Eur J Clin Pharmacol (1985) 28 [Suppl]: 31–34

European Journal of Clinical Pharmacology

β-Blockers and Psychometric Performance

The Clinical Relevance of Psychometric Testing

A.H. Mann

Academic Department of Psychiatry, The Royal Free Hospital, Pond Street, London, NW3 2QG, U.K.

Summary. Tests of performance emphasizing alertness, psychomotor speed, and reaction time are widely used in testing the effect of β-blockers on the central nervous system. However, this aspect of brain function is only one of many that are amenable to psychometry. This paper is presented from the standpoint of psychiatry and provides an overview of psychometric testing that highlights the points of contact with general medicine and psychiatry, and with research in both these areas. The applications of psychometry in defining relatively stable individual differences such as intelligence or personality, and in distinguishing morbid from normal states, are discussed. The use of psychometry in detecting performance changes relevant to pharmacological research is reviewed. Interpretation of all test results depends upon an understanding of the structure of the test itself. The confidence that may be placed on interpreting results depends on adequate reliability. For any inference to be drawn there must be both normative data and criterion validity, i.e. discovery of the score point at which there are demonstrable changes in day-to-day function other than those detected within the confines of the test. Unfortunately many psychometric tests fall short on these data and, therefore, inferences from their use should be made cautiously.

Key words: β-blockers, psychometric tests; CNS, motivation, normative data, personality, questionnaires, socio-cultural effects, validity co-efficient

Assessment of part function of the brain to determine whether or not damage is present, and the measurement of intelligence are two examples of psychometry for which clinical relevance is now established. For other widely used measures, such as tests that assess personality, the relevance is more in their prediction of lifestyle or occupational choice. Therefore, it is unfortunate that when assessing the effect on cerebral function of β-blocking drugs the measures most commonly used concern performance skills which require tests of alertness, speed and accurate fine movement. Other measurements concern symptoms that might be produced by β-blockers, such as drowsiness, depression, hallucination or nightmare. These two measurements, of performance skills and side-effects, are more recent developments and have been less subject to scrutiny than have measurements of the more stable conditions of intelligence and personality.

The clinical relevance of the tests used to assess CNS side-effects is hard to determine at the present time. Therefore, the purpose of this paper is to summarize what must be understood about psychometric measures, before their results can be interpreted with confidence and translated into clinical practice. Two important questions should be answered: (1) How good is a test? (2) What factors should be taken into account when interpreting the specific use of a test? Further background information can be gained from reviews of psychological testing (Anastasi 1976), psychomotor functions and psychoactive drugs (Hindmarch 1980), and performance and impaired performance (Nicholson 1976).

How Good is a Test?

There are three prerequisites for a good test: (1) high reliability; (2) good validity; and (3) normative data.

Reliability

Reliability refers to the consistency of scores obtained in a test by the same individual on two occasions. High reliability also implies that an individual

consecutively completing two tests containing equivalent items will obtain very similar results. Knowledge of reliability enables experimental score differences to be evaluated as true differences or to be attributed to fluctuations in the test itself. Reliability can be assessed by test-retest, which entails repeating the measurement on two occasions. Test-retest is the best method for assessing reliability when the measurement is of a stable trait such as intelligence or personality. The alternative method of testing reliability involves consecutively conducting two tests of identical length containing equivalent items. This is the best method of assessing the reliability of measures of transient states. The learning/practice effect should however be taken into account. The reliability of questionnaires is probably best assessed by the split-half method in which the items are divided into two groups for the purpose of scoring. Each half should produce a similar total if the test is reliable. Failure to produce similar scores implies that the questionnaire consists of unrelated items and will be unreliable.

There are several important ways of improving reliability. The administration of the test itself should be standardized, attention being paid to defining the equipment, the setting and the instructions. Of particular importance is a definition of the amount of help given by the test supervisor. Distraction for the subject and the test supervisor should be minimal if reliability is to be maintained. The scoring of any test should be standardized, perhaps employing computerized scoring to eliminate observer error. Usually it is better to determine whether a performance test is primarily concerned with the number of errors made by the individual or with the time taken to complete the test. Scoring becomes more complex if the total score presents both error and time components. It is important to control the practice effect because scores may well improve until a plateau is reached. It is important that this plateau is achieved before the scoring of a performance test is started. However, it is possible to use some tests that measure the effect of a drug upon the time taken to achieve this plateau. Clear instructions are essential.

Questionnaires should be constructed carefully, particular attention being placed upon the wording. It is important to remember that even simple words, such as 'anxiety' and 'depression', have quite a different meaning for a clinician and a lay respondent. Many questionnaires designed to assess CNS side-effects refer to dreams, nightmares and hallucinations. Each of these words will have idiosyncratic meanings and it is helpful to provide a glossary for the respondent and/or the administrator of a questionnaire.

Validity

Validity refers to the ability of a test to measure what it is supposed to measure. Expressed in another way, validity measures the relationship between the test result and independently obtained information on the behaviour or state that is under consideration. The most important component of validity is criterion validity, the checking of the test against independent criteria of behaviour. With concurrent validity the criterion to be assessed independently is occurring at that time, whereas predictive validity implies that the result of the test will predict some future event. Thus the concurrent criterion of a performance test might mean the ability to drive a car or pilot an aeroplane, whereas predictive validity might be output at work over three months while taking the drug. Some criterion data are readily obtainable, others much less so.

A self-report questionnaire can be validated by comparing the questionnaire result with data obtained by independent interview. This exercise only validates self-report versus an interviewer's attempt to elicit the same information. A true validation would involve a clinical observation, for example, effect of mood on sleep and nightmares, an unpractical exercise. Because of this anomaly many psychopharmacologists use visual analogue scales instead of questionnaires. According to their proponents, these scales are the most flexible method for a subject to self-report on mood states. They avoid the respondents having to judge their states in artificial categories such as 'mild' or 'severe' depression. The relationship of test result to criterion can be expressed either as a validity coefficient or as a sensitivity/specificity ratio.

Normative Data

Apart from good validity and high reliability which can both be expressed as coefficients, interpretation of a test result is meaningless without normative data. A score involving six errors may represent an excellent response on one test but a poor response in another. Each test score should be presented in relation to a standard for that test, standards determined from the use of a test in a sample of the population at large. The standardized or Z-score is obtained by dividing the difference between an individual score and the mean of the scores of the normative data, by the standard deviation of the scores in the normative data. The standard score therefore indicates how far from the mean an individual respondent lies. Standard scores allow two individuals to be compared in their use of one test, or allow the same individual's

performance to be compared on several tests measuring different aspects of function. An alternative, but less satisfactory expression of relative position of an individual respondent to normative data is as a percentile. Percentiles are more readily understood but their disadvantage is that they ignore the size of the difference between the scores in the normative data.

Normative data for a symptom questionnaire come from the investigation of the responses of a healthy population. How many of the experiences said to be associated with a β-blocker, such as nightmares, are experienced by individuals taking no drugs and with no particular complaints at all? Certainly many surveys of a general population have indicated that, at any point in time, over half the population have occasional symptoms of headache, anxiety or irritability. Similarly the discovery of abnormal brain appearances by computerized tomography in many psychiatric patients were exciting until investigations with normal healthy populations indicated that abnormal appearances were by no means uncommon during and beyond middle age. Normative data are essential in order to show the sort of scores or changes in scores which are likely to indicate disturbance.

What Factors Affect Test Results?

Most experiments in which the CNS side-effects of β-blockers are investigated will be performed on small groups selected because of their illness status or because of volunteering. Therefore errors might occur from the use of a test in a particular population, although the test is shown to be reliable and valid. These errors, called moderator variables, will now be considered.

Age and Sex of the Subject Population

Age and sex affect both performance test on self-report and mental state. Psychomotor speed slows normally with age and there is a greater report of neurotic symptoms among women than men. Unless normative data are available, classified by age and sex, it is necessary either to select a study population homogeneous for these variables or to ensure that analysis of the results will take age and sex into account.

Personality

Much psychological research has investigated the effect upon performance, specifically sensory-motor coordination, of personality (Eysenck 1952). Patients with a high score of neurotic traits tend to perform least well on the performance test whereas introverts become less bored and are more persistent in tests than extraverts. Personality variables might be particularly relevant in a design in which patients with a disease state are being compared with normal volunteers. Membership of each of these groups might be influenced by differing personality traits. Another reason for assessing personality is that a drug could affect introverts, extraverts or other personality types differently, and local effects might be masked in an overall non-significant finding.

Interest and Motivation

Many psychological tests are boring and many jobs are interesting; the opposite might also be true. Psychological tests that simulate car driving or aeroplane piloting may be much more absorbing than routinely driving a London Transport bus. To what extent therefore will motivation and interest overcome the minor side-effects of sedation in certain circumstances? Subjects might try hard and would not notice the sedation while being tested but would complain of it when at work. This type of error can be examined by the introduction of a known sedative into the experimental situation. If there is no sedative effect after a dose of amylobarbitone, then this indicates that the test is too stimulating and is of little value in detecting a sedative side-effect.

Socio-Cultural Effects

Individual upbringing and culture are influential in determining whether or not certain symptom states are noteworthy and therefore merit complaint. The culture of a volunteer group may differ from that of the general population implying that certain effects would or would not be more likely to be reported in a clinical trial.

Conclusions

From a review of published trials it is clear that many experimentors have devised their own questionnaires and tests of psychomotor performance. The reliability, validity and normative data, let alone the effects of moderator variables, are rarely reported and considered. In view of the wide varieties of tests on different groups in different cultures, it is very difficult to compare one experiment with another. There is an urgent need for a standardized battery of performance tests and questionnaires for use in this field. Preliminary work would be necessary to deter-

mine reliability and validity coefficients before use of these tests in general populations, so that age and sex normative data are available. Once such tests and data are accessible, as is the case with intelligence tests, then the results of performance tests in psychopharmacology could be evaluated with greater confidence than is possible at present.

References

1. Anastasi A (1976) Psychological Testing, 4th edition. Macmillan, New York
2. Eysenck HJ (1952) The Scientific Study of Personality. Routledge, Keegan and Paul, London
3. Hindmarch I (1980) Psychomotor functions and psychoactive drugs. Br J Clin Pharmacol 10: 189-209
4. Nicholson AN (1976) Editorial: Performance and impaired performance. Br J Clin Pharmacol 3: 521

Dr. A. H. Mann
Academic Department of Psychiatry
The Royal Free Hospital
Pond Street
London NW3 2QG, U.K.

Group Discussion

C. T. Dollery

With drugs that have primary CNS action, such as the opiates, benzodiazepines, and clonidine, their central action is reflected by the behavioural effects and symptoms which are caused when the drug is withdrawn. Similarly, with β-blockers, peripheral effects such as tachycardia on standing and increased incidence of angina are apparent on withdrawal. Is there evidence of clear CNS effects when β-blockade is withdrawn suddenly?

A. H. Mann

It is difficult to distinguish between a pharmacological/physiological withdrawal effect and psychological dependence.

C. T. Dollery

There is also controversy about whether the anxiolytic effect of the β-blockers is central or whether it represents the reduction of somatic manifestations by a peripheral action.