Available online at www.sciencedirect.com
ScienceDirect
NeuroToxicology 27 (2006) 951–969
NeuroToxicology

Development and validation of a test battery to assess subtle neurodevelopmental differences in children

Philip W. Davidson [a,*], Bernard Weiss [a], Christopher Beck [a], Deborah A. Cory-Slechta [b], Mark Orlando [a], David Loiselle [a], Edna Carter Young [a], Jean Sloane-Reeves [a], Gary J. Myers [a]

[a] *University of Rochester School of Medicine, Rochester, NY, USA*
[b] *The Environmental and Occupational Health Sciences Institute, The Robert Wood Johnson Medical School, NJ, USA*

Received 28 March 2005; accepted 31 March 2006
Available online 28 April 2006

Abstract

There is increasing concern over the impact of low-dose exposures to environmental chemicals on children's neurobehavioral function. To determine subtle alterations in children's function, it is necessary to move beyond global measures such as IQ and employ tests that can detect small, subtle neurodevelopmental effects across a broad array of behavioral domains. We investigated the sensitivity and specificity of a battery of 63 neurodevelopmental tests or tasks designed to detect outcomes representing the type of subtle neurodevelopmental deficits caused by exposure to neurotoxicants in school-aged children. We studied Neonatal Intensive Care Unit (NICU) graduates, a population known to be at risk for both major and mild anomalies in perception, motor functioning, learning, memory and cognition. This population served as a surrogate to evaluate the capacity of these tests and tasks to predict such deficits. The subjects' histories of previous exposures to any environmental neurotoxicants was not ascertainable, but exposures to elevated levels was not suspected. Over one-third of the 63 measures proved capable of detecting pre-diagnosed lower IQ, the presence of a learning disability (LD) or a neonatal risk profile with at least 70% sensitivity and specificity. Some tests were differentially sensitive and specific, depending upon the presence or absence of one or more of several covariates such as gender, age, hearing status, or familiarity with computers. Tests were also eliminated from the battery if they were affected by too many covariates. We propose calling the final battery of tests that are specific and sensitive to subtle neurodevelopmental changes the Rochester test battery (RTB). Further studies are needed to confirm the capability of the RTB to detect subtle changes associated with neurotoxic exposures.

Keywords: Behavior; Neurotoxicity; Child development

There is a considerable literature on approaches to assessing neurotoxicity (e.g., Fiedler, 1996; Guillette, 2000; Jacobson and Jacobson, 1996). Various strategies have been utilized to determine test efficacy in identifying neurobehavioral effects, including paper and pencil tests and computerized batteries (Amler et al., 1996; Anger et al., 1994; Krasnegor et al., 1994; Williamson, 1996). Much of the work in this field has focused on adults with toxic exposure in the work place; but some of these techniques have been successfully used to assess children's exposure (Baker et al., 1985; Amler and Gilbertini, 1996; Dahl et al., 1996).

The general approach in developing these tools, whether for children or adults, begins with the assumption that deficient performance on tests results from the neurotoxicant in question. The test, in turn, is assumed to accurately measure the effect of that specific neurotoxicant. This approach is not consistent with basic psychometric principles in that it implies attribution of a predicted effect without evidence that the effect can actually be distinguished from a non-effect by the endpoint measure. Although this approach is widely employed and may be acceptable when gross effects are being assessed, it is questionable when the subtle effects of low-level exposures are being sought (Davidson et al., 1995; White et al., 1993). Absence of detectable effects on most test batteries cannot be interpreted as an indication of no risk and the presence of detectable effects must meet certain criteria before a causal relationship can be established.

* Correspondence to: SCDD, Box 671, URMC, 601 Elmwood Avenue, Rochester, NY 14642, USA. Tel.: +1 585 275 6626; fax: +1 585 275 3366.
E-mail address: phil_davidson@urmc.rochester.edu (P.W. Davidson).

0161-813X/$ – see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuro.2006.03.025

An alternative strategy to address this difficulty is to select tests based on their capability to accurately detect more severe behavioral deficits known to be associated with high exposures to environmental chemicals then to confirm these tests' are capable of detecting small or subtle variations. Using this strategy, subjects should be selected because they already display the target deficits independent of any neurotoxic exposure. Thus, the tests in the battery would be known to be capable of detecting the specific outcomes expected to result from a toxic exposure.

There are several parameters used to characterize the capabilities of a psychometric test. Of these, the ones that seem most relevant to the problems facing investigators studying low-dose toxic exposures are *sensitivity* (the conditional probability of detecting a true positive) and *specificity* (the conditional probability of detecting a true negative (Hrudey and Leiss, 2003). Thus, if a test were both sensitive and specific to one or more behaviors expected to result from a neurotoxic exposure, the test might be useful for detecting that characteristic in an actually exposed group of children. For this purpose, both sensitivity and specificity are of equal importance, even though specificity could not actually be determined without testing in an exposed population.

In human research on neurotoxicological endpoints, most behavioral test batteries have failed to determine sensitivity and specificity. However, sensitivity and specificity have been recommended as the best parameters with which to gauge the ability of a test or task to detect a rare event (Hrudey and Leiss, 2003). White et al. (1994) underscored the importance of considering both sensitivity and specificity in test selection for neurotoxicogical studies. Fiedler (1996) points out that more refined study of sensitive and specific behavioral measures are needed.

The aim of the present study was *to evaluate the sensitivity and specificity of a battery of 63 neurodevelopmental tests and tasks to detect outcomes that represent the types of subtle neurodevelopmental deficits expected to occur in children after exposure to low dosages of common developmental neurotoxicants*. Those tests and tasks that were most sensitive and specific for the high risk population studied were then combined into a battery that we call the Rochester test battery (RTB). The tests and tasks fall into three categories: (1) neuropsychological tests with already established psychometric properties not previously used to study developmental neurotoxicity; (2) electrophysiological and behavioral tests of sensory and motor function spanning a broader range of indices than those used in prior investigations; (3) adaptations of performance tasks used previously only in animals or not yet applied to assess developmental neurotoxicity in children.

In the present study we used a combination of animal and human data to develop, refine and extend the scope of tests suitable for humans, and then assessed their sensitivity and specificity in populations already identified to be at risk for the presumed deficits. The resulting data could then be used to assist investigators and assessors in selecting specific tests and tasks with empirically determined detection capacities which might then be used for neurotoxicity research and risk assessment. We focused mainly on non-verbal specific developmental domains, including *overall cognition*, memory, *auditory* and *visual information processing* (Otto and Fox, 1993; Rice, 1996; Al-Damluje, 1976; Harada, 1977; Ino and Mizukoshi, 1977; Amin-Zaki et al., 1974; Brenner and Snyder, 1980; Crofton et al., 1994; Fechter et al., 1998; Murata et al., 1997, 2004; Tsubaki and Irukayama, 1977; Wu et al., 1985; Tsubaki and Takahashi, 1986; Grandjean et al., 1997; Fox et al., 1997; Kremer et al., 1999; Ishikawa and Miyata, 1980; Dementi, 1994; Boyes et al., 1994; Geller et al., 1998), *somatosensory functions* (McConnell et al., 1994; Beach et al., 1996; Broadwell et al., 1995), *fine motor control* (Newland and Weiss, 1991; Anger, 1990), and *complex perceptual motor functions* (Rahill et al., 1996). Verbal ability was tested as a part of several of the tests and tasks included in the battery.

The literature indicates that low birth weight children (<1500 g) are at high risk for deficits in the sensory, perceptual and cognitive domains and could serve as a useful population for determining test sensitivity and specificity (Blumsack et al., 1997; Bruininks and Bruininks, 1977; Goyen et al., 1998; Hack et al., 1994; Hung et al., 1987; Kendrick and Hanten, 1980; Maio-Feldman, 1994; O'Brien et al., 1988; Powls et al., 1997; Saigal et al., 1991; Swanson, 1983). NICU graduates are at high risk for a wide range of deficits that are similar to the various profiles of effects caused by exposures to different neurotoxicants. Their diverse array of deficits is a desirable characteristic for a reference population, and is similar to the profile of effects seen in children exposed to neurotoxicants.

1. Design and methods

1.1. Subjects

The subjects consisted of 293 children recruited from the Golisano Children's Hospital at Strong Neonatal Continuing Care Program (NCCP). The NCCP is a program designed for follow-up of infants hospitalized in the NICU and at high risk for neurodevelopmental deficits. All NCCP patients spent time in the NICU as newborns. There were 1016 infants enrolled in the NCCP between 1987 and 1993 (the relevant years for the target age group). Nearly all of these children resided in the Rochester metropolitan area and were available for recruitment. We selected subjects who ranged in age from 9 to16 years ($X = 12.03$).

We used the NCCP clinical database to determine each eligible child's diagnosis and demographic information including name, address, and primary care pediatrician. Children with a diagnosis of Down syndrome or an intracranial hemorrhage of Grade IV were excluded. Children were then sorted by the pediatric practice providing their primary health care and each pediatrician was contacted by mail with a list of his or her eligible patients. Any child whose pediatrician decided that testing would not be advisable was excluded. If parents gave oral consent, the child was given an appointment for formal consenting and initial screening. The clinic secretary contacted parents for the screening. At the screening appointment, parental consent and child assent were obtained.