Current approaches to neurobehavioural testing in occupational health

Anne Spurgeon

Abstract

In recent years neurobehavioural tests have been used increasingly in occupational settings to identify changes in cognitive function associated with exposure to neurotoxicants. Potential applications of these methods in occupational health include research, diagnosis, and screening. Applications in cross sectional research studies, involving the comparison of the performance of exposed and control groups, are well established. However, the use of such methods requires attention to factors other than exposure which may influence test performance. The validity of adapting existing test batteries for diagnosis or screening is questionable. Well developed techniques exist for diagnosis but this requires lengthy and skilled test administration and interpretation and cannot be accomplished with research batteries. The application of neurobehavioural methods for screening currently presents several difficulties. Current problems and future directions in this field are discussed.


Keywords: neurobehavioural tests; research; diagnosis; screening

During the past 25 years neurobehavioural methods have been used increasingly in human studies to investigate effects on the central nervous system of exposure to neurotoxic chemicals. These methods assess the behavioural outcomes of toxic insult by measuring effects on aspects of cognitive functioning such as visuomotor skills, the ability to sustain attention, and higher level information processing abilities such as learning, memory, and logical reasoning. From the point of view of occupational health they offer several advantages. They are non-invasive and therefore generally acceptable to workers, portable, and relatively inexpensive. Most importantly they provide a powerful tool for assessing the more subtle changes which may occur at lower levels of exposure in the absence of overt clinical signs.

By far the most common application to date has been in the area of research; thus a large section of this paper is devoted to a discussion of research methods. Neurobehavioural tests have been used extensively either in laboratory based studies designed to identify acute reversible effects of neurotoxicant exposure and the levels at which these effects occur, or in large scale epidemiological type studies to investigate the chronic effects of long term exposure. However, partly as a result of research findings and hence an increased awareness of the potential effects of relatively low levels of exposure to neurotoxicants, an increasing interest is now developing in the use of psychological methods in diagnosis. The requirements of research and diagnosis of patients are somewhat different, however, and this has sometimes led to problems of misapplication and misinterpretation. Finally, there are now suggestions that neurobehavioural techniques might usefully be used for screening and health surveillance. This raises further questions about the nature of existing tests and how far they can realistically be applied to different areas of occupational health practice. Each area requires a particular approach which differs from the others both in the techniques of measurement and the interpretation of results. It is important to be very clear, therefore, about the requirements of each, and these I discuss separately.

Research

ACUTE EFFECTS

The use of neurobehavioural methods to investigate the acute reversible effects of neurotoxicants usually involves an experimental design in which people complete psychological tests after different levels or duration of exposure. In an experimental situation exposure usually takes place within an exposure chamber which allows maximum control of both exposure and a range of situational factors such as temperature, humidity, noise, and time of day which have the potential to affect test performance. Also, subjects may be selected as a homogenous group for age, sex, educational level, and lifestyle factors. Effects on performance of different test situations are usually controlled by balancing the order of the tests.

The data derived from this type of study have been used more often to set standards than have those derived from studies of long term exposure, in which dose is more difficult to define and confounding factors more difficult to control.

In practice most experimental studies have
been carried out to investigate the acute effects of exposure to different organic solvents, effects which are often of the prenarcotic type and which are readily identifiable by tests of visuomotor skills, or tests which focus on the ability to control attention.

Criticism of these types of studies has centred on the use of volunteers who are not habitually exposed to the substance in question and may not therefore be representative of a workforce at risk. It has been suggested, for example, that real workers may have a combination of acute and chronic effects of exposure which would not be evident in volunteers. Some studies have attempted to rectify this by carrying out quasieperimental studies of acute effects in the workplace. However, such studies are fraught with methodological and practical difficulties which make it extremely difficult to achieve the definition of exposure and the control of confounding variables which are such positive features of experimental studies.

**CHRONIC EFFECTS**

Unlike experimental studies, large scale epidemiological type studies tend to be concerned with investigating irreversible damage to the central nervous system resulting from long term exposure. This has implications both for the study design and for the type of neurobehavioural measures used.

Most studies have been cross sectional in design and have compared the test performance of a group of people who are regularly exposed to a particular chemical with that of an unexposed control group. As cognitive performance is represented by scores on a continuum—for example, response speed in milliseconds—an effect of exposure is assessed by the size of the difference between the scores of the two groups. In particular, confidence in the results is strengthened if a measure of dose is included and a dose-effect relation can be shown.

In this context, no attempt is made to define "normal" or "impaired" people or to analyse results by the number of cases in each group. Scores on tests of cognitive function are heavily dependent on situational, educational, and pre-existing intellectual factors and universally applicable reference criteria are thus impossible to find. In studies of this type, therefore, very careful attention must be paid to a range of factors, unrelated to exposure, which may exert an influence on test performance. Some of these may be taken care of at the recruitment stage of a study, others may be controlled for within the design of the study or dealt with by statistical adjustment at the analysis stage.

It is normal, for example, to exclude subjects who have pre-existing disease or injury affecting the nervous system or a history of substance misuse. Further, exposed and control subjects are usually matched for sex, educational level, and cultural background. More problematically, factors such as the influence of intelligence level before exposure (usually termed premorbid ability) and the possible effects of lifetime alcohol consumption and smoking may require an assumption that they are likely to exert an interactive effect on test performance. These factors are difficult to measure accurately, particularly that of intelligence. In the past this has usually been carried out with a verbal test such as reading ability or vocabulary, on the basis of evidence that scores on such tests correlate highly with intelligence quotient (IQ), when conventionally measured, and are relatively resistant to nervous system insult. More recently, however, this view has been questioned and investigators are increasingly favouring educational level as a more valid indicator of premorbid functioning.

In common with other epidemiological investigations careful consideration should also be given to the influence of potential confounders—that is, factors which may vary with exposure—and hence lead to results which may falsely provide evidence of an effect. The most obvious confounder in this context is that of age. Confounders should be dealt with at the recruitment or the analysis stage of the study.

As noted above, many of the substances investigated with neurobehavioural tests—for example, a wide range of organic solvents—have short term, prenarcotic effects. When chronic effects are under investigation it is important to establish that subjects are tested at a time removed from the effects of recent exposure or the contamination by acute effects. This presents considerable difficulties as for many substances data on the persistence of effects after exposure do not exist. In many investigations pragmatic decisions have to be taken which assume, for example, that an exposure free period of 12 hours avoids the possibility of identifying acute rather than chronic effects.

In research major emphasis is placed on standardisation of test procedure and on unambiguous quantitative scoring methods. Also important is the control of time of testing because of the potential influence of circadian rhythms on performance. This is often complicated if factory workers are involved, as many work on rotating shifts.

Longitudinal neurobehavioural studies are rare and usually involve either repeated testing of both exposed and control groups over months or years or long term follow up of previously diagnosed cases to assess the reversibility or otherwise of identified effects. Both approaches are problematic in a neurobehavioural context. As already mentioned, there are numerous difficulties associated with the identification of a case which have led some investigators to question the results of a few studies adopting this approach. Further, designs which involve comparison of neurobehavioural test scores before exposure with those at subsequent intervals over time assume that test techniques will remain unchanged (an unrealistic and not necessarily desirable situation) and that test conditions including all potential influences on performance will be consistent after a period of years. Also, in the
Current approaches to neurobehavioural testing in occupational health

absence of normative data (discussed later) it is difficult to allow for the effects of normal aging over time. As a result of these difficulties and also the considerable costs involved in repeated testing of many people most investigators have preferred to use cross sectional designs to investigate chronic effects despite the accepted limitations of these.

TEST BATTERIES FOR RESEARCH

For research purposes the current requirements are for administration and scoring of tests to be standardised and so to be valid and reliable and to sample rapidly a range of cognitive functions. Further, the tests should be acceptable to subjects who are, to all outward appearances, healthy, and do not perceive themselves to be patients.

In recent years several test batteries have emerged in an attempt to fulfil these requirements. Some were developed by particular groups to service their own programme of research. Most of these were in Scandinavian countries, reflecting the origins of neurobehavioural work which took place mainly in those countries during the 1970s. They include in particular two batteries developed in Sweden, the TUFF battery and the Swedish performance evaluation system (SPES). The TUFF battery draws heavily on non-automated neuropsychological tests established over several years in clinical diagnosis. They were selected with the objective of defining a battery which could be used both for research and diagnosis. Data provided by the developers seem to indicate that most tests in the battery are sufficiently sensitive to discriminate both between occupationally exposed workers and controls, and between patients with a clinical diagnosis of organic solvent syndrome and non-patients. However, in a research context the battery is time consuming and labour intensive.

There is now an increasing trend towards the use of computer administered tests for research as this fulfils both the need for speed and for standardisation of administration and scoring. The SPES battery was designed in this way and represents a combination of some well established neuropsychological tests and others derived from techniques developed within a framework of cognitive psychology. The battery is used entirely for research purposes.

Although the TUFF and the SPES were originally developed for in house use by particular research groups some batteries were developed primarily with the aim of international distribution and application. Of these the most widely used are the computer administered neurobehavioural evaluation system (NES) developed in the United States and the non-automated World Health Organisation neurobehavioural core test battery (WHO-NCTB) developed in Italy. The WHO-NCTB has recently been produced in an automated version under the title of the Milan automated testing system (MANS). Both contain the facility for translation into other languages and have a network of support and information.

Most of the batteries now in common use drew on existing tests from two sources for their development. These were neuropsychological tests, principally intended for diagnosis of patients with diseases or injury of the nervous system, and psychometric tests of intellectual ability designed to assess potential for future attainment. Both types of tests are associated with banks of normative data developed from long experience of their use and the temptation inevitably exists to apply these research norms to identify affected people. In the process of adapting such tests to provide for automated administration and an unambiguous scoring system they were inevitably changed. It cannot be emphasised too strongly therefore that the application of existing norms in this context is entirely inappropriate.

Because of their common origins there is considerable overlap in the tests used in different batteries, although details of presentation, response, and scoring may vary. Most batteries will include, for example, a test of simple or choice reaction time, a test of short term memory and longer term learning, and a mesure of more complex reasoning ability. Outcomes are usually measured in terms of accuracy or speed of response.

CURRENT PROBLEMS AND FUTURE DIRECTIONS

Although most existing test batteries tend to fulfil the practical criteria for research purposes they remain somewhat limited in their capacity to define the scientific effects of any particular neurotoxicant. For example, the use of neuropsychological tests in a group situation, isolated from the diagnostic skills of the neuropsychologist who necessarily relies on qualitative as well as quantitative information, may provide at best a limited and at worst an inaccurate assessment. Essentially many of these tests have been extracted from a diagnostic situation and placed in a research context for which they were never intended.

Increasingly there is a recognition that neuropsychological tests are a unique area of study requiring its own purpose developed techniques that may draw on, but not simply borrow, methods developed in other areas for other purposes. Recent thinking in the field therefore has seen a gradual movement away from heavy reliance on neuropsychologically based tests, towards tests based on the theories and techniques developed to explore patterns of normal cognitive functioning.

Such techniques allow the analysis of aspects of performance which reflect the approach or strategy of the worker in performing the task, and the identification of subtle shifts from the expected or normal pattern. These shifts may occur within or between particular tests which are designed, for example, to assess performance at different levels of difficulty, under different conditions of presentation or response, or with different types of test material. Potentially, data derived from test batteries developed along these lines should allow researchers to define the effects of particular chemicals more precisely in psychologi-
cal terms. Further, reference to a common theoretical base should facilitate comparisons between results of different studies which have not necessarily used exactly the same tests. Finally, an important added advantage of this approach is that individual exposed workers who attempt to fake their results are more readily identifiable. Currently test development along these lines is at a relatively early stage. Examples of batteries which place more emphasis on this type of approach are the information processing and performance test battery (IPPTB) and the automated cognitive test (ACT) battery.

In summary, neurobehavioural tests used for research purposes seem to be in a transitional phase. Original batteries were formed in response to an urgent demand to consider questions about the more subtle effects on populations of occupational exposure to neurotoxins. It might be argued that the initial demand to answer these questions was to some extent outpaced by the ability of the researchers to investigate them. Thus although many of the early methodological difficulties in this field have now been overcome and existing tests have been able to provide some answers, there remains a need to develop more refined, theoretically based techniques which can provide more precise information about particular chemicals at group level.

Diagnosis
The difficulty of defining a case in neurobehavioural terms for research purposes has already been noted. Because of the wide range of factors which may affect test performance there are considerable problems in developing a normal range of scores which would be applicable across different populations. This problem necessarily restricts the type of research design which can successfully be used, and also creates problems for practitioners faced with the challenge of diagnosis of patients. The preponderance of neuropsychological tests in research batteries has led some investigators to assume that such batteries may also be used for diagnosis. This is entirely inappropriate, a fact which has been repeatedly underlined by those involved in the development of the batteries. Even when batteries were developed with diagnosis as well as research in mind, it has been emphasised that diagnosis cannot be arrived at with reference to test scores alone. Such scores simply provide complementary data to be placed alongside clinical and exposure history. This draws attention to the essential nature of diagnosis in this field, which relies heavily on the skill and experience of the practitioner and is invariably a relatively lengthy process.

Clearly, therefore, although test data play an important part in the diagnostic process, the data required will be both qualitative and quantitative. Neuropsychological assessment in this field is essentially similar to assessment after other forms of nervous system insult such as head injury, degenerative disease, or cardiovascular incident. Neurotoxic exposure as a causal or contributory factor is therefore determined by consideration of exposure history and with reference to previous case reports compiled by other clinical neuropsychologists, although it must be said that information on the specific patterns of deficits associated with particular types of neurotoxicant exposure is currently very limited.

As an example of the diagnostic process Hartman describes a diagnostic decision tree, typical of a sequence of clinical evaluations for neurotoxic syndromes used at a major hospital in Chicago. Tracing through this flow chart it becomes clear that the diagnostic process is a multidisciplinary effort involving neurologists and hygienists as well as psychologists. The input of the psychologists necessarily involves the use of some well established neuropsychological batteries, perhaps the most often used being the Halstead-Reitan battery, the Luria—Nebraska battery, and the Weschler adult intelligence and memory scales. Other batteries tend to contain tests drawn from these three basic reference points. As well as tests of intellectual functioning it is customary to include in the diagnostic process the evaluation of mood and personality changes assessed either by clinical interview or by questionnaire. Perhaps the most well established questionnaire is the Minnesota multiphasic personality inventory (MMPI) although a range of standardised culture specific measures exist for assessing aspects of psychiatric status such as anxiety and depression. It should be emphasised throughout, however, that diagnosis is a long and complex process which should be undertaken by a team of skilled practitioners. It cannot be accomplished by the use of short test batteries designed to identify differences between the performance of groups.

Screening
The growing acceptance of neurobehavioural tests as a means of assessing neurotoxicity has led to suggestions that such techniques could usefully be used as initial screening devices to identify people at risk or requiring further investigation or for the purposes of early detection and routine health surveillance. Neurobehavioural batteries contain several features which suggest that they may be ideally suited for this purpose. They are, as has been noted, cheap in time and resources and highly acceptable to workers. They are also sensitive to subtle and thus it may be argued, early changes although it must be said that much remains to be discovered about the nature and progress of neurotoxic effects. Hence the predictive value of such methods for early detection remains controversial.

The main remaining difficulty to be overcome before such an approach can be successfully implemented is the difficulty of defining a cut off point for use as an indicator of impairment.

The application of neurobehavioural tests in research has underlined the difficulty with research designs which require case definition,
as the distribution of scores for any one population will depend on a range of demographic and lifestyle factors. With the development and international use of the NES and the WHO-NCTB there was initial optimism that normative data might be developed by pooling of results, particularly those relating to control groups. In recent years, however, there has been increasing acceptance that reference data in the form of a range of normal scores are likely to be population and context dependent. Without such data, however, definition of risk and subsequent assessment of levels of sensitivity and specificity remain problematic.

Conversely, those tests which have the potential to identify a case are long and complicated, requiring a high level of input from a skilled practitioner, which is clearly incompatible with the demands of a screening programme. Some neurobehavioural test batteries are in fact entitled screening batteries. However, these seem to be essentially research batteries and cannot be said to consider fully the problems discussed above. Those screening instruments which have been successfully used in the workplace to date, and which come under the broader definition of neurobehavioural methods, are based on questionnaires— for example, the Orebro Q-1624 designed for the detection of neurotoxicity related to solvents. The questionnaire contains 16 questions relating to early symptoms of neurotoxicant exposure and offers a cut off score to identify people who require further, more detailed investigation. This questionnaire, which has been used extensively and successfully in Scandinavian countries for screening purposes, currently seems to represent the most useful approach in this area. Although a full validation procedure has not been carried out current indications are that the questionnaire has reasonably high sensitivity although rather low specificity.24

Conclusion
Neurobehavioural testing has several present and future potential applications to occupational health practice. It has been used most extensively in research, where it currently represents the most reliable approach to the assessment of subtle, subclinical effects of neurotoxicant exposure. Several questions remain, however, relating to the precise nature of identified effects, their association with particular types and levels of exposure, and the link, if any, between early changes and later development of more severe effects on cognitive functioning. The refinement of neurobehavioural tests necessary to consider these questions is likely to involve a shift in emphasis away from neuropsychological approaches towards the application of techniques derived from the study of normal cognitive functioning. In the diagnosis of individual patients existing approaches to assessment have been shown to be appropriate in the hands of skilled practitioners. There remains, however, a need to develop a larger bank of reference information from individual case histories to assist in future assessments. Perhaps the greatest challenges lie in the area of neurobehavioural screening. The potential contribution which neurobehavioural tests could make to this area make it a priority for future research effort. In all cases, however, it is of major importance that users of neurobehavioural tests are clear about their objectives and the particular application in which they are involved if they are to avoid the inappropriate use of tests and the consequent misinterpretation of data which has sometimes occurred in the past.