown & Steer, 1988). They have been further examined among 393 outpatients with anxiety disorders including 5 large samples of panic disorder with and without agoraphobia, social phobia, OCD and GAD (Beck & Steer, 1990); two samples of outpatients with anxiety disorders (Fydrich, Dowdell & Chambless, 1990); adult inpatients and outpatients with mixed psychiatric disorders (Beck, Epstein, et al., 1988; Steer, Beck, Brown, & Beck, 1993; Steer, Ranieri, Beck & Clark, 1993; Steer, Rissmiller, Ranieri & Beck 1993); among patients with panic disorder and agoraphobia (De-Beurs, Wilson, Chambless, Goldstein & Feske, 1997); older psychiatric outpatients (Kabacoff, Segal, Hersen & Van-Hasselt, 1997); adolescent psychiatric patients (Kumar, Steer & Beck, 1993) and non-clinical samples of undergraduates, medical students and non-students (Creamer, Foran & Bell, 1995).

The BAI has consistently been found to have high levels of internal consistency among patients with anxiety disorders ($\alpha = .92-.93$; De-Beurs et al., 1997; Fydrich et al., 1990; Beck & Steer, 1990). Test-retest reliability among patients with panic disorder and agoraphobia was also high (.83) over a 5-week period (De-Beurs et al.). Factor analysis of the original validation sample indicated a two-factor solution – somatic aspects of anxiety and subjective or panic-related aspects of anxiety (Beck, Epstein, et al., 1988). Other studies have replicated this two factor structure among psychiatric inpatients (Steer, Rissmiller, et al.,

1993); older psychiatric outpatients (Kabacoff et al., 1997); a non-clinical sample of undergraduate students (Creamer et al., 1995) and adolescent psychiatric outpatients (Steer, Kumar, Ranieri & Beck, 1995). A further large scale study of 367 outpatients with anxiety disorders by Beck and colleagues generated a four factor solution – neurophysiologic, subjective, panic and autonomic symptoms of self-reported anxiety (Beck & Steer, 1991b). The BAI factors showed adequate levels of internal consistency, convergent and discriminant validity. Two studies have replicated this four-factor structure among community dwelling adults (Osman, Barrios, Aukes, Osman & Markway, 1993) and disadvantaged older primary care patients (Wetherell & Arean, 1997). Recent research has suggested the BAI may be characterised by both a 2- and 4-factor structure in clinical samples and non-clinical samples (Beck, Epstein, et al., 1988; Beck & Steer, 1991b; Osman, Kopper, Barrios, Osman & Wade, 1997).

Evidence for the concurrent validity of the BAI has been demonstrated by its substantial significant correlations with anxiety diaries, clinically rated anxiety and other self-report measures of anxiety across a number of studies. The BAI has demonstrated moderate positive correlations with anxiety diaries (Fydrich et al., 1992); the Hamilton Anxiety Rating Scale (Beck, Epstein, et al., 1988) and with the state and trait subscales of the STAI (Fydrich et al., 1992; Kabacoff et al., 1997); and the anxiety subscale of the SCL-90-R (Steer et al., 1993).

However, the construct validity of the BAI has been criticised for its lack of divergent validity with measures of depression. Across studies with both clinical and non-clinical samples the BAI has been found to correlate significantly with clinically rated (e.g. Hamilton Psychiatric Rating scale for Depression) and self-report measures (e.g., BDI) of depression (Beck, Epstein, et al., 1988; Fydrich et al., 1992). However, these correlations tend to be lower than those with anxiety measures and related to the common symptoms of depression and anxiety measures (De-Beurs et al., 1997).

Only one study examining the sensitivity and specificity of BAI scores for detecting clinical anxiety among older psychiatric patients has been conducted (Kabacoff et al., 1997). However, no single optimal cut-off score emerged that was acceptable regarding sensitivity and specificity. Therefore, the ranges provided for interpreting the BAI total score act only as guidelines until these studies are performed.

The sensitivity of the BAI to change was supported by findings of significant change in a clinical population concordant with other anxiety measures among patients in treatment for panic disorder with agoraphobia (De-Beurs et al., 1997).

Suitability for special populations

The BAI has demonstrated validity for use with adult, adolescent and older psychiatric patients and non-clinical samples. A French and Turkish version of the BAI have demonstrated adequate psychometric properties (Freeston, Ladouceur, Thibodeau, Gagnon & Rheume, 1994; Ulosoy, Sahin & Erkmen, 1998). There are some cautions for interpreting BAI scores with women and younger populations. BAI scores were found to be significantly related to age and gender. Women with anxiety disorders have been found to score on average 4 points higher than males with anxiety disorders in clinical samples (Beck & Steer, 1990) and also score higher among non-clinical samples (Osman et al., 1997). BAI scores were also found to be inversely related to age, with younger patients with anxiety disorders reporting more anxiety than older patients with anxiety disorders (Beck & Steer, 1990).

A recent study compared the psychometric properties of the BAI and STAI among older psychiatric outpatients (Kabacoff et al., 1997). Both scales had a high level of internal consistency, however, the BAI had better discriminant and factorial validity than the STAI in this population. Another study found the BAI correlated with trait anxiety then state anxiety in a non-clinical sample (Creamer et al., 1995). However, other studies among clinical and a non-clinical sample which controlled statistically for symptoms of depression found the BAI was more highly correlated with state anxiety than trait anxiety as one would expect (Fydrich et al., 1990; Osman et al., 1997). Further studies comparing the psychometric properties of the BAI and STAI in other populations, are required to further explore these findings.

Administration and Scoring

The BAI takes between 5 and 10 minutes to administer. Ratings are based on how much the respondent was bothered by each symptom in the last week. Thirteen items assess physiological symptoms, five describe cognitive aspects and three cover both somatic and cognitive symptoms. The BAI is scored on a 4-point scale – 0 (Not at all); 1 (Mildly; it did not bother me much); 2 (Moderately; it was very unpleasant, but I could stand it); 3 (Severely; I could barely stand it). Scores range from 0 to 63. The BAI total score is the sum of the ratings of the 21 symptoms. Total scores are interpreted according to the following guidelines– “Minimal” (0-7); “Mild” (8-15); “Moderate” (16-25) and “Severe” (26-63) levels of anxiety. It should be noted that this 1993 revision of the BAI represents a change in the diagnostic ranges and descriptive labels for interpreting the scale.

Availability

The BAI is copyright protected and may not be reproduced without permission. It can only be purchased by a Registered Psychologist. A copy may be purchased from The Psychological Corporation, PO Box 9959, San Antonio, TX 78204-0959 or via URL: <http://www.psychcorp.com>. Scoring and interpretation must be supervised by a Registered Psychologist. Computer software for the administration, scoring and interpretation of the BAI is available.

Impact of Events Scale (IES)

Key reference: Horowitz, M., Wilner, N. & Alvarez, W. (1979). Impact of event scale: A measure of subjective stress, *Psychosomatic Medicine*, 41 (3), 209-219.

Zilberg, N. J., Weiss, D. S. & Horowitz, M. J. (1982). Impact of event scale: A cross-validation study and some empirical evidence supporting a conceptual model of stress response syndrome, *Journal of Consulting and Clinical Psychology*, 50 (3), 407-414.

Summary

The Impact of Event Scale (IES) is a 15 item self-report inventory of the current degree of subjective stress experienced as a result of a specific event. It is one of the most widely used measures of PTSD-related symptoms of avoidance and intrusiveness. It does not provide a diagnosis of PTSD. The IES has been found to have adequate reliability, factor structure, discriminant and concurrent validity and sensitivity to change. It has also demonstrated reliability, content, construct and criterion related validity amongst adolescents. The advantage of the IES is that unlike the majority of PTSD measures it is applicable across different types of trauma, although its psychometric properties amongst different trauma types requires further investigation.

Description and development of the IES

The IES was developed to provide a measure of the current degree of subjective stress experienced as a result of a specific event (Horowitz, Wilner & Alvarez, 1979). It was developed to provide an overall measure of subjective stress that incorporated items reflecting the two most commonly reported traumatic responses to stressful events - intrusion (intrusively experienced ideas, images, feelings or bad dreams) and avoidance (consciously recognised avoidance of certain ideas, feelings or situations). Over a period of several years, a number of versions of the instrument were administered to patients and non-patients who had been exposed to stressful life events. During this time the wording and format of the instrument was revised several times. Items were originally rated for both frequency and intensity over a one week period.

A 20-item instrument was initially administered to a sample of 66 adults with stress response syndromes in reaction to a stressful life event. This indicated all IES items were endorsed frequently and cluster analysis indicated two clinically derived clusters (intrusion and avoidance). Based on these findings, a number of revisions were made to the instrument. The number of items was reduced to 15, based on items that empirically clustered and had significant item-subscale correlations. Due to the high degree of similarity between ratings of intensity and frequency, the dual response format was also deemed unnecessary and only the frequency variable was retained.

The IES demonstrated adequate reliability in the original validation sample. Both subscales exhibited good internal consistency (α : intrusion = .78; avoidance = .82) and split half reliability of the total scale was high ($r = .86$). Test-retest reliability among a sample of non-patients recently exposed to a traumatic event was also acceptable (total score $r = .87$; intrusion $r = .89$; avoidance $r = .79$).

The sensitivity of the IES was supported by findings of significant change in a clinical population concordant with clinical impressions and its ability to discriminate between patients and non-patients responses to discrete life events of varied magnitude (Horowitz et al., 1979).

Reliability and validity of the IES

As the IES was published prior to the DSM-III, which first incorporated the diagnostic category of Post Traumatic Stress Disorder (PTSD), the psychometric properties of the IES were examined among patient and non-patient samples of bereaved individuals (Zilbert, Weiss & Horowitz, 1982). Reliability analysis indicated that the IES had high internal consistency ($\alpha = .79-.92$) among both samples across time (0-13 months). The IES also demonstrated sensitivity in significantly discriminating between patient and non-patient samples across time and detecting change over time, thereby supporting its suitability as a measure of treatment outcome.

Factor analysis confirmed the empirically derived 2-factor structure of the scale – intrusion and avoidance (Zilberg et al., 1982). A number of studies have identified a third factor, comprising of emotional numbing, in addition to the avoidance and intrusion factors (Joseph, Williams, Yule & Walker, 1992; McDonald, 1997; Schwarzwald, Solomon, Weisenberg & Mululiner, 1987).

The IES has demonstrated concurrent validity with another measure of emotional distress the GHQ-28 (Hodgkinson & Joseph, 1995). Both the intrusion and avoidance subscales of the IES exhibited positive significant correlations with the GHQ-28.

Administration and scoring

The specific life event and the date of its occurrence are recorded at the top of the page. The respondent then indicates whether or not each item had been experienced within the past 7 days. If the experience is recalled it is rated on a 3 point (rarely, sometimes, often) or 4 point (not at all, rarely, sometimes, often) frequency scale. Subscale scoring is as follows - Intrusion (1, 4, 5, 6, 10, 11, 14) and Avoidance (2, 3, 7, 8, 9, 12, 13, 15). A copy of the IES is provided below.

Use with special populations

The IES has been used among a variety of traumatised groups including Vietnam veterans (Kulka et al., 1990; Scott & Dua, 1999); combat victims (Schwarzwald et al., 1987); parental bereavement (Zillberg et al., 1982), female bank armed robbery victims (Hodgkinson & Joseph, 1995), shipping disaster victims (Joseph et al., 1992) and victims of rape (Burge, 1988).

The psychometric properties of the IES for assessing the effects of trauma in adolescents has been recently examined, due to increasing interest in the effects of exposure to war and/or natural disasters on children and adolescents (see Saigh, Fairbank, & Yasik, 1998 for a review). Yule, Ten Bruggencate & Joseph (1994) examined the IES's psychometric properties among 334 adolescent survivors of the *Juniper* cruise ship disaster. Unlike adult samples, principal components analysis yielded a 3-factor solution comprised of avoidance, intrusion and emotional numbing (61.1% of total variance). Sack, Seeley, Him & Clarke (1998) later investigated the IES's psychometric properties among 180 Khmer refugee youth. They found the IES had a high level of internal consistency ($\alpha = .92$). Confirmatory

factor analysis produced a three-factor solution, replicating that found by Yule et al. (1994). All but two items of the IES discriminated between Cambodian youth diagnosed with PTSD from those not diagnosed with PTSD. ROC analysis indicated 19 to be the optimal cut-off score. Of those who screened positive, 30% were true cases and of those who screened negative 90% were non-cases.

The IES is easy to administer and has been used across a variety of sites and samples. However, it only provides a measure of the avoidance and intrusive symptoms and does not produce a diagnosis of PTSD.

Availability

The IES is reproduced below and can be used with acknowledgement of the authors.

Impact of Events Scale

On _____ you experienced _____
 (date) (life event)

Below are some statements made by people after a stressful life event. With respect to your traumatic event **CIRCLE THE NUMBER** that best describes how frequently these comments were true for **YOU** in the **LAST 7 DAYS**. If they did not occur during that time, please mark the “not at all” column.

Statement	Not at all	Rarely	Sometimes	Often
I thought about it when I didn't mean to	0	1	3	5
I avoided letting myself get upset when I thought about it or was reminded of it	0	1	3	5
I tried to remove it from memory	0	1	3	5
I had trouble falling asleep or staying asleep, because pictures or thoughts about it came into my mind	0	1	3	5
I had waves of strong feelings about it	0	1	3	5
I had dreams about it	0	1	3	5
I stayed away from reminders of it	0	1	3	5
I felt as if it hadn't happened or it wasn't real	0	1	3	5
I tried not to talk about it	0	1	3	5
Pictures about it popped into my mind	0	1	3	5
Other things kept making me think about it	0	1	3	5
I was aware that I still had a lot of feelings about it, but I didn't deal with them	0	1	3	5
I tried not to think about it	0	1	3	5
Any reminder brought back feelings about it	0	1	3	5
My feelings about it were kind of numb	0	1	3	5

Eating Attitudes Test (EAT)

Key references: Garner, D.M., Olmstead, M.P., Bohr, Y and Garfinkel, P.E. (1982). The eating attitudes test: Psychometric features and clinical correlates. *Psychological Medicine*, 12, 871-878

Garner, D.M. (1997). Psychoeducational principles in treatment. In D.M.Garner & P.E. Garfinkel (Eds.) *Handbook of treatment for eating disorders*, New York: Guilford Press.

Associated Webpage: URL: <http://www.eatingdisorders-toledo.com/pretest.html>.

Summary

The Eating Attitudes Test (EAT) is a screening test that detects disturbed eating patterns. It is reliable and valid and has been used widely with women and girls. A version for assessing disordered eating in young girls has been developed (the Children's Eating Attitudes Test; CHEAT). Given the high prevalence of eating disorders amongst substance misusers, the EAT is a useful screening instrument when developing a treatment plan, although there is a dearth of studies investigating the psychometric properties of eating disorder measures in substance abusing populations.

Description and development of the Eating Attitudes Test (EAT)

The Eating Attitudes Test (EAT) is a screening instrument that is useful for detecting the presence of disturbed eating patterns in populations at high risk for eating disorders. It is possibly the most widely used screening measure of eating disorder symptomatology. It was initially developed and validated with two independent groups of female anorexia nervosa patients and female normal control subjects. The 40 items of the EAT were obtained by a series of administrations and analysis inclusion of only those items that reliably discriminated anorexic patients from normal controls (Garner and Garfinkel, 1979). More recently, the EAT-26 has been developed in which the 26 items were found to load onto 3 factors. Factor I was labelled "Dieting" and reflects pathological avoidance of fattening foods and preoccupation with body shape. factor II was labelled "bulimia and food preoccupation" and was positively related to bulimia and a heavier body weight; factor III was labelled "oral control" and consisted of items reflecting self control about food and social pressure to gain weight (Garner et al., 1982).

Reliability and validity of the EAT

Unlike measures of substance problems, which were initially validated on men, measures of disordered eating have typically been validated with women and girls. This is due primarily to the disproportional number of eating disorders occurring with women (approximately only 10% of all diagnosed eating disorders occur in men; DSM IV). A search of the literature showed that while the measures reviewed have been used with male participants, little validation within this population has occurred (e.g., Rathner & Rumpole, 1994; Lee, Lee, Leung, & Yu, 1997). As such, the following psychometric findings refer primarily to samples of women and girls. Further validation of these measures for male clients is still required.

The initial reliability study found the EAT-40 to have good internal consistency ($\alpha = .79$ for the clinical sample; .94 for a pooled sample of both the clinical and normal control group;

Garner and Garfinkel, 1979) and good test-retest reliability ($r = .84$; Williamson, Anderson, Jackman & Jackson, 1995).

The EAT correlates highly with the three EDI (see below) symptom scales, notably with the Drive for Thinness subscale ($r = .81$; Gross, Rosen, Leitenberg, & Willmuth, 1986). Although the EAT was developed to assess characteristics of anorexia, most psychometric research has used anorexic women with and without bulimia, and has compared the EAT with other measures of bulimia. For instance, the EAT correlates moderately with self-monitoring records of frequency of bingeing ($r = .66$) and purging ($r = .54$) and established measures of bulimia, such as the BULIT ($r = .67$) and the BITE ($r = .70$). Likewise, while the EAT discriminates between anorexic and non-anorexic women (Garner & Garfinkel, 1979), and between binge-eating women and non-binge-eating women (Prather & Williamson, 1988), it fails to distinguish anorexic from bulimic individuals (Williamson et al., 1995). Notwithstanding, it appears to be sensitive to therapeutic interventions in both bulimic (Williamson et al., 1989) and anorexic women (Garner & Garfinkel, 1979).

Following the development of the original 40-item EAT, a preliminary factor analysis found 3 factors that accounted for 40% of the variance. The 14 items not loading on any of these 3 factors were eliminated leaving a total of 26 items. The EAT-26 correlated highly with the EAT-40 indicating that the EAT-26 may be reliably used as a screening instrument for disordered eating. While the authors caution against equating an elevated score with a diagnosis of anorexia, a cut-off score of 20 on the EAT-26 correctly identified 84% of the subjects as either anorexic or controls (Garner et al., 1982), while a cut-off score of 30 for the EAT-40 correctly classified 85% of anorexic women with and without bulimic features from non-eating disordered college women (Garner et al., 1982).

On the whole, the EAT appears to best identify individuals likely to engage in clinically significant disordered eating, but lacks the sensitivity to distinguish between individuals meeting criteria for anorexia from those meeting criteria for bulimia. Furthermore, the authors noted that while the EAT is a psychometrically sound measure of eating disorder symptomatology, it does not assess more general dysfunctional attitudes and related psychopathology. Accordingly Garner, Olmstead, & Polivy (1983) developed the Eating Disorders Inventory (EDI).

Suitability for special populations

High rates of eating disorders, notably bulimia, are found in substance abusing populations (studies find between eight and 40% of substance misusing women report a current or past history of an eating disorder; see Holderness, Brooks-Gunn, & Warren, 1994 for a review). However, women who present with comorbid alcohol problems in alcohol treatment centres typically do not have their eating disorder detected nor treated (Wilson, 1993). Furthermore, there are a number of behavioural overlaps between alcohol misusers and eating disordered women. For instance, both may engage in vomiting behaviour and/or have a loss of appetite, but for differing reasons, e.g., for fear of gaining weight in eating disordered women, but due to the consequences of alcohol misuse in alcohol abusing women. As such, measures of eating disorders administered in this population must be able to differentiate between behaviours due to substance misuse from those due to the eating disorder (Black & Wilson, 1996). To date, while the EAT has been used in studies of substance misusing women, the robustness of the EAT's psychometric properties has not been investigated in this population.

The applicability of this measure to Indigenous Australians is questionable. Cross-cultural studies of eating disorders indicate that these disorders occur mainly in Westernised countries and less so in other cultures. The prevalence of eating disorders among Indigenous Australians has not been systematically investigated to the authors' knowledge. Furthermore, all self-report measures of eating disorders have been developed and validated in Europe and North America and are based on Western ideas of abnormal eating and body shape/weight (Fedoroff & McFarlane, 1998). Consequently, even though the EAT and the EDI (see next section) have been translated into a variety of languages, when administered to non-Western samples, many items are not able to be translated appropriately, some concepts such as shape concerns and binge-eating are unfamiliar, and some behaviours, such as cutting food into small pieces or fasting are in accord with social and religious norms, resulting in poor validity within these populations (Fedoroff & McFarlane, 1998).

A modified version, the CHEAT (Children's EAT) was developed to assess disordered eating attitudes in pre-pubescent girls (Maloney, McGuire & Daniels, 1988).

Administration and scoring

The EAT takes less than 10 minutes to administer, is available in a variety of languages and requires a 5th grade reading level (Williamson et al., 1995).

Availability

This instrument is copyright protected, but may be used free of charge and with due acknowledgement of the source. It is available to download or to complete "online" @ URL: <http://www.eatingdisorders-toledo.com/pretest.html>. A copy of the EAT may also be obtained from: Garner, D.M. (1997). Psychoeducational principles in treatment. In D.M.Garner & P.E. Garfinkel (Eds.) *Handbook of treatment for eating disorders*, New York: Guilford Press.

Eating Disorders Inventory II (EDI)

Key reference: Garner, D. (1991). *Eating Disorders Inventory 2. Professional Manual*. Odessa, FL: Psychological Assessment Resources.

Associated Web pages:

URL: [http:// www.parinc.com/product.cfm?ProductID=201](http://www.parinc.com/product.cfm?ProductID=201);

URL: <http://www.eatingdisorders-toledo.com/EDI-2.html>.

Summary

The Eating Disorders Inventory (EDI) is a screening test that detects disturbed eating patterns and associated psychopathology. It is reliable and valid and has been used widely with women and girls in clinical practice and research. Both pen and paper and computerised versions are available.

Description and development of the Eating Disorders Inventory (EDI)

The EDI-2 is a 91-item scale that measures the severity of disordered eating symptomatology (Garner, 1991). It derives three primary subscales - Drive for Thinness (DT), Bulimia, and Body Dissatisfaction (BD) - and 8 additional scales measuring associated psychopathology: Ineffectiveness, Interpersonal Distrust, Perfectionism, Maturity Fear, Interoceptive Awareness, Impulse Regulation, Social Insecurity and Asceticism.

The DT scale measures respondents' concerns regarding weight gain, and preoccupation with weight and dieting. Extreme DT is a core feature in both anorexia and bulimia. The Bulimia scale assesses bulimic attitudes and behaviours such as bingeing and purging. The BD scale measures dissatisfaction with areas of the body that are typically of greatest concern with those with eating disorders (see Garner, 2001 for detailed descriptions of all 8 subscales).

Each of the subscales measures an independent trait and is considered continuous, with higher scores indicating a greater manifestation of that particular trait. The EDI is not a diagnostic instrument, and is best used for identifying those who engage in eating behaviours and/or report attitudes that indicate a high risk for developing disordered eating symptoms. Accordingly, no studies have determined an optimal cut-off point for distinguishing eating disordered people from non-eating disordered people. However, the manual recommends a cut-off of 14 on the DT scale as indicative of high risk disordered eating.

Overall, the EDI is best considered as a screening tool for high risk of disordered eating, for general clinical screening and for assessing treatment outcome rather than as a diagnostic tool.

Reliability and validity of the EDI

The EDI-2 manual reports good internal consistency (Cronbach's α for subscales range from .80 to .92) and good test-retest reliability ($r = .72$ in a non-patient sample) over a 12-month period. The 3 symptom scales correlate strongly with the Eating Attitudes Test (EAT-26) and with self-reported dieting and purging (Gross et al., 1986).

The EDI demonstrates slightly greater ability to discriminate between subgroups of eating disordered women than the EAT. Like the EAT, the 3 symptom scales discriminated between bulimic women and non bulimic women (Gross et al., 1986) and between anorexic and non-anorexic women (Garner et al., 1983). Further, the Bulimia subscale also discriminated between anorexic women with bulimic symptoms from restricting anorexics (Garner et al., 1983), and between bulimic and non-bulimic women (Wilson & Smith, 1989). The EDI failed, however, to discriminate bulimic women from dieting, weight concerned, non-bulimic college women (Wilson & Smith, 1989). Scores on the Drive for Thinness scale have been found to predict the development of bulimic behaviour over a ten-year period (Joiner, Heatherton, Rudd, & Schmidt, 1997). Further psychometric details are presented in the manual and @ URL: <http://www.eatingdisorders-toledo.com/EDI-2.html>.

Suitability for special populations

The EDI has been translated in numerous languages including Chinese, German, Arabic and Hebrew. Cross-cultural validation is currently under study (Garner, 2001). Notwithstanding, issues noted above regarding the use of the EAT with non-Caucasian, non-Westernised populations should also be considered when using the EDI. To the authors' knowledge, no studies have examined the utility of the EDI with Indigenous Australian women. Likewise, as with the EAT, the EDI is frequently used, but has not been validated within substance abusing populations.

Administration and scoring

The EDI can be administered in less than 20 minutes, requires a 5th grade reading level, is available in a number of languages and measures specific psychological features associated with anorexia and bulimia in addition to symptoms (Williamson et al., 1995). A computerised version is also available from the publishers.

Availability

This instrument is copyright protected and should be purchased from Psychological Assessment Resources Inc. 16204 N. Florida Ave., Lutz, Florida, 33549 (phone 1- 813-968-3003; fax 1-813-968-2598) or @ URL: <http://www.parinc.com>.

Bulimic Investigatory Test – Edinburgh (BITE)

Key reference: Henderson, M. & Freeman, C.P.L. (1987). A self-rating scale for bulimia: The BITE. *British Journal of Psychiatry*, 150, 18-24.

Summary

The BITE is a 33-item self-report measure that assesses cognitive and behaviour aspect of bulimia, focusing strongly on binge-eating behaviour. The BITE derives a symptom subscale and severity subscale, which may be used separately or combined to create a global score. A cut-off of 25 on the global score has been suggested as identifying those likely to engage in clinically significant disordered eating, although this cut-off point has been debated. In general the BITE is a promising measure of binge-eating behaviour, which discriminates between categories of eating disorders.

Description and development of the BITE

The Bulimic Investigatory Test - Edinburgh (BITE; Henderson & Freeman, 1987) was developed to measure cognitive and behavioural aspects of bulimia. The BITE was initially devised to fill a gap in self-report measures for assessing bulimic behaviour, in that previous methods of assessing the prevalence of this disorder failed to define definitive features of bulimia, such as binge-eating. The aim for developing the BITE was to create a highly sensitive instrument for identifying bulimic individuals. Based on the symptoms of bulimia described by Palmer (1979), Bruch (1975) and Russell (1979; all cited in Henderson and Freeman, 1987), the authors of the BITE created a list of questions, which were subsequently administered to a small sample of binge-eaters and normal controls. Following the removal and rewording of ambiguous items and confusing items, the remaining 40-items were tested for reliability and validity and was further refined to create the final 33-item version. The BITE is divided into two subscales: a symptom subscale and a severity subscale, which may be added together to derive a total score.

Reliability and validity of the BITE

Internal consistency for both subscales has been found to be adequate ($\alpha = .96$ symptoms and $.62$ for severity) and while 1-week test-retest reliability with controls was good ($r = .86$), 15-week temporal stability for a small bulimic sample was less strong ($r = .68$; Waller, 1992).

The BITE generally correlates moderately with other self-report measures of bulimia such as the BULIT-R ($r = .90$; Welch et al., 1993) and Bulimia subscale of the EDI ($r = .69$; Henderson & Freeman, 1987). The BITE correlates less strongly with measures of general eating disorders such as the EAT ($r = .70$) and the Drive for Thinness scale of the EDI ($r = .59$). Additionally, both the symptom and severity subscales discriminate non-eating disordered women from eating disordered women, and between women with disparate

eating disorder diagnoses; specifically bulimic women scored higher than anorexic women with bulimic symptoms, who in turn scored higher than anorexic women without bulimic symptoms (Waller, 1992).

Although a total score of 25 is deemed as clinically significant, the utility of this cut-off for diagnostic purposes has been debated. Waller (1992), for instance, notes that while the BITE adequately identifies and classifies normal weight bulimics, it tends to misclassify low-weight binge-eaters. As such, Waller (1992) argues that the BITE may be more useful as a tool for detecting normal weight binge-eating women than in low-weight binge-eating women. As the BITE assesses binge-eating specifically, this measure may be useful in measuring the proposed Binge Eating Disorder.

Suitability for special populations

To date the BITE has been used with a sample of elderly people living in residential care, with Afro-Caribbean women and with women and girls in Spain, Italy, Japan, Brazil and Australia (e.g., Ampollini et al., 1999; Bartlett, Shrimanker, & Ballard, 2000; Nakai, Hamagaki, & Takagi, 1998; Reiss, 1996).

Although bulimic symptoms tend to occur more often with substance abusing populations the BITE has yet to be validated for its ability to discriminate eating disordered and non-eating disordered women in this population. This is important as substance abusers show overlapping symptoms and behaviours (e.g., frequent vomiting, disrupted eating routines, weight fluctuations, etc).

Administration and scoring

Although the scoring of the BITE was designed to be performed by the Statistical Package for the Social Sciences (SPSS), the BITE can be hand-scored. The authors advise that when the BITE is used as a screening instrument or in research, that respondents focus on their behaviour and feelings over the previous 3 months. When used to measure treatment progress, the time frame is the past month. The BITE requires only a 4th grade reading level and takes less than 10 minutes to administer. An optional data sheet is included to assess demographic information (Williamson et al., 1995).

All YES responses are scored one point except for items 1, 13, 21, 23, and 31, where NO responses are scored one point. All items, bar those marked with an asterisk (items 6, 7 & 27), are summed to form the Symptom Scale. Items 6, 7 & 27 are summed to form a Severity Scale. A global score can be derived by summing all items.

A score on the symptom scale of >20 is suggestive of highly disordered eating and possibly indicates the presence of binge-eating behaviour. Those scoring between 10 and 19 are likely to be engaging in disordered eating but may not meet DSM criteria for bulimia. The authors suggest respondents scoring 15 should be followed up by a diagnostic interview. Scores ≤10 suggest the absence of disordered eating.

Similarly, a score ≥ 5 on the Severity Scale suggests a clinically significant level of disordered eating and should be followed up by an interview, regardless of the Symptom Scale score. A global score ≥25 is indicative of an eating disorder.

Availability

The BITE is in the public domain and is reproduced below. It may be used without cost but with due acknowledgment of the source.

BULIMIC INVESTIGATORY TEST, EDINBURGH (BITE)**Optional front data sheet**

1. What is your sex? MALE 1 FEMALE 2 (Please circle number)
 2. Are you:
MARRIED 1 SINGLE 2 DIVORCED 3 SEPARATED 4 WIDOWED 5
 3. What is your occupation? _____
 4. If married, what is your spouse's occupation? _____
 5. What is your age? _____ years
 6. What is your height? _____ cm
 7. What is your weight? _____ kg
 8. What is the most you have ever weighed? _____ kg
 9. What is the least that you have weighed at your present height? _____ kg
 10. What would your ideal weight be if you could choose it? _____ kg
 11. Do you feel yourself to be VERY OVERWEIGHT 5
 OVERWEIGHT 4 (Please
 AVERAGE 3 circle
 UNDERWEIGHT 2 number)
 VERY UNDERWEIGHT 1
 12. Do you have regular periods? (if applicable) YES 1 NO 2
 13. How often, on average, do you eat the following meals?

	EVERY DAY	5/7 DAYS	3/7 DAYS	1/7 DAYS	NEVER	
BREAKFAST	1	2	3	4	5	(Please circle number)
LUNCH	1	2	3	4	5	
DINNER	1	2	3	4	5	
BETWEEN MEAL SNACKS	1	2	3	4	5	

 14. Have you ever consulted someone in a professional capacity for advice on dieting/eating YES 1 NO 2
 15. Have you ever been a member of a slimming club? YES 1 NO 2
 16. Have you ever suffered from any type of eating disorder? YES 1 NO 2
- if yes , please give details over:

BULIMIC INVESTIGATORY TEST, EDINBURGH (BITE)

1. Do you have a regular eating pattern? YES NO
2. Are you a strict dieter? YES NO
3. Do you feel a failure if you break your diet once? YES NO
4. Do you count the calories of everything you eat, even when not on a diet YES NO
5. Do you ever fast for a whole day? YES NO
- *6. If yes, how often is this?
EVERY SECOND DAY 5 2-3 TIMES A WEEK 4
ONCE A WEEK 3 NOW AND THEN 2 HAVE ONCE 1

BULIMIC INVESTIGATORY TEST, EDINBURGH (BITE) (continued)

*7. Do you do any of the following to help you lose weight:

	NEVER	OCCASIONALLY	ONCE A WEEK	2-3 TIMES A WEEK	DAILY	2-3 TIMES A DAY	5+ TIMES A DAY
TAKE DIET PILLS	0	2	3	4	5	6	7
TAKE DIURETICS	0	2	3	4	5	6	7
TAKE LAXATIVES	0	2	3	4	5	6	7
MAKE YOURSELF VOMIT	0	2	3	4	5	6	7

8. Does your pattern of eating severely disrupt your life? YES NO
9. Would you say that food dominated your life? YES NO
10. Do you ever eat and eat until you are stopped by physical discomfort? YES NO
11. Are there times when all you can think about is food? YES NO
12. Do you eat sensibly in front of others and make up in private? YES NO
13. Can you always stop eating when you want to? YES NO
14. Do you ever experience *overpowering* urges to eat and eat and eat? YES NO
15. When you are feeling anxious do you tend to eat a lot? YES NO
16. Does the thought of becoming fat *terrify* you? YES NO
17. Do you ever eat large amounts of food rapidly (not a meal)? YES NO
18. Are you ashamed of your eating habits? YES NO
19. Do you worry that you have no control over how much you eat? YES NO
20. Do you turn to food for comfort? YES NO
21. Are you able to leave food on the plate at the end of a meal? YES NO
22. Do you deceive other people about how much you eat? YES NO
23. Does how hungry you feel determine how much you eat? YES NO
24. Do you ever binge on large amounts of food? YES NO
25. if yes, do such binges leave you feeling miserable? YES NO
26. If you do binge, is it only when you are alone? YES NO
- *27. If you do binge, how often is this? YES NO

HARDLY EVER	1	ONCE A WEEK	3
DAILY	5	ONCE A MONTH	2
2-3 TIME S AWEEK	4	2-3 TIMES A DAY	6

28. Would you go to great lengths to satisfy an urge to binge? YES NO
29. If you overeat do you feel *very* guilty? YES NO
30. Do you ever eat in secret? YES NO
31. Are your eating habits what you would consider to be normal? YES NO
32. Would you consider yourself to be a compulsive eater? YES NO
33. Does your weight fluctuate by more than 2kg in a week? YES NO

Part V

Recommendations for future research

General Comments

The aim of the present review was twofold: (i) to identify screening and diagnostic instruments that could be used to detect alcohol and other drug problems in individuals in the general population and those with psychiatric disorders and (ii) to identify relevant psychiatric instruments that could be used to screen or diagnose psychiatric problems in individuals with a substance misuse disorder. While there has been increasing acknowledgment of the potential role of standardised instruments in the addictions research literature, such instruments are generally used by a small proportion of those working in the drug and alcohol field in Australia (Dawe and Richmond, 1997); with no information on their use in other mental health settings. However, although the use of these instruments is strongly recommended, it is also important to emphasise that a screening or diagnostic instrument, no matter how well validated, cannot replace a clinical interview. Individuals who score in the clinical range on a screening instrument should be further assessed to determine their needs and develop a treatment plan.

Screening instruments, overall, need to be administered and interpreted without comprehensive training; many of the instruments reviewed in this document fall into this category. Therefore the AUDIT, for example, can be used by a range of clinicians in primary health care settings, specialist inpatient and outpatient settings and work settings. Similarly, the GHQ is also an instrument that can be administered in a range of settings in which individuals with substance misuse problems are treated.

Instruments that lead to a formal diagnosis, such as the CIDI, require specialist training as does the BPRS, an instrument that is clinician-rated. They are included in the present review, however, because it is helpful for all clinicians who are working in the mental health field to have some familiarity with a range of instruments, and an understanding of the principles upon which they are based.

Before administering any instrument, the client's capacity to complete such an instrument must be determined, and sensible clinical judgement should be exercised. Information on time required to administer the instrument and scoring guidelines are included whenever possible. Therefore, the clinician is positioned to make an informed judgement about the appropriateness of administration of the instrument. Unless the information obtained from the instrument can be acted upon, there is little point in asking the client to complete it. For example, asking a distressed, psychotic patient to complete an AUDIT on entry to hospital may not be feasible for the individual and may add little to the immediate treatment plan. Following admission, and at a point at which the patient is less distressed, an AUDIT may contribute important information that should be considered in developing a management plan.

Recommendation for future research

It is recommended that future research focus on at least the following areas:

- (a) The appropriateness of newer screening and diagnostic instruments for substance misuse in psychiatric populations.

Considerable literature is available on the utility of instruments such as the MAST and the ADS in patients with psychiatric illnesses. Recently, the AUDIT has been assessed

in this population. However, many other instruments that hold promise require further testing in individuals with psychiatric problems. Given the prevalence of alcohol and substance misuse problems in these populations, this is an area of research that warrants continued investigation.

(b) The relevance of screening and diagnostic instruments for Indigenous Australians.

There has been little investigation of the relevance of screening and diagnostic instruments with this population. In part, this reflects the enduring concerns regarding the usefulness of psychological testing at all. However, this has been recognised in certain Indigenous groups to be clearly disadvantaging (E. Hunter, 1996, personal communication). There are now several studies in which the prevalence of drug and alcohol use has been documented; with drug and hazardous alcohol use disproportionately higher in Indigenous compared with non-Indigenous samples (e.g., Hunter et al., 1991; Perkins et al., 1994). Further research is a priority, to enable the adequate planning of both substance misuse and psychiatric services and extension of current treatment approaches (e.g., Brady, Dawe and Richmond, 1998). As a first step in this process, investigation of the appropriateness of general screening instruments and diagnostic nomenclature is essential.

Regarding problem substance use, the issue of solvent abuse in the form of petrol sniffing has only been addressed in the context of the DAST's appropriateness as a screening measure. Given the concern expressed about the use of this substance and proposed strategies advocated to reduce this (Burns et al., 1995) it would seem sensible and timely that instruments or methods for routine screening for petrol sniffing be developed as part of a national strategy to extend current treatment options.

(c) The appropriateness of current screening instruments in the detection of harmful and hazardous drinking in women.

Another area that would appear to warrant further investigation in an Australian context is the sensitivity of alcohol screening instruments in detecting at risk drinking in women. This is an issue that has been of growing concern both in the UK and the USA (see Saltz and Ames, 1996; Ames et al., 1996 for a discussion). There are major problems associated with all of the instruments that have been developed, with the exception of the T-ACE and TWEAK. While these may appear to overcome some of the difficulties relating to quantity and/or the focus on male behaviour, they have not been evaluated outside of the USA. Even the AUDIT, which has been investigated in American women, has been used but not validated with Australian women. Further, the use of alcohol by young Indigenous women is also cause for concern. Whether these instruments are equally applicable to non-Indigenous and Indigenous women needs further research.

(d) The impact of screening instruments in treatment.

Finally, the barriers to screening for alcohol/drug and psychiatric disorders by mental health and drug/alcohol professionals requires research to determine whether the availability of screening instruments does lead to better detection and management.

References

References

- Adelekan, M., Green, A., DasGupta, N., Tallack, F., Stimson, G. V. & Wells, B. (1996). Reliability and validity of the opiate treatment index among opioid users in the United Kingdom. *Drug and Alcohol Review*, 15, 261-270.
- Adelekan, M., Metrebian, N., Tallack, F., Stimson, G. V., & Shanahan, W. (1996). Who should collect Opiate Treatment Index data in opiate treatment outcome monitoring: Clinic staff or researchers? *Drug and Alcohol Review*, 15(1), 65-71.
- Aertgeerts, B., Buntinx, F., Ansoms, S., & Fevery, J. (2001). Screening properties of questionnaires and laboratory tests for the detection of alcohol abuse or dependence in a general practice population. *British Journal of General Practice*, 51, 206-217.
- Aertgeerts, B., Buntinx, F., Bande Knops, J., Vandermeulen, C., Roelants, M., Ansoms, S., & Fevery, J. (2000). The value of CAGE, CUGE, and AUDIT in screening for alcohol abuse and dependence among college freshmen. *Alcoholism: Clinical and Experimental Research*, 24(1), 53-57.
- Aertgeerts, B., Buntinx, F., Fevery, J., & Ansoms, S. (2000). Is there a difference between CAGE interview and written CAGE questionnaires? *Alcoholism: Clinical and Experimental Research*, 24(5), 733-736.
- Aguilar, G.I., Hidalgo, M. D., Cano G., R., & Lopez, J. C. (1995). Prospective study of hopelessness in first-episode psychotic patients: Psychometric characteristics of the Beck Hopelessness Scale in this group. *Anales de Psiquiatria*, 11(4), 121-125.
- Albrecht, S. A., Cornelius, M. D., Braxter, B., Reynolds, M. D., Stone, C., & Cassidy, B. (1999). An assessment of nicotine dependence among pregnant adolescents. *Journal of Substance Abuse Treatment*, 16(4), 337-344.
- Allen, J.P. Fertig, J.B., Towle, L.H. and Altshuler, V.B. (1994). Psychometric analyses of the alcohol dependence scale among United States and Russian clinical samples. *International Journal of the Addictions*, 29, 71-87.
- Allen, J. P., Litten, R. Z., Fertig, J. B., & Babor, T. (1997). A review of research on the Alcohol Use Disorders Identification Test (AUDIT). *Alcoholism: Clinical and Experimental Research*, 21(4), 613-619.
- Allen, J. P., Litten, R. Z., Fertig, J. B., & Sillanaukee, P. (2001). Carbohydrate-Deficient Transferrin: An aid to early recognition of alcohol relapse. *American Journal on Addictions*, 10(Suppl 1), 24-28.
- American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*. Washington, D.C.: American Psychiatric Association.
- American Psychiatric Association (Eds.). (2000). *Handbook of psychiatric measures*. Washington, DC: Author.
- Ames, G., Schmidt, C., Knee, L. & Saltz, R. (1996). Combining methods to identify new measures of women's drinking problems. Part I: The ethnographic stage. *Addiction*, 91, 829-844.
- Ampollini, P., Marchesi, C., Gariboldi, S., Cella, P., Bertacca, S. G., Borghi, C., & Maggini, C. (1999). The Parma high school epidemiological survey: Eating disorders. *Journal of Adolescent Health*, 4(3), 158-159.
- Ananth, J., Vandewater, S., Kamal, M., Brodsky, A., Gamal, R., & Miller, M. (1989). Missed diagnosis of substance abuse in psychiatric patients. *Hospital and Community Psychiatry*, 40, 297-299.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, N.J.: Prentice Hall.
- Andrews, G. & Peters, L. (2000) The CIDI-Auto: A computerised diagnostic interview for psychiatry. Sydney, NSW. Retrieved February 1, 2001, from the World Wide Web:
[URL:http://www.unsw.edu.au/crufad/cidi/discuss.htm](http://www.unsw.edu.au/crufad/cidi/discuss.htm)

- Andrews, G., Peters, L. & Teeson, M. (1994). *The Measurement of Consumer Outcome in Mental Health: A Report to the National Mental Health Information Strategy Committee*. Sydney, AGPS.
- Annis, H.M., Sobell, L.C., Ayala-Velazquez, Rybakowski, J.K., Sandahl, C., Saunders, B., Thomas, S., & Ziolkowski, M. (1996). Drinking-related assessment instruments: Cross-cultural studies. *Substance Use and Misuse*, 31, 1525-1546.
- Appleby, L., Dyson, V., Altman, E., & Luchins, D. J. (1997). Assessing substance use in multiproblem patients: reliability and validity of the Addiction Severity Index in a mental hospital population. *Journal of Nervous and Mental Disease*, 185(3), 159-165.
- Arnau, R. C., Meagher, M. W., Norris, M. P., & Bramson, R. (2001). Psychometric evaluation of the Beck Depression Inventory-II with primary care medical patients. *Health Psychology*, 20(2), 112-119.
- Australian Institute of Health and Welfare (1999). *1998 National Drug Strategy Household Survey: First results*. Canberra: Author.
- Australian Institute of Health and Welfare (2000). *1998 National Drug Strategy Household Survey: Detailed findings*. Canberra: Author.
- Australian Psychological Society (1996) *Guidelines for the Provision of Psychological Services for and the Conduct of Psychological Research with Aboriginal and Torres Strait Islander People of Australia*. Melbourne: Author.
- Babor, T., De La Fuente, J. R., Saunders, J., & Grant, M. (1992). *AUDIT: The Alcohol Use Disorders Identification Test: Guidelines for use in Primary Health Care* (WHO/PSA/92.4). Geneva: World Health Organization.
- Babor, T. F., Steinberg, K., Anton, R., & Del Boca, F. (2000). Talk is cheap: Measuring drinking outcomes in clinical trials. *Journal of Studies on Alcohol*, 61(1), 55-63.
- Baillie, A. J., & Mattick, R. P. (1996). The Benzodiazepine Dependence Questionnaire: Development, reliability and validity. *British Journal of Psychiatry*, 169(3), 276-281.
- Barry, K. L., & Fleming, M. F. (1993). The Alcohol Use Disorders Identification Test (AUDIT) and the SMAST-13: Predictive validity in a rural primary care sample. *Alcohol and Alcoholism*, 28(1), 33-42.
- Bartko, J.J. (1976). On various intraclass correlation reliability coefficients. *Psychological Bulletin*, 83, 762-765.
- Bartlett, S., Shrimanker, J., & Ballard, C. (2000). Prevalence of abnormal eating amongst elderly people in residential care. *International Journal of Geriatric Psychiatry*, 15(3), 283-284.
- Beck, A. T., Brown, G., Berchick, R. J., & Stewart, B. L. (1990). Relationship between hopelessness and ultimate suicide: A replication with psychiatric outpatients. *American Journal of Psychiatry*, 147(2), 190-195.
- Beck, A. T., Brown, G. K., Steer, R. A., Dahlsgaard, K. K., & Grisham, J. R. (1999). Suicide ideation at its worst point: A predictor of eventual suicide in psychiatric outpatients. *Suicide and Life-Threatening Behavior*, 29(1), 1-9.
- Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology*, 56(6), 893-897.
- Beck, A. T., Kovacs, M., & Weissman, A. (1975). Hopelessness and suicidal behavior: An overview. *JAMA: Journal of the American Medical Association*, 234(11), 1146-1149.
- Beck, A. T., Kovacs, M., & Weissman, A. (1979). Assessment of suicidal intention: The Scale for Suicide Ideation. *Journal of Consulting and Clinical Psychology*, 47(2), 343-352.
- Beck, A.T. & Steer, R.A. (1978). *Beck Depression Inventory Manual*. The Psychological Corporation Harcourt Brace Jovanovich, Inc. San Antonio.
- Beck, A.T. & Steer, R.A. (1987). *Beck Hopelessness Scale Manual*. The Psychological Corporation. Harcourt Brace, Jovanovich, Inc. San Antonio.
- Beck, A.T. & Steer, R.A. (1991a). *Beck Scale for Suicide Ideation: Manual*. The Psychological Corporation Harcourt Brace Jovanovich, Inc. San Antonio.
- Beck, A. T., & Steer, R. A. (1991b). Relationship between the Beck Anxiety Inventory and the Hamilton Anxiety Rating Scale with anxious outpatients. *Journal of Anxiety Disorders*, 5(3), 213-223.

- Beck, A.T. & Steer, R.A. (1990). *Beck Anxiety Inventory Manual*. The Psychological Corporation Harcourt Brace Jovanovich, Inc. San Antonio.
- Beck, A.T., Steer, R.A., & Brown, G.K. (1996). *Beck Depression Inventory II Manual*. The Psychological Corporation, San Antonio.
- Beck, A.T., Steer, R.A. & Garbin, M.G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty five years of evaluation. *Clinical Psychology Review*, 8, 77-100.
- Beck, A. T., Steer, R. A., Kovacs, M., & Garrison, B. (1985). Hopelessness and eventual suicide: A 10-year prospective study of patients hospitalized with suicidal ideation. *American Journal of Psychiatry*, 142(5), 559-563.
- Beck, A. T., Steer, R. A., & Ranieri, W. F. (1988). Scale for suicide ideation: Psychometric properties of a self-report version. *Journal of Clinical Psychology*, 44(4), 499-505.
- Beck, A. T., Weissman, A., Lester, D., & Trexler, L. (1974). The measurement of pessimism: The Hopelessness Scale. *Journal of Consulting and Clinical Psychology*, 42(6), 861-865.
- Beckman, H. T., & Lueger, R. J. (1997). Readability of self-report clinical outcome measures. *Journal of Clinical Psychology*, 53(8), 785-789.
- Benishek, L. A., Hayes, C. M., Bieschke, K. J., & Stoffelmayr, B. E. (1998). Exploratory and confirmatory factor analyses of the Brief Symptom Inventory among substance abusers. *Journal of Substance Abuse*, 10(2), 103-114.
- Benowitz, N.L., Hall, S.M., Herning, R.I., Jacob, P., Jones, R.T. & Osman, A.L. (1983). Smokers of low-yield cigarettes do not consume less nicotine. *New England Journal of Medicine*, 309, 139-142.
- Biron, L. L., Brochu, S., & Desjardins, L. (1995). The issue of drugs and crime among a sample of incarcerated women. *Deviant Behavior*, 16(1), 25-43.
- Bisson, J., Nadeau, L., & Demers, A. (1999). The validity of the CAGE scale to screen for heavy drinking and drinking problems in a general population survey. *Addiction*, 94(5), 715-722.
- Black, C.M.D. & Wilson, G.T. (1996). Assessment of eating disorders: interview versus questionnaire. *International Journal of Eating Disorders*, 20, 43-50.
- Blanke, R.V. (1986). Accuracy in urinalysis. In Hawks, R.L. & Chiang, C.N. (Eds.), *Urine Testing for Drugs of Abuse*, NIDA Research Monograph 73, National Institute of Drug Abuse, Rockville, Maryland.
- Blanton Lacy, M. D. (1997). The validity of the Beck Depression Inventory, Scale for Suicide Ideation, and Hopelessness Scale in an African-American college population. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 58(1-B), 0409.
- Blow, F. C., Brower, K. J., Schulenberg, J. E., Demo-Dananberg, L. M., Young, K. J., & Beresford, T. P. (1992). The Michigan Alcoholism Screening Test: Geriatric version (MAST-G): A new elderly-specific screening instrument. *Alcoholism: Clinical and Experimental Research*, 16, 176.
- Bohn, M.J., Babor, T., & Kranzler, H.R. (1991). validity of the Drug Abuse Screening Test (DAST-10) in inpatient substance abusers. *Problems of Drug Dependence*, 119, 233-235
- Bonicatto, S., Dew, A. M., & Soria, J. J. (1998). Analysis of the psychometric properties of the Spanish version of the Beck Depression Inventory in Argentina. *Psychiatry Research*, 79(3), 277-285.
- Bonyng, E. R. (1993). Unidimensionality of SCL-90—R scales in adult and adolescent crisis samples. *Journal of Clinical Psychology*, 49(2), 212-215.
- Boulet, J., & Boss, M. W. (1991). Reliability and validity of the Brief Symptom Inventory. *Psychological Assessment*, 3(3), 433-437.
- Bradley, K. A., Boyd Wickizer, J., Powell, S. H., & Burman, M. L. (1998). Alcohol screening questionnaires in women: A critical review. *JAMA: Journal of the American Medical Association*, 280(2), 166-171.
- Bradley, K.A., Bush, K.R., McDonnell, M.B., Malone, T., & Fihn, S.D. (1998). Screening for problem drinking: Comparison of the CAGE and AUDIT. *Journal of General Internal Medicine*, 13, 379-388.
- Brady, M. (1992). *Heavy Metal: The social meaning of petrol sniffing in Australia*. Canberra: Aboriginal Studies Press.

- Brady, M., Dawe, S. & Richmond, R. Expanding knowledge among Aboriginal service providers on treatment options for excessive alcohol use. *Drug and Alcohol Review*, 17, 69-76.
- Breakey, W. R., Calabrese, L., Rosenblatt, A., & Crum, R. M. (1998). Detecting alcohol use disorders in the severely mentally ill. *Community Mental Health Journal*, 34(2), 165-174.
- Brown, R.A., Burgess, E.S., Sales, S.D., Whiteley, J.A., Evans, D.M., & Miller, I.W. (1998). Reliability and validity of a smoking timeline follow-back interview. *Psychology of Addictive Behaviours*, 12, 101-112.
- Brown, S., Inskip, H., & Barraclough, B. (2000). Causes of the excess mortality of schizophrenia. *British Journal of Psychiatry*, 177, 212-217.
- Buchsbaum, D. G., Buchanan, R. G., Welsh, J., Centor, R. M., & Schnoll, S. H. (1992). Screening for drinking disorders in the elderly using the CAGE questionnaire. *Journal of the American Geriatric Society*, 40(7), 662-665.
- Burge, S. K. (1988). Post-traumatic stress disorder in victims of rape. *Journal of Traumatic Stress*, 1(2), 193-210.
- Burgess, P.M., Stripp, A.M., Pead, J. & Holman, C.P. (1989). Severity of opiate dependence in an Australian sample: Further validation of the SODQ. *British Journal of Addiction*, 84, 1451-1459.
- Burns, C.B., D'Abbs, P. & Currie, B.J. (1995). Patterns of petrol sniffing and other drug use in young men from an Australian Aboriginal community in Arnhem Land, Northern Territory. *Drug and Alcohol Review*, 14, 159-169.
- Burvill, P.W., Knuiman, M.W. (1983). Which version of the GHQ should be used in community samples? *Australian and New Zealand Journal of Psychiatry*, 17, 237-242.
- Butler, S. F., Budman, S. H., Goldman, R. J., Newman, F. J., Beckley, K. E., Trotter, D., & Cacciola, J. S. (2001). Initial validation of a computer-administered Addiction Severity Index: The ASI-MV. *Psychology of Addictive Behaviors*, 15(1), 4-12.
- Byrne, B. M., Baron, P., & Balev, J. (1996). The Beck Depression Inventory: Testing for its factorial validity and invariance across gender for Bulgarian non-clinical adolescents. *Personality and Individual Differences*, 21(5), 641-651.
- Byrne, B. M., Baron, P., Larsson, B., & Melin, L. (1996). Measuring depression for Swedish nonclinical adolescents: Factorial validity and equivalence of the Beck Depression Inventory across gender. *Scandinavian Journal of Psychology*, 37(1), 37-45.
- Cahalan, D., Cisin, C. & Crossely, H.M. (1967). American drinking practises: A national survey of behaviour and attitudes related to alcoholic beverages. Report No. 3. Washington, D.C.: Social Research Group, The George Washington University cited in NIAAA Treatment Handbook Series: Alcoholism treatment assessment research instruments. D.J. Lettieri, J.E. Nelson, M.A. Sayers (eds) US Department of Health and Human Services 2.
- Caraveo Anduaga, J. J., Martinez, N. A., Saldivar, G., Lopez J, J. L., & Saltijeral, M. T. (1998). Performance of the GHQ-12 in relation to current and lifetime CIDI psychiatric diagnoses: GHQ-12 in relation to CIDI diagnoses. *Salud Mental*, 21(4), 1-11.
- Carey, K. B., Cocco, K. M., & Correia, C. J. (1997). Reliability and validity of the Addiction Severity Index among outpatients with severe mental illness. *Psychological Assessment*, 9(4), 422-428.
- Carey, K. B., & Correia, C. J. (1998). Severe mental illness and addictions: assessment considerations. *Addictive Behaviours*, 23(6), 735-748.
- Carey, K. B., & Simons, J. (2000). Utility of collateral information in assessing substance use among psychiatric outpatients. *Journal of Substance Abuse*, 11(2), 139-147.
- Carey, K. B., & Teitelbaum, L. M. (1996). Goals and methods of alcohol assessment. *Professional Psychology: Research and Practice*, 27(5), 460-466.
- Carney, M.-A., Tennen, H., Affleck, G., del-Boca, F.-K., & Kranzler, H.-R. (1998). Levels and patterns of alcohol consumption using timeline follow-back, daily diaries and real-time "electronic interviews." *Journal of Studies on Alcohol*, 59(4), 447-454.
- Carpenter, K. M., & Hittner, J. B. (1995). Dimensional characteristics of the SCL-90-R: evaluation of gender differences in dually diagnosed inpatients. *J Clin Psychol*, 51(3), 383-390.
- Chan, A.W.K. (1991). Biochemical markers for alcoholism. In: Miller, N.S.ed., *Comprehensive Handbook of Drug and Alcohol Addiction*. New York: Marcel Dekker.

- Chan, A. W. K. (1993). Recent developments in detection and biological indicators of alcoholism. *Drugs and Society*, 8(1), 31-67.
- Chan, A.W.K., Pristach, E.A. & Welte, J.W. (1994). Detection of alcoholism in three populations by the Brief - MAST. *Alcoholism Clinical and Experimental Research*, 18, 695-701.
- Chan, A. W. K., Pristach, E. A., Welte, J. W., & Russell, M. (1993). Use of the TWEAK test in screening for alcoholism/heavy drinking in three populations. *Alcoholism: Clinical and Experimental Research*, 17(6), 1188-1192.
- Chan Pensley, E. (1999). Alcohol-Use Disorders Identification Test: a comparison between paper and pencil and computerized versions. *Alcohol and Alcoholism*, 34(6), 882-885.
- Chang, G. (1997). Primary care: Detection of women with alcohol use disorders. *Harvard Review of Psychiatry*, 4, 334-337.
- Chang, G., Goetz, M. A., Wilkins Haug, L., & Berman, S. (1999). Identifying prenatal alcohol use: Screening instruments versus clinical predictors. *American Journal on Addictions*, 8, 87-93.
- Chang, G., Wilkins Haug, L., Berman, S., & Goetz, M. A. (1999). The TWEAK: Application in a prenatal setting. *Journal of Studies on Alcohol*, 60(3), 306-309.
- Chang, G., Wilkins Haug, L., Berman, S., Goetz, M. A., Behr, H., & Hiley, A. (1998). Alcohol use and pregnancy: improving identification. *Obstetrics and Gynecology*, 91(6), 892-898.
- Chantarujikapong, S. I., Smith, E. M., & Fox, L. W. (1997). Comparison of the Alcohol Dependence Scale and Diagnostic Interview Schedule in homeless women. *Alcoholism: Clinical and Experimental Research*, 21(4), 586-595.
- Cherpitel, C. J. (1995). Analysis of cut-off points for screening instruments for alcohol problems in the emergency room. *Journal of Studies on Alcohol*, 56(6), 695-700.
- Cherpitel, C. J. (1997). Comparison of screening instruments for alcohol problems between black and white emergency room patients from two regions of the country. *Alcoholism: Clinical and Experimental Research*, 21(8), 1391-1397.
- Cherpitel, C. J. (1998). Differences in performance of screening instruments for problem drinking among Blacks, Whites and Hispanics in an emergency room population. *Journal of Studies on Alcohol*, 59(4), 420-426.
- Cherpitel, C. J. (1999). Screening for alcohol problems in the U.S. general population: A comparison of the CAGE and TWEAK by gender, ethnicity, and services utilization. *Journal of Studies on Alcohol*, 60(5), 705-711.
- Chung, T., Colby, S. M., Barnett, N. P., Rohsenow, D. J., Spirito, A., & Monti, P. M. (2000). Screening adolescents for problem drinking: Performance of brief screens against DSM-IV alcohol diagnoses. *Journal of Studies on Alcohol*, 61(4), 579-587.
- Churchill, A.C., Burgess, P.M., Pead, J. & Gill, T. (1993). Measurement of the severity of amphetamine dependence. *Addiction*, 88, 1335-1340.
- Claussen, B., & Aasland, O. G. (1993). The Alcohol Use Disorders Identification Test (AUDIT) in a routine health examination of long-term unemployed. *Addiction*, 88(3), 363-368.
- Clough, A. R., Burns, C. B. & Mununggurr, N. (2000). Kava in Arnhem Land: a review of consumption and it's social correlates. *Drug and Alcohol Review*, 19, 319 -328.
- Cocco, K. M., & Carey, K. B. (1998). Psychometric properties of the Drug Abuse Screening Test in psychiatric outpatients. *Addiction*, 94(2), 269-281.
- Cochrane Brink, K. A., Lofchy, J. S., & Sakinofsky, I. (2000). Clinical rating scales in suicide risk assessment. *General Hospital Psychiatry*, 22(6), 445-451.
- Comfort, M., & Kaltenbach, K. A. (1996). The psychosocial history: an interview for pregnant and parenting women in substance abuse treatment and research. In E. R. R. (ED.) (Ed.), *Treatment for drug-exposed women and their children: Advances in research methodology. NIDA Research Monograph 165* (pp. 123-142). Rockville, MD: U.S. Department of Health and Human Services.
- Comfort, M., Zanis, D. A., Whiteley, M. J., Kelly Tyler, A., & Kaltenbach, K. A. (1999). Assessing the needs of substance abusing women: Psychometric data on the psychosocial history. *Journal of Substance Abuse Treatment*, 17(1-2), 79-83.
- Commonwealth Department of Health and Aged Care (1999). *National Recommendations for the Clinical Management of Alcohol Related Problems in Indigenous Primary Care Settings*, Canberra, ACT: AGPS.

- Conigliaro, J., Kraemer, K., & McNeil, M. (2000). Screening and identification of older adults with alcohol problems in primary care. *Journal of Geriatric Psychiatry and Neurology*, 13(3), 106-114.
- Conigrave, K. M., & Elvy, G. (1998). An Australian modification of the popular AUDIT questionnaire. *Drug and Alcohol Review*, 17(2), 223-224.
- Conigrave, K. M., Hall, W. D., & Saunders, J. B. (1995). The AUDIT questionnaire: Choosing a cut-off score. *Addiction*, 90(10), 1349-1356.
- Conigrave, K. M., Saunders, J. B., & Reznik, R. B. (1995). Predictive capacity of the AUDIT questionnaire for alcohol-related harm. *Addiction*, 90(11), 1479-1485.
- Cooper, L., Peters, L., & Andrews, G. (1998). Validity of the Composite International Diagnostic Interview (CIDI) psychosis module in a psychiatric setting. *Journal of Psychiatric Research*, 32(6), 361-368.
- Corse, S. J., Hirschinger, N. B., & Zanis, D. (1995). The use of the Addiction Severity Index with people with severe mental illness. *Psychiatric Rehabilitation Journal*, 19(1), 9-18.
- Cottler, L. B., Grant, B. F., Blaine, J., Mavreas, V., Pull, C., Hasin, D., Compton, W. M., Rubio Stipeck, M., & Mager, D. (1997). Concordance of the DSM-IV alcohol and drug use disorder criteria and diagnoses as measured by AUDADIS-ADR, CIDI and SCAN. *Drug and Alcohol Dependence*, 47(3), 195-205.
- Creamer, M., Foran, J., & Bell, R. (1995). The Beck Anxiety Inventory in a non-clinical sample. *Behaviour Research and Therapy*, 33(4), 477-485.
- Crook, G.M., Oei, T.P.S. & Young, R.M. (1994). Structure of the MAST with an Australian sample of alcoholics. *Drug and Alcohol Review*, 13, 41-46.
- Cyr, J. J., McKenna Foley, J. M., & Peacock, E. (1985). Factor structure of the SCL-90-R: Is there one? *Journal of Personality Assessment*, 49(6), 571-578.
- Daepfen, J. B., Yersin, B., Landry, U., Pecoud, A., & Decrey, H. (2000). Reliability and validity of the Alcohol Use Disorders Identification Test (AUDIT) imbedded within a general health risk screening questionnaire: Results of a survey in 332 primary care patients. *Alcoholism: Clinical and Experimental Research*, 24(5), 659-665.
- Darke, S. (1998). Self-report among injecting drug users: A review. *Drug and Alcohol Dependence*, 51(3), 253-263.
- Darke, S., Hall, W., Wodak, A., Heather, N. & Ward, J. (1992). Development and validation of a multi-dimensional instrument for assessing outcome of treatment among opiate users: The Opiate Treatment Index. *British Journal of Addiction*, 87, 733-742.
- Darke, S., Ross, J., & Hall, W. (1996). Overdose among heroin users in Sydney, Australia: I. prevalence and correlates of non-fatal overdose. *Addiction*, 91(3), 405-411.
- Darke, S., Ward, J., Hall, W., Heather, N. & Wodak, A. (1991). *The Opiate Treatment Index (OTI) manual*. NDARC Monograph 11. Sydney: National Drug and Alcohol Research Centre.
- Davison, M. L., Bershadsky, B., Bieber, J., Silversmith, D., Maruish, M. E., & Kane, R. L. (1997). "Development of a brief, multidimensional, self-report instrument for treatment outcomes assessment in psychiatric settings: Preliminary findings": Erratum. *Assessment*, 4(4), 419.
- Davidson, R. (1986). Assessment of the alcohol dependence syndrome: A review of self-report screening questionnaires. *British Journal of Addiction*, 26, 243-255.
- Davidson, R. & Raistrick, D. (1986). The validity of the Short Alcohol Dependence Data (SADD) questionnaire: A short self-report questionnaire for the assessment of alcohol dependence. *British Journal of Addiction*, 81, 217-222.
- Davidson, R. Bunting, B. & Raistrick, D. (1989). The homogeneity of the alcohol dependence syndrome: A factorial analysis of the SADD questionnaire. *British Journal of Addiction*, 84, 907-915.
- Dawe, S., Harnett, P.H. Rendalls, V. & Staiger, P. (2001) *Behavioural Family Intervention in methadone maintained families: Results of a pilot study*. Manuscript submitted for publication.
- Dawe, S. & Richmond, R. (1997) The use of brief interventions for controlled drinking in Australian alcohol agencies. *Journal of Substance Misuse and Treatment*, 14, 1-6
- Dawe, S., Seinen, A., & Kavanagh, D. (2000). An examination of the utility of the AUDIT in people with schizophrenia. *Journal of Studies on Alcohol*, 61(5), 744-750.

- de Beurs, E., Wilson, K. A., Chambless, D. L., Goldstein, A. J., & Feske, U. (1997). Convergent and divergent validity of the Beck Anxiety Inventory for patients with panic disorder and agoraphobia. *Depress Anxiety*, 6(4), 140-146.
- Deering, D. E. A., & Sellman, J. (1996). An inter-rater reliability study of the Opiate Treatment Index. *Drug and Alcohol Review*, 15(1), 57-63.
- Degenhardt, L.J., Conigrave, K.M., Wutzke, S.E., & Saunders, J.B. (2001). The validity of an Australian modification of the AUDIT questionnaire. *Drug and Alcohol Review*, 20, 143-154.
- DeHart, S. S., & Hoffman, N. G. (1997). Screening and diagnosis: Alcohol Use Disorders in older adults. In A. M. Gurnack (Ed.), *Older adults' misuse of alcohol, medicines, and other drugs : research and practice issues* (pp. 25-53). New York: Springer.
- de las Cuevas, C., Sanz, E. J., de la Fuente, J. A., Padilla, J., & Berenguer, J. C. (2000). The Severity of Dependence Scale (SDS) as screening test for benzodiazepine dependence: SDS validation study. *Addiction*, 95(2), 245-250.
- De Paulo, J.R. & Folstein, M.F. (1978). Psychiatric disturbances in neurological patients: Detection, recognition and hospital course. *Annals of Neurology*, 4, 225-228.
- Derogatis, L.R. (1994). *Symptom Checklist-90-Revised: Administration, scoring and procedures manual, 3rd edition*. National Computer Systems, Inc., Minneapolis, MN 55440
- Derogatis, L. R., & Melisaratos, N. (1983). The Brief Symptom Inventory: An introductory report. *Psychological Medicine*, 13(3), 595-605.
- DeWit, D. J., Adlaf, E. M., Offord, D. R., & Ogbourne, A. C. (2000). Age at first alcohol use: a risk factor for the development of alcohol disorders. *American Journal of Psychiatry*, 157, 745-750.
- Dixon, L. Haas, G., Weiden, P.J., Sweeney, J. & Frances, A.J. (1991). Drug abuse in schizophrenic patients: Clinical correlates and reasons for use. *American Journal of Psychiatry*, 148, 224-230.
- Doherty, B. & Webb, M. (1989). The distribution of alcohol dependence severity among in-patient problem drinkers. *British Journal of Addiction*, 84, 917-922.
- Dozois, D. J. A., Dobson, K. S., & Ahnberg, J. L. (1998). A psychometric evaluation of the Beck Depression Inventory-II. *Psychological Assessment*, 10(2), 83-89.
- Drake, R., Osher, F., & Wallach, M. (1989). Alcohol use and abuse in schizophrenia: A prospective community study. *Journal of Nervous and Mental Disease*, 177, 408-414.
- Drake, R. E., McHugo, G. J., & Biesanz, J. C. (1995). The test-retest reliability of standardized instruments among homeless persons with substance use disorders. *Journal of Studies on Alcohol*, 56(2), 161-167.
- Drew, N. (2000). Psychological testing with Indigenous people in Australia. In D. G. H. P. E. P. Dudgeon (Ed.), *Working with Indigenous Australians: A Handbook for Psychologists*. (pp. 325-334.). Perth WA: Gunada Press.
- Dubowski, K.M. (1991). *The technology of breath-alcohol analysis*. Rockville, MD: US Dept of Health and Human Services, National Institute on Alcohol Abuse and Alcoholism.
- Dudgeon, P., Garvey, D., & Pickett, H. (2000). *Working with indigenous Australians : a handbook for psychologists*. Perth, W.A.: Gunada Press.
- Dupont, R.L. & Baumgartner, W.A. (1995). Drug testing by urine and hair analysis: Complementary features and scientific issues. *Forensic Science International*, 70, 63-76.
- Edwards, G. & Gross, M.M. (1976). Alcohol dependence: Provisional description of a clinical syndrome. *British Medical Journal*, 1, 1058-1061.
- Edwards, G., Arif, A. & Hodgson, R. (1981). Nomenclature and classification of drug and alcohol problems, *Bulletin of the World Health Organisation*, 59, 225-242.
- Ee Heok, K., Soo Meng, K., Cheong Sung, T., Yeo, B., Halstead, S. & Mahendran, R. (1990). A follow-up study of Asian problem drinkers. *British Journal of Addiction*, 85, 1261-1264.
- Ellis, T. E., & Ratliff, K. G. (1986). Cognitive characteristics of suicidal and nonsuicidal psychiatric inpatients. *Cognitive Therapy and Research*, 10(6), 625-634.
- Etter, J. F., Vu Duc, T., & Perneger, T. V. (1999). Validity of the Fagerstroem test for nicotine dependence and of the Heaviness of Smoking Index among relatively light smokers. *Addiction*, 94(2), 269-281.

- Ewing, J.A. (1984). Detecting alcoholism: The CAGE questionnaire. *JAMA*, 252, 1905-1907.
- Ewing, J. A., Bradley, K. A., & Burman, M. L. (1998). Screening for alcoholism using CAGE. *JAMA: Journal of the American Medical Association*, 280(22), 1904.
- Fagerstrom, K.O. (1978). Measuring degree of physical dependence to tobacco smoking with reference to individualisation of treatment. *Addictive Behaviors*, 3, 235-241.
- Fagerstrom, K. & Schneider, N.G. (1989). Measuring nicotine dependence: A review of the Fagerstrom Tolerance Questionnaire. *Journal of Behavioural Medicine*, 12, 159-182.
- Fals Stewart, W., O'Farrell, T. J., Freitas, T. T., McFarlin, S. K., & Rutigliano, P. (2000). The Timeline Followback reports of psychoactive substance use by drug-abusing patients: Psychometric properties. *Journal of Consulting and Clinical Psychology*, 68(1), 134-144.
- Fedoroff, I.C., & McFarlane, T. (1998). Cultural aspects of eating disorders. In S.S. Kazarian & D.R. Evans (Eds.). *Cultural Clinical Psychology: Theory, research and practice*. (pp.152-176). New York: Oxford University Press.
- Fergusson, D. M., Horwood, L. J., & Lynskey, M. T. (1995). The prevalence and risk factors associated with abusive or hazardous alcohol consumption in 16-year-olds. *Addiction*, 90(7), 935-946.
- Ferri, C. P., Marsden, J., de Araujo, M., Laranjeira, R. R., & Gossop, M. (2000). Validity and reliability of the Severity of Dependence Scale (SDS) in a Brazilian sample of drug users. *Drug and Alcohol Review*, 19(4), 451-455.
- Finch, E., & Strang, J. (1998). Reliability and validity of self-report: on the importance of considering context. *Drug and Alcohol Dependence*, 51(3), 269.
- Fingerhood, M. (2000). Substance abuse in older people. *Journal of the American Geriatric Society*, 48(8), 985-995.
- Flemenbach, A. & Zimmerman, R.L. (1973). Inter- and intra-rater reliability of the Brief Psychiatric Rating Scale. *Psychological Reports*, 36, 783-792.
- Fleming, M. F., Barry, K. L., & MacDonald, R. (1991). The Alcohol Use Disorders Identification Test (AUDIT) in a college sample. *International Journal of the Addictions*, 26(11), 1173-1185.
- Forensic Science International* (1995), 70, 1-222.
- Forensic Science International* (1997), 84, 1-313.
- Franken, I. H., & Hendriks, V. M. (1999). Predicting outcome of inpatient detoxification of substance abusers. *Psychiatr Serv*, 50(6), 813-817.
- Frederiksen, L.W., Epstein, L.H. & Kosevsky, B.P. (1975). Reliability and controlling effects of three procedures for self-monitoring smoking. *Psychological Record*, 25, 255-264.
- Freeston, M. H., Ladouceur, R., Thibodeau, N., Gagnon, F., & Rheaume, J. (1994). L'inventaire d'anxiété de Beck. Propriétés psychométriques d'une traduction française. [The Beck Anxiety Inventory. Psychometric properties of a French translation]. *Encephale*, 20(1), 47-55.
- Fydrich, T., Dowdall, D., & Chambless, D. L. (1992). Reliability and validity of the Beck Anxiety Inventory. *Journal of Anxiety Disorders*, 6(1), 55-61.
- Gariti, P.W., Alterman, A.I., Ehrman, R.N., & Pettinati, H.M. (1998). Reliability and validity of the aggregate method if determining number of cigarettes smoked per day. *The American Journal of Addictions*, 7, 283-287.
- Garner, D. (1991). *Eating Disorder Inventory - 2 Professional Manual*. Odessa, FL: Psychological Assessment Resources, Inc.
- Garner, D.M. (1997). Psychosocial principles in treatment. In D.M. Garner & P.E. Garfinkel (Eds.). *Handbook for treatment for eating disorders*. New York: Guilford Press.
- Garner, D. M. (2001). *Eating Disorders Inventory-2 (EDI-2)*. Retrieved April 26, 2001, from the World Wide Web: <http://www.eatingdisorders-toledo.com/EDI-2.html>
- Garner, D.M. & Garfinkel, P.E. (1979). The Eating Attitudes Test: An index of the symptoms of anorexia nervosa. *Psychological Medicine*, 9, 273-279.
- Garner, D.M., Olmstead, M.P., Bohr, Y. & Garfinkel, P.E. (1982). The eating attitudes test: Psychometric features and clinical correlates. *Psychological Medicine*, 12, 871-878.

- Garner, D.M., Olmstead, M.P., & Polivy, J. (1983). Development and validation of a multi-dimensional Eating Disorder Inventory for anorexia and bulimia. *International Journal of Eating Disorders*, 7, 47-55.
- Gavin, G.R., Ross, H.E. & Skinner, A. (1989). Diagnostic validity of the Drug Abuse Screening Test in the assessment of DSM-III drug disorders. *British Journal of Addiction*, 84, 301-307.
- George, T. P., Zeidonis, D. M., Feingold, A., Pepper, W. T., Satterburg, C. A., Winkel, J., Rounsaville, B. J., & Kosten, T. R. (2000). Nicotine transdermal patch and atypical antipsychotic medications for smoking cessation in schizophrenia. *American Journal of Psychiatry*, 157, 1835-1842.
- Glanz, L. M., Haas, G. L., & Sweeney, J. A. (1995). Assessment of hopelessness in suicidal patients. *Clinical Psychology Review*, 15(1), 49-64.
- Goldberg, D. P., Gater, R., Sartorius, N., Ustun, T. B., Piccinelli, M., Gureje, O., & Rutter, C. (1997). The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychological Medicine*, 27(1), 191-197.
- Goldberg, D. & Williams, P. (1988). *A Users's Guide to the General Health Questionnaire*. NFER-NELSON Publishing Co. Ltd. Windsor, Berkshire, UK.
- Gomberg, E. S. L. (1990). Drugs, alcohol, and aging. In L. T. Kozlowski & H. M. Annis (Eds.), *Research advances in alcohol and drug problems, Vol. 10* (pp. 171-213). New York, NY, US: Plenum Press.
- Gorelick, D.A., Irwin, M.R., Schmidt-Lackner, S. & Marder, S.R. (1990). Alcoholism among male schizophrenic inpatients. *Annals of Clinical Psychiatry*, 2, 19-22.
- Gossop, M., Best, D., Marsden, J., & Strang, J. (1997). Test-retest reliability of the Severity of Dependence Scale. *Addiction*, 92(3), 353.
- Gossop, M., Darke, S., Griffiths, P., Hando, J., Powis, B., Hall, W. & Strang, J. (1995). The Severity of Dependence Scale (SDS): Psychometric properties of the SDS in English and Australian samples of heroin, cocaine and amphetamine users. *Addiction*, 90, 607-614.
- Gotlib, I. H., & Cane, D. B. (1989). Self-report assessment of depression and anxiety. In P. C. Kendall & D. Watson (Eds.), *Anxiety and depression: Distinctive and overlapping features* (pp. 131-169). San Diego, CA, US: Academic Press Inc.
- Grant, B., & Dawson, D. A. (1997). Age at onset of alcohol use and its association with DSM-IV alcohol abuse and dependence: results from the National Longitudinal Alcohol Epidemiologic Survey. *Journal of Substance Abuse*, 9, 103-110.
- Green, S. & Whichelow (1994). Longitudinal validity of the CAGE questionnaire. *Addiction Research*, 2, 195-201.
- Gross, J., Rosen, J.C., Leitenberg, H., & Willmuth, M.E. (1986). Validity of the Eating Attitudes Test and the Eating Disorders Inventory in bulimia nervosa. *Journal of Consulting and Clinical Psychology*, 54(6), 875-876.
- Gureje, O., Obikoya, B. & Ikluesan, B.A. (1992). Alcohol abuse and dependence in an urban primary care clinic in Nigeria. *Drug and Alcohol Dependence*, 30, 163-167.
- Gust, S., Hughes, J.R. & Pechacek, T. (1986). Prevalence of tobacco dependence and withdrawal. In L. Harris (Ed), *Problems of drug dependence 1985: Proceedings of the 47th annual scientific meeting of the Committee on Problems of Drug Dependence* (NIDA Research Monograph No 67). Washington, DC:U.S. Government Printing Office.
- Hafkenschied, A. (1991). Psychometric evaluation of a standardised and expanded Brief Psychiatric Rating Scale. *Acta Psychiatrica Scandinavica*, 84, 294-300.
- Hall, W. (1992). The Australian debate about the legalisation of heroin and other illicit drugs, 1988-1991. *Journal of Drug Issues*, 22, 563-577.
- Hamdi, N., Nizam, A.H. & Sabir, A.T. (1988). The factorial structure, reliability and validity of an Arabic version of the Beck Depression Inventory. *Dirasat*, 15, 30-40 (Abstract available in English).
- Hamid, R., Deren, S., Beardsley, M., & Tortu, S. (1999). Agreement between urinalysis and self-reported drug use. *Subst Use Misuse*, 34(11), 1585-1592.
- Hanley, J. & McNeil, B.J. (1982). The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. *Diagnostic Radiology*, 143, 29-36.

- Hardy, G. E., Shapiro, D. A., Haynes, C. E., & Rick, J. E. (1999). Validation of the General Health Questionnaire-12: Using a sample of employees from England's health care services. *Psychological Assessment, 11*(2), 159-165.
- Hawks, R.L. & Chiang, C.N. (1986). Examples of specific drug assays. In Hawks, R.L. and Chiang, C.N. (Eds.), *Urine Testing for Drugs of Abuse*, NIDA Research Monograph 73, National Institute of Drug Abuse, Rockville, Maryland.
- Hays, R.D., Merz, J.F. & Nicholas, R. (1995). Response burden, reliability and validity of the CAGE, Short MAST and AUDIT alcohol screening measures. *Behaviour Research Methods, Instruments and Computers, 27*, 277-280
- Heatherston, T. F., Kozlowski, L. T., Frecker, R. C., & Fagerstrom, K. O. (1991). The Fagerstrom Test for Nicotine Dependence: A revision of the Fagerstrom Tolerance Questionnaire. *British Journal of Addiction, 86*(9), 1119-1127.
- Helander, A., & Tabakoff, B. (1997). Biochemical markers of alcohol use and abuse: Experiences from the pilot study of the WHO/ISBRA collaborative project on state and trait markers of alcohol. *Alcohol and Alcoholism, 32*(2), 133-144.
- Henderson, M. & Freeman, C.P.L. (1987). A self-rating scale for bulimia: The BITE. *British Journal of Psychiatry, 150*, 18-24.
- Henningfield, J. E. (1990). Understanding nicotine addiction and physical withdrawal process. *Journal of the American Dental Association, Suppl*2S-6S.
- Hides, L., Dawe, S., Kavanagh, D., & Young, R. (2001). *The reliability and validity of measures of cannabis abuse in schizophrenia*. Manuscript submitted for publication.
- Hodgins, D. C., & Lightfoot, L. O. (1989). The use of the Alcohol Dependence Scale with incarcerated male offenders. *International Journal of Offender Therapy and Comparative Criminology, 33*(1), 59-67.
- Hodgkinson, P., & Joseph, S. (1995). Factor analysis of the Impact of Events Scale with female bank staff following an armed raid. *Personality and Individual Differences, 19*(5), 773-775.
- Hoffer, A., & Foster, H. D. (2000). Why schizophrenics smoke but have a lower incidence of lung cancer: Implications for the treatment of both disorders. *Journal Orthomolecular Medicine, 15*, 141-144.
- Holderness, C. C., Brooks Gunn, J., & Warren, M. P. (1994). Co-morbidity of eating disorders and substance abuse review of the literature. *International Journal of Eating Disorders, 16*(1), 1-34.
- Horowitz, L.M., Rosenberg, S.E., Baer, B.A., Ureno, G. & Villasenor, V.S. (1988). Inventory of interpersonal problems: Psychometric properties and clinical applications. *Journal of Consulting and Clinical Psychology, 56*, 885-892.
- Horowitz, M. J., Wilner, N., & Alvarez, W. (1979). Impact of Event Scale: A measure of subjective stress. *Psychosomatic Medicine, 41*(3), 209-218.
- Hulse, G. K., Saunders, J. B., Roydhouse, R. M., Stockwell, T. R., & Basso, M. R. (2000). Screening for hazardous alcohol use and dependence in psychiatric in-patients using the AUDIT questionnaire. *Drug and Alcohol Review, 19*(3), 291-298.
- Hunter, E. (1993). *Aboriginal health and history: Power and prejudice in remote Australia*. Cambridge: Cambridge University Press.
- Hunter, E., Hall, W. & Spargo, R. (1991). *The distribution and correlates of alcohol consumption in a remote aboriginal population*. NDARC Monograph 12. Sydney: National Drug and Alcohol Research Centre.
- Hunter, E. E., Powell, B. J., Penick, E. C., Nickel, E. J., Othmer, E., & DeSouza, C. (2000). Development and validation of a mania scale for the Symptom Checklist 90. *Journal of Nervous and Mental Disease, 188*(3), 176-178.
- Indran, S.K. (1992). Quantity Frequency (Consumption Index) Versus "Cage" in the Detection of Alcoholism, *Australian and New Zealand Journal of Psychiatry, 26*, 493-501.
- Ivanoff, A., Jang, S.J., Smyth, N.J. & Linehan, M.M. (1994). Fewer reasons for staying alive when you are thinking of killing yourself: The Brief Reasons for Living Inventory. *Journal of Psychopathology and Behavioral Assessment, 16*, 1-13.
- Jacob, K. S., Bhugra, D., & Mann, A. H. (1997). The validation of the 12-item General Health Questionnaire among ethnic Indian women living in the United Kingdom. *Psychol Med, 27*(5), 1215-1217.

- Jarvis, M., Tunstall-Pedoe, H., Feyerabend, C., Vesey, C., & Saloojee, Y. (1984). Biochemical markers of smoke absorption and self-reported exposure to passive smoking. *Journal of Epidemiology and Community Health*, 38, 335-339.
- Joiner, T.-E., Jr., Heatherton, T.-F., Rudd, M. D., & Schmidt, N.-B. (1997). Perfectionism, perceived weight status, and bulimic symptoms: Two studies testing a diathesis-stress model. *Journal of Abnormal Psychology*, 106(1), 145-153.
- Jorge, M.R. & Masur, J. (1985). The use of the short-form alcohol dependence data questionnaire (SADD) in Brazilian alcoholic patients. *British Journal of Addiction*, 80, 301-305.
- Joseph, S. A., Williams, R., Yule, W., & Walker, A. (1992). Factor analysis of the Impact of Events Scale with survivors of two disasters at sea. *Personality and Individual Differences*, 13(6), 693-697.
- Kabacoff, R. I., Segal, D. L., Hersen, M., & Van Hasselt, V. B. (1997). Psychometric properties and diagnostic utility of the Beck Anxiety Inventory and the State-Trait Anxiety Inventory with older adult psychiatric outpatients. *Journal of Anxiety Disorders*, 11(1), 33-47.
- Kaminer, Y., Bukstein, O., & Tarter, R. E. (1991). The Teen-Addiction Severity Index: rationale and reliability. *International Journal on Addictions*, 26(2), 219-226.
- Kaminer, Y., Wagner, E., Plummer, B., & Seifer, R. (1993). Validation of the Teen Addiction Severity Index (T-ASI): Preliminary findings. *American Journal on Addictions*, 2(3), 250-254.
- Kan, C. C., Breteler, M. H., Timmermans, E. A., van der Ven, A. H., & Zitman, F. G. (1999). Scalability, reliability, and validity of the benzodiazepine dependence self-report questionnaire in outpatient benzodiazepine users. *Comprehensive Psychiatry*, 40(4), 283-291.
- Kaplan, R., & Saccuzzo, D. (2001). *Psychological Testing: Principles, Applications, and Issues* (5th ed.). Pacific Grove, CA: Brooks/Cole.
- Karno, M., Granholm, E., & Lin, A. (2000). Factor structure of the Alcohol Use Disorders Identification Test (AUDIT) in a mental health clinic sample. *Journal of Studies on Alcohol*, 61(5), 751-758.
- Kavanagh, D. J., McGrath, J. J., & Jenner, L. (2000). Substance use in psychotic disorders: Results from the Australian survey of mental health and wellbeing. *Acta Psychiatrica Scandinavica*, 102, 5.
- Kavanagh, D. J., Saunders, J. B., Young, R., White, A., Jenner, L., Clair, A., & Wallis, J. (1999). *Evaluation and brief intervention for substance abuse in early psychosis: A report to AUSEINET*. Brisbane: Department of Psychiatry, University of Queensland.
- Keller, F and Wolfersdorf, M. (1993). Hopelessness and the tendency to commit suicide in the course of depressive disorders. *Crisis*, 14, 173-177.
- Khavari, K.A. & Farber, P.D. (1978). A profile instrument for the quantification and assessment of alcohol &. *Journal of Studies on Alcohol*, 39, 1525-1539.
- Kilic, C., Rezaki, M., Rezaki, B., Kaplan, I., Ozgen, G., Sagduyu, A., & Ozturk, M. O. (1997). General Health Questionnaire (GHQ12 & GHQ28): psychometric properties and factor structure of the scales in a Turkish primary care sample. *Soc Psychiatry Psychiatr Epidemiol*, 32(6), 327-331.
- Killen, J. D., Ammerman, S., Rojas, N., Varady, J., Haydel, F., & Robinson, T. N. (2001). Do adolescent smokers experience withdrawal effects when deprived of nicotine? *Experimental and Clinical Psychopharmacology*, 9(2), 176-182.
- Killen, J. D., Fortmann, S. P., Davis, L., & Varady, A. (1997). Nicotine patch and self-help video for cigarette smoking cessation. *Journal of Consulting and Clinical Psychology*, 65(4), 663-672.
- Kilpatrick, B., Howlett, M., Sedgwick, P., & Ghodse, A. H. (2000). Drug use, self report and urinalysis. *Drug and Alcohol Dependence*, 58(1-2), 111-116.
- King, M. (1986). At risk drinking among general practice attendees: Prevalence, characteristics and alcohol-related problems. *British Journal of Psychiatry*, 148, 533-540.
- King, M. (1986). At risk drinking among general practice attenders: Validation of the CAGE questionnaire. *Psychological Medicine*, 16, 213-217.
- Kivlahan, D.R., Sher, K.J. and Donovan, D.M. (1989). The alcohol dependence scale: A validation study among inpatient alcoholics. *Journal of Studies on Alcohol*, 50, 170-175.
- Kleinman, P.H., Miller, A.B., Millman, R.B., Woody, G.E., Todd, T., Kemp, J., & Lipton, D. (1990). Psychopathology among cocaine abusers entering treatment. *Journal of Nervous and Mental Disease*, 178, 442-447.

- Kozlowski, L.T., Porter, C.Q., Orleans, C.T. & Pope, M.A. (1994). Predicting smoking cessation with self-reported measures of nicotine dependence: FTQ, FTND and HSI. *Drug and Alcohol Dependence*, 34, 211-216.
- Kulka, R. A., Schlenger, W. E., Fairbank, J. A., Hough, R. L., Jordan, B. K., Marmar, C. R., & Weiss, D. S. (1990). *Trauma and the Vietnam war generation: Report of findings from the National Vietnam Veterans Readjustment Study*. Philadelphia, PA, US: Brunner/Mazel, Inc.
- Kumar, G., Steer, R. A., & Beck, A. T. (1993). Factor structure of the Beck Anxiety Inventory with adolescent psychiatry inpatients. *Anxiety, Stress and Coping: An International Journal*, 6(2), 125-131.
- Lancet (1980). Screening tests for alcoholism. November 22, 1980. 1117-1118.
- Laranjeira, R. (1995). The need for better measurement of the subjective component of the ADS. *Addiction*, 90, 1041-1042.
- Larson, A., Shannon, C., & Eldridge, C. (1999). Indigenous Australians who inject drugs: Results from a Brisbane study. *Drug and Alcohol Review*, 18(1), 53-62.
- Lasser, K., Boyd, J. W., Woolhandler, S., Himmelstein, D. U., McCormick, D., & Bor, D. H. (2000). Smoking and Mental Illness: A Population-Based Prevalence Study. *JAMA*, 284(20), 2606.
- Layton, C. (1986). Note on test-retest characteristics of the General Health Questionnaire. *Perceptual and Motor Skills*, 62, 221-222.
- Lee, S., Lee, A. M., Leung, T., & Yu, H. (1997). Psychometric properties of the Eating Disorders Inventory (EDI-1) in a nonclinical Chinese population in Hong Kong. *International Journal of Eating Disorders*, 21(2), 187-194.
- Leigh, G and Skinner, H. (1988). Physiological assessment. In *Assessment of addictive Behaviors*, (Eds.) D.M. Donovan and G.A. Marlatt. The Guildford Press: New York.
- Lennings, C. J. (1999). An evaluation of the Leeds Dependence Questionnaire. *Journal of Child and Adolescent Substance Abuse*, 8(3), 73-87.
- Leonhard, C., Mulvey, K., Gastfriend, D. R., & Shwartz, M. (2000). The Addiction Severity Index: A field study of internal consistence and validity. *Journal of Substance Abuse Treatment*, 18, 129-135.
- Liu, Z., Lian, Z., Zhou, W., Mu, Y., Teng, Y., Zhu, R., Zhu, G., Chen, J., Cao, J., & Cai, Z. (2000). A preliminary study on the severity of drug addiction in drug abusers. *Chinese Mental Health Journal*, 14(4), 231-234.
- Loxton, N. J. (1998). *Disordered eating and alcohol misuse in adolescent girls : a preliminary investigation of biopsychological influences*. Unpublished honours thesis, Griffith University, Brisbane.
- Loxton, N. J., & Dawe, S. (2001). Alcohol abuse and dysfunctional eating in adolescent girls: The influence of individual differences in sensitivity to reward and punishment. *International Journal of Eating Disorders*, 29(4), 455-462.
- MacKenzie, D. M., Langa, A., & Brown, T. M. (1996). Identifying hazardous or harmful alcohol use in medical admissions: A comparison of AUDIT, CAGE and brief MAST. *Alcohol and Alcoholism*, 31(6), 591-599.
- Maisto, S. A., Conigliaro, J., McNeil, M., Kraemer, K., & Kelley, M. E. (2000). An empirical investigation of the factor structure of the AUDIT. *Psychological Assessment*, 12(3), 346-353.
- Maloney, M. J., McGuire, J. & Daniels, S. R. (1988). Reliability testing of a children's version of the Eating Attitude Test. *Journal of the American Academy of Child and Adolescent Psychiatry*, 27, 541-543.
- Manno, J.E. (1986). Interpretation of urinalysis results. In Hawks, R.L. & Chiang, C.N. (Eds.), *Urine Testing for Drugs of Abuse*, NIDA Research Monograph 73, National Institute of Drug Abuse, Rockville, Maryland.
- Marsh, A., Evans, M.B. and Strang, J. (1995). Radioimmunoassay of drugs of abuse in hair: Part 2: The determination of methadone in the hair of know drug users. *Journal of pharmacological and Biomedical Analysis*, 13, 829-839.
- Martin, C. S., & Winters, K. C. (1998). Diagnosis and assessment of alcohol use disorders among adolescents. *Alcohol Health and Research World*, 22(2), 95-105.

- Martino, S., Grilo, C.M., & Fehon, D.C. (2000). Development of the Dug Abuse Screening Test for adolescents (DAST-A). *Addictive Behaviors, 25*, 57-70.
- Mattick, R.P. & Jarvis, T. (1993). *An outline for the management of alcohol problems*. Monograph Series No 20. National Drug Strategy, Canberra, AGPS.
- Mattick, R.P., Oliphant, D.A., Bell, J. & Hall, W. Psychiatric morbidity in methadone maintenance patients: Prevalence, effect on drug use and detection. submitted for publication
- McDonald, A. S. (1997). Factor structure of the Impact of Events Scale in a non-clinical sample. *Personality and Individual Differences, 23*(3), 419-424.
- McGorry, P.D., Copolov, D.L., & Singh, B.S. (1990). *Royal Park Multidiagnostic Instrument for Psychosis: Part I. Rationale and Review, Schizophrenia Bulletin, 16*: 501-515.
- McGorry, P.D., Mihalopoulos, C., Henry, L., Dakis, J., Jackson, H.J., Flaum, M., Harrigan, S., McKenzie, D., Kulkarni, J., & Karoly, R. (1995). Spurious precision: Procedural validity of diagnostic assessment in psychotic disorders. *American Journal of Psychiatry, 152*, 220-223.
- McGorry, P.D., Singh, B.S., Copolov, D.L., Kaplan, I., Dossetor, C.R., & van Riel, R.J. (1990). *Royal Park Multidiagnostic Instrument for Psychosis: Part II. Development, reliability, and validity. Schizophrenia Bulletin, 16*: 517-535.
- McHugo, G.J., Paskus, T.S. and Drake, R.E. (1993). Detection of alcoholism in schizophrenia using the MAST. *Alcoholism Clinical and Experimental Research, 17*, 187-191.
- McKelvie, G. C., J. (2000). Alcohol and substance misuse amongst Indigenous Australians. In P. Dudgeon & D. Garvey & H. Pickett (Eds.), *Working with Indigenous Australians: A Handbook for Psychologists*. (pp. 499-515). Perth WA: Gunada Press.
- McLellan, A. T., Kushner, H., Metzger, D., Peters, R., Smith, I., Grissom, G., Pettinati, H., & Argeriou, M. (1992). The Fifth Edition of the Addiction Severity Index. *Journal of Substance Abuse Treatment, 9*(3), 199-213.
- McLellan, A. T., Luborsky, L., Woody, G. E. & O'Brien, C. P. (1980). An improved evaluation instrument for substance abuse patients: The Addiction Severity Index. *Journal of Nervous and Mental Disease, 168*, 26-33.
- McMurran, M., & Hollins, C. R. (1989). The Short Alcohol Dependence Data (SADD) Questionnaire: Norms and reliability data for male young offenders. *British Journal of Addiction, 84*, 315-318.
- McPherson, T. L., & Hersch, R. K. (2000). Brief substance use screening instruments for primary care settings: A review. *Journal of Substance Abuse Treatment, 18*, 193-202.
- Meehan, J.P., Webb, M.G.T., & Unwin, A.R. (1985). The severity of alcohol dependence questionnaire in a sample of Irish problem drinkers. *British Journal of Addiction, 80*, 57-63.
- Midanik, L. (1982). The validity of self-reported alcohol consumption and alcohol problems: A literature review. *British Journal of Addiction, 77*, 357-382.
- Mieczkowski, T., Barzelay, D., Gropper, B., & Wish, E. (1991). Concordance of three measures of cocaine use in an arrestee population: Hair, urine, and self-report. *Journal of Psychoactive Drugs, 23*(3), 241-249.
- Miele, G. M., Carpenter, K. M., Cockerham, M. S., Trautman, K. D., Blaine, J., & Hasin, D. S. (2000a). Substance Dependence Severity Scale (SDSS): Reliability and validity of a clinician-administered interview for DSM-IV substance use disorders. *Drug and Alcohol Dependence, 59*(1), 63-75.
- Miele, G. M., Carpenter, K. M., Cockerham, M. S., Trautman, K. D., Blaine, J., & Hasin, D. S. (2000b). Concurrent and predictive validity of the Substance Dependence Severity Scale (SDSS). *Drug and Alcohol Dependence, 59*(1), 77-88.
- Miller, W.R. (1983). Motivational interviewing with problem drinkers. *Behavioural Psychotherapy, 11*, 147-172.
- Miller, W.R., Crawford, V.L. & Taylor (1979). Significant others as corroborative sources for problem drinkers. *Addictive Behaviors, 4*, 67-70.
- Miller, W.R. & Marlatt, A. (1984). *Manual for the Comprehensive Drinker Profile, Brief Drinker Profile and Follow up Drinker Profile*. Psychological Assessment Resources, Inc. PO Box 98, Odessa, Florida, 33556.
- Minkoff, K., Bergman, E., Beck, A. T., & Beck, R. (1973). Hopelessness, depression, and attempted suicide. *American Journal of Psychiatry, 130*, 455-459.

- Mischke, H.D. & Venneri, R.L. (1987). Reliability and validity of the MAST, Mortimer Filkins questionnaire and CAGE in DWI assessment. *Journal of Studies on Alcohol*, 48, 492-501.
- Morton, J. L., Jones, T. V., & Manganaro, M. A. (1996). Performance of alcoholism screening questionnaires in elderly veterans. *American Journal of Medicine*, 101(2), 153-159.
- Murphy, J.M., Berwick, D.M., Weinstein, M.C., Borus, J.F., Budman, S.H. & Klerman, G.L. (1987). Performance of screening and diagnostic tests: Application of receiver operating characteristic analysis. *Archives of General Psychiatry*, 44, 550-555.
- Nakai, Y., Hamagaki, S., & Takagi, R. (1998). The validity of Bulimic Investigatory Test, Edinburgh (BITE) and the Survey of Bulimia Nervosa. *Seishin Igaku Clinical Psychiatry*, 40(7), 711-716
- National Health and Medical Research Council (2000). Australian Drinking Guidelines: Consultation draft. Canberra, ACT: Author.
- Niles, B.L. & McCrady, B.S. (1991). Detection of alcohol problems in a hospital setting. *Addictive Behaviours*, 16, 223-233.
- Nilssen, O. & Cone, H. (1994). Screening patients for alcohol problems in primary health care settings. *Alcohol Health and Research World*, 18, 136-139.
- Nimeus, A., Traeskman Bendz, L., & Alsen, M. (1997). Hopelessness and suicidal behavior. *Journal of Affective Disorders*, 42(2-3), 137-144.
- Nunnally, J. C. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- O'Hare, T., & Sherrer, M. V. (1999). Validating the Alcohol Use Disorder Identification Test with college first-offenders. *Journal of Substance Abuse Treatment*, 17(1-2), 113-119.
- O'Hare, T., & Tran, T. V. (1997). Predicting problem drinking in college students: Gender differences and the CAGE questionnaire. *Addictive Behaviors*, 22(1), 13-21.
- Olincy, A., Young, D. A., & Freedman, R. (1997). Increased levels of the nicotine metabolite cotinine in schizophrenic smokers compared to other smokers. *Biological Psychiatry*, 42, 1-5.
- Osman, A., Barrios, F. X., Aukes, D., & Osman, J. R. (1993). The Beck Anxiety Inventory: Psychometric properties in a community population. *Journal of Psychopathology and Behavioral Assessment*, 15(4), 287-297.
- Osman, A., Kopper, B. A., Barrios, F. X., Osman, J. R., & Wade, T. (1997). The Beck Anxiety Inventory: Reexamination of factor structure and psychometric properties. *Journal of Clinical Psychology*, 53(1), 7-14.
- Overall, J.E. & Gorham, D.R. (1962). The Brief Psychiatric Rating Scale. *Psychological Reports*, 10, 799-812.
- Pan, P. C., & Goldberg, D. P. (1990). A comparison of the validity of GHQ-12 and CHQ-12 in Chinese primary care patients in Manchester. *Psychological Medicine*, 20(4), 931-940.
- Parsons, K. J., WallBrown, F. H., & Myers, R. W. (1994). Michigan Alcoholism Screening Test: Evidence supporting general as well as specific factors. *Educational and Psychological Measurement*, 54(2), 530-536.
- Perkins, J. L., Sanson-Fisher, R. W., Blunden, S., Lunnay, D., Redman, S., & Hensley, M. J. (1994). The Prevalence of drug use in urban Aboriginal communities. *Addiction*, 89, 1319-1331.
- Peters, R. H., Greenbaum, P. E., Steinberg, M. L., Carter, C. R., Ortiz, M. M., Fry, B. C., & Valle, S. K. (2000). Effectiveness of screening instruments in detecting substance use disorders among prisoners. *Journal of Substance Abuse Treatment*, 18(4), 349-358.
- Peters, L., Morris-Yates, A., & Andrews, G. (1994). Computerised diagnosis: The CIDI, Auto. *Computers and Mental Health*: 1. 103-107. 550-555.
- Pevalin, D. J. (2000). Multiple applications of the GHQ-12 in a general population sample: An investigation of long-term retest effects. *Social Psychiatry and Psychiatric Epidemiology*, 35(11), 508-512.
- Peveler, R.C. & Fairburn, M.G. (1990). Measurement of neurotic symptoms by self-report questionnaire: Validity of the SCL-(0)-R. *Psychological Medicine*, 20, 873-879.
- Phillips, G.T., Gossop, M.R., Edwards, G., Sutherland, G., Taylor, C. & Strang, J. (1987). The application of the SODQ to the measurement of the severity of opiate dependence in a British sample. *British Journal of Addiction*, 82, 691-699.

- Pokorny, A. D., Miller, B. A., & Kaplan, H. B. (1972). The brief MAST: A shortened version of the Michigan Alcoholism Screening Test. *American Journal of Psychiatry*, 129(3), 342-345.
- Politi, P. L., Piccinelli, M., & Wilkinson, G. (1994). Reliability, validity and factor structure of the 12-item General Health Questionnaire among young males in Italy. *Acta Psychiatrica Scandinavica*, 90(6), 432-437.
- Pols, R.G. & Hawks, D. (1992). *Is there a safe level of daily consumption of alcohol for men and women: Recommendations regarding responsible drinking behaviour*. Canberra, AGPS.
- Powell, J. E., & McInness, E. (1994). Alcohol use among older hospital patients: Findings from an Australian study. *Drug and Alcohol Review*, 13(1), 5-12.
- Prather, R.C. & Williamson, D.A. (1988). Psychopathology associated with bulimia, binge-eating, and obesity. *International Journal of Eating Disorders*, 7, 177-184.
- Pre, J., & Vassy, R. (1992). Plasma thiocyanate and cigarette-smoking status. *Medical Science Research*, 20, 671-672.
- Raistrick, D., Bradshaw, J., Tober, G., Weiner, J., Allison, J. & Healey, C. (1994). Development of the Leeds Dependence Questionnaire (LDQ): A questionnaire to measure alcohol and opiate dependence in the context of a treatment evaluation package. *Addiction*, 89, 563-572.
- Raistrick, D., Dunbar, G. & Davidson, R. (1983). Development of a questionnaire to measure alcohol dependence. *British Journal of Addiction*, 78, 89-95.
- Rajendran, R. & Cheridan, R.R. (1990). Tamil adaption of the Alcohol Dependence Scale. *Journal of the Indian Academy of Applied Psychology*, 16, 26-28.
- Rathner, G., & Rumpold, G. (1994). Convergent validity of the Eating Disorder Inventory and the Anorexia Nervosa Inventory for Self-Rating in an Austrian nonclinical population. *International Journal of Eating Disorders*, 16(4), 381-393.
- Rauter, U. K., Leonard, C. E., & Swett, C. P. (1996). SCL-90—R factor structure in an acute, involuntary, adult psychiatric inpatient sample. *Journal of Clinical Psychology*, 52(6), 625-629.
- Reiss, D. (1996). Abnormal eating attitudes and behaviours in two ethnic groups from a female British urban population. *Psychological Medicine*, 26(2), 289-299.
- Rey, J.M., Morris-Yates, A. & Staanslaw, H. (1992). Measuring the accuracy of diagnostic tests using receiver operating characteristics (ROC) analysis. *International Journal of Methods in Psychiatric Research*, 2, 1-11.
- Richter, P., Scheurer, H. & Krober, H.J. (1991). Measuring depression in criminal offenders. *Zeitschrift für Klinische Psychologie*, 39, 383-388.
- Richter, P., Werner, J., Heerlein, A., Kraus, A., & Sauer, H. (1998). On the validity of the Beck Depression Inventory: A review. *Psychopathology*, 31(3), 160-168.
- Riskind, J. H., Beck, A. T., Brown, G., & Steer, R. A. (1987). Taking the measure of anxiety and depression. Validity of the reconstructed Hamilton scales. *Journal of Nervous and Mental Disease*, 175(8), 474-479.
- Roche, A. M., & Watt, K. (1999). Drinking and university students: From celebration to inebriation. *Drug and Alcohol Review*, 18(4), 389-399.
- Rojas, N.L., Killen, J.D., Haydel, K.F., & Robinson, T.N. (1998). Nicotine dependence among adolescent smokers. *Archives of Pediatric and Adolescent Medicine*, 152, 151-156.
- Room, R. (1990). Measuring alcohol consumption in the United States: Methods and rationales. In (Ed.) L. Kozłowski. *Research Advances in Alcohol and Drug Problems, Vol 10*. New York: Plenum Press.
- Rosen, C. S., Drescher, K. D., Moos, R. H., Finney, J. W., Murphy, R. T., & Gusman, F. (2000). Six- and ten-item indexes of psychological distress based on the Symptom Checklist-90. *Assessment*, 7(2), 103-111.
- Rosen, C. S., Henson, B. R., Finney, J. W., & Moos, R. H. (2000). Consistency of self-administered and interview-based Addiction Severity Index composite scores. *Addiction*, 95(3), 149-425.
- Rosenberg, S. D., Drake, R. E., Wolfoprd, G. L., Mueser, K. T., Oxman, T. E., Vidaver, R. M., Carrieri, K. L., & Luckoor, R. (1998). Dartmouth Assessment of Lifestyle Instrument (DALI): a substance use disorder screen for people with severe mental illness. *American Journal of Psychiatry*, 155(2), 232-238.

- Ross, H.E., Gavin, D.R. & Skinner, H. (1990). Diagnostic validity of the MAST and the Alcohol Dependence Scale in the assessment of DSM-III alcohol disorders. *Journal of Studies on Alcohol*, 51, 506-513.
- Ross, H.E. & Glaser, F.B. (1989). Psychiatric screening of alcohol and drug patients: The validity of the GHQ-60. *American Journal of Drug and Alcohol Abuse*, 15, 429-442.
- Rounsaville, B. J. (1979). Detecting depressive disorders in drug abusers: A comparison of screening instruments. *Journal of Affective Disorders*, 1(4), 255-267.
- Russell, M. (1994). New assessment tools for risk drinking during pregnancy. *Alcohol Health and Research World*, 18, 55-61.
- Russell, M. & Bigler, L. (1979). Screening for alcohol-related problems in an outpatient obstetric-gynecologic clinic. *American Journal of Obstetrics and Gynaecology*, 134, 4-12.
- Russell, M., Martier, S.S., Sokol, R.J. and Mudar, P. (1994). Screening for pregnancy risk-drinking. *Alcoholism Clinical and Experimental Research*, 18, 1156-1161
- Russell, M., Martier, S. S., Sokol, R. J., Mudar, P., Jacobson, S., & Jacobson, J. (1996). Detecting risk drinking during pregnancy: a comparison of four screening questionnaires. *American Journal of Public Health*, 86(10), 1435-1439.
- Russell, M. & Skinner, J.B. (1988). Early measures of maternal alcohol misuse as predictors of adverse pregnancy outcomes. *Alcoholism: Clinical and Experimental Research*, 12, 824-830.
- Ruz, F. I., Gonzalez, S. F., & Ruiz A. F. (1998). Drug dependent patients in the methadone maintenance program: evaluation in primary care of psychosocial and organic severity. *Aten Primaria*, 21(6), 384-388.
- Sachs, H., & Kintz, P. (1998). Testing for drugs in hair. Critical review of chromatographic procedures since 1992. *J Chromatogr B Biomed Sci Appl*, 713(1), 147-161.
- Sack, W. H., Seeley, J. R., Him, C., & Clarke, G. N. (1998). Psychometric properties of the Impact of Events Scale in traumatized Cambodian refugee youth. *Personality and Individual Differences*, 25(1), 57-67.
- Saigh, P. A., Fairbank, J. A., & Yasik, A. E. (1998). War-related posttraumatic stress disorder among children and adolescents. In T. W. Miller (Ed.), *Children of trauma: Stressful life events and their effects on children and adolescents* (pp. 119-140). Madison, CT, US: International Universities Press Inc.
- Salaspuro, M. (1994). Biological state markers of alcohol abuse. *Alcohol Health and Research World*, 18(2), 131-135.
- Salaspuro, M. (1999). Carbohydrate-deficient transferrin as compared to other markers of alcoholism: a systematic review. *Alcohol*, 19(3), 261-271.
- Saltstone, R., Halliwell, S., & Hayslip, M. A. (1994). A multivariate evaluation of the Michigan Alcoholism Screening Test and the Drug Abuse Screening Test in a female offender population. *Addictive Behaviors*, 19(5), 455-462.
- Saltz, R. & Ames, G. (1996). Combining methods to identify new measures of women's drinking problems. Part II: The survey stage. *Addiction*, 91, 1041-1051.
- Saunders, J.B. (1993). Physical complications of excessive alcohol consumption. In Mattick, R.P. and Jarvis, T. (eds.) *An outline for the management of alcohol problems*. Monograph Series No 20. National Drug Strategy, Canberra, AGPS.
- Saunders, J. B., Aasland, O. G., Amundsen, A., & Grant, M. (1993). Alcohol consumption and related problems among primary health care patients: WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption: I. *Addiction*, 88(3), 349-362.
- Saunders, J. B., Aasland, O. G., Babor, T. F., De La Fuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption: II. *Addiction*, 88, 791-804.
- Schmitz, N., Kruse, J., Heckrath, C., Alberti, L., & Tress, W. (1999). Diagnosing mental disorders in primary care: the General Health Questionnaire (GHQ) and the Symptom Check List (SCL-90-R) as screening instruments. *Soc Psychiatry Psychiatr Epidemiol*, 34(7), 360-366.
- Schmitz, N., Kruse, J., & Tress, W. (1999). Psychometric properties of the General Health Questionnaire (GHQ-12) in a German primary care sample. *Acta Psychiatr Scand*, 100(6), 462-468.

- Schwarzwald, J., Solomon, Z., Weisenberg, M., & Mikulincer, M. (1987). Validation of the impact of event scale for psychological sequelae of combat. *Journal of Consulting and Clinical Psychology, 55*(2), 251-256.
- Scott, W., & Dua, J. (1999). Development of a scale to assess posttraumatic stress disorder. *International Journal of Stress Management, 6*(3), 149-165.
- Searles, J.S., Alterman, A.I. & Purtill, J. (1990). The detection of alcoholism in hospitalised schizophrenics: A comparison of the MAST and the MAC. *Alcoholism Clinical and Experimental Research, 14*, 557-560.
- Sembhi, S., & Lee, J. W. (1999). Cannabis use in psychotic patients. *Aust N Z J Psychiatry, 33*(4), 529-532.
- Seidel, R.W., Gusman, F.D. & Abueg, F.R. (1994). Theoretical and practical foundations of an inpatient post-traumatic stress disorder and alcoholism treatment program. *Psychotherapy, 31*, 67-78.
- Selzer, M.L. (1971). The Michigan Alcoholism Screening Test: The quest for a new diagnostic instrument. *American Journal of Psychiatry, 127*, 1653-1658.
- Selzer, M.L., Vinokur, A. & van Rooijen, L. (1975). A Self Administered Short Michigan Alcoholism Screening Test (SMAST). *Journal of Studies on Alcohol, 36*, 117-126.
- Skinner, H.A. and Horn, J.L. (1984). *Alcohol Dependence Scale (ADS): User's Guide*. Toronto: Addiction Research Foundation.
- Skinner, H. & Goldberg, A.E. (1986). Evidence for a drug dependence syndrome among narcotic users. *British Journal of Addiction, 81*, 479-484.
- Skinner, H. (1982). The Drug Abuse Screening Test. *Addictive Behaviors, 7*, 363-371.
- Skipsey, K., Bureson, J. A., & Kranzler, H. R. (1997). Utility of the AUDIT for identification of hazardous or harmful drinking in drug-dependent patients. *Drug and Alcohol Dependence, 45*(3), 157-163.
- Skowrow, S. & Smith, D.I. (1986). *Survey of homelessness, alcohol consumption and related problems amongst Aboriginals in the Hedland area*. Unpublished manuscript, Western Australian Alcohol and Drug Authority, Perth.
- Smart, R.G., Adlaf, E.M. & Knoke, D. (1991). Use of the CAGE scale in a population survey of drinking. *Journal of Studies on Alcohol, 52*, 593-596.
- Smith, A.R. (1986). Alcoholism and gender: Patterns of diagnosis and response. *The Journal of Drug Issues, 16*, 407-420.
- Sobell, L.C., Agrawal, S., Annis, H., Ayala-Velazquez, H., Echeverria, L., Leo, G.I., Rybakowski, J.K., Sandahl, C., Saunders, B., Thomas, S., & Ziolkowski, M. (2001). Cross-cultural evaluation of two drinking assessment instruments: Alcohol Timeline Followback and Inventory of Drinking Situations. *Substance Use and Misuse, 36*, 313-331.
- Sobell, L. C., Breslin, F. C., & Sobell, M. B. (1997). Substance-related disorders: Alcohol. In S. M. Turner (Ed.), *Adult psychopathology and diagnosis* (3rd ed., pp. 128-158). New York, NY: John Wiley & Sons Inc.
- Sobell, L. C., Brown, J., Leo, G. I., & Sobell, M. B. (1996). The reliability of the Alcohol Timeline Followback when administered by telephone and by computer. *Drug and Alcohol Dependence, 42*(1), 49-54.
- Sobell, L.C. & Sobell (1991). Reliability of alcohol abusers: self-reports of drinking behaviour. *Behaviour Research and Therapy, 17*, 157-160.
- Sobell, L.C. & Sobell, M.B. (1992). Timeline follow-back: A technique for assessing self-reported alcohol consumption. In R. Litten and J. Allen (eds). *Measuring Alcohol Consumption*. Humana Press Inc.
- Sobell, L. C., & Sobell, M. B. (1995). Alcohol consumption measures. In J. P. Allen & M. Columbus (Eds.), *Assessing alcohol problems: A guide for clinicians and researchers* (4 ed., pp. 55-76). Washington D.C.: U.S. Govt Printing Office.
- Sobell, L. C., & Sobell, M. B. (1996). Timeline Followback (TLFB) for alcohol (Version 4.0b) [computer software]. Toronto, Canada: Addiction Research Foundation.

- Sobell, L. C., & Sobell, M. B. (1998). Identification and assessment of alcohol problems. In G. P. Koocher & J. C. Norcross & S. S. Hill (Eds.), *Psychologists' Desk Reference* (pp. 62-67). New York: Oxford University Press.
- Sobell, L.C. & Sobell, M.B. (2000). Alcohol Timeline Followback (TLFB). In American Psychiatric Association (Ed.). *Handbook of psychiatric measures* (pp. 477-479). Washington, DC: American Psychiatric Association.
- Sobell, L. C., Toneatto, T., & Sobell, M. B. (1994). Behavioral assessment and treatment planning for alcohol, tobacco, and other drug problems: Current status with an emphasis on clinical applications. *Behavior Therapy*, 25(4), 533-580.
- Sokol, R.J., Martier, S.S. & Ager, J.W. (1989). The T-ACE questions: Practical prenatal detection of risk-drinking. *American Journal of Obstetrics and Gynaecology*, 160, 863-870.
- Soyka, M., Albus, M., Kathman, N. & Finelli, A. (1993). Prevalence of alcohol and drug abuse in schizophrenic inpatients. *European Archives of Psychiatry and Clinical Neuroscience*, 242, 362-372.
- Spielberger, C.D., Gorsuch, R.L., Lushene, R., Vagg, P.R. & Jacobs, G.A. (1983). *Manual for the State-Trait Anxiety Inventory (Form Y)*, Palo Alto, Consulting Psychologist Press, Inc.
- Spitzer, R.L., Williams, J.B.W. & Gibbon, M. (1986). *Structured Clinical Interview for DSM-III-R: Patient Version (SCID-P)*. New York: New York State Psychiatric Institute (Biometric Research Department).
- Sriram, T. G., Chandrashekar, C. R., Isaac, M. K., & Shanmugham, V. (1989). The General Health Questionnaire (GHQ): Comparison of the English version and a translated Indian version. *Social Psychiatry and Psychiatric Epidemiology*, 24(6), 317-320.
- Stanley, M. A., Beck, J. G., & Zebb, B. J. (1996). Psychometric properties of four anxiety measures in older adults. *Behaviour Research and Therapy*, 34(10), 827-838.
- Statistical Package for the Social Science (SPSS) [computer software]. Chicago, IL: SPSS inc.
- Stastny, D. & Potter, M. (1991). Alcohol abuse by patients undergoing methadone treatment programs. *British Journal of Addictions*, 86, 307-310.
- Steer, R. A., Ball, R., Ranieri, W. F., & Beck, A. T. (1997). Further evidence for the construct validity of the Beck depression Inventory-II with psychiatric outpatients. *Psychological Reports*, 80(2), 443-446.
- Steer, R. A., Beck, A. T., Brown, G. K., & Beck, J. S. (1993). Classification of suicidal and nonsuicidal outpatients: A cluster-analytic approach. *Journal of Clinical Psychology*, 49(5), 603-614.
- Steer, R.A., Beck, A.T., Riskind, J. & Brown, G. (1986). Differentiation of depressive disorders from generalised anxiety by the Beck Depression Inventory. *Journal of Clinical Psychology*, 40, 475-478.
- Steer, R. A., & Clark, D. A. (1997). Psychometric characteristics of the Beck Depression Inventory-II with college students. *Measurement and Evaluation in Counseling and Development*, 30(3), 128-136.
- Steer, R. A., Iguchi, M. Y., & Platt, J. J. (1994). Hopelessness in IV drug users not in treatment and seeking HIV testing and counselling. *Drug and Alcohol Dependence*, 34(2), 99-103.
- Steer, R. A., Kumar, G., Ranieri, W. F., & Beck, A. T. (1995). Use of the Beck Anxiety Inventory with adolescent psychiatric outpatients. *Psychological Reports*, 76(2), 459-465.
- Steer, R. A., Kumar, G., Ranieri, W. F., & Beck, A. T. (1998). Use of the Beck Depression Inventory-II with adolescent psychiatric outpatients. *Journal of Psychopathology and Behavioral Assessment*, 20(2), 127-137.
- Steer, R. A., Ranieri, W. F., Beck, A. T., & Clark, D. A. (1993). Further evidence for the validity of the Beck Anxiety Inventory with psychiatric outpatients. *Journal of Anxiety Disorders*, 7(3), 195-205.
- Steer, R. A., Rissmiller, D. J., Ranieri, W. F., & Beck, A. T. (1993). Structure of the computer-assisted Beck Anxiety Inventory with psychiatric inpatients. *Journal of Personality Assessment*, 60(3), 532-542.
- Steer, R.A., Rissmiller, D.J., Ranieri, W.F., & Beck, A.T. (1995). Use of the computer-administered Beck Depression Inventory and Hopelessness Scale with psychiatric patients. *Computers in Human Behavior*, 10, 223-229.

- Steer, R.A. & Schut, J. (1979). Types of psychopathology displayed by heroin addicts. *American Journal of Psychiatry*, 136, 1463-1465.
- Stellman, S. D., & Garfinkel, L. (1986). Smoking habits and tar levels in a new American Cancer Society prospective study of 1.2 million men and women. *J Natl Cancer Inst*, 76(6), 1057-1063.
- Stibler, H. (1991). Carbohydrate-deficient transferrin in serum: a new marker of potentially harmful alcohol consumption reviewed. *Clinical Chemistry*, 37, 2029-2037.
- Stockwell, T., Hodgson, R., Edwards, G., Taylor, C. & Rankin. H. (1979). The development of a questionnaire to measure severity of alcohol dependence. *British Journal of Addiction*, 74, 79-87.
- Stockwell, T., Murphy, D. & Hodgson, R. (1983). The severity of alcohol dependence questionnaire: Its use, reliability and validity. *British Journal of Addiction*, 78, 145-155.
- Stockwell, T., Sitharthan, T., McGrath, D., & Lang, E. (1994). The measurement of alcohol dependence and impaired control in community samples. *Addiction*, 89, 167-174.
- Stoffelmayr, B. E., Mavis, B. E., & Kasim, R. M. (1994). The longitudinal stability of the Addiction Severity Index. *Journal of Substance Abuse Treatment*, 11(4), 373-378.
- Strang, J., Black, J. & Marsh, A. (1993). Hair analysis for drugs: Technological breakthrough or ethical quagmire? *Addiction*, 88, 163-166.
- Sutherland, G., Edwards, G., Taylor, C., Phillips, G.T. and Gossop, M.R. (1988). The opiate dependence syndrome: replication study using the SODQ in a New York clinic, *British Journal of Addiction*, 83, 755-760.
- Sutherland, G., Edwards, G., Taylor, C., Phillips, G., Gossop, M. & Brady, R. (1986). The measurement of opiate dependence, *British Journal of Addiction*, 81, 485-494.
- Svikis, D. S., McCaul, M. E., Haug, N. A., & Boney, T. Y. (1996). Detecting alcohol problems in drug-dependent women of childbearing age. *American Journal of Drug and Alcohol Abuse*, 22(4), 563-575.
- Swan, P. R., B. (1999). *Ways Forward: National Consultancy Report on Aboriginal and Torres Strait Islander Mental Health. Part 1*. Canberra.: Australian Government Printing Service.
- Swett, C., Cohen, C., Surrey, J., Compaine, A. (1991). High rates of alcohol use and history of physical and sexual abuse among women outpatients. *American Journal of Drug and Alcohol Abuse*, 17, 49-60.
- Swift, W., Hall, W., Didcott, P., & Reilly, D. (1998). Patterns and correlates of cannabis dependence among long-term users in an Australian rural area. *Addiction*, 93(8), 1149-1160.
- Swift, W., Maher, L., & Sunjic, S. (1999). Transitions between routes of heroin administration: A study of Caucasian and Indochinese heroin users in south-western Sydney, Australia. *Addiction*, 94(1), 71-82.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston ; London: Allyn and Bacon.
- Tanaka, E., Sakamoto, S., Ono, Y., & Fujihara, S. (1996). Hopelessness in a community population in Japan. *Journal of Clinical Psychology*, 52(6), 609-615.
- Tate, J.C. & Schmidt, J.M. (1993). A proposed revision of the Fagerstrom Tolerance Questionnaire. *Addictive Behaviors*, 18, 135-143.
- Tarter, R.E. (1990). Evaluation and treatment of adolescent substance abuse: A decision tree method. *American Journal of Drug and Alcohol Abuse*, 16, 1-46.
- Teesson, M., Clement, N., Copeland, J., Conroy, A., & Reid, A. (2000). *The measurement of outcome in alcohol and other drug treatment: A review of available instruments* (Technical Report 92). Sydney, NSW: National Drug and Alcohol Research Centre.
- Teesson, M., & Gallagher, J. (1999). Evaluation of a treatment program for serious mental illness and substance use in an inner city area. *Journal of Mental Health*, 8, 19-28.
- Teitelbaum, L. M., & Carey, K. B. (2000). Temporal stability of alcohol screening measures in a psychiatric setting. *Psychology of Addictive Behaviors*, 14(4), 401-404.
- Teitelbaum, L., & Mullen, B. (2000). The validity of the MAST in psychiatric settings: A meta-analytic integration. *Journal of Studies on Alcohol*, 61(2), 254-261.
- Toland, A.M. & Moss, H.B. (1989). Identification of the alcoholic schizophrenic: Use of clinical laboratory tests and the MAST. *Journal of Studies on Alcohol*, 50, 49-53.

- Topp, L., & Mattick, R. P. (1997). Choosing a cut-off on the Severity of Dependence Scale (SDS) for amphetamine users. *Addiction*, *92*(7), 839-845.
- Topp, L., Mattick, R.P. & Lovibond, P.F. (1995). *The Nature of the Amphetamine Dependence Syndrome: Appetitive or Aversive Motivation?* NDARC Technical Report No. 30. Sydney: National Drug and Alcohol Research Centre.
- Torres, C., Hernandez, M. & Oretaga, H. (1991). Validity and reproducibility of Beck's Inventory for Depression in a cardiology hospital. *Salud Mental*, *14*, 1-6 (abstract available in English).
- Uestuen, B., Compton, W., Mager, D., Babor, T., Baiyewu, O., Chatterji, S., Cottler, L., Goegues, A., Mavreas, V., Peters, L., Pull, C., Saunders, J., Smeets, R., Stipek, M. R., Vrsti, R., Hasin, D., Room, R., Van den Brink, W., Regier, D., Blaine, J., Grant, B. F., & Sartorius, N. (1997). WHO Study on the reliability and validity of the alcohol and drug use disorder instruments: Overview of methods and results. *Drug and Alcohol Dependence*, *47*(3), 161-169.
- Ulusoy, M., Sahin, N. H., & Erkmén, H. (1998). Turkish version of the Beck Anxiety Inventory: Psychometric properties. *Journal of Cognitive Psychotherapy*, *12*(2), 153-162.
- United Nations (1998). *Guidelines for testing drugs under international control in hair, sweat and saliva*. New York: Author.
- Varner, R. V., Chen, Y. R., Swann, A. C., & Moeller, F. G. (2000). The Brief Psychiatric Rating Scale as an acute inpatient outcome measurement tool: A pilot study. *Journal of Clinical Psychiatry*, *61*(6), 418-421.
- Vicary, D. (2000). A model of intervention. In P. Dudgeon & D. Garvey & H. Pickett (Eds.), *Working with Indigenous Australians: A Handbook for Psychologists* (pp. 425-430). Perth WA: Gunada Press.
- Waage, H., Silsand, T., Urdal, P., & Langård, S. (1992). Discrimination of smoking status by thiocyanate and cotinine in serum, and carbon monoxide in expired air. *International Journal of Epidemiology*, *21*, 488-493.
- Wagenknecht, L. E., Cutter, G. R., Haley, N. J., Sidney, S., Manolio, T. A., Hughes, G. H., & Jacobs, D. R. (1990). Racial differences in serum cotinine levels among smokers in the Coronary Artery Risk Development in (Young) Adults study. *American Journal of Public Health*, *80*(9), 1053-1056.
- Waller, G. (1992). Bulimic attitudes in different eating disorders: Clinical utility of the BITE. *International Journal of Eating Disorders*, *11*, 73-78.
- Wanberg, K. W., Horn, J. L., & Foster, F. M. (1977). A differential assessment model for alcoholism: The scales of the Alcohol Use Inventory. *Journal of Studies on Alcohol*, *38*(3), 512-543.
- Warner, R., Taylor, D. Wright, J. & Sloat, A. (1994). Substance use among the mentally ill: Prevalence, reasons for use and effects on illness. *American Journal of Orthopsychiatry*, *64*, 30-39.
- Waterson, E.J. & Murray-Lyon, I.M. (1988). Are the CAGE questions outdated? (letter). *British Journal of Addiction*, *83*, 1113-1115.
- Watkins, J. P., Eisele, G. R., & Matthews, K. O. (2000). Occupational medical program alcohol screening. Utility of the CAGE and BMAST. *Journal of Substance Abuse Treatment*, *19*(1), 51-57
- Watson, C. G., Detra, E., Fox, K. L., Ewing, J. W., Gearhart, L. P., & DeMotts, J. R. (1995). Comparative concurrent validities of five alcoholism measures in a psychiatric hospital. *Journal of Clinical Psychology*, *51*(5), 676-684.
- Weiss, R.D., Griffin, M.L. & Mirin, S.M. (1989). Diagnosing major depression in cocaine abusers: The use of depression rating scales. *Psychiatry Research*, *28*, 335-343.
- Welch, G., Thompson, L., & Hall, A. (1993). The BULIT—R: Its reliability and clinical validity as a screening tool for DSM-III—R bulimia nervosa in a female tertiary education population. *International Journal of Eating Disorders*, *14*, 95-105
- Werneke, U., Goldberg, D. P., Yalcin, I., & Uestuen, B. T. (2000). The stability of the factor structure of the General Health Questionnaire. *Psychological Medicine*, *30*(4), 823-829.
- Westermeyer, J., Tucker, P. & Nugent, S. (1995). Comorbid anxiety disorder among patients with substance abuse disorders. *American Journal on Addictions*, *4*, 97-106.
- Westaway, M.S. & Wolmarans, L. (1992). Depression and self esteem: Rapid screening for depression in Black, low literacy, hospitalised tuberculosis patients. *Social Science Medicine*, *35*, 1311-1315.

- Wetherell, J. L., & Arean, P. A. (1997). Psychometric evaluation of the Beck Anxiety Inventory with older medical patients. *Psychological Assessment, 9*(2), 136-144.
- Wetzel, R. D. (1977). Factor structure of Beck's suicide intent scales. *Psychological Reports, 40*(1), 295-302.
- White, H.R. & LaBouvie, E.W. (1989). Towards the assessment of adolescent problem drinking. *Journal of studies of Alcohol, 50*, 30-37.
- Whisman, M. A., Perez, J. E., & Ramel, W. (2000). Factor structure of the Beck Depression Inventory—Second Edition (BDI-II) in a student sample. *Journal of Clinical Psychology, 56*(4), 545-551.
- Williams, R. J., Connor, J. P., & Ricciardelli, L. A. (1998). Self-efficacy for refusal mediated by outcome expectancies in the prediction of alcohol-dependence amongst young adults. *Journal of Drug Education, 28*(4), 347-359.
- Williamson, D.A., Anderson, D.A., Jackman, L.P. & Jackson, S.R. (1995). Assessment of eating disordered thoughts, feelings, and behaviours. In D.B. Allison (Ed.). *Handbook of assessment methods for eating behaviours and weight-related problems: Methods, theory and research.* (pp. 347-386). Thousand Oaks, CA: Sage.
- Williamson, D.A., Prather, R.C., Bennett, S.M., Davis, C.J., Watkins, P.C., & Grenier, C.E. (1989). An uncontrolled evaluation of inpatient and outpatient cognitive behavior therapy for bulimia nervosa. *Behavior Modification, 13*, 340-360.
- Wilson, G.T. (1993). Binge eating and addictive disorders. In C.G Fairburn. & G. T. Wilson (Eds.). *Binge eating: nature, assessment and treatment.* (pp. 97-122). New York: Guilford Press.
- Wilson, G.T., & Smith, D. (1989). Assessment of Bulimia Nervosa: An evaluation of the Eating Disorders Examination. *International Journal of Eating Disorders, 8*, 173-179.
- Winefield, H. R., Goldney, R. D., Winefield, A. H., & Tiggemann, M. (1989). The General Health Questionnaire: Reliability and validity for Australian youth. *Australian and New Zealand Journal of Psychiatry, 23*(1), 53-58.
- Winters, K. (1994). *Assessment of Adolescent Alcohol and Drug Abuse: A Handbook.* Western Psychological Services, Los Angeles, California, 90025-1251.
- Wish, E. D., Gray, T., Sushinsky, J., Yacoubian, G. S., Jr., & Fitzgerald, N. (2000). An experiment to enhance the reporting of drug use by arrestees. *Journal of Drug Issues, 30*(1), 55-76.
- Wittchen, H.U. (1994). Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI): A critical review. *Journal of Psychiatric Research, 28* 57-84.
- Wolford, G. L., Rosenberg, S. D., Drake, R. E., Mueser, K. T., Oxman, T. E., Hoffman, D., Vidaver, R. M., Luckoor, R., & Carrieri, K. L. (1999). Evaluation of methods for detecting substance use disorder in persons with severe mental illness. *Psychology of Addictive Behaviors, 13*, 313-326.
- World Health Organisation (1992). *ICD-10, Chapter V, Mental and Behavioural Disorders (including disorders of Psychological development) - Diagnostic Criteria for Research.* Geneva: World Health Organisation.
- World Health Organisation (1992). *Schedules for Clinical Assessment in Neuropsychiatry (SCAN).* Geneva: World Health Organisation.
- World Health Organisation (1993). *The Composite International Diagnostic Interview - Version 1.1.* Geneva: World Health Organisation.
- Worsley, A., & Gribbin, C.C. (1977). A factor analysis study of the twelve item GHQ. *Australian and New Zealand Journal of Psychiatry, 11*, 269-273.
- Wright, M. (2000). An alternative model of practice: Aboriginal community based Inreach service for Aboriginal people with serious mental health issues. In P. Dudgeon & D. Garvey & H. Pickett (Eds.), *Working with Indigenous Australians: A Handbook for Psychologists* (pp. 439-445). Perth WA: Gunada Press.
- Xu, J (1991). Clinical applications and evaluation of the Beck Depression Inventory. *Psychological Science China, 4*, 32-34 (abstract available in English).
- Yacoubian, G. S., Jr. (2000). Reassessing the need for urinalysis as a validation technique. *Journal of Drug Issues, 30*(2), 323-334.

- Yule, W., Ten Bruggencate, S., & Joseph, S. A. (1994). Principal components analysis of the Impact of Events Scale in adolescents who survived a shipping disaster. *Personality and Individual Differences, 16*(5), 685-691
- Zanis, D. A., McLellan, A. T., & Corse, S. (1997). Is the Addiction Severity Index a reliable and valid assessment instrument among clients with severe and persistent mental illness and substance abuse disorders? *Community Mental Health Journal, 33*(3), 213-227.
- Zilberg, N. J., Weiss, D. S., & Horowitz, M. J. (1982). Impact of Event Scale: A cross-validation study and some empirical evidence supporting a conceptual model of stress response syndromes. *Journal of Consulting and Clinical Psychology, 50*(3), 407-414.
- Zung, B.J. (1982). Screening alcohol problems in psychiatric inpatients. *Psychiatric Forum, 11*(1), 32-37.