Environmental Health Perspectives
Vol 90, pp. 195-198, 1991

Use of Computerized Test Batteries for Quantifying Neurobehavioral Outcomes

by Richard Letz*

Neurobehavioral testing provides for noninvasive assessment of the functional integrity of the nervous system. Neurobehavioral tests have been used as quantitative outcome measures in a number of epidemiologic investigations of the potential effects on the nervous system from exposure to organic solvents, heavy metals, and pesticides. Because of the functional complexity of the nervous system, sets of tests assessing a range of functions have been used, with inconsistency from one study to another. Although there has been recent progress in standardizing a core set of tests for use in occupational epidemiology, major consensus on testing methods has not emerged. Standardization of test methods is essential to provide a consistent database for risk analysis. Automation of data collection provides not only standardization, but also improved time efficiency, of data collection and analysis. The computerized Neurobehavioral Evaluation System (NES) has been developed to address the need for standardized, efficient data acquisition of a range of neurobehavioral variables. Examples of application of NES in epidemiologic studies of workers exposed to solvents are summarized. The need for use of NES as a tool for surveillance and in prospective epidemiologic investigations is emphasized.

Environmental Neurotoxicity As a Problem

Certainly, there are substantial environmental exposures to potential neurotoxicants, particularly in the workplace. Neurotoxic diseases are among the list of top ten leading work-related diseases and injuries prepared by the U.S. National Institute for Occupational Safety and Health (NIOSH) (*1*). Of the more than 60,000 chemicals in commerce, at least 750 have been reported to have adverse effects on the nervous system. Sixty-five chemicals of these 750 are also on the list of 200 chemicals having exposure to more than 1 million U.S. workers (*2*).

It is well known that clinically overt effects occur from high exposures to a number of solvents, heavy metals, and pesticides. Indeed, these effects are what determines that the exposures are high. There are also many reports of neurobehavioral dysfunction in people exposed to concentrations of these agents lower than those producing clinically overt symptoms. Of the 91 NIOSH criteria documents, 36 cite effects on the nervous system, often at concentrations lower than those required to produce effects on other organ systems (*2*).

It is commonly believed that subtle neurobehavioral deficits caused by lower level exposure can, with prolonged exposure, progress to more severe effects. This belief implies that an opportunity exists that early detection can allow remedial action to be taken before dysfunction progresses to irreversible damage. Neurobehavioral dysfunction can be considered, then, in the language of some, a marker for neurotoxic disease.

*Division of Environmental and Occupational Medicine, Mount Sinai School of Medicine, 1 Gustave Levy Place, Box 1057, New York, NY 10029.

Impediments to Assessing Risks of Neurotoxicity

There are theoretical and practical impediments to assessing neurotoxic risks in humans. The largest procedural impediment to quantitative assessment of risk from exposure to neurotoxicants is the absence of a large, consistent database in humans. A large database in humans does not exist because measuring neurologic or neurobehavioral outcomes is difficult and expensive, and what data do exist are difficult to interpret because they have been generated by a wide variety of testing methods. Theoretic and (interrelated) procedural reasons for use of so many methods include: the nervous system is complex, with many separate functions to be tested; there is no explicit, generally accepted functional model of the nervous system; there is no general neurotoxicologic model that generates explicit behavioral hypotheses to be tested; there is no standardized, commonly accepted tool to generate the data; investigators are rewarded for application of novel, rather than standard, techniques.

To create the needed database, work will be required at the interface of four fields: neurology, psychology, toxicology and epidemiology (Fig. 1). Advances and consensus on theory will be needed from both neuropsychology and neurotoxicology so that practical methods from psychology and epidemiology can be optimized.

Need for Advancements in Theory

There is great need for major consensus on theoretical models at the neuropsychologic and neurotoxicologic levels. At the

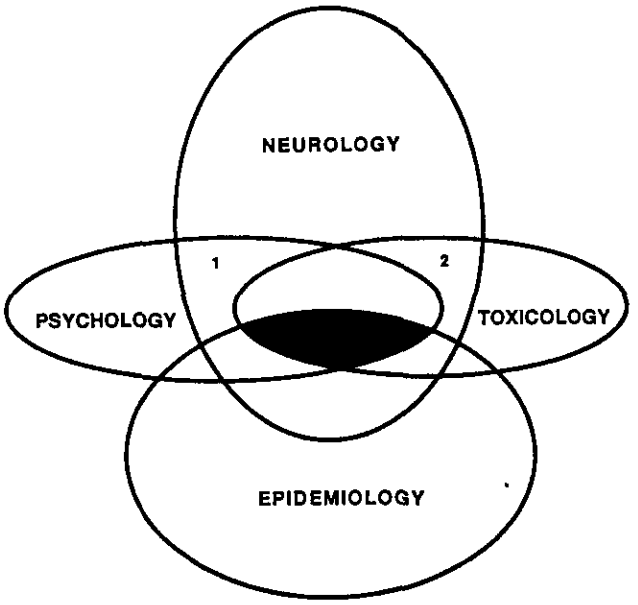

FIGURE 1. The interface of four fields: behavioral testing in neurotoxicologic epidemiology (shaded area); neuropsychology (1); and neurotoxicology (2).

neuropsychological level, agreement must be reached on an *explicit model of the entire range of neuropsychological* functions. Then the neuropsychological functions tested by existing behavioral tests could be specified. (No behavioral test assesses a single function; all tap some blend of sensory, motor, and cognitive functions.) Neuropsychologists currently use implicit neurobehavioral models, but these must be made more explicit and become more universally accepted. Perhaps information processing theory (*3*) can provide a unifying theme, as Williamson has suggested (*4*). In addition to providing a context for specification of neurobehavioral tests, a consensus neuropsychological model would provide some context from which to view any subclinical deficits in performance found in epidemiologic investigations.

Agreement at the neuropsychological level would also allow development of new behavioral tests, tailored to individual neuropsychologic functions more specifically than existing composite tests. This approach has been taken by Eckerman et al. (*5*) to develop a computer-based neurobehavioral testing system from a theoretic framework provided by Carroll (*6*). Unfortunately, neither the theoretic framework nor the computerized system has found widespread acceptance in neurotoxicology or neuroepidemiology.

In addition, more explicit neurotoxicologic theories are needed than those currently available. The range of behavioral functions that may be affected by exposure to a toxic agent is extremely wide, and for this reason investigators typically use sets, or batteries, of tests. In practice, no study of a particular exposure situation can realistically sample very many of these functions, so neurotoxicologic theory should tell us where to concentrate our testing efforts and which neurobehavioral tests to choose.

Need for Practical Tools

In addition to the needs for development of new theoretic tools and building of consensus on them, there is great need for development of new practical tools and standardization of them. The World Health Organization (WHO), and its European Office in particular, has had a major impact by sponsoring meetings on some of these needs, both in terms of neurotoxic disease definition (*7*) and standardization of neurobehavioral methods (*8–10*).

In selecting which behavioral tests to include in a test battery, there are hundreds of behavioral tests from which to choose. In addition, such tests are often modified *ad hoc* either through ignorance of protocol details or to improve them for particular studies. As a result, divergent results from apparently similar studies are hard to reconcile. One instance of addressing this problem of lack of test standardization and providing some consensus has emerged from a WHO-NIOSH conference held in Cincinnati in 1983 (*10*). At that meeting, a set of seven behavioral tests that has become known as the WHO Neurobehavioral Core Test Battery (WHO-NCTB) was recommended to be used in all epidemiologic investigations of workers exposed to potential neurotoxicants. Manually administered tests were selected so that they might be used in developing countries as well. It was intended that investigators would supplement this core with other tests as time and equipment available for testing permitted. A major study is currently underway in eight countries to determine norms and explore cultural differences in performance for the seven WHO-NCTB tests.

Computerized Neurobehavioral Testing

The general topic of computerized psychological assessment has been summarized recently (*11*). Most of the effort has been devoted to clinical concerns where the commercial market resides. Such applications include computerized scoring of questionnaires, particularly the Minnesota Multiphasic Personality Inventory, computerized patient report generation, and cognitive rehabilitation software. Computerized neurobehavioral testing for clinical purposes is not well developed, probably due to the need for intense clinician-patient interaction in that setting.

Computerized neurobehavioral testing offers several advantages in the epidemiologic situation. The primary advantages are rigid standardization of the testing protocol through the computer program and efficiency of data collection. The data collected are objective and quantitative. Since conclusions in this context are on a group basis, study objectives can be met by less intense testing of a greater number of individuals than in the clinical testing situation. Thus, with computerized tests, sample sizes can be increased easily from the 20 to 50 subjects typical of past studies of this type. Larger sample sizes will allow better modeling of covariates, reduced sampling bias, and increased statistical power.

There are some disadvantages of computerized testing. The most obvious problem involves potential fear of the computer by the subject. This apparent problem can be controlled by proper hardware (using simple responses, covering unused keys, etc.) and software design (simple, consistent instructions and smooth program flow), and keeping the nature of the behavioral tests

simple and obvious to the subjects. Perhaps the most significant criticism of currently available computerized tests is that they tap only a limited range of the full behavioral repertoire. Visual presentation of stimuli and manual responding are emphasized in most currently available tests. It is fortuitous that, in the two best-studied exposure areas, lead and solvents, deficits in psychomotor and visuomotor functions are among those reported most often.

Only a few computerized test systems that assess a wide range of behavioral functions have been developed for environmental epidemiologic applications. Many investigators have applied individual computerized neurobehavioral tests. For example, special-purpose laboratory reaction time tests have been implemented on general purpose microcomputers, although hidden technical difficulties may often be ignored. The computerized test systems that contain a number of tests and have been developed for, or used in, epidemiologic applications include: MicroTox System by Eckerman et al. (*5*); Neurobehavioral Evaluation System (NES) by Baker et al. (*12*); Swedish Performance Evaluation System (SPES) by Iregren et al. (*13*); Milan Automated Neurobehavioral System (MANS), a computer implementation of many of the tests in the WHO-NCTB, by Camerino et al. (*14*); and Cognitive Scanner, a commercially available system from Denmark, developed by Laursen et al. (*15*). Only NES and SPES have been used in multiple studies appearing in the peer-reviewed literature. SPES appears to be optimized for laboratory investigations and NES for field epidemiologic investigations.

Neurobehavioral Evaluation System

The computerized neurobehavioral system that has been used most widely for application in epidemiologic studies is NES (*12*). It consists of over 15 computerized neurobehavioral tests and questionnaires that tap the broad functional domains of psychomotor speed and control, perceptual speed, learning and memory, attention and affect. A subset of these tests is chosen for each study situation, depending upon the toxic agent in question, the study design, and the time available for testing. The test instructions have been translated from English into eight other languages, and more than 50 investigators have joined the NES Users' Group and obtained the NES software (*16*). At this point, more than 10 laboratory and epidemiologic studies that used NES have been published.

A brief review of the studies in which NES tests have been applied has recently been provided (*17*). NES tests have been used in studies of more than 5000 subjects. Groups exposed to potential neurotoxicants that have been studied include painters (*18,19*), pesticide applicators (*20*), and mercury-exposed workers (*21*). Other epidemiologic studies of painters, floorlayers, and dry cleaners have been completed but are not yet published. Some NES tests are also being used in the Third National Health and Nutrition Evaluation Survey (NHANES-III).

Neurobehavioral Data Analysis

For most neurobehavioral tests, the major modifier of performance is age, which may account for 5 to 40% of the total variance, depending upon the particular test and age range of the sample. Education may account for up to 30% of the total variance. These effects should be controlled for in all epidemiologic investigations employing neurobehavioral tests. Other effects such as gender and time of day are commonly thought to be important factors and may occasionally be statistically significant predictors for some neurobehavioral tests, but they rarely account for more than 5% of the total variance in test scores. All these effects combined, including age and education, never account for more than 50% of the total variance of scores for a particular test. The other major source of variance in neurobehavioral test scores is within-subjects error, which may be 10 to 50% of the total variance. Of this, much may be actual instrumental error, again depending on the test. However, human performance is always subject to a degree of noninstrument noise.

Alcohol intake is commonly considered to have a major impact on neurobehavioral test performance. Certainly, acute alcohol intoxication and chronic alcoholism resulting in nutritional deficit should be considered criteria for exclusion from data analysis. However, the utility of self-reported alcohol consumption as a predictor variable in regression analyses of neurobehavioral outcomes is not clear. Although negative effects of moderate drinking have been reported (*22*), others have found no such effects on a variety of neurobehavioral measures in neurotoxically exposed, but otherwise healthy, populations (*19,23*). In addition, paradoxical positive effects of greater alcohol consumption are found on occasion (*18*).

The most troublesome source of variance in neurobehavioral test performance (computerized or manual) is the motivational state of the subject at the time of testing. Malingering is a potentially important cause of suboptimal performance, although its frequency of occurrence in epidemiologic studies is unknown. It is possible that the frequency of malingering will increase as the use of neurobehavioral tests in neurotoxic exposure situations and the number of litigations increases. While techniques exist for detecting malingering (*24*), their true efficacy is unknown. Such techniques have not been implemented in currently available computerized test systems.

Individuals with excessive within-test reproducibility can be identified and excluded from the data analysis as a way of handling subjects with submaximal effort. This procedure assumes that submaximal effort results in increased variability. However, individuals with true neurotoxic impairment may also have more variable performance, and excluding them would bias analyses toward the null hypothesis (*25*). Finally, poor effort may be the result of depression, which itself may or may not be caused by exposure. This issue is a difficult one for both traditional behavioral tests and for computerized tests, but one that may be handled better in the traditional testing situation where there is more direct subject-examiner interaction.

The reliability of computerized neurobehavioral outcomes is moderately high and generally comparable to that of manually administered neurobehavioral tests. For example, the average reliability for NES tests in field epidemiologic investigations is about 0.7 (*17*). The reliability of computerized neurobehavioral test outcomes can be improved in laboratory investigations by increasing the amount of training for subjects and increasing the length of the tests.

Since more than 50% of the total variance in performance on most neurobehavioral tests is due to between-subjects factors, a

within-subjects (test-retest, cross-over) study design should usually be employed, if the critical hypotheses can be addressed by such a design. This is easily accomplished in investigations of effects of acute exposures. The implication for studies of effects of chronic exposure is that large sample sizes will be necessary to observe subtle effects in studies with between-subjects designs, or that pre-exposure baseline performance should be assessed in prospective studies. The practical efficiency of computerized neurobehavioral outcomes is helpful in both of these instances, i.e., testing large numbers of subjects or implementing routine testing. The efficiency considerations of obtaining pre-exposure baseline performance would be a bonus to the usual advantages of prospective study design (e.g., increased epidemiologic validity). In addition, collection of baseline behavioral performance information would allow greater power for detecting effects of accidental exposures as well as allow greater confidence in making decisions about changes in the performance of individuals (*16*).

Concluding Remarks

Several computerized systems capable of generating neurobehavioral outcomes in epidemiologic studies of the effects of exposure to potential neurotoxicants have been developed and are being applied. NES is the most widely used system and has been found useful in a variety of exposure situations.

There is still much work to be done. Advances in theory are needed. Theoretic advances will trigger additional test system development and provide a context for assessing the meaning of test outcome deficits that are found. Prospective studies are needed. They will allow determination of the biological significance of subtle deficits in neurobehavioral test outcomes and allow prevention of disease. In addition, such studies will provide a necessary database for estimating effect modifiers and external comparison groups to other studies. Consensus on optimal outcome measures and other statistical methods is needed. Finally, standardization of at least a few computerized tests among all test systems should facilitate comparison of results from diverse studies.

This work was supported in part by a center grant from the National Institute of Environmental Health Sciences (ES00928-15). Many members of the NES Users' Group have contributed much to the ideas and data presented here. Special thanks go to F. Gerr and D. Hershman for their critical comments.

REFERENCES

1. Centers for Disease Control. Neurotoxic disorders. Morbid. Mortal. Weekly Rep. 35(8): 537–542 (1986).
2. Anger, W. K. Workplace exposures. In: Neurobehavioral Toxicology (Z. Annau, Ed.), The Johns Hopkins University Press, Baltimore, MD, 1986, pp. 331–347.
3. Wickens, C. D. Engineering Psychology and Human Performance. Charles Merrill, Columbus, OH, 1984.
4. Williamson, A. M., and Teo, R. K. C. Neurobehavioral effects of occupational exposure to lead. Br. J. Ind. Med. 43: 374–380 (1986).
5. Eckerman, D. A., Carroll, J. B., Foree, D., Gullion, C. M., Lansman, M., Long, E. R., Waller, M. B., and Wallsten, T. S. An approach to brief field testing for neurotoxicity. Neurobehav. Toxicol. Teratol. 7: 387–393 (1985).
6. Carroll, J. B. Individual Difference Relations in Psychometric and Experimental Cognitive Tasks. Thurstone Laboratory Report No. 163, NTIS Doc. AS-A086057, The University of North Carolina, Chapel Hill, NC, 1980.
7. WHO. Organic Solvents and the Central Nervous System. World Health Organization Environmental Health Document 5, Copenhagen, 1985.
8. Gilioli, R., Cassitto, M. G., and Foa, V. Neurobehavioral Methods in Occupational Health. Pergamon Press, New York, 1983.
9. WHO. Neurobehavioral Methods in Occupational and Environmental Health. World Health Organization Environmental Health Document 3, Copenhagen, 1985.
10. Johnson, B. L. Prevention of Neurotoxic Illness in Working Populations. John Wiley and Sons, New York, 1987.
11. Butcher, J. N., Ed. Computerized Psychological Assessment: A Practitioner's Guide. Basic Books, New York, 1987.
12. Baker, E. L., Letz, R., and Fidler, A. T. A computer-administered neurobehavioral evaluation system for occupational and environmental epidemiology. J. Occup. Med. 27: 206–212 (1985).
13. Iregren, A., Gamberale, F., and Kjellberg, A. A microcomputer-based behavioral testing system. In: Neurobehavioral Methods in Occupational and Environmental Health, WHO Environmental Health Document 3, World Health Organization, Copenhagen, 1985.
14. Camerino, D. Presentation, Description and Preliminary Evaluation of M.A.N.S. Institute of Occupational Health, University of Milan, Milan, Italy, 1987.
15. Laursen, P., and Jorgensen, T. Computerized neuropsychological test system. In: Neurobehavioral Methods in Occupational and Environmental Health, WHO Environmental Health Document 3, World Health Organization, Copenhagen, 1985.
16. Letz, R. Occupational screening for neurotoxicity: automated techniques. Toxicology 49: 417–424 (1988).
17. Letz, R. The Neurobehavioral Evaluation System: an international effort. Toxicology 8: 303–309 (1989).
18. Fidler, A. T., Baker, E. L., and Letz, R. The neurobehavioral effects of occupational exposure to organic solvents among construction painters. Br. J. Ind. Med. 44: 292–308 (1987).
19. Baker, E. L., Letz, R., Eisen, E., Pothier, L., Larson, M., and Wolford, R. Neurobehavioral effects of solvents in construction painters. J. Occup. Med. 30: 116–123 (1988).
20. Maizlish, N., Schenker, M., Weisskopf, C., Seiber, J., and Samuels, S. A behavioral evaluation of pest control workers with short-term, low-level exposure to the organophosphate diazinon. Am. J. Ind. Med. 12: 153–172 (1987).
21. Albers, J. A., Kallenbach, L. R., Fine, L. J., Langolf, G. D., Wolfe, R. A., Donfrion, P. D., Allessi, A. G., Stolp-Smith, K. A., Bromberg, M. B., and the Mercury Workers Study Group. Neurological abnormalities associated with remote occupational elemental mercury exposure. Ann. Neurol. 24: 651–659 (1987).
22. Parker, D. A., Parker, E. S., Brody, J. A., and Schoenberg, R. Alcohol use and cognitive loss among employed men and women. Am. J. Public Health 73: 521–526 (1983).
23. Ryan, C. M., Morrow, L. A., Bromet, E. J., and Parkinson, D. K. Assessment of neuropsychological dysfunction in the workplace: normative data from the Pittsburgh Occupational Exposures Test Battery. J. Clin. Exp. Neuropsychol. 9: 665–679 (1987).
24. Rogers, R., Ed. Clinical Assessment of Malingering and Deception. The Guilford Press, New York, 1988.
25. Eisen, E. A., Letz, R., Wegman, D. H., Baker, E. L., and Pothier, L. Defining measurement precision for effort-dependent tests: the case of three neurobehavioral tests. Int. J. Epidemiol. 17: 292–298 (1988).