Pergamon

Neurotoxicology and Teratology, Vol. 16, No. 5, pp. 511-524, 1994

0892-0362/94 $6.00 + .00

0892-0362(94)00040-9

Criteria For Progressive Modification of Neurobehavioral Batteries

ROBERTA F. WHITE,* FREDRIC GERR,‡ ROGER F. COHEN,† ROBERT GREEN,** MURIEL D. LEZAK,§ JEFFREY LYBARGER,|| JAMES MACK,** ELLEN SILBERGELD,# JOSE VALCIUKAS,## WILLARD CHAPPELL### AND LESLIE HUTCHINSON||

Departments of Neurology and Environmental Health, Boston University School of Medicine, Boston, MA 02118
†Mentor Clinical Care, Brookline, MA 02780
‡Occupational and Environmental Health, Emory University School of Publilc Health, Atlanta, GA 30329
***Department of Neurology, Emory University School of Medicine, Emory University, Atlanta, GA 30329*
§Department of Neurology, Oregon Health Sciences University, Portland, OR 97201
||Agency for Toxic Substances & Disease Registry, Atlanta, GA 30333
***Department of Neurology and Psychiatry, Case Western Reserve University, Cleveland, OH 44106*
#Department of Epidemiology and Preventive Medicine, Johns Hopkins University, Baltimore, MD 21201
##Forensic Psychiatry Clinic, New York, NY 10029

Received and Accepted 1 June 1994

WHITE, R. F., F. GERR, R. F. COHEN, R. GREEN, M. D. LEZAK, J. LYBARGER, J. MACK, E. SILBERGELD, J. VALCIUKAS, W. CHAPPELL AND L. HUTCHINSON. *Criteria for progressive modification of neurobehavioral batteries.* NEUROTOXICOL TERATOL **16**(5) 511-524, 1994.—Six specific issues affecting the progressive modification of neurobehavioral test batteries used in field studies of populations exposed to neurotoxicants are discussed and test review recommendations are provided addressing each issue. The issues include: (a) general test review standards, (b) comprehensive assessment, (c) tailored batteries, (d) incorporation of new tests and techniques, (e) personnel and mechanisms for review, and (f) development of a battery assessing peripheral nervous system function.

Neurobehavioral tests Neurotoxicology

THE AGENCY FOR Toxic Substances and Disease Registry (ATSDR) convened the Workshop on Neurobehavioral Testing in Atlanta, GA, September 11-13, 1991. The general purpose of the workshop was to evaluate methods and strategies for ATSDR use in determining whether neurobehavioral effects are associated with exposure to hazardous substances (in humans). The charge given to us was to develop criteria and methods for the progressive evaluation and modification of test batteries for use in field studies assessing nervous system effects of exposure to known or suspected neurotoxicants.

Six areas of interest were identified as foci for the workshop:

1. Development of general standards to be employed in progressive evaluation of test battery components to determine if individual tests should remain in the battery;
2. Progressive modification of adult and pediatric test batteries so that each battery approximates as comprehensive an evaluation of nervous system function as possible;
3. Tailoring of test batteries for use when assessing the effects of specific toxicants as the range of possible effects becomes known from previous studies;
4. Review of newly developed neurobehavioral assessment methods for possible inclusion in adult and pediatric test batteries;
5. Recommendations regarding personnel and mechanisms to be established to accomplish the task of progressively reviewing standard batteries;
6. Development of a battery for assessment of peripheral nervous system (PNS) function and criteria for review of a PNS battery.

[1] Requests for reprints should be addressed to Roberta White, Ph.D., Department of Veterans Affairs Medical Center, 150 South Huntington Avenue, Boston, MA 02130.

Neurobehavioral test batteries aimed at assessing the effects of neurotoxicant exposure are generally administered with one of two objectives, either (a) epidemiologic assessment of health effects of exposure in a group of individuals or (b) clinical assessment of individual subjects to render diagnostic conclusions about exposure effects (i.e., determine if the subject meets criteria for diagnosis of toxic encephalopathy or peripheral neuropathy). In some field studies these two goals are combined: Subjects are assessed individually and assigned appropriate diagnoses and group data are also analyzed epidemiologically (28,30,43,54). When clinical diagnosis is one of the aims of assessment, the criteria for test inclusion and progressive evaluation are different from those applicable to the epidemiologic setting (75).

In the epidemiologic setting, psychometric concerns are critical and have greater importance than they do in clinical assessment and diagnosis. For example, a test that is passed by virtually all subjects except those with brain damage affecting a particular part of the brain may be extremely informative in the clinic but adds no information at all in an epidemiologic behavioral neurotoxicology field study (57). This type of test suffers from the problem of restriction of score range, which should generally be avoided in epidemiologic settings unless a focussed hypothesis involving cerebral localization of toxic effects is being addressed. Likewise, qualitative findings are quite important in clinical settings but are difficult to score and observe reliably or summarize in the epidemiologic setting. Finally, epidemiologic settings almost always place greater constraints on time available for testing. Whereas a clinical battery might require 3 to 4 h or more of testing to produce an accurate diagnosis (74), the investigator rarely has more than 1 h of a subject's time available for behavioral testing during field studies. A 1-h battery is frequently all that can be administered in an epidemiologic setting. In prior studies, 1-h batteries have provided a fairly comprehensive assessment in which one or more theoretical hypotheses have been tested but individual clinical diagnosis was not possible (4,18,72,70,77).

The validity of epidemiologic testing is undisputed, but individual diagnosis might be appropriate and essential under certain circumstances: (a) exploring potential outcomes, assessing clusters of cases or meeting the concerns of exposed citizens; (b) forming a set of hypotheses about exposure effects based on casework, especially in novel or unusual exposure situations; (c) differentiating toxic effects from social or psychological effects, or in investigating exposure effects on particularly susceptible individuals or subgroups of exposed individuals. Furthermore, epidemiologic studies may be essential for detecting subtle deficits caused by exposure. Statistically significant differences between exposed cases and nonexposed controls can be found in well designed and performed epidemiologic studies that would not be detected using only clinical evaluation and diagnosis as outcome measures (77).

Recommendation. Intensive investigation of nervous system function using assessment techniques that produce clinical diagnoses of individual subjects may be informative under particular circumstances. These include investigations of clusters, assessment of subgroups of especially susceptible or vulnerable subjects, or in exploring the effects of novel exposures or unusual subject groups. In addition, intensive investigation of a small sample of subjects could be undertaken at the beginning of a large-scale study to explore the occurrence of specific symptoms and nervous system deficits. It can be argued that random or stratified subsamples of subjects should be extensively studied by a team of sophisticated clinicians in any large study.

I. GENERAL STANDARDS FOR EVALUATION OF TEST BATTERY COMPONENTS

The two previous articles offer a neuropsychological test battery for adults and the functional areas to be assessed for children in environmental research. The decision rules for test inclusion apply not only to developing new test batteries but also to the ongoing evaluation of test battery components (individual tests) to determine whether the tests should be retained in the battery. For this reason, general standards for evaluation of tests included in assessment batteries designed for field testing of subjects exposed to known or suspected neurotoxicants are briefly described next.

The evaluation of test battery components using these standards requires careful investigation of and control for both subject and exposure variables. Some of the most pertinent subject variables considered by the group include developmental stage(s) or age at the time of exposure (which in environmental studies is often different than age at the time of testing), age at the time of testing, education, occupation, socioeconomic status of origin, handedness, gender, language background, ethnicity, learning history, medical history, premorbid psychiatric adjustment, individual neurogenetic and neurochemical susceptibility, social/political/forensic issues affecting subjects, and exposure to neurotoxicants other than those under investigation (self-administered drugs including alcohol, medications, other environmental or occupational exposures). The ability to predict the variance in test performance due to subject variables (often treated as covariates in behavioral neurotoxicology field studies) reduces the total error variance when analyzing test data. Relevant exposure variables include type of toxicant(s), route(s) of entry, dosage, biological markers of exposure, and duration and chronicity of exposure.

Evaluation standards are divided into general evaluation criteria and validation criteria.

General Criteria

Efficiency and scoring. Each test included in a screening battery should be as efficient as possible to maximize information obtained about each area of cognitive function (see section II for a listing of domains). Many tests assess more than one cognitive domain, and it is possible to score some tests so that the contribution of each domain (or many of the domains) can be captured. For example, if a symbol-digit substitution task is included in the battery, separate scores can measure motor speed (time to complete task), accuracy of scanning (error scores), spatial orientation (number of rotated symbols + omissions of responses occurring in a series of filled in cells), learning (incidental recall of symbols and/or digit-symbol associations), and consistency of responding over time (fatigue or acceleration effects seen when comparing number of responses made as the task progresses at set timed intervals, e.g., every 5 s). All of these scores can be obtained without adding time to test administration while simultaneously maximizing information derived from the task (40).

Within the rubric of testing efficiency, several parameters are essential in reviewing existing test battery components. Most of these relate to scoring. They include: ease and objectivity of scoring, dimensions of scoring, consistency measures of scoring rules, the use of raw scores versus scaled scores,

and the appropriate comparison groups to be used in score evaluation. Scoring issues (especially as they relate to consistency of scoring and test administration) are especially important in field settings where multiple examiners, at least some of whom may be relatively naive about psychometric theory and standardization rules, may carry out testing.

The use of summary test scores is an issue of concern. Valuable information is sometimes lost in studies when summary scores are used rather than component scores. This may result in negative conclusions when positive conclusions are more accurate or may even result in a distortion in interpretation when summary scores prove insensitive. For example, if grip strength measures on a hand dynamometer are measured for each hand and then combined, the finding of a "normal" score may obscure a weakness in one hand accompanied by above average strength in the other.

A major concern regarding summary scores involves the use of IQ scores. The finding of low IQ scores might be interpreted to mean that exposure produced a lowering of general intelligence when, in fact, the IQ score decrements could be explained on the basis of significant impairment on tests measuring only attention or visuospatial function. Validity of the construct of IQ or general intelligence is disputed. The impact of the conclusions of the Needleman et al. studies (51,52) regarding the effects of childhood lead exposure on public health policy, which used the construct of intelligence and IQ, serves as an example of the controversial use of the IQ score.

Although many support the use of IQ summary scores, the opposition to such scores deserves explanation, described next. Whereas summary test scores for specific research purposes can make obvious trends that would have remained subtle in certain instances, the general use of summary score [e.g., IQ, Memory Quotient (MQ), "Impairment Indices"], is both inappropriate and unscientific. These scores were all developed more than 60 years ago (more than 75 for the IQ scores), when notions about mental processes on which these scores are based represented the best armchair speculations of their day. By the 1990's, knowledge about brain function and current research have demonstrated not just the inutility of these summary scores but how, in obscuring relevant data, they may actually cover up important performance differences between groups. Moreover, the naive reliance on summary scores by persons and groups who make important financial, educational, and vocational training decisions about individuals has resulted in innumerable personal tragedies such that some learning disabled or culturally deprived children have been declared retarded, some severely cognitively impaired persons have been deprived of benefits they deserve, and so on (Lezak, in press). So long as these scores are in the records of participants in epidemiological field studies, these persons are vulnerable to this kind of inappropriately based decisions. This argument does not mean that many valuable tests of mental abilities or cognitive functions from which summary scores, such as IQs, as often derived are not useful. However, the data from the individual tests should be handled as discrete variables in initial analyses.

A final scoring issue is the use of qualitative observations of test performance (e.g., spatial neglect, perseverations) in epidemiologic settings. Such measures are often unreliable to score, too low in frequency of occurrence to be useful, and cumbersome in data analysis.

Recommendations. Scoring methods should be easy and objective. Objectivity of scoring can be determined by assessing inter-scorer reliability.

Scoring procedures should capture important features of each performance dimension the test is intended to assess.

The different dimensions of cognitive functioning assessed by a given test should be scored, analyzed, and reported separately when possible.

In most cases, raw scores from case and control groups (not scaled scores based on national normative data) should be used in data analyses. Whenever normative data are used, their appropriateness for the subjects under study should be carefully considered.

Summary scores such as IQ, MQ, and Impairment Indices should not be used or used only with great caution and for well defined reasons.

Functional specificity. This concept refers to the capacity of tasks to tap specific types of cognitive processing and to differentially measure processing deficits within a designated cognitive domain. For example, digit-symbol tasks, as noted, tap many types of cognitive processing, while the Continuous Performance Test is most clearly a measure of attention. Although several authors felt that relatively "pure" assessment of specific cognitive processes would enhance the interpretation of study results at a processing level and in regard to brain-behavior relationships, some felt that purity of cognitive measures was not extremely important.

Recommendations. The contributing cognitive processes assessed by each task in the battery should be operationally defined (and scored) as precisely as possible.

Repeatability. Because environmental field studies often involve use of longitudinal study designs to assess chronicity of effects, recovery, or delayed effects of exposure, it is important to consider the advisability of repeating the use of screening battery tests (or alternate forms of the tests) in longitudinal assessments of individual subjects. Particular problems here include: (a) familiarity with test stimuli (which can be somewhat overcome using alternate test forms); (b) learning how to perform the task on its first or subsequent presentation (inflating all future performances of the task); (c) restriction of range of scores at the upper end of the outcome range (so that repeated testing produces perfect scores); (d) variability in test performance from one period of testing to the next based on some task parameter causing the test to be unstable or unreliable in ways that cannot be defined or operationalized and that are not related to exposure.

Recommendations. Tests should not be subject to the practice effects induced when the test-taker is aware of task parameters. Tests should not be subject to practice effects to the extent that improved scores approach ceiling effects. Tests should be as impervious as possible to the effects of subjects communicating with one another regarding test parameters, stimuli, or correct responses. It may be necessary to query subjects about information they had learned from other subjects about the tasks or to ask field examiners about comments from subjects indicating advance familiarity with task demands. Inclusion of instructions asking subjects not to discuss individual tests and questions may have to be added to the test protocol or emphasized in it. Test-retest reliability of task performance should be carefully considered.

Task acceptability to subjects. Some tests are acceptable or even useful based on important criteria (e.g., they are robust psychometrically or are sensitive exposure indicators) but are nonetheless noxious to subjects. This is especially true of tests involving discomfort such as nerve conduction testing. However, subjects sometimes also object to behavioral tests if they are difficult, boring, or seem irrelevant, especially if they do

not have obvious face validity as a measure of memory or other cognitive function. If the measure is an especially sensitive one, concerns of subject discomfort must sometimes be weighed against the utility of the task. Subjects can certainly be reassured that all of the tests were included for specific, well thought out reasons and completion is much appreciated. However, at other times it may be necessary to delete a task from the battery to ensure continued cooperation, particularly in longitudinal testing situations.

Recommendations. Examiners using the ATSDR screening battery tasks should provide feedback to the Agency on their impressions about task acceptability. If indicated, it may be necessary to interview subjects or to collect systematic data during field studies regarding the acceptability of each test in the battery or specific tests, especially if refusals occur during tasks or at follow-up testing in longitudinal studies.

Reliability. Test reliability must be addressed during test selection and it is best to keep error variance at a minimum. However, adherence to an artificially strict criterion for reliability would not appreciably enhance the power of statistical comparisons that use the tests and could lead to the loss of potentially important information concerning variability in test performance as an exposure outcome. In addition, criteria for adequate reliability may be somewhat dependent upon the functional domain being assessed by a task. For example, measures of vocabulary or word knowledge should be quite stable over short periods of time while measures of attention span or mood may vary appreciably. In some cases, variable performance between and within tasks may itself reveal an important exposure effect. Finally, minimization of examiner effects and of diurnal variations in performance and practice effects are important (especially in longitudinal investigations).

Recommendations. Test-retest reliability should be examined for individual tests included in the battery through review of experimental, clinical, or behavioral neurotoxicologic studies in which such reliability has been examined. If no such studies exist, but a test is included in the screening battery, test-retest reliability studies should be conducted. Extremely high levels of reliability (e.g., >0.80) need not be enforced because measures with lower reliability will not substantially reduce the power to detect significant performance differences between groups in epidemiologic studies. Criteria for adequate levels of reliability will vary depending on the nature of the function being assessed by a test and should be determined with this variability in mind. Reliability within a single testing session can be assessed by comparing performance on tests from the same behavioral or cognitive domain. Variability in performance between and within tasks should be assessed as a possible dependent measure of exposure effects.

Score distribution. To be useful in an epidemiologic battery designed to assess behavioral effects of toxic exposures, a test should have a reasonably wide distribution of possible score outcomes. If there are too few scores possible, spurious positive or negative results may be obtained (exaggeration of small differences in performance or response). If the test has too low a ceiling or too high a floor, false negative findings may reflect test parameters, obscuring actual dysfunction that would be detected on a better designed test. The range of scores for some tests can be improved by adding a speed of response measure to the traditional percent total correct scoring.

Recommendations. Scatter plots of score distributions for each test in the screening battery should be reviewed to examine floor and ceiling effects.

Miscellaneous field experiences. Idiosyncratic field conditions can result in the elimination of specific tests or adjustment of test administration in unforeseen ways. A significant portion of a group of subjects might, for example, prove untestable by the use of computers or might be unable to comprehend test instructions for a visuospatial task.

Recommendations. Field notes on test battery alterations should be systematically maintained and reviewed by ATSDR investigators or advisors. Such historical information may allow investigators to anticipate special field testing demands in specific subject groups prior to beginning field assessment or at least to approach a subject group with an anticipatory flexible assessment plan.

VALIDITY

Specificity Versus Sensitivity

In examining the relationship between tests to be validated and external validating criteria, one must consider the importance of selecting both broadly sensitive tests and tests that are indicative only of specific nervous system dysfunctions. Tests sensitive to a wide variety of validating variables are likely to be useful in determining the presence or absence of a neurobehavioral effect in response to a variety of toxic substances. Performance on a commonly used test such as the Benton Visual Retention Test, for example, reflects several independent cognitive functions, including attention (not memory), planning ability, and visuoperceptual ability (8,64,65,80). Because such tests are representative of multiple underlying cognitive functions, they are broadly sensitive to diverse types of CNS impairment and are thus likely to be validated with respect to a wide variety of criteria.

Recommendations. Sensitivity to the effects of a wide variety of validating criteria is one important feature establishing validity, but specificity—sensitivity to a specific set of validating criteria—may be equally or more important than broad sensitivity in some studies.

Criteria for Establishing Validity of Measures

In behavioral neurotoxicologic field studies, validity of individual tests in a screening battery may be verified by assessing the relationship between test performance and a number of outside criterion measures or variables. These include sensitivity to neurotoxicant exposure, CNS and PNS sensitivity, native intellectual capacity or "mental abilities," specificity of cognitive function, clinical signs and symptoms, diagnosis, patterns of performance among battery tasks, disabilities, and outcome in follow-up studies.

Neurotoxicant Sensitivity

A major goal of studies in behavioral neurotoxicology is to identify tests that are sensitive to exposure to neurotoxicants. This may include neurotoxicants as a general class and/or specific neurotoxicants, specific parameters of exposure such as dosage or chronicity, or the interrelationships between a subject variable (such as age at exposure), neurotoxicant exposure(s) and neurobehavioral outcome measures. For some tests it may be possible to develop expected levels of sensitivity, i.e., minimal differences between exposed groups and controls which confirm dysfunction suggestive of neurotoxic effects in the exposed groups. However, no rigid a priori standards for sensitivity are possible or appropriate at this time.

Recommendations. Each task in the screening battery

should be evaluated each time it is used to assess exposure groups and across investigations in terms of its sensitivity to exposure parameters (type of neurotoxicant, dosage—including dose-effect relationships, chronicity, and other pertinent exposure variables). If a task has no demonstrated sensitivity in repeated studies, a decision to eliminate the task may be necessary unless the task is considered to be valuable as a "control" measure known to be impervious to neurotoxicant effects or as a "marker" measure for some other variable such as premorbid intelligence.

Specificity as well as sensitivity should be examined when exploring the relationships between exposure variables and test performance.

Sensitivity to CNS and PNS dysfunction. The assessment of nervous system dysfunction secondary to exposure to neurotoxicants is most accurately and efficiently accomplished through the use of tests which are known to be valid as indicators of brain or CNS and peripheral nervous system (PNS) function.

Many tests are now available that measure some aspect of behavioral or cognitive function and have been validated as measures of CNS function. Such validation requires the administration of tests to subjects with known brain damage, preferably with specific types of neuropathological or neurochemical abnormality (e.g., demyelination in white matter secondary to multiple sclerosis, neurofibrillary plaques and tangles secondary to Alzheimer's disease, neurotransmitter deficits in Parkinson's disease or Huntington's disease) or with specific sites of cerebral damage (e.g., focal stroke involving the mesial aspects of the left temporal lobe or the right thalamus, surgical removal of a structure such as the amygdala). It may be argued that CNS validation of a test requires investigations in which in-depth evaluations completed by a neuropsychologist, behavioral neurologist and psychiatrist must be combined with imaging studies to verify the criterion neurological diagnosis being used in the study. When a behavioral test with known CNS validity is used in a carefully controlled study of subjects exposed to a known or suspected neurotoxicant, it is usually reasonable to conclude that the impairment in test performances reflects CNS damage secondary to exposure. If, however, toxicant-related impairment is seen in a behavioral test which has not been validated on subjects with known CNS dysfunction, one can safely conclude only that a specific test performance is affected by exposure or (by generalization) that the function measured by the test is affected. If the function measured by the test is similar to a known functional effect of CNS damage, one could reason by analogy that the impairment reflected a neurotoxic effect of exposure. However, this requires a greater logical leap than the same conclusion based on a CNS-validated test (71). A number of tests currently advocated for use in behavioral neurotoxicologic batteries, including the computerized versions of common neuropsychological tasks, have this drawback, though others such as the World Health Organization extended neurobehavioral battery (33) do not (70,75). Some group members considered this drawback to be more limiting than other group members.

A more difficult topic is the identification of appropriate neurologic disorders to be used in CNS validity studies for selected tests. Traumatic brain injury (TBI) is a possible model for neurotoxic effects because attention and executive deficits are common to both TBI and toxic encephalopathies. The utility of models of brain damage in which the neuropathology is similar to that seen in specific types of toxic encephalopathy is another option. Multiple sclerosis is a demyelinating disease of the white matter, which seems to be affected in some solvent exposures such as toluene (59,60), and Parkinson's disease affects the basal ganglia, as do carbon disulfide (54) and carbon monoxide (16).

Tests used to measure PNS function must have independent validation as accurate measures of the specific aspect of PNS functioning being addressed. PNS test measures and their validation are discussed fully in Section VI.

Recommendations. In order to be retained as measures in standard batteries used in ATSDR field studies, tasks that have not been validated as sensitive CNS measures can: (a) be independently validated through separate studies; (b) reviewed periodically to determine if studies carried out by independent investigators have accomplished such validation. CNS validation can be legitimately addressed by establishing relationships between test performance and results of neuropathologic studies on brain tissue, neurologic diagnosis, neuroimaging scans (computerized tomography, magnetic resonance imaging), neurophysiologic and neurometabolic studies (electroencephalogram [EEG], computerized EEG, functional imaging and functional EEG, positron emission tomography, cerebral blood flow, evoked potentials), neuroendocrine or neurochemical assays, or assessing dose-effect relationships following controlled administration of pharmacological agents (drugs and medications) which have known neuropathological or neurochemical effects.

Native intellectual capacity or mental abilities. Some tests are included in screening batteries as control measures. The rationale for inclusion may include that of a "negative control," i.e., a test which is not predicted to be sensitive to exposure to neurotoxicants. However, at other times such measures are included as indicators of premorbid or native abilities. In these cases, scores from the control tests may be used as covariates in data analysis. When completing environmental studies, it is important to be very careful in choosing such measures for at least two reasons raised by workshop members. First, the control measure may actually be an outcome measure. There is particular danger of this occurring when exposure occurred during childhood (even if testing is carried out in adulthood). Thus, many cognitive functions, including language acquisition and vocabulary knowledge, have been associated with developmental exposure to well-described toxicants such as lead; therefore, it is difficult to identify a "safe" indicator of native intellectual capacity in subjects with childhood exposure. The second reason is that the appropriate test might not be the same for all subcultural groups. Vocabulary might be applicable to one but motor function or visuoperceptual skill to another.

Recommendations. Control tests should be validated in appropriate manner specific to their use in any particular study. Sources of validation might include estimates for children based on parental performance, premorbid (i.e., preexposure) school records or test performance, income, vocational status or success, or social position.

Specificity of cognitive function. The specificity of the tests that measure cognitive processing is important in characterizing the functional effects of exposure and in hypothesizing CNS mechanisms and sites of exposure. Factor analytic studies often do not produce the same dimensions of cognitive functioning as are described in cognitive research and repeatedly supported by neuropsychological studies. This implies that factor analytic studies of test batteries and complex omnibus tests (such as WAIS or Stanford-Binet) should be interpreted cautiously.

Recommendations. When evaluating screening battery tasks

for continued retention in the battery, it is recommended that the criterion of functional specificity be evaluated by reviewing the literature (especially in neuropsychology, cognitive neuroscience) to determine whether cognitive specificity has been established in other research and to examine the relationship between performance on different tasks or components of tasks in the screening battery which are thought to measure the same cognitive processing capacity.

Clinical signs and symptoms. The workshop group, though not specifically charged with assessment of subject-perceived symptoms of intoxication, was concerned that such self-report symptoms be carefully elicited to gain cues to behavioral domains of special interest in particular exposed groups and to understand the relationship between test performance and symptoms. An additional concern is the possible omission of executive system symptoms from both formal assessment and self-report measures.

Recommendations. Ongoing evaluation of tests in the screening battery should include evaluation of the relationship between test performance and self-perceived symptoms. If symptoms far outweigh test findings, inspection of symptom clusters may lead to identification of susceptible behavioral domains for which assessment techniques should be included in a tailored or general test battery.

Diagnosis. In behavioral neurotoxicology field studies, information may be available about the relationship between performance on screening battery tests and diagnosis of exposure-related illness. Of particular interest would be CNS or PNS disease, though metabolic, genetic, or endocrinologic disorders affecting behavior are also of potential interest.

Recommendations. Tests reviewed for continued inclusion in a screening battery should be evaluated with regard to diagnostic validity as predictors of neurotoxic illness.

Patterns of interrelationships of test performance. Component tests of the screening battery may cluster together to form patterns of deficit within a cognitive process, cognitive domain or set of processes and domains. These patterns may be important in predicting functional deficit and/or sites of CNS effects of neurotoxicants.

Recommendations. When tests are reviewed in terms of inclusion criteria, patterns of test performance should be examined.

Disabilities. Dysfunction on behavioral tests often predicts certain types of disability in daily living. These may be academic or vocational or they may relate to personal, social, and familial adjustment.

Recommendations. When tests are reviewed with regard to continued inclusion in test batteries, the validation criterion of prediction to daily life disabilities at work, in school, or in personal adjustment should be considered.

Outcome in follow-up studies. Data collected in ongoing field studies can provide useful predictors of findings available only in follow-up studies. For example, children with nervous system function secondary to environmental exposure might show specific types of test deficits that are related to a later manifestation of poor occupational performance and lowered earnings in adulthood (or, similarly, intact performance on some tests might predict to a positive occupational outcome) (70). Another example would be a relationship between a specific pattern of test performance in exposed adults and later development of a progressive dementing disorder.

Recommendations. When follow-up studies are completed on subjects who have previously undergone CNS or PNS testing, studies of the relationship between performance on tests and relevant outcomes should be undertaken.

II. ONGOING EVALUATION OF COMPREHENSIVENESS OF TEST BATTERIES

Review of test batteries to make decisions about retention of existing tests or inclusion of new tasks should consider the comprehensiveness of the overall battery. The use of comprehensive test batteries in epidemiological investigations has a number of distinct advantages. For example, it is often possible for the neuropsychologist associated with a study to develop hypotheses about functional cerebral localization based on the pattern of positive and negative results. It is also often possible to develop hypotheses about deficits in psychological processes associated with exposure to specific neurotoxicants (71). In addition, such testing allows the investigator to discover new functional deficits that may go undetected in highly focussed batteries. Comprehensive testing of behavioral domains is appropriate when invoking principles for testing the full developmental age range which may be seen in environmental studies. Children, for example, often show a much wider range of functional deficits following toxic exposures than do adults (70,76), and it is likely that geriatric exposure may produce different deficits than exposure in younger adults (56). Finally, comprehensive testing can sometimes allow the investigator to compare test scores earned by a subject to a model of likely premorbid abilities. This allows investigation of individual deficit patterns which may be obscured in analyses of group data. For example, a subject may earn superior scores on visuospatial and vocabulary tests but low average scores on tests of learning and memory (in group comparisons the memory scores would not appear as impaired, but the subject's performance differential clearly reflects a relative memory problem). The advantages of comprehensive testing are particularly salient when little is known about the functional or neuropathological effects of the toxicant exposure(s) under investigation. Comprehensive testing may be less appropriate for studies involving well-investigated toxicants.

The behavioral domains in Table 1 are essential for inclusion in comprehensive behavioral sampling. These domains are similar to those published by other groups in the field of behavioral neurotoxicology and neuropsychology (33,39,74). Within each domain, the specific cognitive processes that are starred reflect a consensus of the authors as especially essential to assess.

Recommendations. When reviewing test batteries with regard to the criterion of comprehensiveness, a determination must be made as to whether an individual test: (a) materially contributes to the comprehensiveness of the battery; (b) offers an accurate measure of the functional domain which the test is thought to address.

If there is a question about a test (on the basis of its psychometric properties or apparent insensitivity to exposure or other important independent variables) but its inclusion is thought to have some value or to be necessary for other methodological reasons, an additional test from the same cognitive domain can be added to the battery to increase the sensitivity of the battery.

In addition to the accumulated experience of the Agency from field studies, the experience of other researchers in the field of behavioral neurotoxicology and neuropsychology should be considered when reviewing tests to be used in each cognitive domain.

TAILORED OR FOCUSED BATTERIES

The situations under which tailored or focused batteries could contribute important information to field studies were

TABLE 1
ESSENTIAL BEHAVIORAL DOMAINS

- Attentional Functioning
  - Span*
  - Divided attention*
  - Sustained attention*
  - Focussing of attention
- Executive Functioning
  - Flexibility and shifting
  - Initiation
  - Planning
  - Monitoring
  - Evaluation
- Memory
  - Anterograde, material specific memory
    - Verbal*: Learning and retention, with measurement of delayed recall and recognition
    - Visual-nonverbal*
  - Retrograde memory
  - Skill learning (Procedural memory)
  - Priming
- Language* (for children)
  - Word retrieval
  - Syntactic comprehension
  - Word comprehension
  - Reading comprehension
  - Writing/Spelling
- Conceptual functioning
- Level of premorbid abilities
- Nonverbal/visuospatial skills
  - Complex visual perception
  - Visual organization*
  - Constructional ability
  - Right-left orientation
  - Spatial orientation
- Complex auditory functions
- Somatosensory functions
- Speed of information processing*
  - Simple reaction time
  - Choice reaction time
- Mood
  - States*
  - Range*
- Sensory abilities
- Motor*
- Personality

considered by the workshop in some detail, and recommendations for specific types of test batteries were offered.

Known Neurotoxicants

When the effects of a neurotoxicant or group of neurotoxicants have been well described for the environmental exposure situation, it may be most efficient to employ a battery of known sensitive tests and control tasks rather than a general unfocused screening battery. However, the investigator must be careful in designing the battery if the circumstances and variables affecting exposure effects in prior studies are in any significant way different from those occurring in the case group under investigation. The most obvious of these is developmental stage at the time of exposure. Focused batteries for children should not be based exclusively on findings in adults. One also has to consider exposure variables when designing a focused battery. Exposure variables such as dosage and chronicity of exposure, which may lead to different behavioral outcomes, should also be considered when using prior studies to design a focused battery. Whereas batteries for environmental studies are often based on occupational study methods, this must be done with caution because the behavioral deficits associated with chronic low dose exposure (common in environmental settings) may cluster in different cognitive domains than those associated with the higher dose, acute or intermittent chronic exposure seen in occupational settings.

Whereas the existence of a complete description of the nervous system sequelae of exposure to any neurotoxicant is doubtful, lead represents the toxic substance about which the greatest knowledge has accumulated. Studies completed to date on lead exposure in adults have varied in findings concerning the lowest dosage at which nervous system effects can be detected through the use of CNS and PNS tests (74). However, when effects are detected they are generally confined to the domains of executive function, visuospatial skills, memory, motor function, and mood (4,5,29,79). A focused battery assessing adults with lead exposure would sample these domains heavily, using tests which have been shown to be sensitive in prior investigations. However, such a tailored assessment would be overly restricted if one were assessing children with lead exposure or adults whose history of exposure included exposure during childhood. The many studies of childhood lead exposure suggest that the functional sequelae of such exposure are widespread, extending in some cases to basic intellectual capacity, language/verbal skills, and personality (7,51,52,70). Therefore, a focused battery might be inappropriate unless combined with a comprehensive standard battery.

Potential Neurotoxic Substances

Given the large number of man-made substances with high potential for neurotoxicity, it is likely that there are many substances which are in fact, neurotoxic but have not yet been so identified. When a group of subjects is identified with exposure to substances which are potentially neurotoxic, addition of a focused battery to the standard battery can be helpful. In such situations, focussed batteries can be used to explore a wider range of possible subtle behavioral effects, to use tests which are exquisitely sensitive to subtle deficits (e.g., Paced Auditory Serial Addition Test), or to compare the effects of the new substance to known effects of similar substances.

Explication of Brain-Behavior Relationships

Because neuropsychological tests have been used extensively in the study of patients with brain damage, a great deal is known about the relationship between impaired performance on neuropsychological tasks and focal or diffuse brain damage. Patients with damage to different parts of the brain show different patterns of impaired performance on neuropsychological tests (13,31,39,42,58,68). Knowledge of these functional deficit patterns allows the neuropsychologist to render inferences about sites of neuropathological damage following trauma to the brain (e.g., stroke, tumor, infection).

Such inferential reasoning also allows behavioral neurotoxicologists to develop hypotheses about cerebral sites of damage caused by toxic exposure.

For example, there is neuropathological evidence linking mercury exposure to brain damage involving the white matter (73), the occipital lobes, the brain stem and the cortex diffusely (6). Such evidence allows the investigator to tailor a battery of tests investigating mercury exposure which would include tasks assessing executive, visuospatial, and motor function. These tasks are sensitive to damage to the aforementioned areas of the brain. Tin is another example of a toxin for which brain-behavior relationships could be further explored neuropsychologically. The neuropathology of tin exposure has been found to involve the hippocampal formation (9,20), a structure which in humans is involved in the mediation of memory function. Thus, studies of subjects with tin exposure could be tailored to emphasize memory as a functional domain. Carbon disulfide is known both from neurological and behavioral studies to affect the basal ganglia of the brain, producing a Parkinsonian syndrome among workers with chronic exposure (43,54,55). The expected behavioral deficits associated with basal ganglia dysfunction have been documented in a number of studies (motor, visuospatial, reasoning, new learning, mood) (69). A tailored battery designed for a carbon disulfide-exposed group might include focused tests which are known to be sensitive to Parkinsonian syndromes. Organic solvents are a complicated issue and the neuropathological effects of solvents may vary enormously depending on type of solvent, exposure variables, and individual susceptibilities. However, the occurrence of frontal lobe atrophy following chronic solvent exposure has been documented in neuroimaging studies (47). These findings suggest the importance of including tests of executive function in a focused solvent battery. A final example is that of the solvent toluene. Exposure to toluene among glue sniffers has been found to be associated with white matter abnormalities on magnetic resonance imaging (MRI) scan (21,59). Although this finding awaits confirmation, the inclusion of tests sensitive to white matter pathology is potentially informative in a study of subjects exposed to toluene, perhaps with adjunctive MRI testing.

Cognitive Processing Deficits

Hypotheses regarding underlying cognitive processing deficits occurring as a result of exposure to neurotoxicants can also be explored using primary or adjunctive tailored test batteries in field settings. For example, the finding of basal ganglia dysfunction following carbon disulfide exposure suggests that functions which appear to be at least partially mediated by neurotransmitters such as dopamine may be affected by CS^2 exposure. These functions (which may underlie the ability to carry out other types of cognitive processing) include underarousal and cognitive rigidity, both of which could be specifically addressed in a tailored battery. Dysphoria and other negative affective responses (e.g., fatigue, irritability) are also known to be associated with basal ganglia syndromes (22,36, 45,48,63,69,78) and could be specifically explored in a focused battery.

When studying both children and adults, focused tests can also be used to explore cognitive domains that are now thought to be unaffected by neurotoxic exposure in the absence of anoxia. For example, in studies of adults, language is rarely found to be affected on standard test batteries (1,74). Although there is some evidence to suggest that word list generation may be affected by exposure to solvents, this is most likely secondary to an executive deficit not to primary language dysfunction. Likewise, vocabulary recognition appears to be unaffected, though definitions are sometimes found to be rather concrete. Certainly adults with toxic exposure are not clinically noted to be aphasic (76). However, subtle language deficits (e.g., slowing of retrieval on naming tests, impairment in comprehension of complex communications) have not been carefully explored.

Detailed Assessment of Subject Variables

In individual groups of people with environmental exposures, exploration of highly specific subject variables may be crucial to interpreting neurobehavioral test data. In such circumstances, focused batteries may be used. Specific issues that can be addressed in tailored batteries of tests include motivational factors (including malingering), preexisting learning disabilities, preexisting psychiatric disorder, exposure-related stress conditions, or effects of familial neurologic disorders. There is concern among some authors that psychiatric and social factors might affect performance on cognitive tests so that impaired performance might actually reflect psychosocial factors, not exposure. These factors include concern and fear among subjects about health outcome or desire for compensation for real or imagined adverse health effects.

Special concern was voiced about the subject of subcultural group membership. In some exposure settings or with some subgroups of exposed populations, specialized testing may be necessary to accurately assess cognitive domains and behavioral dysfunction. In addition, it may be necessary to carefully select control groups when subcultural membership is an issue.

Community Concerns About Diagnosable Disease

In some exposed populations, there may be specific concerns among exposed individuals about the development of diagnosable disease entities. Thus, a population with exposure to carbon disulfide might be concerned about the occurrence of a Parkinsonian syndrome or a population of persons with aluminum exposure might be concerned about the possibility of developing Alzheimer's disease. Any exposed group might voice concern about exposure-induced learning disabilities or permanent brain damage in children, or about teratogenic effects of exposure. In these communities, it is important to confirm or clarify the nature of health concerns and to assess potential clusters and other small population effects. An extensive neurological, medical and neuropsychological evaluation may be appropriate for at least some subjects.

Recommendations for Implementation and Review

It is recommended that four levels of test battery be used for field studies: (a) Core screening batteries for broad cross sectional studies; (b) Focused batteries which include clusters of tests to address issues listed above (viz., known neurotoxicants, potential neurotoxicants, brain-behavior relationships, cognitive processing deficits). These tests could be given to all subjects or a subset of subjects in addition to a core screening battery. Examples of focused batteries are provided (see Appendix). Quality control batteries including tests designed to further examine population variables that may affect test results in a specific subject group to be administered to a subset of subjects (see Appendix).

Full neurological and neuropsychological assessment to be administered to a subset of subjects when diagnosis is an issue or the exposure is unusual enough to warrant in-depth evalua-

tion of some subjects (see Appendix). Special test batteries would be subject to the same general criteria for ongoing review as the core screening battery (see Section I).

IV. INCORPORATION OF NEW METHODS IN TEST BATTERIES

New testing methods that measure behavioral and PNS function are constantly being developed and validated (see Appendix). Inclusion of these measures in standard test batteries used in field studies can add important new information to study findings and can serve as a means of exploring new instruments potentially useful in epidemiologic assessment.

The use of new testing measures and the innovative combination of behavioral test measures with other innovative assessment techniques in neurology, neurogenetics, neurochemistry, and neuroimaging were enthusiastically endorsed by workshop participants. Innovative testing strategies such as the use of videotaping were also discussed.

Recommendations. To identify new tests for possible inclusion in screening or tailored batteries, the following steps were recommended:

Identify relevant functional domains for which new assessment techniques are needed and review the literature on that domain. Follow the literature on neuropsychological and PNS assessment and related fields. Maintain communication between investigators carrying out human studies and those engaged in animal research explicating models of neurotoxicity to identify important functional and CNS endpoints to be assessed with appropriate tests. Some participants recommended that ATSDR convene a workshop of these groups of investigators to stimulate interaction and development of test strategies. Promote the development of tests that will facilitate early identification of neurodegenerative effects of exposure. Consider the concomitant use of neuroimaging, neurophysiologic, neuroendocrine, and neurogenetic measures.

New tests should be added under the following circumstances: A previously untapped domain is determined to be important in the core screening battery. A new test from a particular behavioral domain proves through pilot studies to better fulfill the criteria for test inclusion than does an existing battery test. Signs and symptoms are suggestive of neurotoxic illness, but the core screening battery tests are negative.

New tests can be piloted with a fading-in technique in which many subjects receive the new test in addition to the current core battery. If the new test proves to better fulfill the inclusion criteria than the existing test, it should replace the old test (unless the test is to be retained for some other reason, in which case the new test should be added). The continued use of functionally nonspecific tests in the core battery should be reviewed carefully, particularly if there is more than one such task in the battery or it is time-consuming.

V. REVIEW PERSONNEL AND MECHANISMS

Recommendations for test battery review procedures.

It was recommended by workshop participants that a committee be assembled for ongoing review of the core and tailored batteries. Some authors strongly favored inclusion of a community member, but others strongly opposed their inclusion at this level of review. However, the following areas of expertise were listed by workshop members as essential for inclusion on the committee. Note that a single committee member could embody one or several of these areas of expertise:

1. Behavioral neurotoxicology
2. Neuropsychology
3. Neurologic assessment (central and peripheral)
4. Toxicology
5. Epidemiology
6. Biostatistics
7. Psychometrics
8. Neurodevelopment
9. Behavioral/neuropsychological field testing
10. Behavioral neurology

Core screening battery tests should be reviewed at least once a year by the expert committee and consultants. When problems are identified during field testing, relevant committee members should be consulted as needed. It may be important to have an expert experienced in field testing present at the outset of studies to iron out procedural difficulties and assess special aspects of the group under study. Tailored batteries and new tests should be reviewed by relevant experts before and after each study in which they are used.

Statistical analysis issues were discussed and the following general thoughts were voiced:

1. Conclusions about study outcomes should not be based on rigid p values.
2. Because environmental studies often involve subtle effects and small numbers of subjects, a flexible approach to statistical analysis is necessary.
3. Positive but statistically nonsignificant results should not necessarily be interpreted as reflecting an absence of neurotoxic effects.
4. Development of meta-analytic methods for use in analyzing data from more than one exposure site is recommended.
5. Small magnitude effects in environmental studies may be important.
6. Use of nonparametric techniques may be appropriate.
7. Confounders must be carefully considered. Some variables thought to be confounders may actually be outcome measures (e.g., low education level in a population that suffered childhood exposure).
8. Methodology for analysis of clusters of cases may be needed.
9. The absence of significant findings may reflect the limitations of the measures used or the fact that variance in performance (not central tendency) is the manifestation of exposure.
10. Statistical analysis should emphasize statistical power over statistical probability.

VI. PERIPHERAL NERVOUS SYSTEM BATTERY

A battery for assessing peripheral nervous system (PNS) function was seen as an essential component of any comprehensive assessment. The following provides a discussion of PNS responses to toxic insult, a review of currently available methods for its assessment, and a discussion of validation of these methods. This serves as a supplemental information relevant to both workgroups (1,3).

Introduction to Assessment of Peripheral Nervous System Function

The most commonly used methods for objective assessment of the sensorimotor components of PNS function (as opposed to autonomic function) are the neurological physical examination and electrophysiologic evaluation of nerve con-

duction (nerve conduction velocity and amplitude measurement) and motor activity (electromyography). These methods have utility in the clinical and laboratory setting for evaluation of individual patients or subjects. However, they are of limited utility for use in field-based epidemiologic studies of the effects on populations of exposure to neurotoxicants in the environment. Methods of assessing the integrity of the PNS that have greatest utility in environmental epidemiology are those that provide outcomes that are objective, quantitative, reliable, sensitive, standardized, and nonaversive, can be administered rapidly with portable, inexpensive equipment by minimally trained testers, and have readily available normative data.

Nerve Conduction Velocity

Nerve conduction velocity measurement is a physiological procedure that allows measurement of nerve conduction velocity as well as compound action potential amplitude. In combination with needle electromyography it can provide extensive information about the location of a peripheral nerve lesion as well as the specific pathological process (i.e., generalized axonal neuropathy, demyelination, focal compression). Normative data are available (35) and the reliability of the measure is considered to be good, although it has not been studied extensively. Nerve conduction velocity measurement has been performed in field and laboratory studies of the effects of exposure to neurotoxicants (62) as well as in the study of virtually all large fiber neuropathies of nontoxic origin. Because it is a physiological measure, it is not subject to motivational factors as are behavioral measures. Nerve conduction velocity measurement provides information about the function of large myelinated nerve fibers only, however.

Vibrotactile Thresholds

The most well established behavioral measure of peripheral somatosensory function is measurement of cutaneous vibrotactile thresholds. The method involves presentation of cutaneous vibration stimulation to a subject who must make a behavioral response to it. Studies of the reliability of a variety of testing protocols in different populations of interest are available in the literature. The best test-retest correlation coefficients for the method are around 0.9 (24,25). In addition to reliability, the effects of age and other covariates have been studied for this method (24). Vibrotactile thresholds are best correlated with electrophysiologic measures that assess long nerve segments as opposed to short ones with correlations approaching 0.7 (26). Vibrotactile thresholds have been used in field studies of a variety of groups, including organophosphate-poisoned agricultural workers (46), acrylamide-exposed workers (3), and workers experiencing sensorineural symptoms related to use of vibrating power tools (15). Testing equipment is relatively inexpensive, easy to use and transportable.

Thermal Thresholds

Measurement of thermal thresholds is less well studied than is measurement of vibrotactile thresholds. It requires more sophisticated equipment than measurement of vibrotactile thresholds and most testing protocols require more time than does vibrotactile threshold testing. Arezzo et al. (2) present coefficients of variation for repeated measures with a forced-choice technique that range from 8.3% to 47.1%. The mean coefficient of variation for 10 subjects was 19.0% and 26.6% for the upper and lower extremities, respectively.

Thermal thresholds were used in a study of construction painters exposed to organic solvents (14). Analyses indicated some association between exposure and thermal threshold. Significantly elevated thermal thresholds were found in a group of symptomatic vibration exposed workers when compared to unexposed asymptomatic controls (19). Thermal thresholds have also been used to study diabetic small fiber neuropathy (17,38) and neuropathy associated with uremia (41). In summary, a growing body of evidence suggests that thermal thresholds are useful for detection of small fiber dysfunction that might not be measurable with other techniques. Field use is currently somewhat limited by existing instruments that are expensive and barely portable.

Grip Strength Dynamometry

Measurement of motor strength with simple mechanical devices is performed routinely in virtually all physical medicine and rehabilitation programs to assess both degree of impairment and progress of patients. Well established measurement technique and extensive normative data are available. The reliability of the measure has been studied in both children and adults and ranges from good to excellent (44,66). Comparisons with other measures of neurologic function are not available and grip strength dynamometry has rarely if ever been used in a study of the effects of occupational or environmental exposure to neurotoxicants. The method deserves further exploration, as it is rapid, reliable, quantitative, acceptable to subjects, requires inexpensive equipment, and is the only nonaversive measure of motor function currently available.

Postural Stability

Portable equipment has recently become available for the quantitative assessment of postural or standing stability. Assessment of postural stability is especially attractive because of the simplicity of the test protocol. Subjects are instructed to stand upright and remain steady to the best of their ability. Currently, a variety of testing protocols and measurement systems are in use. The average coefficient of variation of repeated measurement with one system ranged from 39% to 42% but did not show systematic change over the course of the repeated measures (12). Normative data are available for at least one test protocol (12). Significant changes in postural stability have been found among styrene-exposed workers (49), patients with solvent related psycho-organic syndrome (37) and in children exposed to lead (11). Experimental studies utilizing administration of ethanol have produced equivocal results (10).

Heart Rate Variability

This technique has been used for detection of autonomic neuropathy primarily in diabetics (53). The equipment required to perform this test is portable and relatively inexpensive. It has also been used in studies of vibration exposed workers (27) and a recent study of lead-exposed workers (50). A standard protocol, consensus summary measures, and normative data are not currently available.

Recommendations

Measures recommended for immediate inclusion in a screening battery[2]:

[2] Measures of vibrotactile threshold in the finger and hand grip strength also were included in the recommendations of Workgroup 1 (1).

1. Vibrotactile threshold measurement; great toe bilaterally
2. Grip and pinch strength dynamometry; upper extremity, bilaterally
3. Thermal threshold measurement; one lower extremity site

These measures are recommended because they (a) target both large and small nerve fiber function, (b) are well validated in studies of groups exposed to neurotoxicants or those at risk of peripheral neuropathy of metabolic origin such as diabetes mellitus, and (c) testing equipment and test protocols are currently available.

Measures recommended for priority development:

1. Heart rate variability
2. Postural stability

Heart rate variability measurement targets a PNS subsystem (the autonomic nervous) system not accessible to measurement with any other testing method. If measurement equipment and a consensus test protocol were currently available, it would be recommended for inclusion at this time. Postural stability may be a sensitive integrated measure of both central and peripheral neurologic functions required for maintenance of standing stability. These measures are considered priority areas for development.

Measures recommended for consideration in tailored batteries:

1. Functional or behavioral measures
   a. Odor identification
   b. Tremor measurement
   c. Color vision testing
   d. Saccade eye movements
2. Electrophysiologic measures
   a. Nerve conduction velocity (NCV)
   b. Quantitative electroencephalography
   c. Evoked potentials

A variety of circumstances preclude recommendation of these measures for inclusion in the screening battery at this time. These measures either involve aversive stimulation (NCV), are unproven in the field setting (evoked potentials, quantitative EEG), or are inadequately tested for sensitivity to neurotoxicant exposure.

VII. MISCELLANEOUS CONSIDERATIONS

We also express considerable concern on four issues:

1. Appropriate selection of tests for distinct subcultural groups;
2. Quality control in test administration;
3. Appropriate interview/questionnaire assessment of symptoms;
4. Combining the use of behavioral with other neurological, genetic, and medical testing.

ACKNOWLEDGEMENTS

This article is based on the deliberations of Workgroup 3 from the Workshop on Neurobehavioral Testing, convened by the ATSDR in Atlanta, Georgia, on September 11-13, 1991. The Workgroup was composed of nine individuals, who were selected based on their training, background, and experience in fields of science that bear directly on ATSDR's charge to the groups.

The article is a consensus product of the authors as members of a working group. The article therefore does not necessarily represent the opinions of the individual authors nor the agencies or institutions by which they are employed, and it does not necessarily represent the policy of the sponsoring organization, the Agency For Toxic Substances & Disease Registry.

REFERENCES

1. Anger, W. K. Worksite behavioral research—Results, sensitive methods, test batteries and the transition from laboratory data to human health. NeuroTox. 11:629–72; 1990.
2. Arezzo, J. C.; Schaumburg, H. H.; Laudadio, C. Thermal sensitivity tester: Device for quantitative assessment of thermal sense in diabetic neuropathy. Diabetes 35:590–592; 1986.
3. Arezzo, J. C.; Schaumburg, H. H.; Petersen, C. A. Rapid screening for peripheral neuropathy: A field study with the Optacon. Neurology 33:626–629; 1983.
4. Baker, E. L.; Feldman, R. G.; White, R. F.; et al. Occupational lead neurotoxicity, a behavioral and electrophysiologic evaluation. I. Study design and year one results. Br. J. Industr. Med. 41:352–361; 1984.
5. Baker, E. L.; White, R. F.; Pothier, L. J. et al. Occupational lead neurotoxicity. II. Improvement in behavioral effects following exposure reduction. Br. J. Industr. Med. 42:507–516; 1985.
6. Battigelli, M. C. Mercury toxicity from industrial exposure. J. Occ. Med. 394–399; 1960.
7. Bellinger, D. et al. Longitudinal analyses of prenatal and postnatal lead exposure and early cognitive development. New Eng. J. Med. 316:1037–43; 1987.
8. Benton, A. L. Constructional praxia the minor hemisphere. Confina Neurologica (Basel). 29:1–16; 1967.
9. Besser, R.; Kramer, G.; Thumler, R. et al. Acute trimethyltin limbic cerebellar syndrome. Neurology 37:945–950; 1987.
10. Bhattacharya, A.; Morgan, R.; Shukla, R.; Ramakrishanan, H. K.; Wang, L. Noninvasive estimation of afferent inputs for postural stability under low levels of alcohol. Ann. Biomed. Engineer 15:533–550; 1987.
11. Bhattacharya, A.; Shukla, R.; Bornschein, R.; Dietrich, K.; Kope, J. E. Postural disequilibrium quantification in children with chronic lead exposure: A pilot study. NeuroTox. 9:327–340; 1988.
12. Black, F. O.; Wall, C.; Rockette, H. E.; Kitch, R. Normal subject postural sway during the Romberg test. Am. J. Otolaryngol. 3: 309–318; 1982.
13. Boller, F.; Grafman, J.; Rizzolatti, G.; Goodglass, H., eds. Handbook of neuropsychology. New York: Elsevier; 1988.
14. Bove, F. J.; Letz, R.; Baker, E. L. Sensory thresholds among construction trade painters: A cross-sectional study using new methods for measuring temperature and vibration sensitivity. J. Occ. Med. 31:320–325; 1989.
15. Cherniack, M. G.; Letz, R.; Gerr, F.; Brammer, A.; Pace, P. Detailed clinical assessment of neurological function in symptomatic shipyard workers. Brit. J. Industr. Med. 47:566–572; 1990.
16. Choi, I. S. Delayed neurologic sequelae in carbon monoxide intoxication. Arch. Neurol. 40:433–435; 1983.
17. Dyck, P. J.; Bushek, W.; Spring, E. M.; et al. Vibratory and cooling detection thresholds compared with other tests in diagnosing and staging diabetic neuropathy. Diabetes Care 10:432–440; 1987.
18. Echeverria, D.; White, R. F.; Sampao, C. A neurobehavioral evaluation of PCE exposure in patients and dry cleaners: A possible relationship between clinical and preclinical effects. Paper presented at 8th International Neurotoxicology Conference, Little Rock, AR; October 1–4; 1990.
19. Ekenvall, L.; Nilsson, B. Y.; Gustavsson, P. Temperature and vibration thresholds in vibration syndrome. Br. J. Industr. Med. 43:825–829; 1986.
20. Feldman, R. G.; White, R. F.; Eriator, I. Trimethyltin encephalopathy. Arch. Neurol. 50:1320–1324; 1993.
21. Filley, C. M.; Heaton, R. K.; Rosenberg, N. L. White matter dementia in chronic toluene abuse. Neurology 40:532–534; 1990.
22. Folstein, S. E.; Abbott, M. H.; Chase, G. A. et al. The association of affective disorders in a case series and in families. Psychol. Med. 13:45–52; 1983.

23. Fowler, C. J.; Carroll, M. B.; Burns, D.; et al. A portable system for measuring cutaneous thresholds for warming and cooling. J. Neurol. Neurosurg. Psychiatry 50:1211–1215; 1987.
24. Gerr, F.; Hershman, D.; Letz, R. Vibrotactile threshold measurement for detecting neurotoxicity: Reliability and determination of age- and height-standardized normative values. Arch. Environ. Health 45:148–154; 1990.
25. Gerr, F. E.; Letz, R. Reliability of a widely used test of peripheral cutaneous vibration sensitivity and a comparison of two testing protocols. Br. J. Industr. Med. 45:635–639; 1988.
26. Gerr, F.; Letz, R.; Hershman, D.; Farraye, J.; Simpson, D. Comparison of vibrotactile thresholds with physical examination and electrophysiological assessment. Muscle and Nerve 14:1059–1066; 1991.
27. Harada, N.; Kondo, H.; Kimura, K. Assessment of autonomic nervous function in patients with vibration syndrome using heart rate variation and plasma cyclic nucleotides. Br. J. Industr. Med. 47:263–268; 1990.
28. Hanninen, H. Psychological picture of manifest and latent carbon disulphide poisoning. Br. J. Industr. Med. 28:374–381; 1971.
29. Hanninen, H.; Hernberg, S.; Tolonen, M.; Martelin, T. Psychological performance of subjects with low exposure to lead. J. Occ. Med. 20:240–255; 1978.
30. Hanninen, H.; Nurimen, M.; Tolonen, M.; Martelin, T. Psychological tests as indicators of excessive exposure to carbon disulphide. Scand. J. Psychol. 19:163–170; 1978.
31. Heilman, K. M.; Valenstein, E., eds. Clinical neuropsychology. New York: Oxford University Press; 1979.
32. Jamal, G. A.; Hansen, S.; Weir, A. I.; Ballantyne, J. P. The neurophysiologic investigation of small fiber neuropathies. Muscle and Nerve 10:537–545; 1987.
33. Johnson, B. L., ed. Prevention of neurotoxic illness in working populations. New York: Wiley; 1987:169–214.
34. Katims, J. J.; Naviasky, E. H.; Ng, L. K. Y.; Rendell, M.; Bleecker, M. L. New Screening device for assessment of peripheral neuropathy. J. Occ. Med. 28:1219–1221; 1986.
35. Kimura, J. Electrodiagnosis in diseases of nerve and muscle: Principles and practice (2nd ed.). Philadelphia, PA: F. A. Davis Co.; 1989.
36. Kurlan, R.; Caine, E.; Rubin, A. et al. Cerebrospinal fluid correlates of depression in Huntington's disease. Arch. Neurol. 45: 445–447; 1988.
37. Ledin, T.; Odkvist, L. M.; Moller, C. Posturography findings in workers exposed to organic solvents. Acta. Otolaryngol. (Stockh). 107:357–361; 1989.
38. Le Quesne, P.; Fowler, C. J.; Parkhouse, N. Peripheral neuropathy profile in various groups of diabetics. J. Neurol. Neurosurg. Psychiatry 53:558–563; 1990.
39. Lezak, M. D. Neuropsychological assessment (2nd ed.). New York: Oxford University Press; 1983.
40. Lezak, M. D. Neuropsychological assessment (3rd ed.). New York: Oxford University Press (in press).
41. Lindblom, U.; Tegner, R. Thermal sensitivity in uremic neuropathy. Acta. Neurol. Scand. 71:290–294; 1985.
42. Luria, A. R. Higher cortical function in man (2nd ed.). New York: Basic Books; 1980.
43. Matthews, C. G.; Chapman, L. J.; Woodard, A. R. Differential neuropsychologic profiles in idiopathic vs. pesticide-induced Parkinsonism. In: Johnson, B. L. ed. Advances in neurobehavioral toxicology. Chelsea, MI: Lewis Press; 1990:323–330.
44. Mathiowetz, M. S.; Weber, K.; Volland, G.; Kashman, N. Reliability and validity of grip and pinch strength evaluations. J. Hand. Surg. 9A:222–226; 1984.
45. Mayeux, R.; Stern, Y.; Cote, R. J.; Leventhal, J. Depression, intellectual impairment and Parkinson's disease. Neurology 31: 645–650; 1981.
46. McConnell, R.; Keifer, M.; Rosenstock, L.; Elevated tactile vibration threshold among workers previously poisoned with methamidophos: Leon, Nicaragua. 23rd International Congress on Occupational Health. Montreal, Canada, September 22–28, 1990 (Abstract).
47. Mikkelsen, S.; Jorgensen, M, Browne, E.; Gyldensted, C. Mixed solvent exposure and organic brain damage: A study of painters. Acta. Neurol. Scand. Suppl. 118, 78:1–143; 1988.
48. Mindham, R. Psychiatric symptoms in Parkinson's. J. Neurol. Neurosurg. Psychiatry 50:1192–1196; 1970.
49. Moller, C.; Odkvist, L.; Larsby, B.; Tham, R.; Ledin, T.; Bergholtz, L. Otoneurological findings in workers exposed to styrene. Scand. J. Work Environ. Health 16:189–194; 1990.
50. Murata, K.; Araki, S. Autonomic nervous dysfunction in workers exposed to lead, zinc, and copper in relation to peripheral nerve conduction: A study of R-R interval variability. Am. J. Indust. Med. 20:663–671, 1991.
51. Needleman, H. L.; Gunnoe, C.; Leviton, A.; et al. Deficits in psychologic and classroom performance of children with elevated dentine lead levels. New Eng. J. Med. 300(13):689–695; 1979.
52. Needleman, H. L.; Schell, A.; Bellinger, D.; Leviton, A.; Allred, E. The long term effects of exposure to low doses of lead in childhood. New Eng. J. Med. 322:83–88; 1990.
53. Pagani, M.; Malfatto, G.; Pierini, S. et al. Spectral analysis of heart rate variability in the assessment of autonomic diabetic neuropathy. J. Autonom. Nervous System 23:143–153; 1988.
54. Peters, H. A.; Levine, L. L.; Matthews, C. G.; Chapman, L. J. Extrapyramidal and other neurological manifestations associated with carbon disulphide fumigant exposure. Arch. Neurol. 45:537–540; 1988.
55. Peters, H. A.; Levine, R. L.; Matthews, C. G.; Sauter, S. L.; Rankin, J. H. Carbon-induced neuropsychiatric changes in grain storage workers. Am. J. Industr. Med. 3:373–391; 1982.
56. Petit, T. L. Memory, synaptic plasticity and neurotoxins. NeuroTox. 11:323–332; 1990.
57. Proctor, S. P.; White, R. F. Psychoneurological criteria for the development of neurobehavioral test batteries. In: Johnson, B. L., ed. Advances in neurobehavioral toxicology. Chelsea, MI: Lewis Publishers; 1990:273–281.
58. Reitan, R. M. Theoretical and methodological bases of the Halstead-Reitan Neuropsychological Test Battery. In: Grant, I.; Adams, K. eds. neuropsychological Assessment of Neuropsychiatric Disorders. New York: Oxford University Press; 1986.
59. Rosenberg, N. L.; Kleinschmidt-DeMasters, B. K.; Davis. K. A.; Dreisbach, J. N.; Hormes, J. T.; Filley, C. M. Toluene abuse causes diffuse central nervous system white matter changes. Ann. Neurol. 23:611–614; 1988.
60. Rosenberg, N. L.; Spitz, M.C.; Filley, C.M.; Davis, K.A.; Schaumburg, H. H. Central nervous system effects of chronic toluene abuse-clinical, brain stem evoked response and magnetic resonance imaging studies. Neurotox. Teratol. 10:489–495; 1988.
61. Schaumburg, H. H.; Spencer, P. S.; Thomas, P. K. Disorders of peripheral nerves. Philadelphia, PA: F. A. Davis Company; 1983.
62. Seppalainen, A. M. H. Neurophysiological approaches to the detection of early neurotoxicity in humans. CRC Critical Reviews in Toxicol. 18:245–298; 1988.
63. Schoenfeld, M.; Myers, R. H.; Cupples, L. A. et al. Increased suicide among patients with Huntington's disease. J. Neurol. Neurosurg. Psychiatry 47:1283–1287; 1984.
64. Silverstein, A. B. Perceptual, motor and memory functions in the Visual Retention Test. Am. J. Mental Deficiency 66:613–617; 1962.
65. Snow, W. G. The Rey-Osterrieth Complex Figure Test as a measure of visual recall. Paper presented at the Seventh Annual Meeting of the International/Neuropsychological Society, New York City; 1979.
66. Stuberg, W. A.; Metcalf, W. K. Reliability of quantitative muscle testing in healthy children and children with Duchenne muscular dystrophy using a hand-held dynamometer. Physical Therapy 68: 977–982; 1988.
67. Tegner, R.; Lindholm, B. Vibratory perception threshold compared with nerve conduction velocity in the evaluation of uremic neuropathy. Acta Neurol. Scand. 71:284–289; 1985.
68. White, R. F., ed. Clinical syndromes in adult neuropsychology: The practitioner's handbook. New York: Elsevier; 1992.
69. White, R. F.; Au, R.; Durso, R.; Moss, M. B. Neuropsychological function in Parkinson's disease. In: White, R.F., ed. Clinical syndromes in adult neuropsychology. New York: Elsevier: pp. 253–286; 1992.

70. White, R. F.; Diamond, R.; Proctor, S. P.; Morey, C.; Hu, H. Residual cognitive deficits 50 years after childhood plumbism. Brit J. Med. 50:613–622; 1993.
71. White, R. F.; Feldman, R. G. Neuropsychological assessment of toxic encephalopathy. Am. J. Industr. Med. 11:395–398; 1987.
72. White, R. F.; Feldman, R. G.; Echeverria, D. E.; Schweikert, J. neuropsychological effects of chronic solvent exposure. Report to NIOSH on Grant #5K01 OH0028-03; 1990.
73. White, R. F.; Feldman, R. G.; Moss, M. B.; Proctor, S. P. Magnetic resonance imaging (MRI), neurobehavioral testing and toxic encephalopathy: Two cases. Environ. Res. 61:117–123; 1993.
74. White, R. F.; Feldman, R. G.; Travers, P. H.; Neurobehavioral effects of toxicity due to metals, solvents and insecticides. J. Clin. Neuropharm. 13:392–412; 1990.
75. White, R. F.; Proctor, S. P. Research and clinical criteria for the development of neurobehavioral test batteries. J. Occ. Med. 34: 140–148; 1992.
76. White, R. F.; Proctor, S. P.; Feldman, R. G. Behavioral effects of toxic exposures. In: White, R. F., ed. Clinical syndromes in adult neuropsychology. New York: Elsevier; pp. 1–52; 1992.
77. White, R. F.; Robins, T. G.; Proctor, S. P.; Echeverria, D.; Rocskay, A. S. Neuropsychological effects of exposure to naphtha among automotive workers. Occ. Envir. Med. (in press).
78. White, R. F.; Vasterling, J.; Koroshetz, W.; Myers, R. Neuropsychology of Huntington's disease. In: White, R. F., ed. Clinical syndromes in adult neuropsychology. New York: Elsevier; pp. 213–252; 1992.
79. Yokoyama, K.; Araki, S.; Aono, H. Reversibility of psychological performance in subclinical lead absorption. NeuroTox. 9:405–410; 1988.
80. Zubrick, A.; Smith, A. Factors affecting BVRT performance in adults with acute focal cerebral lesions. Paper presented at the Sixth Annual Meeting of the International Neuropsychological Society, Minneapolis, MN; 1978.

APPENDIX

EXAMPLES OF FOCUSED BATTERIES

1. Known neurotoxicant (lead in adults): CPT, Grooved Pegboard, WMS Vis Reproductions subtest (Immediate Recall, Delayed Recall, Savings Score), WAIS-R Similarities, Profile of Mood States (POMS), Paced Auditory Serial Addition Test (PASAT);
2. Possible neurotoxicant (new solvent): Trails, CPT, WMS Vis Reproductions, Embedded Figures, POMS;
3. Brain-behavior relationships (TCE, to examine hypothesis of frontal/limbic dysfunction): PASAT, CPT, Trails B, California Verbal Learning Test (CVLT) or Rey Auditory Verbal Learning Test (RAVLT), Peterson Task, POMS;
4. Cognitive processing (executive function): PASAT, Porteus Mazes, Figural Fluency or Design Fluency (WCST was recommended by three clinicians—RC, ML, JM; RW has used the task in field studies where performance on the task was found to be affected by subjects' discussing test parameters and where it was found to be too affected by the subjects' learning of task parameters to be used in longitudinal studies).

Control Variables

1. Learning disabilities: WRAT, Reading Tests;
2. Psychiatric: Schedule of Affective Disorders and Schizophrenia-L, MMPI;

Full Neurologic and Neuropsychologic Evaluation

Neurologic examination, tremor test, eyeblinks, nerve conduction, WAIS/WISC (Info, Simil, Arith, Digit Span, Pic Com, Pic Arr, Block Des, Digit Symbol), CVLT or RAVLT, WMS Visual Reproductions, Peterson Task, Controlled Word Association, Boston Naming Test, Repetition Test, Rey-Osterrieth, Finger Tap, Pegboard, Ravens, MMPI, POMS;

Ellen Silbergeld wrote the following concerning her views on development and incorporation of new test methods.

An important method for identifying and selecting new tests in evaluating neurotoxicity in populations and cases is the set of basic research findings on (a) mechanisms of specific neurotoxins and (b) neurobiological substrates of behavioral function. These two areas of research, taken in combination, can produce the following logic for test selection: identification of specific neurochemical/neuroanatomic mechanisms for neurotoxicants, examination of the role(s) of these mechanisms in neurobiological systems, functional expression of alterations in these systems, neuropsychologic or other tests that can measure changes in functional expression. For instance, a considerable amount of research indicates that low level lead exposure affects GABAergic pathways in globus pallidus (in rats); changes in GABAergic function have been correlated with specific alterations in behavioral state; these states may be measurable by specific tests. Other examples are the mechanistic studies done on certain chlorinated solvents by Fuxe et al: these indicate specific mechanisms involving median eminence dopaminergic neurotransmission. The behavioral and neuroendocrinological consequences of altering these pathways have also been described.

SELECTION OF TESTS PREDICTIVE OF LONG LATENCY NEURODEGENERATIVE DISEASES AND DYSFUNCTION

This selection principle is derived from the current hypothesis that environmental neurotoxicants may be etiologic factors in the induction of long latency degenerative diseases of the nervous system. The drug impurity MPTP is the paradigm example of this hypothesis. Based on this, a rationale for test selection could be those tests that are predictive of the later induction of such diseases. The goal of these tests is to detect the early onset of pathology before overt behavioral or other dysfunction is expressed. This is similar to the rationale for selecting tests sensitive to low level exposures.

Considerable research has been done on developing predictive tests for diseases such as Huntington's chorea, dementia, and Parkinsonism. At present, these tests are not predictive over the long term, that is, when the gap between overt disease expression and time of testing is very great. But this is an arena that could be quite useful for neurotoxicology, particularly if manganese is introduced into the environment on a wide scale.

OTHER MEASURES OF THE NERVOUS SYSTEM

In addition to behavior and electrophysiology, other signals of the nervous system can be measured by biochemical means. These measures have not been considered in this Workshop, but ATSDR should consider their utility as adding to the repertoire of neurotoxicity assessment methods as well as contributing to the validation of neurobehavioral/neurophysiological methods discussed at this Workshop. In the near term, neuroendocrine markers are probably ready for consideration in studies of exposures to neurotoxicants. There are data on effects of solvents and metals on such neuroendocrine markers as growth hormone, prolactin, somatostatin, and

FSH. Other markers available in compartments that can be sampled (e.g., blood, urine, saliva) are: some neuropeptides, pituitary and hypothalamic hormones and releasing factors; neurotransmitter precursors and metabolites.

In the longer term, it is possible that the lymphocyte can provide information on gene-level damage to the nervous system in a manner analogous to the information that is being extracted for the purpose of estimating dose and cellular response to carcinogens (DNA adducts). As we identify neurotoxicant-induced alterations in gene transcription and translation, the lymphocyte may be exploited to provide parallel signals of these events.