

## 1.1 - Contemporary Changes in Psychopathology

In the last three decades Psychiatry has undergone a major change in its approach to psychopathology. Psychopathological classifications used to reflect belief systems that were based on impressionistic clinical similarities. Most were not grounded in quantifiable data, used unrepresentative populations, and were devoid of a coherent theory (Millon, 1987,1991). The validity of the classifications rested on the personal authority of their originators, which often was limited to a single country (Kendler, 1990). During the sixties there was a growing awareness among clinicians and researchers that the absence of an objective and reliable system for describing psychopathology and for making psychiatric diagnoses was limiting scientific progress (Klerman, 1986). The development of quantitative techniques for measuring psychopathology (i.e. psychometrics) and the refining of standardised diagnostic criteria have led to a transformation of this science.

Psychopathology no longer depends on the intuitive artistry of brilliant clinicians and theoreticians who formulated dazzling but unfalsifiable insights. Psychopathology has acquired a solid footing in the empirical methodologies and quantitative techniques used in psychology (Blashfield, 1986; Millon, 1987). A major innovation was the use of diagnostic criteria that were intended to provide operational definitions of the psychiatric diagnosis. Many instruments were developed in response to the need for a better descriptive diagnosis. These new instruments drew upon existing psychometric methodologies, particularly those for educational testing, with multivariate statistics methods (Blashfield, 1986). The term psychopathology was synonymous with descriptive symptomatology; now, according to Millon (1991), can be justly employed to represent 'the science of abnormal behaviour and mental disorder'.

When investigators evaluate how well a new classification system performs, they rely on the concepts of reliability and validity, concepts that are also used to transform the empirical data collected on psychological processes into psychometric tests. The use of these concepts in this context suggests that a classification system is developed in a similar way to a psychological test. Blashfield and Livesley (1991) proposed that the psychiatric classification resembles a psychological test with both structural and analytical parallels. Table 1.1 summarizes these similarities. As can be

seen the general goal of a psychiatric classification system and a psychological test is the same - the measurement of a person by the use of scales and criteria that are designed to represent a construct. According to Blashfield and Livesley (1991) the profitability of this analogy is reflected in the profusion of psychometric literature on measurement issues.

Table 1.1 - Similarities between a Psychological Test and a Psychiatric Classification

|                   | <b>Psychological Test</b>                                                                                                                                                                                                                                           | <b>Psychiatric Classification</b>                                                                                                                                                                  |
|-------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Structure</b>  | <p>-a test is composed of one or more scales that are intended to measure underlying constructs.</p> <p>-each scale contains a number of scorable items, which are the fundamental measurement units of the test.</p>                                               | <p>-diagnostic categories are structural equivalents to scales.</p> <p>-diagnostic criteria are like test items; the basic unit of measurement that are combined to form a diagnostic category</p> |
| <b>Analytical</b> | <p>Reliability assess the degree of measurement error in both systems.</p> <p>Validity in both systems must have some correlational or causal relations with other variables if the scales or categories are to be useful in improving scientific understanding</p> |                                                                                                                                                                                                    |

## 1.2 - Different Traditions of Measurement in Psychiatry and Psychology

Although there are similarities between a psychiatric diagnosis and a psychological test the models of understanding and measurement of psychopathology in psychiatry and clinical psychology have evolved from different traditions. In psychiatry the unstructured clinical interview is the standard method of assessing the patient's presenting problems, history, and mental state. An important goal is to collect information from which a diagnosis can be made. Classification systems specify the type of information that must be elicited during the interview. In clinical psychology the traditional approach to measurement has been the use of psychological tests. Although many items used in psychological tests have been derived from unstructured interviews, tests differ from interviews in that the items are constant for each patient, quantitative scores are generated, and test norms can be developed to compare individual scores (Blashfield and Livesley, 1991).

However, the general goal of psychiatric classification and psychological testing is the same - the measurement of persons by constructs that scales or categories are designed to represent.

Traditionally psychiatry has adopted the categorical model of assessment whereas psychology has opted for a dimensional approach. The categorical view of psychopathology is often linked to an etiological model in which the categories are assumed to represent underlying disease processes that have caused the psychological manifestations (Kendell, 1975; Millon, 1987, 1991). It assumes that the various mental disorders are qualitatively different (McReynolds, 1989). The dimensional model on the other hand, is associated with a continuum view of psychopathology that assumes there is no clear boundary between normal and abnormal (McReynolds, 1989). The debate between dimensional and categorical models has an extensive literature (Kendell, 1975; Frances, 1982; Millon, 1987; Carson, 1991; Widiger and Trull, 1991; Millon, 1991), as well as an often contentious history. Indeed, authors such as Eysenck (1983, 1986) have proposed the abolition of categorical diagnoses altogether.

There are several advantages to the dimensional model (Frances, 1982; Millon, 1991; Widiger and Trull, 1991): 1 - it is associated with a continuum view of psychopathology that suggests there is no clear boundary between the normal and the abnormal; 2 - it combines several clinical attributes in a single configuration; 3 - it reduces the halo effect; 4 - single attributes are not given special significance; 5 - dimensional judgements are more reliable than categorical ones (Heumann and Morey 1990); 6 - it has more flexibility, having different cut-off points for different purposes; 7 - superiority in evaluating change; 8 - it is less restrictive than the categorical model because it is associated with a higher level of data scaling. Despite these advantages, dimensional diagnoses have not fared well in everyday clinical practice. Numerous complications and limitations have been noted in the literature, in particular the lack of agreement among theorists concerning the number of dimensions necessary to represent a psychopathological phenomenon. Moreover a dimensional diagnosis is a less tangible concept than a categorical one and its meaning more difficult to communicate in a clinical setting.

The advantages of the categorical model are also many (Millon, 1991; Widiger and Trull, 1991): 1 - its ease of use for clinicians; 2 - a categorical diagnosis restores unity by integrating the seemingly diverse elements of a patient's psychopathology into a single configuration; 3 - as a concept a categorical diagnosis is easier to communicate than a dimensional one; 4 - it is easier to remember and to report; 5 - like all class concepts it provides directions for practical behaviour. However a major disadvantage is that categorical diagnoses contribute to the fallacious belief that collections of psychopathological processes comprise discrete entities or even diseases, when in fact, they are

merely concepts that help to focus and coordinate our observations (Millon, 1991). Furthermore the process of categorization can distort the data, exaggerating both the homogeneity within groups and the heterogeneity between them (Millon, 1991).

Categorical and dimensional models need not be framed in opposition, or be considered mutually exclusive. Assessments can be formulated that include features of both. Qualitative (i.e. categorical) distinctions can be used to assess those features which best characterize a patient's condition, whilst lists of individual features and measurements of their severity are quantitative (i.e. dimensional) techniques. Dimensions can also be transformed into categories by using different cut-off points. Skinner (1986) elaborated several hybrid models that integrate elements of normally divergent schema. In what he termed the class-quantitative approach, efforts were made to synthesize quantitative dimensions and discrete categories. But which model is more useful or more appropriate? The answer will vary depending on the objective one has in mind. It may also depend on the stage of development that the subject material has reached (Kendell, 1975). As Hempel (1961) pointed out, most sciences start with a typology and dichotomous present/absent distinctions, but later these are replaced by dimensions as more accurate measurement becomes available.

### **1.3 - Two Traditions of Validation in Psychiatry and Psychology**

As evidence was mounting during the 1960s and 1970s that the reliability of psychiatric diagnoses was unacceptably low, much of the emphasis on research in psychiatry during this period promoted the reliability of the psychiatric diagnoses (Blashfield, 1986). There was a belief that if a psychiatric classification was not reliable, the development of a science about mental disorders was not possible. Spitzer and Fleiss (1974) believed that reliability was a constraint on the validity of the psychiatric classification and that high reliability was a necessary prerequisite to attempts at improving the validity. This led to the development and implementation of operational criteria which significantly improved reliability. Recently, many authors have criticised what they called an excessive concern with reliability which leads to increased interdiagnostician agreement but has not led to an enhancement in the rate of scientific progress in the fundamental issues of psychopathology (Grayson et al, 1990; Carson, 1991). Consensus between expert psychiatrists has been used as the criterion to validate the concept of a case; yet validity should be established by evidence, not by agreement (Eysenck, 1986). These criticisms have contributed to the demand that

all concepts constituting a nosology should be empirically anchored. Spitzer and Williams (1989) proposed that 'the way to improve classification is on a broader empirical basis'.

The ideas of validity used in the psychiatric literature have developed around two main topics. On the one side the introduction of structured interviews made it necessary to validate both the diagnostic instruments and the underlying constructs. A valid diagnostic instrument is one that accurately measures the condition it is designed to measure; Spitzer (1983) called this procedural validity. The other more substantive topic of validity has developed according to several different theoretical assumptions about the nature of the psychiatric diagnosis. One important aspect of these influences was in considering the psychiatric diagnosis as a category that has a common etiology and outcome. In an influential paper Robins and Guze (1970) proposed a method for achieving diagnostic validity with five phases: clinical description, laboratory study, exclusion of other disorders, follow-up study, and family study. The goal was to develop a set of mutually exclusive, discrete syndromes with established empirical validity (Widiger and Trull, 1991). This approach sees a clinical syndrome as consisting of two elements: a group of correlated symptoms and a more or less distinctive natural history. This paper is considered by many as a landmark in the history of psychiatry (Cloninger, 1989). Its importance to the conceptualization and validation of diagnosis has been compared to the impact of Cronbach and Meehl (1955) on assessment (Widiger and Trull, 1991).

The influence of this paper on the ideas about validation of the psychiatric diagnosis can still be found in the views expressed more recently by several authors which emphasise the aspects of the validity of the delimitation of one disorder from another, the elucidation of the etiology and specific outcomes. Feighner and Herbststein (1987) state that validity refers to the accuracy with which diagnostic criteria define and differentiate a disease from other diseases. Robins and Guze (1970) considered 'delimitations from other disorders' as an important aspect in the validation process because this method could establish the unity of a categorial model of psychopathology. Kendell (1989) in a review of the validation of the psychiatric diagnosis proposes that the most effective way of establishing the validity of a clinical syndrome is to elucidate its etiology. He suggests four kinds of clinical research to help clarify validity issues: prospective follow-up studies, therapeutic trials, family studies and twin studies. The emphasis is on the importance of the outcome and therefore on predictive validity, where 'in the context of clinical psychiatry, statements about diagnostic validity are essentially statements about predictive power, and hence

about practical utility'.

As the aim of this view of validation was to achieve well-established categories of diagnosis several authors proposed that an implicit goal of the process of validation of a diagnosis was to create definitions with high content validity (Blashfield, 1986). Applied to psychiatric diagnosis, content validity refers to the extent that the criteria of a disorder represent the domain of the symptoms associated with that disorder.

Blashfield and Livesley (1991) argued that content validity is the most fundamental type of validity when applied to psychiatric classification. It can be viewed as a method of assessing how well a psychiatric classification represents the consensus of opinion among mental health professionals. These ideas and strategies of validation have usually been associated with a categorial system of classification, which assumes that psychopathology comprises discrete disorders (Cloninger, 1989). One important limitation of this approach to validation issues is that by eschewing any explicit theoretical model of psychiatric disorders, it ignores the hypothetical-deductive method of construct validation (Cloninger, 1989).

In the psychological literature the importance of the concept of validity started earlier than in psychiatry and has changed substantially over the years (Anastasi, 1986; Angoff, 1988; Messick, 1989). An early focus during the 1940s was on defined validity in purely operational terms, where the prediction of specific criteria were the most important, 'in a very general sense, a test is valid for anything with which it correlates' (Messick, 1989). In 1954 there was a major effort to introduce some kind of order into the chaotic state of test construction procedures by the American Psychological Association (Messick, 1989). There was a recommendation that validity be broken into four types: content, predictive, concurrent, and construct validity. In 1966 the APA reduced these to three types by amalgamation of the predictive and concurrent into criterion-related validities. Both editions of the APA Technical Recommendations drew a link between the type of validity to be used and the aims of the particular test. The 1974 edition re-emphasized this link, but also saw the different types of validity as 'interdependent kinds of inferential interpretation' and 'aspects of validity...inter-related operationally and logically' (Messick, 1989).

These three types of validity were widely accepted until the late 1970s when several authors started to criticise the tendency of the studies to use only one category of evidence as sufficient for the validity of a particular test (Angoff, 1988; Messick, 1989). At the same time the work of Anastasi

(1986) and Cronbach (1988, 1990) moved towards the recognition of validity as a unitary concept where an ideal validation includes several types of evidence, which span all three of the traditional categories, obtaining the combinations of evidence that optimally reflect the value of a test. Cronbach (1988) summarized this view when he said that 'all validation is one'. The 1985 Standards and Recommendations stated that validity was a unitary concept and referred to the 'appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores'. The APA text no longer refers to types of validity but to categories of validity evidence: content, criterion, and construct.

The logic of this unitary view of validity in psychology is that because content and criterion validity contribute to the score meaning, they have come to be recognized as aspects of construct validity. In a sense this leaves only one category of validity: construct. Construct validation requires many lines of evidence, such that as a process it can never be completed (Cronbach, 1990). Validation of the construct framework is a form of hypothesis testing and uses all the philosophical and empirical means by which scientific theories are evaluated including statistical, experimental, and rational methods of marshalling evidence. Validation requires a hypothetical-deductive approach; it is a scientific enquiry (Messick, 1989). Validity is an overall evaluative judgment, founded on empirical evidence and theoretical rationales (Messick, 1989).

The unitary view of validity maintains that it is necessary to build the precursors of validity into the system from the outset. Construct validation has focused attention on the role of psychological theory in test construction and on the need to formulate hypotheses that can be proved or disproved in the validation process. Psychometricians have agreed that validating a psychological test cannot be viewed as an enterprise separate from exploring the validity of the psychological theory associated with the test. Attempts to create atheoretical tests have proved fruitless. Central to the process of construct validation is a sound construct theory. This theory should specify the internal structure of the construct, how it manifests itself in other indicators, and how it relates to other variables. Cronbach and Meehl (1955) maintained that when we examine a psychological trait or construct, presumed to be measured by a test, we bring about an interaction between the scores obtained on the test and the theory underlying the construct. In this way the theoretical conception of the construct dictates the nature of the data that are collected both to validate the scores and to interpret the results. In turn, the data resulting from the test administration are used to validate, reject or revise the theory itself. Viewed this way, we see that all data flowing from the theory, including concurrent and predictive data, are useful for construct validity.

Messick (1989) reviewed the state of the art of the ideas of validity in psychology. The dominant idea is that validity of a concept is an evolving property and validation is a continuing process. Because evidence is always incomplete, validation is essentially a matter of making the most reasonable base from which to guide the current use of the concept and research into advancing its understanding. Indeed, there is an appeal to multiple perspectives with respect to validity theory and validation methods. It is not the type of validity but the relation between the evidence and the inferences drawn that should determine the validation focus. The major concern of validity is not to explain any single isolated event, behaviour, or item response, because they almost certainly reflect a confounding of multiple determinants. Rather, the intent is to account for consistency in behaviour, which frequently reflects distinguishable determinants. To validate a concept is to ascertain the degree to which multiple lines of evidence are consonant with the inference, while establishing that alternative inferences are less well supported. An important aspect of the unified view of validity is the shift from prediction to explanation as the fundamental validity focus. The idea being that utility and relevance of the prediction cannot be appraised in the absence of sound empirically grounded interpretation of the concept being measured (Messick, 1989).

Although validation in psychiatry and psychology has evolved from different traditions and is at different stages of development there is a tendency towards a common ground. For example, the idea of different categories of validity has been absorbed in relation to the validation of the psychiatric diagnosis (Spitzer and Williams, 1985; Blashfield, 1986; Nelson-Gray, 1991). However, despite the fact that it has become the norm to look at psychiatric diagnosis considering three classical 'types' of validity and in particular content validity, there seems to be a different emphasis rather than a different view of validity between these two traditions. These differences are related to the different strategies of the focus of the validation process. On the psychiatric side there is, as discussed before, a tendency of focus on etiology and outcome (Kendell, 1989). The unified view of validation in psychology, on the other hand, has its emphasis on the explanatory power of the construct validation and takes a strong stand against the notion of validity types: 'the implication that validities come in different types leads to confusion and, in the face of confusion, oversimplification' (Pedhazur and Schmelkin, 1992).

#### **1.4 - Influence of Psychiatry and Psychology on the Alcohol Dependence Syndrome (ADS) and**



## its validation

The alcohol dependence syndrome, as described by Edwards and Gross (1976), follows traditions from both psychiatry and psychology. Recurring sets of symptoms are seen in a number of patients which have a medical component represented by the clinical syndrome. The notion that the presence of covarying symptoms might signify 'disease entities' can be traced to the seventeenth century writings of Thomas Sydenham (Blashfield, 1984; Millon, 1987). He argued that a careful observer of patients would note that certain sets of symptoms tended to co-occur. If these co-occurring sets of symptoms were seen in a large enough number of patients, this observation would suggest that the syndrome may represent more than a chance collection of symptoms. These ideas have been assimilated in the psychiatric literature. Millon (1987) summarised the characteristics of a syndrome in psychiatry as the clustering of a set of signs or symptoms that frequently co-occur and covary. Not all the signs and symptoms are likely to be immediately observed but the presence of a subset suggests that other features may be uncovered on closer examination. Not only is there a waxing and waning in the salience of its components, but only a few of its typical indices are likely to be manifest at any one time. Kendell (1989) similarly described a clinical syndrome as consisting of a cluster of related symptoms which in a psychiatric context may be abnormal behaviours, abnormal or distressing subjective experiences or a mixture of the two. For him a psychiatric syndrome has also a characteristic time course.

When Edwards (1977) described the Alcohol Dependence Syndrome he specifically stated the meaning of the word syndrome as an observable coincidence of phenomena and that not all of them need always be present. Moreover, the early description of the ADS (Edwards and Gross, 1976; Edwards, 1977) specify that there was logic and evidence for believing that there was some altered behavioral, subjective and psychobiological state represented by the elements of the ADS clustered together (Table 1.2). That is, the elements not only covary frequently, but make sense as a coherently organized and reasonably distinctive group of characteristics. The ADS was proposed as an idea to be tested, where its diverse phenomena must be subject to formal identification, differentiation and quantification procedures. Justification for the syndrome was mainly pragmatic and communicative and it was hoped that future research would give a more detailed understanding of the 'latent' processes at work that produced the covariance of signs, symptoms, course, prognosis, and response to treatment. For Edwards (1977): 'The obvious challenge however is to get beyond the stage of observation to that of explanation.'

Table 1.2 - Elements of the Alcohol Dependence Syndrome

|                                             |
|---------------------------------------------|
| Narrowing of drinking repertoire            |
| Saliency of drink-seeking behaviour         |
| Increase tolerance to alcohol               |
| Repeated withdrawal symptoms                |
| Relief-avoidance of withdrawal              |
| Subjective awareness of compulsion to drink |
| Reinstatement after abstinence              |

The influence of psychology can be found in several levels of the ADS. One important aspect is that the syndrome was proposed as a dimensional diagnosis, varying in quantitative severity from person to person and in time within the same person. The more explicit influence from psychology came from learning theory where the ADS was proposed as a synthesis of both general learning theory and specific conditioning models of dependence (Edwards, 1986; Babor et al, 1987a). The ADS has also benefited from the methodological principles of psychology, in particular in its measurement and at the experimental level which attempts to define the construct in behavioural terms using well-controlled experimental situations (Hodgson and Stockwell, 1985).

One of the most important aspects of the ADS was the theoretical accuracy which characterized the description of its elements. It was not only a group of symptoms but a description of a series of elements that were theoretically bound. Unlike previous models of alcoholism that had observational elements but no theoretical input the elements of the ADS have a precise theoretical meaning. One example is the description of drinking behaviour known as 'narrowing of drinking repertoire'. The emphasis here is not only on the description of the quantity of alcohol consumed, but how drinking behaviour related to the severity of dependence and the need to pace the drinks at regular and predictable intervals throughout the day so as to avoid withdrawal symptoms.

Discussing theoretical accuracy Hempel (1965) wrote that the characteristic that distinguishes a latent scientific classification is its success in grouping its elements according to theoretically consonant propositions. He also argued that in the course of scientific development a classification system defined by reference to manifest, observable characteristics will tend to give way to systems based on theoretical concepts:

"the development of a scientific discipline may often be said to proceed from an initial

`natural history' stage...to subsequent more and more `theoretical' stages... The vocabulary required in the early stages of this development will be largely observational...The shift toward theoretical systematization is marked by the introduction of new, `theoretical' systematization is marked by the introduction of new, `theoretical' terms...more or less removed from the level of directly observable things and events... These terms have a distinct meaning and function only in the context of a corresponding theory. (pp. 139-140)"

The ADS was proposed as a conceptual model, which attempts to set forth the relationships among factors that influence excessive drinking. This has to be distinguished from its later use in the form of definitions used in the internationally used classification systems ICD-10 (WHO, 1992), DSM-III-R (APA, 1987) and DSM-IV (APA, 1993) which are essentially operational definitions (Jaffe, 1992). The purpose of the model was not only identification but mainly conceptualization. Bridgman (1927) pointed out that the meaning of an operationally defined concept becomes synonymous with how we measure it, not with what we say about it. Burdock (1961) referred to the process of conceptual definition as involving a process of abstraction (in Zubin, 1961):

"There are two kinds of abstractions: abstractions from actuality, and abstraction from possibility. When we refer to something as highly abstract, it is usually the latter that we mean. Abstraction from actuality is what we mean by an operational definition; whereas abstraction from possibility describes inference of necessary properties from a model."

There has been a substantial body of evidence supporting the concept of the ADS, summarised by Edwards (1986). The strategies that have been used in the validation process of the ADS have relied on an array of methods and procedures that reflects the influences of different views of the syndrome and its validation. There are many ways in which the analyses of this evidence can clarify the stage of the validation. The difficulty that arises is that there is not a single research program testing the ADS and its validation but a diversity of ideas and influences that represent the heterogeneity of the professionals involved. The positive side is that this diversity of research reveals that the ADS construct has been accepted by a large number of experts and as a consequence it is open to a rather different interpretation of the original version and to a different emphasis on its validation. The next sections will discuss the evidence of the ADS using the general framework of an unified view of the validation process. Each section will group the evidences according to similar analysis and methodology used by several studies which have relied for their design on the ADS construct. At the end of the chapter one particular aspects of the validation process that will be examined further in the thesis will be proposed.

## 1.5 - Content Validity and Internal Consistency Analysis of the ADS

In the unified view of validity, content validity is inseparable from the validation of the construct itself (Messick, 1989; Pedhazur and Schmelkin, 1992). From the perspective of classical test theory the items in a test are assumed to be a sample of all possible items that could be used to measure the construct under investigation. Thus, content validity can be defined in terms of the sampling adequacy of the items representing the construct. The more representative of the target behaviour a sample of items is, the more the instrument lends itself to inferential analysis. Two approaches have been used to study the content validity of the ADS. One strategy has been the creation of several questionnaires specifically designed to measure the syndrome; the second has been the adoption of the operational definition of the ADS used by the main psychiatric classification systems (DSM-III-R, DSM-IV and ICD-10) to make standardized diagnoses.

The studies shown in Table 1.3 used several similar operational definition of the ADS and provide strong evidence that it has what, in psychometric terms, is called domain clarity. Domain clarity is said to occur when different researchers working independently choose items to measure a construct, from the same general pool of domain specifications, which when tested produce comparable results (Messick, 1989). There seems to have been an agreement about a core group of behaviours that represents the ADS. However, going a little further in this examination of the construct, Messick (1989) argues that there are many ways in which the measurement of a construct or a test may be imprecise. Measurements are not only imprecise by virtue of random errors but also because the items chosen to make up the construct can never be a perfect sample. Items that should have been included may be omitted ('construct underrepresentation') or included when they should have been left out ('construct irrelevant test variance'), or both. When compared with the original description of Edwards and Gross (1976) the studies in Table 1.3 have used groups of items with different levels of 'construct underrepresentation' and 'construct irrelevant test variance'. For example on the side of underrepresentation, the SADQ (Stockwell et al, 1979) has only items related to alcohol withdrawal, relief drinking, level of consumption and reinstatement. On the side of irrelevancy to the original version of the ADS the SADD (Davidson et al, 1989) has an item representing black-outs that was not in the original version.

Table 1.3 - Studies with clinical population measuring the factor structure of the Alcohol Dependence Syndrome (ADS)

| <b>Instrument /Author</b>                       | <b>Item Content</b>                                                                                                                                                                                                   | <b>Analysis</b>              | <b>Variance</b> |
|-------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------|-----------------|
| <b>SADQ/<br/>Stockwell et al<br/>(1979)</b>     | Physical withdrawal symptoms; affective symptoms of withdrawal; relief drinking; level of alcohol consumption and rapidity of reinstatement after withdrawal                                                          | Factor Analysis              | 53%             |
| <b>EADS/<br/>Chick (1980a)</b>                  | Withdrawal symptoms; subjective need; aspects of salience; relief drinking; awareness of compulsion to drink; increased tolerance                                                                                     | Principal Component Analysis | 24.6%           |
| <b>Rand Report/<br/>Polich et al<br/>(1981)</b> | Tremors; morning drinking; loss of control; black-outs; missing meals; continuous drinking                                                                                                                            | Principal Component Analysis | 52%             |
| <b>ADS/ Skinner (1-981)</b>                     | Loss of behavioural control; psychoperceptual and psychophysical withdrawal symptoms; obsessive compulsive drinking style                                                                                             | Factor Analysis              | 28%             |
| <b>LSMDQ/<br/>Hesselbrook et al<br/>(1983)</b>  | Salience of drink-seeking behaviour; increase tolerance to alcohol; repeated withdrawal symptoms; relief-avoidance of withdrawal; compulsion to drink                                                                 | Factor Analysis              | 23%             |
| <b>DSM-III-R/<br/>Kosten et al<br/>(1987)</b>   | Relief use; withdrawal; preoccupation; rapid restart; continue despite problems; give up non-alcohol activities; impaired in daily activity; tolerance; inability to cut down use; use more than intended             | Factor Analysis              | 56%             |
| <b>SADD/<br/>Davidson et al<br/>(1989)</b>      | Salience of drink-seeking behaviour; narrowing of drinking repertoire; relief drinking; continue use despite problems; inability to stop; withdrawal symptoms; black-outs; subjective awareness of a need for alcohol | Factor Analysis              | 44%             |

A systematic way to assess the content related evidence of a construct it by means of Internal Consistency Analysis (George et al, 1989; Messick, 1989). It refers to the degree of

intercorrelation between a set of indicators intended to measure the same phenomenon. The analysis tells us if the relationship among indicators is sufficiently strong to suggest that they are tapping the same underlying phenomenon. In terms of psychopathology, internal consistency is the degree to which symptoms cluster to form empirically, theoretically, and clinically meaningful syndromes (George et al, 1989). This is relevant because the degree of homogeneity in a construct should be commensurate with the degree of homogeneity theoretically expected. The logic of internal consistency analysis is that a set of indicators proposed to measure one diagnostic construct are examined in terms of dimensionality, distinctiveness, and/or homogeneity. Not every relevant indicator must be included in the measurement tool, rather the total universe of indicators must be representatively sampled.

A wide variety of correlation analyses are used to measure internal consistency (Messick, 1989); the two leading methods are Factor Analysis/Principal Component Analysis and Internal Consistency Reliability (Cronbach's alpha). Traditionally, internal consistency was treated as one form of reliability assessment, with reliability referring to the degree to which measurement is free from random error. More recently, however, psychometricians are increasingly viewing reliability and validity in a less rigid way. Reliability and validity are now seen as inseparable. Because internal consistency analysis is based on issues of sampling content homogeneity, and the dimensional structure of sets of indicators, it is relevant to validity assessment and hypothesis testing, as well as to estimating reliability (George et al, 1989). The next three sections describes the evidence of internal consistency of the ADS in different populations.

### **1.5A - Clinical Population Studies**

The study of the internal consistency of the ADS in clinical population is the area of the validation process that has been most extensively studied. The reason for this was that the clinical identification of the syndrome and its delimitation from other disorders was an important theoretical issue just after the description of the ADS in 1976 and remains important today. From the early operational description by Stockwell et al (1979) to a more recent version of the DSM-IV and ICD-10 there has been a remarkable similarity amongst the studies showing a unidimensional structure. Most of the studies used Exploratory Factor Analysis or Principal Component Analysis to analyze the item contents of an interview or a questionnaire measuring the ADS. Table 1.3 shows the main instruments used, their item contents and factor structure. All these studies

showed that one factor was responsible for the largest single amount of variance. These studies had different levels of 'construct underrepresentation' and 'construct irrelevant test variance', but their similarity in terms of factor structure is a robust finding that suggests the dependence construct has a consistent unidimensional structure.

Several other studies have used these same instruments in different populations and settings and found similar results in terms of the factor structure (Meehan et al, 1985; Kivlahan et al, 1989; Drummond, 1991; Stockwell et al, 1994). One important aspect of all these studies evaluating the ADS is that besides having used different diagnostic criteria they also used different methods of data collection (questionnaires and structure interviews). This diversity of criteria and data collection techniques demonstrate that the ADS construct is not uniquely tied to any particular method of measurement and that the different measures show convergence. The consistency of these findings has recently been confirmed by a study of Rounsaville et al (1993) who compared the DSM-III-R, DSM-IV and ICD-10 and assessed the factor structure of the three systems using interviews based on the Composite International Diagnostic Interview (CIDI). All three exploratory factor analyses yielded a one-factor solution with loadings of 0.84 or above for all items related to alcohol dependence.

Studies using Cronbach's alpha coefficient as a measure of internal consistency have recently received more attention (Table 1.4). There has been an impressive agreement between studies, with very high levels of Cronbach's alpha in different versions of the ADS. These studies have used different populations and methods of data collection. Skinner & Allen (1982), Hesselbrock et al (1983), Kivlahan et al (1989), Caetano (1992a), Stockwell et al (1994); Raistrick et al (in press) used questionnaires in clinical populations. Kosten et al (1987), Babor et al (1987a), Rounsaville et al (1993) and Hall et al (1993) used interviews with clinical populations, whilst Grant et al (1992) used interviews with a sample of the general population. They have also used differed item contents, with studies using diagnostic criteria based on DSM-III-R, DSM-IV and ICD-10. Babor et al (1988), Caetano (1992b) and Rounsaville et al (1993) compared two or more sets of items assessing a slightly different version of the ADS and found practically the same values of alpha.

Table 1.4 - Internal Consistency Analyses of the ADS using Cronbach's alpha.

|  |
|--|
|  |
|--|

| Studies                    | Cronbach's alpha |
|----------------------------|------------------|
| Skinner & Allen (1982)     | 0.92             |
| Hesselbrock et al (1983)   | 0.82             |
| Kosten et al (1987)        | 0.91             |
| Babor et al (1987b)        | 0.91             |
| Babor (1988)               | 0.62 to 0.91     |
| Kivlahan et al (1989)      | 0.85             |
| Grant et al (1992)         | 0.82             |
| Caetano (1992)             | 0.95, 0.96       |
| Rounsaville et al (1993)   | 0.85, 0.82, 0.76 |
| Hall et al (1993)          | 0.63 to 0.95     |
| Stockwell et al (1994)     | 0.98             |
| Raistrick et al (in press) | 0.94             |

### 1.5B - General Population Studies

Patients in treatment tend to have symptoms that are relatively coherent and persistent over time, but that is not always the case in the general population. Fillmore and Midanik (1984) found that some symptoms of alcohol dependence had much lower correlations in community survey data. Therefore, studies of the ADS using data from community surveys can be used as a stringent test of the coherence of the construct.

Several recent studies have relied on data from a sample of 43,809 respondents of a nationwide representative survey of the noninstitutionalized population of the United States aged 18 years and older, the National Health Interview Survey (NHIS88). In the first study of the series which analyzed the coherence of the syndrome Grant et al (1992) studied the degree of heterogeneity of the DSM-III-R alcohol dependence criteria observed in this population. The DSM-III-R has nine criteria for alcohol dependence and at least three are needed in order to make a diagnosis. The diagnosis can be achieved by a combination of any three or more of these criteria, so that theoretically there are 466 permutations. They found that in the general population only 189 subtypes (40.6%) were observed. This indicates that the alcohol dependence syndrome, as seen in the general population, is heterogeneous but not as heterogeneous as is theoretically possible. Symptoms of physiological dependence and impaired control over drinking were identified as playing a key role, appearing in over 80% of all reported subtypes regardless of age, race or sex.



Multhen et al (1993a) used the ICD-10 dependence syndrome criteria, with 20 symptom items, in order to test the dimensionality of the ADS. Factor analysis was performed on the results from 17,465 individuals who were current drinkers. The large number of current drinkers in the sample enabled the researchers to use two subsamples, obtained by random division of the full sample, in order to analyze the results by cross-validation. Impaired control, tolerance and withdrawal were identified as being part of one dimension. In another study based on the same sample but using DSM-III-R and DSM-IV dependence and abuse criteria Muthen et al (1993b) assessed whether more than one dimension was necessary to identify abuse and dependence. They found by means of Factor Analysis a two factors one representing abuse and the other dependence. The alcohol dependence factor contained items of tolerance, withdrawal and relief drinking.

Hasin et al (1993) based their study on a previous national survey of alcohol problems in the USA with a sample that consisted of 3212 adults who had consumed any alcoholic beverages during the twelve months preceding the interview. The items used represented several of the elements of the ADS and also of a complex of alcohol related problems (narrowing of repertoire, salience, tolerance, withdrawal, withdrawal avoidance, compulsion, and several problems related to the use of alcohol). They found, using confirmatory and exploratory factor analysis, that a single factor explained the structure of the data better than a two-factor model.

Few studies have compared general and clinical populations in relation to the structure of the ADS. Caetano (1991) compared a general population and a treatment sample from the same state in the USA and found that alcohol dependent individuals in the treatment sample were more severely dependent than their counterparts in the general population. The definition of ADS was based on a series of 18 items covering 8 of the 9 indicators of dependence from DSM-III-R. The number of dependence indicators reported by respondents in the two samples varied substantially, although the results from the regression analysis suggested that differences in severity of dependence between the two populations were due to differences in sociodemographic characteristics of the sample. Two studies compared the structure of the ADS between the two populations and found different results. Mohan et al (1992) used DSM-III-R criteria to compare the dimensional differences between a sample of patients from a treatment centre and another from the community. They found a one factor solution for the treatment sample but a two factors solution ('withdrawal' and 'social') for the community data. Stockwell et al (1994) compared the factor structure of a form of the SADQ for community samples of drinkers (SADQ-C) in a large sample

of attenders at a controlled drinking clinic and a random survey of Western Australian households. Both samples revealed a single major factor identified by Principal Component Analysis, accounting for 71.7% and 69.1% of the variance, respectively.

### **1.5C - Cross-cultural Studies**

One important aspect of the validation of the ADS is whether or not it shows constancy across cultures. Some of the critics of the ADS concept argues that it is a culturally specific expression of alcohol-related problems that arises in Anglo-Saxon societies in which the disease model of the Alcoholics Anonymous has been influential (Hall, 1993). Several studies have compared the structure of the ADS between two or three countries. Babor, Lauerman and Cooney (1987c) compared the responses of alcoholics in treatment in the USA and France on a common set of items dealing with dependence and the consequences of drinking. Analysis of the internal consistency of the dependency scales for both samples gave high Chronbach's alphas. Factor analysis of each sample revealed two orthogonal dimensions, distinguishing dependence symptoms from adverse consequences of drinking. In another study Babor et al (1988) made a secondary analysis of measures of dependency in patients from USA, Britain, and France which produced Cronbach's alpha ranging from 0.62 to 0.91. Allen et al (1993) using items from the Composite International Diagnostic Interview tested the dimensionality of the ADS in clinical samples in the United States and Russia. The Confirmatory Factor Analysis showed in the two groups that a single-factor model provided a high degree of goodness of fit. In another study comparing the same american and russian samples Allen et al (1994) found that the Alcohol Dependence Scale had a single dimension in both samples.

More recently a large WHO Collaborative study (Hall et al, 1993) tested the cross-cultural validity of the ADS in six countries (Australia, Bulgaria, Kenya, Mexico, Norway, and the United States). Cronbach's alpha varied from 0.63 to 0.95. Principal Components Analyses were performed on the 13 symptoms of alcohol dependence in each centre, and the degree of agreement between the results was assessed by calculating coefficients of congruence between the item loadings on the first principal component. The evidence for the unidimensionality was strong and consistent with all the centres having one first factor accounting for most of the variance. The Coefficients of Congruence all had values of 0.98 or more while the Root Mean Square differences ranged between 0.08 and 0.18, with an overall average difference of 0.13 between item loadings.

### 1.5D - Conclusion of the Internal Consistency Studies

The study of the internal consistency of the ADS is the area of the validation process that has received stronger support over the past two decades. Three main aspects contribute to the strength of the evidences in favour of the unidimensionality of the syndrome. Firstly, the ADS has been studied in a considerable variety of different populations and settings. The majority of the early studies with the ADS were with clinical populations, but over the years the syndrome started to be identified also in the general population and across several different cultures. Secondly, there has been a great diversity of methods of measurement (questionnaires and interviews) using slightly different operational criteria. This diversity of methods shows that the ADS can be independently defined by independent researchers and keeping its internal structure in terms of homogeneity of the sample content of the concept. Thirdly, several methods of analysis have been employed in the assessment of the structure of the ADS. Principal Component Analysis and Exploratory Factor Analysis have been the procedure more commonly used over the years, but more recently Cronbach's alpha and Confirmatory Factor Analysis have also contributed to the methods used. Despite the strong support for the unidimensional structure of most of the operational definitions of the ADS some inconsistencies have been found. Chick (1980a) found that 'Impaired Control', 'Narrowed Repertoire' and 'Salience' loaded in different small factors, presenting a challenge to the measurement of the syndrome. Caetano (1990), using exploratory factor analysis, found a four factor structure for items from DSM-III-R and for the ICD-10 a four factor solution among men and a three factor solution among women. In another study Caetano (1992a) also found when using different operational definition based on DSM-III-R that 2 to 5 factors represented the structure, depending on the operational definition used. Mohan et al (1992) also found a one factor solution for the treatment sample but a two factors for the community data.

These discordant results reveal important discrepancies that have yet to be explained. One possible explanation could lie with the relative impurity of some of the symptoms used to form the ADS construct. Some of the elements of the syndrome are difficult to operationalize and could be subject to idiosyncratic interpretation on the part of both researcher and subjects (Davidson, 1987). It is quite possible that additional dimensions could be a function of variance introduced by the nature of the inquiry rather than a confirmation of the multifactorial nature of the syndrome. Such anomalies should be followed up empirically because they are frequently the source of new insights

into construct validation (Messick, 1989). The repeated occurrence of discordant loadings on a factor might indicate that the construct is broader than originally conceived; that ostensibly separate constructs should be unified or that a higher order construct should be invoked to systematize the findings (Messick, 1989; Pedhazur and Schmelkin, 1992).

### **1.6 - Convergent and Discriminant Evidence of the Measurement of the ADS - Cross-Structural Analysis**

Internal Structure Analysis is necessary to determine whether there is a relationship between the structure of a set of indicators and the construct they are supposed to reflect. However, evidence from internal structural analysis is necessary, but in itself is insufficient to lend support to the validity of the construct. The reason is that a given internal structure may be consistent with different definitions of the construct (Messick, 1989). Moreover, any method of measurement can be distorted by a host of variables related to the subject, the researcher, and the setting in which the measurements are made. When a single method is used to measure a construct, it is not possible to determine to what extent subjects' responses are influenced by factors such as: response set, avoidance of extreme responses, giving socially desirable responses and reactivity to researcher's expectations (Messick, 1989).

In order to overcome these problems of bias in the measurement of a construct, various authors have advocated the use of multiple methods of measurement (Campbell and Fiske, 1959; Messick, 1989; Pedhazur & Schmelkin, 1991). The idea is to use several different forms of measurement (questionnaires and/or interviews) answered simultaneously by the same group of patients and assess their inter-correlation. Campbell and Fiske (1959) proposed the concept of convergent and discriminant validity. Convergent validity refers to a convergence among different methods designed to measure the same construct. Discriminant validity refers to the distinctiveness of constructs, demonstrated by the divergence of methods designed to measure different constructs (Pedhazur and Schmelkin, 1991). Messick (1980) supported the idea that convergent and discriminant validity were not a special form of validity. They should be considered data collection and data analysis strategies to be used for testing the conceptual connections between different forms of measurement and the construct.

Several studies have evaluated simultaneously different methods of measuring alcohol dependence. Stockwell et al (1979) compared an independent rating of alcohol dependence with the total score of the SADQ and found a Pearson correlation coefficient of 0.84. Meehan et al (1985) and Drummond and Chalmers (1986) also found high levels of concordance among a rater and the SADQ. Davidson and Raistrick (1986) compared the scores of the SADD with other questionnaire (SADQ) and a standardized interview (EADS). They found that the total scores of the SADQ and SADD had a  $\rho=0.83$  ( $p<.01$ ) and between the SADD and EADS a  $\rho=.51$  ( $p<.05$ ). Cooney et al (1986) studied four scales that measured the ADS: the Rand Dependence scale, the Severity of Alcohol Dependence Questionnaire (SADQ), the Last Month of Drinking Withdrawal Scale, and the Last Six Months of Drinking Impaired Control and Dependence scale. Using a Factor Analysis all four scale scores loaded on one factor, indicating that the four scores were highly inter-correlated. Alcohol consumption and psychosocial problems scales did not load on the dependence factor, but each did load on two other factors.

Two studies explored the correlation between the ability of two questionnaires to identify alcohol dependence and its the degree of severity. Jorge et al (1986) compared the SADD and the ADS, and found a significant correlation between the scores obtained on the two scales ( $r=0.61$ ,  $p<0.01$ ). However, analysis of how patients were classified in terms of severity of dependence, revealed a lack of agreement between the two questionnaires ( $Kappa =0.20$ ). In a sample of problem drinkers on an in-patient unit, Doherty and Webb (1989) found a strong correlation between the SADQ and SADD scores ( $r=0.806$ ). Further analysis revealed a strong agreement between the two questionnaires in respect of the assessment of low, middle range and high levels of dependence ( $Chi-Square=92.23$ ,  $p<0.0001$ ).

More recently greater emphasis has been given to the assessment of the different operational criteria of the ADS used in the psychiatric classification (DSM-III-R, DSM-IV and ICD-10). In a study that explored the correlation between two questionnaires measuring alcohol dependence, Caetano (1990) found a Pearson product-moment correlation of 0.93 between an index with 20 items, representing the DSM-III-R concept, and an index composed of 16 items, representing the ICD-10 concept. In a later study, Caetano (1992) assessed the impact of two different operationalizations of the ADS according to DSM-III-R. He created two sets of measures, one using traditional items to define the various elements of the concept and a second using questions worded as closely as possible to those used in DSM-III-R. He found low Phi correlations among

individual items, with 60% of these correlations being between 0.21 and 0.40. However, when these two sets of items were transformed into scales the Pearson correlation coefficient between the two measures rose to 0.68 ( $p < .01$ ).

Cottler et al (1991) used the Composite International Diagnostic Interview (CIDI), a standardised diagnostic interview for the assessment of psychiatric disorder which allows classification according to DSM-III-R and ICD-10 criteria. They interviewed 590 patients from 18 sites around the world in a field trial designed to test the cross-cultural acceptability and reliability of the questions. The Kappa value for overall agreement between the two classifications of the alcohol dependence category was 0.81. Rounsaville et al (1993) studied the cross system agreement of 3 sets of criteria for Substance Dependence (DSM-IV, DSM-III-R and ICD-10) in a heterogeneous sample of 521 adults. Agreement for dependence was generally high with Kappa values above 0.85 for all substances (alcohol, cocaine, opiate, stimulant, sedative, marijuana) and 0.84 for alcohol dependence alone. Rapaport, Tipp and Schuckit (1993) as part of the DSM-IV field trials interviewed 100 patients to ascertain substance use diagnosis according to ICD-10 and DSM-III-R. The Kappa values comparing the concordance of diagnosis between the two systems for alcohol dependence was .79. Cottler et al (1993) compared DSM-III-R and two versions of the ICD-10 in a study from the DSM-IV Substance Use Disorders Field Trials. She found that these three versions of the ADS gave similar rates of prevalence in a 887 subjects from the general population. Grant (1993) using a representative sample of the United States general population compared the DSM-III-R and the DSM-IV formulations of alcohol dependence. The 1-year prevalence rates of alcohol abuse and dependence for each classification were remarkably similar. Concordance between diagnostic categories of dependence presented a Kappa value of 0.76.

Only one study was found that used discriminant analysis to compare the ADS with other related constructs (Gorman et al, 1989). Two questionnaires measuring dependence (SADQ and SADD), one measuring alcoholism (MAST), and one problem drinking (ARP) were compared. There was substantial agreement between the two measures of alcohol dependence with an intraclass correlation coefficient of 0.81. There was rather less agreement between the MAST and the SADQ and SADD with coefficients of 0.52 and 0.49 respectively. Even lower correlations were found between the ARP and the SADQ and SADD (0.26 and 0.38 respectively).

In summary, studies of cross-structural analysis of the ADS have provided evidence that similar

operational criteria of the construct are conceptually connected. Similarly to the studies of internal consistency the studies of convergent analysis have used a variety of data collection and statistical analyses that contributes to the strength of the findings. These studies have used several forms of measurement (interviews and questionnaires), with several operational definitions (DSM-III-R, DSM-IV, ICD-10, SADQ, SADD, etc), with several types of population (clinical, general) and several different statistical analyses (correlation, factor analysis, kappa, intra-class correlation).

### **1.7 - Distribution of the Severity of the ADS**

The internal structure and the cross-structure analysis of the ADS have been consistently found to represent a single dimension. In the original description of the ADS emphasis was also given to the fact that the syndrome existed along a continuum of degree of severity (Edwards and Gross, 1976). Empirically there is considerable evidence that individuals who are misusing alcohol can be ordered along a dimension of severity. The severity of the ADS has been assessed in different populations.

Skinner (1982, 1990) showed that patients attending an alcohol clinic had scores on the ADS that conform quite closely to a normal distribution. Meehan et al (1985) analyzed the total scores of the SADQ and found that it varied along a continuum. They could not identify a normal distribution of the total scores but there were two groups of patients identified around a cut-off point of 23 points in the SADQ. Davidson and Raistrick (1986) compared the SADQ and the SADD and found that the SADQ had total scores skewed to the left indicating that it may have been somewhat less able to discriminate between individuals in the mild to moderate range of dependence. However, the SADD was able to identified with a good degree of spread with Shapiro-Weisberg normality statistic approaching significance, indicating a good spread of responses. Doherty and Webb (1989) studied in-patients that were assessed by the SADQ and SADD. They showed that both questionnaires had their scores unimodally distributed along a degree of severity.

Woody et al (1993) in a study of the DSM-IV field trials studied the severity of dependence on alcohol using the CIDI-SAM in a sample 1100 subjects from the general and clinical population. The severity ratings were defined by the number of positive DSM-IV criteria (version used during the Spring of 1992) reported: mild 3-4, moderate 5-6, and severe 7-9. Severity correlated

reasonably well with measures of quantity and frequency of use. A study of general population has also provided support for the identification of the severity of the ADS. Using data from a US national survey, Hasin & Glick (1992) studied 4000 respondents who met criteria for DSM-III-R alcohol dependence, in terms of severity of dependence. Severity was defined by the number of positive DSM-III-R criteria reported: mild 3-4, moderate 5-6, and severe 7-9. They found a gradient of severity with 73.7% classified as mild, 17% as moderate and 9% as severe.

### **1.8 - Test-Retest Reliability and Temporal Stability of the ADS**

The studies discussed in the previous sections have showed that there is evidence for an internal consistency, cross-structure and degree of severity of the ADS. Another aspect that was implied in the original description of the ADS is that it is a condition that has its clinical course presented with a consistent clinical patterns over a reasonable period of time. Temporal stability is sometimes evoked to assess the validity of psychiatric diagnoses (Beiser, Iacono and Erickson, 1989; Blashfield and Livesley, 1991; Nelson-Gray, 1991). Short-term stability must be expected if a diagnosis is to be clinically significant. The assumption is that, the more stable the diagnosis, the more likely it is to reflect a psychopathological process. However, temporal stability is expected for the diagnostic criteria of some diagnostic categories but not for others (Blashfield and Livesley, 1991). For the ADS, temporal stability of the diagnosis is expected to be stable over at least the short term.

One way of assessing temporal stability is through the consistency of diagnosis over a period of time. Inconsistency in diagnosis across time can be attributed to either poor reliability of the instrument or poor validity of the construct. Rice et al (1992) discussed the sources of disagreement between ratings in test-retest reliability studies and maintained that they were either due to error in the measuring instruments or to true change in the state itself. In the temporal stability paradigm, it is assumed that some of the error components will be uncorrelated between assessments, and this reflects true change in the state. Accordingly, we use the term stability rather than long term reliability to indicate this. Thus, there is a distinction between assessing the repeatability of the diagnostic instrument (reliability) and the use of multiple measures to assess the validity of the underlying constructs (stability).



Several studies have showed evidence of the stability of the diagnosis of alcohol dependence. Stockwell et al (1979) asked 45 subjects to complete the SADQ and then to do so again 2 weeks later. They found a Pearson correlation coefficient of 0.85 between the overall total scores. McMurrin and Hollin (1989) re-administrated the SADD from 19 to 40 days later and found a correlation of 0.88 ( $p < 0.001$ ). Cottler et al (1989) using the CIDI-SAM interviewed patients twice, one week apart. For the alcohol dependence diagnosis, using DSM-III-R criteria, they found a kappa value of 0.92 and Yule's Y of 0.89. Williams et al (1992) used the Structured Clinical Interview (SCID) in a multi-site test-retest reliability study, with the second interview administered between one day and two weeks later. The diagnosis of alcohol dependence had a Kappa of 0.75 for the whole sample and 0.83 for a group of patients in the researchers' own Substance Abuse Treatment Unit. Raistrick et al (in press) using a new questionnaire, the Leeds Dependence Questionnaire (LDQ), calculated a test-retest reliability over an interval of 2 to 5 days with a total score retest reliability of 0.95

In summary, although the studies discussed in this section have not been designed to assess temporal stability but rather to assess test-retest reliability, they are able to show that the diagnosis of the ADS is stable over short periods of time. The advantage of using the emphasis of the temporal stability paradigm is that it seems that it is less prone to error of measurement because it assumes that agreement reflects the component corresponding to true clinical state (Rice et al, 1992) and in the context of the ADS it offers more important information in terms of its validation.

### **1.9 - Criterion-related Validity and the ADS**

Validity always requires one or more external criteria that can be applied to the issue under study. The external component of validity refers to the extent to which the measures of the construct relate to other indicators implied by the underlying theory. Criterion-related validity is the evidence that demonstrates that scores on a test are related to some defined criterion measure of interest, not just any criterion, but those that make most sense on theoretical grounds (Messick, 1989; Suen, 1990). A distinction used to be made between two types of criterion-related validity: predictive and concurrent. There is little substantive difference between these two types of criterion-related validity. The only difference is in the exact time when the criterion measurement is made (Suen, 1990). Predictive validity is the extent to which the test score can be used to predict

the score from a criterion measurement procedure that will take place at some future point in time.

Concurrent validity is the relationship between the scores on the test and the criterion measure taken at the same time. The selection of a specific criterion is determined largely by the values and goals of a particular study.

Three main groups of studies can be identified in relation to the approach used in criterion-related validity research of the ADS construct. The first group of studies assess behaviour that occurs in association with alcohol dependence over a period of time in clinical and general population settings. The second group of studies assess behaviours associated with alcohol dependence under laboratory conditions. The third and larger group of studies assess the outcome of patients after treatment for drinking problems and its association with the severity of alcohol dependence. These three groups of studies are discussed in the following sections.

### **1.9A - Correlational Studies**

Research in clinical populations have shown that alcohol dependence is associated with a number of medical and social consequences. One important area of study has been the relationship between alcohol dependence and alcohol related problems. Implicit in the concept of the alcohol dependence syndrome is the view that alcohol-related problems constitute a dimension which is conceptually separate from dependence. Drummond (1990) found a positive correlation ( $r=0.63$ ;  $p<0.001$ ) between problems caused by drinking, as measured by the Alcohol Problems Questionnaire (APQ), and the severity of dependence, measured by the SADQ. Dependence was particularly associated with problems related to friends, as well as physical, affective, financial, marital and work problems. Furthermore, this correlation existed independently of the average quantity of alcohol consumed. Allan (1991) also found, in a sample of clients attending counselling for alcohol problems, that psychological symptoms measured by the General Health Questionnaire (GHQ) and total SADQ scores correlated at 0.4 ( $p<0.001$ ).

Caetano (1993) found that severity of dependence, measured using DSM-III-R indicators, was one of the most powerful predictors of medical consequences in a sample of alcoholic patients. Clients reporting 7-9 indicators had a higher prevalence of medical and social consequences compared to clients reporting 3-6 or 0-2 indicators. The predictive power of the number of dependence indicators on the number of medical consequences reported was independent of sociodemographic characteristics. Multiple regression showed that the number of dependence

indicators were one of the most powerful predictors of all medical consequences.

Another important criterion related to the ADS is the amount of alcohol consumed over a period of time. Grant and Harford (1990) evaluated the risk of alcohol dependence at different levels of alcohol intake. The data used were drawn from a multistage probability survey designed to yield a representative sample of 5,221 adults from the U.S.A. Using linear logistic regression analyses to separate the association between average daily alcohol consumption and DSM-III-R alcohol dependence, they found that the risk of dependence increased with higher consumption. Frequency of heavy drinking has also been associated with degree of dependence. Caetano (1991) in a general population study and using multiple linear regression found that frequency of drinking 5 or more drinks per occasion was the only significant predictor of alcohol dependence. He also found similar results in a clinical population. Dawson and Archer (1993) in a study based on the NHIS88 also found that the relative frequency of heavy drinking had a strong positive association with the risk of past-year alcohol dependence, even after adjusting for the potentially confounding effect of average daily ethanol intake.

### **1.9B - Experimental Studies**

Several studies have looked at dependence using experimental procedures, including ones that have measured the response of an alcohol dependent subjects to a challenge dose of alcohol (Rankin et al, 1979; Hodgson et al, 1979; Stockwell et al, 1982; and Kaplan, 1983). These studies have changed previous ideas about the influence of cognitive factors. The results of these papers suggested that alcohol dependent subjects were more disposed to drink after alcoholic drinks than after soft drinks, irrespective of whether they believed that the priming drink contained alcohol. Cognitive factors assumed greater importance in the drinking behaviour of less dependent subjects.

Dependence has been found to be related to cue reactivity in several studies (Kaplan, 1983; Monti et al, 1987). Alcoholics with a greater urge to drink in response to a series of alcohol related role-play scenes drank more alcohol during the six months following treatment (Rohsenow et al, 1991). Rohsenow (1992) tested the factors that predicted an increase in the urge of alcoholics to drink in high risk situations, such as exposure to their customary alcoholic beverage. In two studies she found that cue reactivity was greatest among those with more severe alcohol dependent histories.

Changes in salivation and urge to drink were both significantly associated with higher ADS scores ( $r=0.65$ ,  $p<0.01$ ).

In a recent study Glautier and Drummond (in press) studied a group of 35 severely dependent alcoholic patients undergoing a cue exposure treatment programme. Their approach to the magnitude of cue reactivity was to examine the relationship between individual elements of the dependence construct and responsiveness to cues. They created a measure of responsiveness based on physiological and subjective responses to drinking cues which were measured on the first day of the exposure programme. Principal component loadings were used to construct a single measure of responsivity. This multivariate measure of responsiveness was found to correlate significantly with SADQ scores, but was independent of levels of alcohol consumption. In particular, experience of affective withdrawal symptoms, craving for alcohol and drinking to relieve withdrawal symptoms were most strongly correlated with cue responsiveness ( $R=0.56$ ,  $p=0.001$ ).

### **1.9C - Outcome Studies**

Outcome studies have proved to be one of the most important and widely applicable methods for assessing the criterion of validity in psychiatry (Kendell, 1989). This happened because the ability to predict the future course of events, and to alter them if necessary, has been a primary function of medicine (Kendell, 1989). Indeed, a more or less distinctive natural history has always been inherent to the concept of a syndrome. The outcome studies of the ADS have showed that it has predictive utility in terms of the natural history of drinking (Chick, 1985). These studies have however, evolved over the years to show the exact role of dependence in the patterning of the outcome.

Early outcome studies reported that there was a relationship between dependence and outcome in terms of drinking behaviour in groups of treated patients. Orford, Oppenheimer and Edwards (1976a) reported on the progress of 100 problems drinkers two years after attendance in a treatment trial. A seven-item scale, concerned with morning drinking, tremors, nausea, loss of control, passing out when drunk and hallucinations given prior to attendance predicted very closely the type of drinking outcomes. Polich et al (1980) also studied the relationship between patients achieving good outcome and levels of dependence. They found that the ability to maintain non-problem drinking decreased as the severity of dependence increased. In terms of the overall

outcome, alcohol dependence symptoms related to an unfavourable prognosis, being associated with adverse consequences of drinking, continued dependence, and alcohol-related death. The analysis, using logit analysis, revealed several variables that affected relapse rates, including severity of dependence, age and marital status. Osejo (1981) reported a 15 year longitudinal study of a community sample of 96 alcoholic men from Sweden. Symptoms of dependence were found to have predictive power in terms of remission. At fifteen years 51% of 'abusers' were in remission compared to only 14% of 'addicts' (62% were still dependent). Similar results have been obtained in other studies using different populations (Vaillant et al, 1982).

More recently outcome studies have started to examine the importance of dependence in association with other variables in the patterning of the outcome. This happened because evidence was accumulating that outcome could not be conceived as a simple unitary dimension, but as a process in which the alcohol dependence variable had an important role in association with other variables. McLellan et al (1983) evaluated male alcoholics six months after treatment in rehabilitation programmes. They found that both severity of dependence and psychiatric symptoms were good predictors of outcome, but that psychiatric symptoms alone were more robust as outcome criteria. Rounsaville et al (1987) performed a one-year follow-up study of 266 alcoholics who had received extensive psychiatric assessment using DSM-III criteria. Degree of dependence, measured at intake, was significantly correlated with a poorer one-year outcome for 10 out of 13 ratings for men. Two other factors were also found to predict outcome: psychiatric diagnosis and a global assessment of psychopathology as measured by the MMPI.

Taylor et al (1986) in a ten year follow-up study collected information on 68 male alcoholics as a basis for exploring patterns of outcome using multivariate analysis. A Principal Component Analysis of several outcome variables showed that degree of dependence was intrinsic to the pattern of outcome but that there were several possible relationships between dependence and outcome. Outcome was represented by two factors that accounted for 40% of the variance. Dependence, measured by the SADQ, had a 0.38 loading on the first factor and 0.72 on the second. The authors described that dependence had a 'Janus' effect, whereby high dependence was related to both good and bad outcome in different circumstances. Babor et al (1987b), in an effort to find the most efficient way to measure treatment response, evaluated the relationships between a large variety of outcome variables using factor analysis. They followed up 321 alcoholics, with DSM-III-R criteria for alcohol dependence, for one year after discharge. The

analysis of the structure of follow-up variables revealed one factor that was most clearly defined in terms of intensity of drinking (average ounces per drinking day), alcohol dependence and alcohol-related problems. A second factor was defined by measures of psychological and social functioning and frequency of drinking.

Using a different approach, Babor et al (1992) initially studied an empirically derived typology of alcoholism and later tested response to treatment. A clustering technique was applied to data obtained from 321 male and female alcoholics to identify homogeneous subtypes. One group, designated type A alcoholics, was characterized by later onset, fewer childhood risk factors, less severe dependence, fewer alcohol related problems, and less psychopathological dysfunction. The other group, termed type B alcoholics, was characterized by childhood risk factors, familial alcoholism, early onset of alcohol-related problems, greater severity of dependence, more chronic involvement in treatment, and greater psychopathological dysfunction. In a development of the previous study Litt et al (1992) assessed the outcome of seventy nine male alcoholics (of both A and B types) that were randomly assigned two different kinds of treatment (coping or interactional therapy). Analysis of outcome indicated that type A alcoholics fared better in interactional treatment, whilst type B alcoholics had better outcomes with the coping- skills treatment. Differences in treatment response were maintained for two years.

In a recent longitudinal study of a sample of the general population, Hasin et al (1990) used criteria of dependence based on DSM-III-R to explore the differentiation of alcohol abuse from dependence in terms of natural history. A 4-year follow-up of the 71 men who initially had only initial indicators of alcohol abuse, 50 (70%) continued to report only indicators of alcohol abuse (N=17) or remission (N=33), and the remaining 30% (N=21) reported indicators of alcohol dependence. In contrast, of the 109 men with initial indicators of alcohol dependence, 50 (46%) still reported indicators of dependence 4 years later, and 59 (54%) reported indicators of abuse only (N=16) or were in remission (N=43). The differences in outcome between the two groups of abusers and dependent drinkers were statistically significant (Chi-Square=4.12, df=1, p<0.05).

In summary, the criterion related validation of the ADS offers a multifaceted quantity of evidence supporting the construct. The most unequivocal evidence concerns the correlational studies which showed that alcohol dependence is associated with different measures of alcohol related problems in different populations. The laboratory studies have also showed clearly that alcohol dependence

is associated with different forms of cue reactivity. The results of the outcome studies are less obvious to interpret. There has been an evolution in these studies from the early days when dependence was expected to determine the outcome, to a more recent focus where dependence is considered as one of the variables that influence outcome. However, there is still much to be explored in the interaction of dependence with other variables in the patterning of outcome. The utility and relevance of outcomes cannot be fully appraised in the absence of a sound empirically grounded interpretation of the concept of ADS and other variables (Messick, 1989). The unified view of validation has shifted the focus from prediction to explanation. It can be argued that a better prediction of the role of the ADS in the outcome will only be achieved with a better measurement of its relationship with the other variables.

### **1.10 - Measurement of dependence on other drugs**

In 1980 a WHO Scientific Group (Edwards et al, 1981) suggested the possibility of extending the dimensional dependence model of the ADS to a spectrum of other drugs. This extension makes information related to studies of these drugs theoretically relevant to the validation of the ADS construct. In the last seven years most of the studies of the more general drug dependence syndrome have examined its dimensionality. In one of the first of these studies Skinner & Goldberg (1986) assessed 105 polydrug users with the Drug Abuse Screening Test (DAST). This instrument has items covering perceptions of a drug abuse problem, dependence symptoms, and various consequences related to drug abuse. They found a five factor Varimax solution which accounted for 55% of the variance. One of the factors was related to the dependence syndrome with three items loading very highly : inability to stop drug use, inability to get through the week without drugs, and withdrawal symptoms when drug use stopped. These items loaded on a factor distinct from other factors relating to problems associated with drug abuse.

A more recent strategy has been to assess the dimensionality of the syndrome in respect of specific drug. One group of drugs that has been extensively studied is the opiates. Previous work with the SADQ was the starting point for early studies of drug dependence in opiate users. Several studies (Sutherland 1986, Phillips 1987, Sutherland 1988, Burgess 1989) using different populations (outpatients in New York, DDU patients in London and Australia) contributed to the development of the Severity of Opiate Dependence Questionnaire (SODQ). It was designed to be comparable with the SADQ and bears a close resemblance to that instrument. A high degree of stability has

been found for the psychometric properties of the SODQ across the different samples studied. The factor structure has shown a strong first factor accounting for 40% of the variance, with items conceptually related to a dimensional model of the opiate dependence syndrome. In another study of opiate users, Stripp et al (1990) used DSM-III-R criteria to investigate the unidimensionality of the syndrome. They used a latent trait modelling framework to analyze the data and the findings also supported a unidimensional concept of the drug dependence syndrome.

Bryant et al (1991) evaluated in great detail the dependence syndrome for cocaine in terms of syndrome coherence and the continuum of severity. They based the diagnosis on DSM-III-R criteria derived from structured clinical interviews with 399 cocaine users and analyzed the data using confirmatory factor analysis. They showed that a single factor provided a good representation for most of the nine criteria (as measured by a coefficient of fit). Amphetamine dependence has been studied by Churchill et al (1993) who created a questionnaire similar to the SADQ and the SODQ. They found, using factor analysis, a unidimensional structure similar to the studies with alcohol and opiates.

Two studies used DSM-III-R criteria for substance abuse in a population of drug users without a drug-specific diagnosis. Kosten et al (1987) examined the structure of the dependence syndrome for various substances in a group of predominantly polydrug users. Factor analyses indicated support for the unidimensional concept for opiates, cocaine, and alcohol, but not for sedatives, hallucinogens, stimulants and cannabis. However, Cronbach's alpha was higher than 0.83 for all drugs studied. Hasin et al (1988) studied drug-using subjects from an alcohol rehabilitation unit and explored the unidimensionality of the drug dependence syndrome. Principal Component Analysis showed that for cocaine, opiates, tranquilizers, barbiturates, and stimulants one factor represented the items very well, accounting for between 30% and 50% of the variance. Feelings of dependence on a drug, unsuccessful attempts to cut down, and tolerance, loaded highly on each of the drug factors.

Cross-structure analysis of the drug dependence syndrome was studied by Rousanville et al (1993) in a sample of 521 subjects from different treatment settings. They assessed the measures of internal consistency of the dependence criteria from DSM-III-R, DSM-IV and ICD-10 using the structured interview CIDI for cocaine, opiates, alcohol, marijuana, stimulants and sedatives. High levels of agreement were found across the three diagnostic systems and across most categories of



drugs ( $> .80$ ), with marijuana consistently associated with lowest levels (.70 - .81). Exploratory factor analysis was employed to assess the factor structure of the three diagnostic systems and across all drug types. All factor analyses yielded a one-factor solution with loadings of 0.64 or above for all items.

There has also been evidence in terms of the criterion-related validity of the drug dependence syndrome. Gossop et al (1992) investigated the relation of severity of dependence upon heroin, cocaine and amphetamines and a series of criteria. Route of administration of the drug, dose and duration of drug use and previous attendance at a drug treatment agency were associated with severity of dependence measured by the Severity of Dependence Scale (SDS). Kosten et al (1992) in a 1-year follow-up found that dependence syndrome severity predicted treatment success for a group of users.

In summary, there is strong evidence that the concept of the dependence syndrome can be extended to others drugs of abuse. The process of validation of the drug dependence syndrome is following a similar pattern and showing similar strengths than the ADS. The evidence for a unidimensional structure is quite convincing for most of the drugs. Others aspects of validation have not yet been so thoroughly examined but they seem to be comparable to the ADS.

### **1.11 - Conclusion and objectives of the thesis**

The process of validation of the Alcohol Dependence Syndrome has progressed steadily over the past seventeen years. The situation is now far advanced compared to the early days following the ADS description, when even its existence as a construct had to be justified and defended (Hodgson, 1980). Over the years one aspect of the ADS that has been particularly studied is its internal structure. These studies have followed a hidden research programme which considered the verification of the dimensionality of the syndrome as the main goal. They have shown a remarkable similarity in terms of the coherence and dimensionality of the syndrome. The findings are particularly impressive because of the diversity of methods and populations examined. Other areas of the validation process have received proportionally less attention, but the evidence seems to suggest that the ADS has become a sound construct. Therefore, the question arises : how should the validation of the ADS be carried forward ?

One possible answer is to make progress in the direction of improving the relationships with others related constructs. In the psychological literature Cronbach and Meehl (1955) pointed out that the meaning of a construct increases when its relation with other constructs improves in a meaningful way. They stated that a construct is defined by the network of associations that they called a nomological network. The basic notion of nomological validity is that the theory of the construct being measured provides a rational basis for deriving empirically testable links between the measurement of the construct and measures of other constructs. Construct validation ultimately rests on studying relations between the construct in question and other constructs or variables in a theoretical context (Pedhazur and Schmelkin, 1992). Very few studies discussed before in this chapter could be said to have increased the nomological validity of the ADS. The studies that have looked at the association of the ADS and alcohol-related problems, and in many respect the studies of outcome of treatment and ADS increased the nomological network because they improved the theoretical relation between the two concepts. However, the difficulty with this approach to the improvement of the validation of a construct is that it needs a mature construct, a construct with a well advanced level of measurement. Messick (1989) recommended that for this reason testing a nomological network is more appropriate in a mature construct validation program than in a beginning one.

The other more feasible option is to improve the internal validity of the ADS. Internal validity in this context is related to the ways the different elements of the syndrome relate to each other. It is different from internal consistency, which relies on the sample adequacy of the items of a construct. The idea behind internal validity is that high indicators on the construct should score highly on other presumed indicators of that construct. This leads one to expect a 'convergence of indicators' across the several aspects of the construct being measured (Messick, 1989). Babor (1986) has also specifically suggested that the validity of the ADS would benefit from clarification of the relationship between cognitive, behavioural and physiological elements.

Very few studies have looked at the internal validity of the ADS. Stockwell et al (1983) examined the correlation between an interview that measured what they defined as narrowing of drinking repertoire and the SADQ score. Using multiple analysis of variance, with the SADQ score as the dependent variable, they found that narrowness, both in terms of limited variability between and within heavy drinking days, contributed significantly to the variance ( $p < 0.001$ ) in each instance. In

their categories of drinking: 'mainly continuous' and 'mainly binge', drinkers tended to have higher SADQ scores than 'occasional abstainers'.

Two studies have looked specifically at the speed of reinstatement in relation to the severity of dependence. Topham (1983) in a prospective study of 48 patients undergoing treatment for alcoholism found that, after six months, 19 of these patients had by their own definition relapsed. Symptoms such as sweating, shaking and craving were related to the degree of dependence at  $p < 0.001$  level, and to the length of time it took for morning drinking to return ( $p < 0.01$ ). Babor et al (1987a) studied the reinstatement of dependence following a period of abstinence in a group of 321 alcoholic patients. They tested the hypothesis that the greater the degree of dependence at admission, the more likely the syndrome would be reinstated once drinking was initiated following a period of abstinence. They found that among alcoholic patients, both recent and lifetime alcohol dependence measures were moderately predictive of reinstatement of alcohol dependence at 1-year follow-up. Furthermore the severity of dependence, as measured by the Last 6 Months of Drinking Questionnaire, correlated with reinstatement in males (coefficient = 0.47,  $p < 0.01$ ).

The main aim of the thesis is to develop measurements of several elements of the ADS and assess their relationship. Four elements of the ADS were chosen to be measured: narrowing of drinking repertoire, subjective awareness of compulsion to drink, salience of drink-seeking behaviour and repeated withdrawal symptoms. Except for the repeated withdrawal symptoms which already has a questionnaire suitable for its measurement, the other three elements had new forms of assessment especially created for the project.

The structure of the thesis is as follows: the next four chapters describe the development of the questionnaires designed to measure elements of the ADS. In chapter 2 the evolution of ideas related to the subjective aspects of drinking is discussed and a new questionnaire specifically designed to measure the subjective awareness of a compulsion to drink is proposed. In chapter 3 the evolution of the ideas of the description and classification of patterns of drinking behaviour is discussed and a new method of measuring drinking repertoire, based on a structured interview and on a series of questionnaires, is proposed. In chapter 4 the factors that reduce or stop drinking will be reviewed and a questionnaire designed to measure salience of drink-seeking behaviour in the form of modifiers of drinking behaviour will be presented. In chapter 5 the ideas contributing to the measurement of the alcohol withdrawal symptoms are presented; it also describes a questionnaire widely used to measure the ADS, the Severity of Alcohol Dependence

Questionnaire (SADQ). The structure of these chapters is very similar, they start with a historical evolution of the main ideas that contributed to the identification and measurement of the element, discuss in more detail its importance in terms of the ADS construct and propose a model on which its measurement will be based. In chapter 6 the design of the survey of a group of patients attending clinical facilities for treatment of drinking problems is described. In chapter 7 the results of the whole survey are presented followed by a statistical analysis of each questionnaire. In chapter 8 a general discussion of the contribution of these new questionnaires for the validation of the ADS is presented.