The Implications of Limitations in Hydrocarbon Research for Neuropsychological Assessment

Paul R. Lees-Haley and Christopher W. Williams
Lees-Haley Psychological Corporation, Woodland Hills, CA

As neuropsychologists are involved increasingly in the neuropsychological assessment of individuals reporting both acute and long-term exposure to hydrocarbons, the value of research for providing guidance also increases. Research findings inform neuropsychologists about likely neuropsychological deficits given certain intensity and duration of exposure and about neuropsychological tests most appropriate for detecting hydrocarbon-induced deficits. Although high-dose events are associated with neuropsychological deficits, methodological weaknesses in the existing research impede a consensus on the implications of low-dose exposures in humans. Three flaws inherent in research on low-dose hydrocarbon exposures and their implications for neuropsychological assessment of individuals are discussed: Selection bias in recruitment of research participants, biased recall of research participants, and between-study variation in neuropsychological tests and results. Because they undermine the reliability and validity of existing research, lack of awareness of these weaknesses may interfere with unbiased assessment of individual patients suspected of hydrocarbon-induced neuropsychological deficits. Examples from the social-cognitive psychological literature contribute to understanding how misleading research findings can lead experts to form expectancies that bias assessment of individuals. © 1997 National Academy of Neuropsychology

Hydrocarbons are organic compounds made up of carbon atoms usually in combination with hydrogen (Harte, Holdren, Schneider, & Shirley, 1991). Derived primarily from petroleum distillation (e.g., Geehr & Saluzzo, 1992), hydrocarbon components are contained in numerous products, including gasoline, common solvents, and oil-based drilling fluids used in oil and gas exploration. Because hydrocarbon exposure is ubiquitous in the United States, neuropsychologists are called upon increasingly to assess the neuropsychological status of persons reporting exposure to toxic levels.

Individuals reporting exposure to a variety of hydrocarbons often are administered neuropsychological test batteries to determine the extent, if any, of hydrocarbon-induced deficits. Although the temptation to attribute decrements in neuropsychological functioning to hydrocarbon exposure is strong in high-dose cases, the informed clinician is wise to beware of premature conclusions when the patient has limited acute exposures or chronic

Address correspondence to: Paul R. Lees-Haley, 21331 Costanso Street, Woodland Hills, CA 91364. E-mail: plh@ix.netcom.com
low-dose exposures. This caution is called for partly because of the youth of the field in general, and more specifically because methodological problems in the existing human research undermine conclusions that a causal relation exists between low-dose hydrocarbon exposures and neuropsychological deficits (Lees-Haley & Williams, in press; Williams & Lees-Haley, 1996). The purpose of this article is to provide an overview of some of the methodological flaws in hydrocarbon neurobehavioral research that have implications for neuropsychological evaluation of patients exposed to hydrocarbons. Knowledge of these research problems can inform clinicians and help them make more accurate assessments.

Background

In contrast to well-documented detrimental but usually reversible effects of acute exposure to hydrocarbons (Klassen, 1996; Kulig, 1990; Sullivan & Krieger, 1992), the long-term or irreversible neuropsychological effects of exposure to low doses of most hydrocarbons remain controversial. The hydrocarbons that have been most studied from a neuropsychological point of view are organic solvents. These include the aliphatic hydrocarbons derived from petroleum and mineral spirits and the aromatic hydrocarbons made from coal tar (Hartman, 1995). Allegations of injury include impairments of concentration and memory, psychomotor slowing or lack of coordination, gait disturbances, disorientation, personality deterioration, and other complaints. Whereas some studies have purported to document chronic effects (for example, in house painters, whose occupation brings them into daily contact with organic solvents) (Arlien-Soborg, Bruhn, Gyldensted, & Melgaard, 1979; Lindstrom, 1973; Mikkelsen, Jorgensen, Browne, & Gyldensted, 1988), other studies have failed to show such effects (Bolla, Schwartz, Agnew, Ford, & Bleecker, 1990; Maizlish et al., 1985; Spurgeon, Glass, Calvert, Cunningham-Hill, & Harrington, 1994; Triebig et al., 1988). A few studies have reanalyzed data from earlier investigations and showed that the data originally reported as evidence of injury did not support the conclusions previously offered (Gade, Mortensen, & Bruhn, 1988; Grasso, Sharratt, Davies, & Irvine, 1984). Reviews systematically applying generally accepted research standards to this literature have challenged the conclusions offered in earlier literature (e.g., see Rebert & Hall, 1994, discussing styrene, for a recent example).

Among the problems discussed in this review are selection bias in recruitment of research participants, overreliance on subjective recall in determining levels and duration of exposure, overreliance on subjective recall in assessing premorbid health status and cognitive functioning, between-study variability in tests used to assess neuropsychological functioning, between-study variability of exposures examined, and lack of dose-response documentation. These problems, present in many studies, cast doubt on findings supporting a causal link between continuous low-level exposure to hydrocarbons and permanent toxic encephalopathy.

Overview of Limitations

Conclusions about hydrocarbons are confounded because most studies have dealt with exposures to multiple chemicals with no clear rationale for reaching general conclusions about all of the chemicals to which the subjects were exposed. When one or more of the substances in the mixture is a known toxin that plausibly explains the effects associated with the mixture, it is unreasonable to generalize to other substances in the mixture without first accounting for the known toxins. As the saying goes, “When you hear hoofbeats, don’t think of zebras.”
When considering the hypothesis that low dose hydrocarbon exposures cause neuropsychological deficits, it is important to realize how basic are the continuing contradictions and ambiguities in this literature. For example, Baker (1994) suggests that psychomotor functioning and short-term memory are the most well-documented areas of loss due to solvent CNS damage. Yet Hanninen et al. (1991) administered tests of psychomotor functioning and visual memory on which exposed and unexposed monozygotic twins showed no differences. In fact, Hanninen’s exposed subjects outperformed unexposed subjects on a test of psychomotor performance.

Because of the youthfulness of this field, it is hardly surprising to find controversy, but it is important to maintain a tentative posture. For example, Kelly and Filley note that neurobehavioral toxicology is “a discipline in its early infancy” and observe that “the clinical data on neurobehavioral syndromes related to toxins are fragmentary and largely inconclusive” (1992, p. 154). Kelly and Filley also point out that the field of neurobehavioral toxicology suffers from several significant problems, including ethical limits on research, inapplicability of animal studies to humans, disproportionate reliance on data from acute exposures with little data for low level chronic exposure, incomplete data concerning the premorbid functioning of the exposed individuals, the fact that most toxic exposures involved mixtures of substances, the variability of physiologic correlates of neurobehavioral effects, the methodological diversity of the different neuropsychological tests employed in these studies, and the uncertain validity and reliability of the neuropsychological tests used in neurobehavioral toxicology. These limitations of the field in general apply specifically to research on neuropsychological implications of hydrocarbon exposures.

One pervasive confound, selection bias in recruitment of research participants, is evident in numerous cross-sectional studies. Exposed subjects were selected not randomly from exposed groups, but from symptomatic medical clinic patients, the rolls of workers seeking disability payments, and symptomatic groups already diagnosed with chronic toxic encephalopathy. These nonrandom sampling procedures systematically excluded those members of the hydrocarbon-exposed groups not experiencing neuropsychological difficulties, thereby skewing results toward the impression of a causal link between hydrocarbon exposure and dysfunction. However, many of the articles based on this selection procedure suggested that an association between exposure and pathology had been found for the population from which the nonrandom sample was selected.

Problems with reliability and bias associated with reliance on subjective recall are important to many essential aspects of neuropsychological assessment, including exposure histories, alternative explanations for problems, and retrospective estimates of premorbid cognitive and emotional functioning (Taubes, 1995). Estimates of the level and duration of exposure have been based on subjective recall, often covering years or even decades of exposure in varying environments, with variable levels of exposure, while engaged in a variety of activities, during which time there were exposures to other relevant substances. “Most manufacturing processes change considerably over time, as do the exposures experienced by the workforce” (Tomenson & Brown, 1995, p. 1063). To assume that such estimates are reliable, or that dose-calculations based on these memories are valid, is unreasonable. Yet vague measurements are used as a basis for concluding that low doses of hydrocarbons cause neuropsychological deficits. The notorious unreliability of recall in estimating important alternative explanations for the data, such as drug and alcohol consumption, further erodes confidence in the literature. Blind blood testing of patients seen for a variety of medical reasons has confirmed that many individuals with no admitted or suspected alcohol or drug history are currently using alcohol or drugs or both (Zweben, 1996).
Finally, between-study variation in tests used to evaluate the neuropsychological status of hydrocarbon exposed workers, and variation in results, is so great in this particular literature as to impede a consensus on which tests detect neuropsychological deficits, and which deficits, if any, are indicative of hydrocarbon poisoning. Certainly it is desirable and appropriate to use a variety of reasonably correlated tests to converge on a valid conclusion. However, the extent of variation in this international literature is so great that it is unclear if tests with similar names or characterizations measure the same functions. Many studies find no association between exposure and the deficits suggested by other reports, which suggests the need to consider other causal factors as well as other explanations for the results.

One goal of this article is that awareness of selection bias, biased recall, and variability in research findings will aid in diminishing schema-based expectations or preconceived assumptions by experts. Awareness of the problems in this literature may inspire clinicians to maintain a skeptical, scientific attitude during neuropsychological assessment and to avoid confirmatory strategies that bias collection and interpretation of data (see Chapman & Chapman, 1967, 1969; Golding & Rorer, 1972; Salovey & Turk, 1991; Snyder, 1981; Snyder & Thomsen, 1988; Starr & Katkin, 1969). Following is a brief critical review of the research illustrating the effects of selection bias, biased recall, and between-study variation in tests and results, with discussion of their implications for neuropsychological assessment.

"Painter's Syndrome"

Many of the early studies were conducted in Scandinavian countries where a "painters' syndrome" could lead to premature retirement and disability benefits. Tens of thousands of persons in Scandinavia have been compensated for such complaints during the last three decades, and both the symptoms and the potential availability of benefits is widely known among the potential claimant community. In one study (Gregersen et al., 1978), "chronic cerebral painters' syndrome" was reportedly found among a group of 35 painters ranging in age from 27 to 58 years. Findings from this and subsequent studies popularized the view that long-term exposure to organic solvents caused a painter's syndrome. Providing a link between exposure and behavioral deficits was the overriding concern of the early studies. One specific goal was to assemble a battery of tests sensitive enough to detect subclinical neuropsychological effects of occupational exposure to neurotoxins (Hanninen, 1971; Ryan, Morrow, Bromet, & Parkinson, 1987; Russell, Flattau, & Pope, 1990).

Painters' syndrome is not universally accepted in the medical or neuropsychological community. For example, painter's syndrome is not a diagnosis included in DSM-III-R (APA, 1987), DSM-IV (APA, 1994), ICD-9 (Med-Index, 1992), or the ICD-10 Classification of Mental and Behavioral Disorders published by the World Health Organization (1992), although the latter does note intoxication from and dependence on volatile solvents, a volatile solvent-induced amnesic syndrome, volatile solvent-induced psychosis, and volatile solvent-induced mental and behavioral disorder. Despite the criticisms in the literature, the notion of "painter's syndrome" has crept into medical, toxicological, and neuropsychological literature through a process of repetition leading to reification without empirical validation or consistent definition of the so-called syndrome. However, as Hartman (1995) notes, in the United States experts are still debating whether solvent syndrome exists. Hartman draws our attention to the critically important fact that "solvent mixture researchers often consider solvent effects as part of a unitary 'solvent syndrome'" but researchers studying individual solvents attest to differences in the neurotoxic potential of common industrial solvents, and experimental studies indicate wide variation in acute solvent neurotoxicity (p. 191). Along similar lines, Snyder and Andrews (1996) recognize that:
The effects that have in the past been described as the general effects of solvents are quite complex and difficult to define and measure in humans. Only through the expansion of research in this area can we anticipate a better understanding of these effects and a more rational approach to the use of the developing literature in the proper control of exposures to solvent vapors. (1996, p. 740)

Based on literature addressing these problems as a unitary syndrome, neuropsychologists evaluating individuals suspected of low-dose exposure have rendered judgments that poor performance on neuropsychological tests was due to low-dose hydrocarbon poisoning. However, a critical review of the research found several errors in methodology casting doubt on the original findings leading to the widespread belief in the neurotoxicity of these substances (Errebo-Knudsen & Olsen, 1986). Included in the weaknesses was lack of a control group matched to the exposed group on confounding variables related to the same kinds of health problems as those linked to hydrocarbons (e.g., alcohol use, health status, educational level, age). Absence of nonexposed control groups matched to exposed individuals on critical variables has been a chronic failing of research linking exposure with neuropsychological dysfunction. There are other shortcomings that are as serious a threat to the integrity of these studies as lack of a control group.

Selection Bias

In cross-sectional studies that compared neuropsychological test scores of exposed and nonexposed workers, subjects presumably affected by toxic exposure were, for the most part, selected not at random from the population of interest, but from those individuals presenting themselves for assessment and treatment of neuropsychological symptoms. From the population of workers with hydrocarbon exposure, this method sampled only those with symptoms. Because of the overrepresentation of symptomatic individuals in samples of exposed workers, it is not surprising that many studies found reliable differences between exposed and nonexposed workers on tests of neuropsychological functioning. Some experts concluded that a consensus had emerged that long-term, low-grade exposure to hydrocarbons caused irreversible neuropsychological problems.

A number of early studies illustrate the biasing effect of selecting research participants from the subset of hydrocarbon-exposed workers with neuropsychological complaints. In a prototypical investigation, Linz and colleagues (1986) selected a sample of industrial painters from an occupational health clinic. These workers complained of shortness of breath, disorientation, nausea, irritability, and other symptoms. They were given a number of neuropsychological tests and their scores were compared to average normative group scores. Results showed that exposed workers complaining of symptoms scored one or more standard deviations (SD) below the normative mean on a number of neurobehavioral functions, including simple motor speed, hand strength, and auditory sensitivity. Compared to a sample of nonexposed workers, the personality profile of this sample of painters revealed elevations in depression, anxiety, and preoccupation with somatic concerns. Clinicians unaware of the sampling problems could conclude this study provided strong support for assertions that neuropsychological difficulties arise from hydrocarbon exposure.

In another study, subjects with a history of exposure to solvents were selected from among referrals to an occupational medicine clinic (Morrow, Ryan, Goldstein, & Hodgson, 1989). Their MMPI scores revealed increased elevations on a number of scales with response profiles indicating a high rate of somatic disturbances, anxiety, depression, social isolation, and fear of losing control. Additionally, the authors reported that the exposed group had clinical profiles similar to those of former prisoners of war with posttraumatic stress disorder.
In both of these examples, there was no random selection of subjects from the population of painters; only those with symptoms were used for comparison with nonexposed samples. The suggestion that mostly low-dose exposure to solvents at work causes a “distinct pattern of personality disturbance” comparable to the experience of being a prisoner of war with posttraumatic stress disorder illustrates the lack of skepticism that characterizes this literature more generally.

In several follow-up studies, the same problematic selection procedure was present. In one, patients diagnosed as having chronic intoxication caused by exposure to solvents were selected (Juntunen, Antti-Poika, Tola, & Partanen, 1982). These patients were evaluated on average 5 years after initial diagnosis. At time of follow-up, subjective signs of impairment tended to diminish (e.g., headache, tiredness) whereas clinical signs, including cerebellar functions, gait and station, and fine motorics, showed deterioration. House painters diagnosed with chronic toxic encephalopathy were selected in another follow-up study (Bruhn, Arlien-Soborg, Gyldensted, & Christensen, 1981). At time of reassessment (2 years), subjects reported slight improvements in headache and dizziness but either stability or some deterioration in neuropsychological functioning. Studies whose samples are affected by unintentional selection bias create the impression that a causal link has been demonstrated, when in fact the selection virtually assures illusory correlation between the hypothesized cause and effects.

In other studies of various designs, exposed workers were selected from those either receiving disability benefits, presenting with various symptoms, or diagnosed with chronic toxic encephalopathy or “painters’ syndrome” (Arlien-Soborg, Henriksen, Gade, Gyldensted, & Paulson, 1982; Callender, Morrow, Subramanian, Duhon, & Ristovv, 1993; Flodin, Edling, & Axelson, 1984; Gregersen, Klausen, & Elsnab, 1987; Juntunen, Hupli, Hernberg, & Luisto, 1980; Kelafant, Berg, & Schleenbaker, 1994; Lindstrom, 1973, 1980; Lindstrom, Antti-Poika, Tola, & Hyytainen, 1982; Ryan, Morrow, & Hodgson, 1988; van Vliet et al., 1989).

**Biased Recall**

Although this literature includes many uncontrolled studies, other researchers attempted to control for factors other than the hypothesized toxic exposure that could account for symptoms. Competing etiological factors included alcohol consumption, use of medication or drugs, pre-existing psychological or neuropsychological dysfunction, general health problems, life stress, involvement in litigation, work history, and so on. Persons exceeding a predetermined level of alcohol consumption, having a history of significant head injury, epilepsy, perinatal trauma, encephalitis, or meningitis were excluded from a sample of solvent-exposed workers in a study by Arlien-Soborg and colleagues (1982). Bleecker and colleagues (Bleecker, Bolla, Agnew, Schwartz, & Ford, 1991) attempted to control statistically for a number of potentially confounding factors including alcohol intake and smoking. Husman (1980) controlled for alcohol intake by matching exposed and referent groups on this variable. Linz and colleagues (1986) sought to control competing factors such as unemployment, financial stresses, and pending litigation by including a comparison group with similar stressors. In a case-controlled study, van Vliet and colleagues (1989) matched painters with construction workers on awards of disability benefits due to neuropsychiatric disorders.

These studies relied on self-reports based on subjective recall to measure alcohol intake, premorbid functioning, health status, and other confounding factors. Although self-report data are routinely relied upon in clinical interviews, they are notoriously problematic in epidemiological studies of neuropsychological issues (e.g., see Taubes, 1995). Self-report
The problem with such extensive reliance on recall is that information retrieved from memory is biased in many ways (see e.g., Fiske & Taylor, 1991). For example, a good deal of research documents the tendency of people to seek information that confirms current understanding of their situation (Swann, 1983). They commonly see more self-confirmatory information than actually exists (Swann, 1987), view neutral evidence as confirmatory (Turk & Salovey, 1988), attend more to self-confirmatory feedback, and recall preferentially information that is consistent with beliefs about themselves (Swann & Read, 1981). Specifically on point for this literature, “Recall bias . . . is especially problematic in case-control and retrospective cohort studies, since both exposure and disease have already occurred at the time participants enter into the study” (Hennekens & Buring, 1987, p. 274). Information recalled from memory that is ambiguous is often interpreted as supportive of current beliefs (Anderson & Pichert, 1978; Cantor & Mischel, 1977; Hastie, 1981; Loftus, Miller, & Burns, 1978; Snyder & Uranowitz, 1978; Taylor & Crocker, 1981) and autobiographical histories may be revised to fit present status (Ross, 1989). “Recall bias may arise because individuals with a particular exposure or adverse health outcome are likely to remember their experiences differently from those who are not similarly affected” (Hennekens & Buring, 1987, p. 35). These tendencies may manifest themselves during assessment by downplaying pre-exposure problems such as the frequency and dose of alcohol or drug consumption, for example, or by elevating levels of premorbid intellectual and emotional functioning, because they are consistent with beliefs that current symptoms were caused by exposure to hydrocarbons.

Aside from the general unreliability of memory for such measures, these studies were confounded by reliance on memory in the context of external incentives, such as obtaining disability benefits. Contrary to proper forensic procedure, many of these studies reported data from persons in litigation without assessing for malingering. The assumption that unassisted recall of autobiographical information is reliable despite substantial differential reinforcement is tenuous to say the least, and contrary to a vast literature on reinforcement of behavior. When reviewing findings from investigations based largely on self-report data, it is essential for neuropsychologists to be aware of the role biased recall plays in measuring potentially confounding variables, such as alcohol consumption, preexisting medical conditions, premorbid cognitive functioning, pre-exposure psychological problems, and the like. Whenever possible, objective sources of information such as medical records should be used.

Use of self-report data to establish the dosage of hydrocarbons to which subjects were exposed is another area in which biased recall is evident. Establishing the level and duration of exposure is critically important in exploring hypothesized relations between exposure and neuropsychological functioning (Baker, 1988; Baker & White, 1985; Brooks et al., 1995; Fidler, Baker, & Letz, 1987; Russell et al., 1990). Quantitative evidence of exposure must be established to link such exposure to behavioral deficits, and inability to demonstrate the nature and extent of exposure limits our ability to confirm dose-response relationships. In clinical evaluations also, data regarding the intensity of hydrocarbon exposure should be gathered. Ideally, in both epidemiological investigations and clinical assessments, precise measurements of intensity and duration of exposure would be obtained by continuous monitoring of exposures by known routes for individuals. These measurements would include the amount of material inhaled or absorbed through the skin. Controlled chamber studies offer the opportunity to measure exposure directly (e.g., see Nihlen, Walinder, Lof, & Johanson, 1994; Prah, Goldstein, Devlin, Otto, Ashley, House, Cohen, & Gerrity, 1994). However, most studies used self-report to estimate these quantities. Unfortunately, self-report data have been relied upon in attempts to estimate levels and duration of exposure.
To estimate exposure, investigators conducting invalid studies relied on retrospective assessments of workers and employers regarding occupational history, work environment, job duties, estimates of daily exposure, quantities of materials used, episodes of acute exposure, and so forth. Various methods have been used to obtain estimates, including data from questions on working conditions and exposure to create an index characterizing exposure (Gregersen et al., 1987); subjects' responses on pension applications (Lindström, Riihimaki, & Hanninen, 1984); a semi-structured occupational and environmental chemical exposure questionnaire administered by an occupational health specialist (Morrow, Robin, Hodgson, & Kamis, 1992); self-reported estimates of number of hours worked per week using solvents (Rasmussen, Jeppesen, & Sabros, 1993); a detailed, self-reported lifetime job history questionnaire (Callender et al., 1993); occupational history including working materials, method, and conditions (Gregersen et al., 1987); history of exposure provided by the subject, the employer, or by occasional hygienic measurements in the workplace (Lindström, 1980); number of years working around solvents (Arlien-Soborg et al., 1979); description of working conditions provided by worker and by employer, inspection of workplace, environmental measurements, and biological tests (Juntunen et al., 1982); and interviews (Ryan et al., 1988). Some studies have used biological monitoring of blood, urine, and exhaled breath to quantify absorption of selected substances (Lauwerys, 1983), but these methods have not been utilized widely.

In sum, the problem of biased recall applies to measurement of exposures as well as to confounds. This is especially true when interviewers and interviewees are not blind to hypotheses of studies. Expectations that people with certain symptoms are likely to have been exposed to particular concentrations of hydrocarbons for certain periods of time logically leads interviewers to behave toward subjects in ways that yield the expected finding, regardless of its objective accuracy. Self-fulfilling prophecy has been documented in numerous social psychological studies (Andersen & Bem, 1981; Fazio, Effrein, & Falender, 1981; Rosenthal & Jacobson, 1968; Snyder & Swann, 1978; Zanna & Pack, 1975), and its influence in clinical evaluations is suspected highly (Sarbin, Taft, & Baily, 1960; Salovey & Turk, 1991; Turk & Salovey, 1988).

There is an abundance of research indicating that the assumptions brought to the evaluation exert a profound influence on the conclusions of both diagnosticians or examiners and the patients or examinees (e.g., Arkes, 1981; Arkes, Wortmann, Saville, & Harkness, 1981; Chapman & Chapman, 1967, 1969; Golding & Rorer, 1972; Salovey & Turk, 1991; Snyder, 1977). For example, in an investigation examining the influence of disease labels on clinical judgment, Temerlin and Trousdale (1969) had psychiatrists read a benign interview of an adult, and write a brief report. The psychiatrists interpreted data elicited during the interview as signs of mental illness, apparently because prior to reading the interview they were told the person was psychotic.

It is imperative that experts evaluating patients exposed to low doses of hydrocarbons be aware of the false assumptions they are likely to bring to the evaluation if they rely on a noncritical review of the existing literature, or upon a review of the abstracts or conclusions without a critical examination of the underlying data. Specifically, by entering the evaluation with the assumption that low doses of hydrocarbons have the impact suggested by the flawed literature, an examiner may presume injury and causation that are unfounded. Furthermore, patients primed by such assumptions implicit in the clinicians' approach to the evaluation can recall and interpret information that supports such beliefs, thus appearing to provide support for what is in fact an illusory correlation. "... in studies of exposures widely believed to be detrimental ... individuals with a particular disease or other adverse health outcome might ... tend either to exaggerate their actual level of exposure if they believe it to have caused their illness, or to minimize their exposure to appear more acceptable to the interviewers or
investigators” (Hennekens & Buring, 1987, p. 35). Memories regarding not only exposure history but also prior psychological and cognitive functioning, alcohol intake, lifestyle, and other information necessary for aiding in assessment of neurotoxic effects, may be distorted by beliefs that one’s symptoms were caused by exposure to hydrocarbons. External incentives, such as disability benefits, retirement, escape from noxious work settings, avoidance of aversive responsibilities, and other factors may create perceptual biases in the patient’s information processing strategy. Failure to disclose the fact that research subjects are claimants is a fatal omission. Litigation and various forms of compensation are fertile soil for growth of response biases.

Variation in Tests and Findings

Between-study variation in assessment instruments and results is so great in this literature that it compromises the usefulness of the existing research for informing neuropsychologists. Anger (1990) reported that 250 tests were used in 185 studies of various workplace neurotoxicants. Most of the tests used have minimal, outmoded, or nonexistent norms. Because of this diversity, a database of consistent, reliable findings that can be referred to by neuropsychologists simply does not exist. Variation in assessment procedures and findings makes problematic comparisons of earlier with later studies. Test selection has been a continual, major concern to researchers (Amler et al., 1994; Anger et al., 1994; Baker et al., 1983; Russell et al., 1990). Hanninen (1990) found that among 8 studies selected for review, 38 different tests were used. The diffuse testing is paralleled by ambiguity in what symptoms to assess. Symptoms vary from one study to the next and a diversity of labels is used to characterize the effects of long-term exposure to hydrocarbons. Organic brain syndrome, toxic encephalopathy, organic solvent intoxication, painters’ syndrome, presenile syndrome, and psycho-organic syndrome, among others, have been used as diagnostic categories without specifying a common or consistent definition of terms. “Painters’ syndrome” is a good example of a label that is often used without definition, as if there was a universal agreement as to the criteria for such a designation, yet the studies that do define it do so inconsistently (e.g., see Arlien-Soborg et al., 1979; Fidler et al., 1987; Maizlish et al., 1985; Spurgeon et al., 1994). Variation in test selection and findings has characterized hydrocarbon neurotoxicity research from the beginning. In an early investigation, Hanninen and colleagues (Hanninen, Eskelinen, Husman, & Nurminen, 1976) assessed the effects of long-term exposure to mixed solvents using Similarities, Picture Completion, Block Design, and Digit Span from the Wechsler Adult Intelligence Scale (WAIS); Logical Memory from the WMS; Associate Learning; the Benton test for visual reproduction and retention; Santa Ana Dexterity Test; Finger tapping; Reaction times; choice reaction time; and the Mira test. Compared to nonexposed controls, exposed car painters showed impairments in verbal and visual intelligence, and in verbal memory.

Morrow and colleagues (1992) tested exposed persons using the Pittsburgh Occupational Exposures Test Battery (POET; Ryan et al., 1987) consisting of Wechsler Intelligence Subtests, Visual Reproduction, Verbal Associative Learning, Symbol-Digit Learning, Incidental Memory, Recurring Words, Boston Embedded Figures, Mental Rotation, Trail Making, and Grooved Pegboard. Results from Experiment 1 showed that in comparison to controls, exposed subjects: recalled significantly less information on measures of verbal and symbol-digit-paired associate learning, but there were no reliable differences following a delay; recalled less information on Logical Memories Immediate and Delayed Recall, again with no differences in percent of information recalled following a delay; and exhibited differences for all intervals on the Four-Word Short-Term Memory Test.
Lindstrom (1980) compared the performance of exposed workers to controls on subtests of the Wechsler Adult Intelligence Scale (similarities, digit span, digit symbol, picture completion, and block design), and on other performance tests such as Visual Memory, Benton Visual Retention, Symmetry Drawing, Santa Ana Dexterity, and the Mira. Declines in visuomotor performance and decreased freedom from distractibility characterized the solvent exposed group. A relation was found between duration of solvent exposure and poor visuomotor performances.

Maroni and colleagues (Maroni, Bulgheroni, Casitto, Merluzzi, & Gilioli, 1977) used a test battery including Raven Progressive Matrices, Picture Completion and Block Design from the WAIS, Pauli test and Symbol-Number Association, to test for differences between exposed and nonexposed, female factory workers. No significant differences were observed between these groups.

Spurgeon and colleagues (1994) used the Neurobehavioral Evaluation System (NES; Baker & Letz, 1986), which included Paired associate learning, Hand-eye coordination, Symbol-digit substitution, Digit-span, Pattern memory, Continuous performance, Color word vigilance, and Associate recall. No dose-effect relations were found for paintmakers exposed to varying levels of solvent. There were differences between paintmakers and controls on the continuous performance and color word vigilance tests for workers in medium and high exposure groups.

To address test selection that tended to be idiosyncratic (Hanninen, 1990), several workshops were convened to develop systematic strategies for investigating the chronic effects of long-term occupational and residential exposure to toxic substances. Among these were the Workshop on Neurobehavioral Testing convened by the Agency for Toxic Substances and Disease Registry in 1991 and the World Health Organization meeting in 1985. From the 1991 Workshop on Neurobehavioral Testing came recommendations that a core and a secondary set of functions be tested (Amler et al., 1994; Anger, 1990; Anger et al., 1994). Core functions included cognitive (learning and memory, coding, sustained attention, intelligence), motor (strength, coordination, speed), sensory (vision, somatosensory), and affect (mood). The domains tested in the Neurobehavioral Evaluation System (NES; Letz & Baker, 1986) included psychomotor performance, perceptual ability, memory and learning, cognitive, and affect. The NES has been applied internationally for research screening purposes, and has not been generally adopted by neuropsychologists in clinical settings. There has not yet been time to evaluate the more recently proposed test batteries. As with other areas of neuropsychology, for example, the assessment of traumatic brain injury, there is no general consensus on a battery of tests best suited for evaluating the neuropsychological consequences of long-term exposure to hydrocarbons.

**SUMMARY AND CONCLUSIONS**

The equivocal nature of findings from research on the neuropsychological impact of hydrocarbon exposure should give pause to neuropsychologists as they assess individual patients. Reliance on the conclusions expressed in the existing literature may lead experts to confuse speculation with empirically founded attributions of neuropsychological deficits to low dose hydrocarbon exposure. For the neuropsychologist seeking to draw reasonable conclusions, a number of flaws in research detract from this effort. Included in the shortcomings are selection bias in recruitment of research subjects, overreliance on recall in determining levels and duration of exposure, between-study variability in substances examined and in tests used to assess neuropsychological functioning, and differences in reported findings. The combination of design flaws and inconsistency of results compels skepticism.
of the claim that acute or long-term, low-level exposure to hydrocarbons causes chronic neuropsychological deficits.

Persons suspected of chronic, low-level exposure to solvents have too often been drawn selectively and nonrandomly from medical clinics, from the rolls of workers seeking to win or maintain disability benefits, or from plaintiff populations. These selection procedures deselected those exposed workers without symptoms, thus skewing the exposed sample in the direction of dysfunction. Biased selection also discarded the protection that random sampling provides against systematic confounds that may have existed in persons visiting medical clinics, suffering from symptoms, on disability, or seeking compensation.

To establish dose-response relationships, accurate quantitative evidence of exposure to target materials is essential. Because this empirical evidence was not available, many studies have relied on worker self-reports of exposure history to estimate exposure levels. A major source of error is that information retrieved from memory often is biased, especially in circumstances confounded by incentives or cognitive biases.

Between-study variation in assessment procedures, instruments, and definitions complicates efforts to replicate findings or develop a base of consistent findings. One researcher reported that in 185 studies exploring workplace exposure to chemicals (e.g., solvents, mercury, lead), 250 different tests were administered to exposed workers (Anger, 1990). Additionally, differential results have been reported. Many studies have reported psychological and neuropsychological deficits but the deficits reported are inconsistent. Others have not found these deficits. Studies that have reported deficits often differ in the kind and intensity of dysfunction.

The confluence of methodological weaknesses reviewed here compel one to question sharply whether the weight of research evidence supports a causal link between acute or repeated, low-level exposure to hydrocarbons and neuropsychological dysfunction. We recognize the difficulty of researching this area, but until methodological issues like those outlined herein are addressed adequately in the future, uncertainty will continue to cloud this important topic. The controlled studies in which no association was observed are especially important in suggesting that we consider alternative explanations for positive findings.

How can we improve? There are several straightforward improvements that should be available to researchers with the same resources as those required to conduct studies discussed herein. Selection bias can be substantially corrected by studying randomly selected members of the exposed population of interest instead of only those members complaining of symptoms. If low doses of hydrocarbons cause cognitive deficits, then a comparison of randomly selected exposed petroleum workers, painters, service station employees, and so on with cognitively comparable nonexposed workers should show differences.

Another avenue of improvement is for investigators to test and discuss the logic of their own assumptions — particularly assumptions that have been challenged in critical reviews such as this. For example, investigators who publish measures based on subjective recall are in effect asserting that subjective recall is a reliable and valid method for making publishable research measurements. These same investigators apparently believe that greater doses of hydrocarbons are associated with greater neuropsychological impairment. If these two assumptions are true, then future articles should contain comparisons of the scores of high, moderate, and low dose groups, and the high dose groups should have the lowest scores.

Articles should provide more data about the meaning and value of the results. Mere significance testing offers little to applied practitioners, for example. Articles should report point estimates with confidence intervals and should report the overlap between experimental and control groups on important variables. Much of the literature on hydrocarbon neurotoxicity deals with such subtle differences that so-called "injured" and "normal" populations overlap more than they diverge. The strength of the association should be reported, not
simply the fact that there appears to be a statistically significant association. Hit rates and error rates should be reported in validation studies which purportedly offer the reader tests or batteries to apply to the neuropsychological evaluation of hydrocarbon-exposed patients.

Studies addressing the most common complaints, for example, memory, should be replicated using the identical test, and should be conducted with tests that examine memory function in more detail, instead of brief screening tests, for example, the Wechsler Memory Scale-Revised (Wechsler, 1987) or Memory Assessment Scales (Williams, 1991), to see if the same results are found. Instead of relying on the unreliable — subjective recall of exposures and dosage — more studies are needed in which air monitoring and blood sampling measures are integrated with neuropsychological testing conducted real time. As Kelly and Filley concluded:

It is clear that extensive clinical studies, using standardized psychological testing, carefully selected control groups, and neuropathological information whenever possible, are needed to clarify the prevalence, nature, severity, and pathophysiology of toxic neurobehavioral syndromes. Neurologic, neuropsychologic, electrophysiologic, and neuroimaging studies will all play a prominent role in these investigations. (1992, p. 154)

REFERENCES


