HSE
Health & Safety Executive

The validity and interpretation of neurobehavioural data obtained in studies to investigate the neurotoxic effects of occupational exposure to mixtures of organic solvents

The feasibility of a benchmarking approach to interpretation

Prepared by
The Institute of Occupational Health
The University of Birmingham
for the Health and Safety Executive

CONTRACT RESEARCH REPORT
355/2001

HSE
Health & Safety Executive

The validity and interpretation of neurobehavioural data obtained in studies to investigate the neurotoxic effects of occupational exposure to mixtures of organic solvents

The feasibility of a benchmarking approach to interpretation

Dr Anne Spurgeon
The Institute of Occupational Health
The University of Birmingham
Edgbaston
Birmingham
B15 2TT
United Kingdom

Concerns continue to be expressed about the validity and significance of neurobehavioural data in determining the neurotoxic effects of occupational exposure to organic solvents. The present report considered (i) the strength of the evidence for neurobehavioural effects, (ii) the feasibility of evaluating the significance of effects by reference to the size of effects observed following exposure to other agents, (benchmarking). In Section I 45 studies were selected and evaluated according to pre-determined criteria. Approximately 80% of all studies reported statistically significant effects of long-term, low-level solvent exposure on some aspects(s) of cognitive functioning. Approximately 50% of studies also reported an exposure-effect relationship. Percentages were similar when methodologically better studies were considered separately. However a number of interpretative issues relating to the pattern and level of effects and the comparability of data between studies were identified. Section II considered the feasibility of using each of the following agents to derive data for benchmarking purposes; carbon monoxide exposure, alcohol consumption, sleep deprivation, normal ageing and head injury. Few current data relating to these agents would provide useful reference points for solvent-related effects. However, data sets could in principle be developed for this purpose in relation to alcohol consumption, head injury and normal ageing.

This report and the work it describes were funded by the Health and Safety Executive (HSE). Its contents, including any opinions and/or conclusions expressed, are those of the author alone and do not necessarily reflect HSE policy.

HSE BOOKS

First published 2001

ISBN 0 7176 2060 3

This report incorporates the outcome of the workshop entitled 'The validity and interpretation of neurobehavioural data obtained in studies to investigate the neurotoxic effects of occupational exposure to mixtures of organic solvents' held at the University of Birmingham on the 1 and 2 February 2001.

As such it consists of the original report compiled by Dr Anne Spurgeon, amended and updated to include the comments, conclusions and recommendations of the workshop participants.

Workshop Participants

Dr Len Levy	(Chair Day 1)
Prof. Malcolm Harrington	(Chair Day 2)
Dr Gus Baker	
Penny Barker	
Julie Bates	
Dr Sally Clapp	
Dr Richard Elliott	
Dr Alastair Hay	
Craig Jackson	
Peter Kelly	
Dr Colin Mackay	
Dr Anne Spurgeon	
Bal Sreenivasan	
Martin Stear	
Dr Brian Stollery	
Carole Sullivan	
Dr Michael Topping	
Kath Travis	

Summary

Psychological performance tests have been used for a number of years in largescale studies to investigate the neurotoxicity of solvents in occupational settings. Concerns have been expressed about the methodological quality of many of these studies and about the validity and significance of the data reported. The present report addresses two issues in relation to these concerns. Section I investigates the strength of the evidence that long-term, low-level occupational exposure to solvents can result in neurobehavioural effects, which are less severe than those which constitute chronic toxic encephalopathy (CTE). Section II discusses the feasibility of evaluating the significance of any demonstrated effects by reference to the size of effects observed following exposure to other agents. This approach is referred to as benchmarking. In Section I 45 studies were selected and evaluated according to pre-determined criteria and their results assessed. Approximately 80% of studies reported statistically significant effects of long-term, low-level solvent exposure on some aspects(s) of cognitive functioning. Approximately 50% of studies also reported an exposure-effect relationship. These percentages were slightly higher when the methodologically better studies were considered separately. However, there was no consistent psychological pattern in terms of the effects observed and only limited evidence that studies of workers with longer or higher exposures reported effects more frequently than did other studies. A wide range of tests and scoring methods were used making it difficult to compare results between studies. In addition, reference to clinical criteria for psychological significance in some studies cast doubt on the validity of the conclusions reached by researchers. Overall, the quality of current evidence is suggestive of neurobehavioural effects, but many questions remain about the nature of those effects and their biological/psychological significance. Section II considered the feasibility of using each of six agents to derive data for benchmarking

purposes. These were, acute/chronic effects of carbon monoxide exposure, chronic effects of alcohol consumption, acute effects of alcohol consumption, acute effects of sleep deprivation, normal ageing effects and a clinical classification system for the severity of head injury. Examination of existing data relating to each of these agents indicated that data on alcohol consumption, head injury and normal ageing effects might yield useful reference scores for benchmarking purposes. Currently few of these data are available in a form which would allow evaluation of solvent-related effects. However, there are indications that useful data sets could in principle be developed for this purpose if considered appropriate.

Contents

1. Background

Psychological performance tests (neurobehavioural tests) have been used since the mid 1960s as a method of determining the neurotoxicity of chemicals in occupational and environmental health. In particular such tests have been used to assess the effects of occupational exposure to organic solvents, usually in cross-sectional studies designed to compare the scores of groups of exposed workers with those of matched controls. It has been noted however that many of the tests employed in such studies were originally developed for use in a clinical diagnostic setting and were not intended for use in epidemiological research. In the occupational field, although they were originally employed in a clinical context, to assess possible cognitive impairment in individuals who had been exposed to neurotoxicants, their use was subsequently extended, perhaps inappropriately, to large scale studies which investigated group effects of certain occupational exposures, notably organic solvents. The validity of the data which emerged from these studies has therefore come under scrutiny partly because of methodological concerns and also because of difficulties of interpretation. The assessment of the biological or 'real-life' significance of statistically significant differences in scores between exposed and control groups is not straightforward and thus the application of the data to regulatory toxicology is somewhat problematic.

A discussion recorded in the minutes of WATCH (February 1997) considered a number of approaches to address the question of interpretation and general support was given to the proposal that a form of benchmarking should be explored as a possible solution. It was suggested that the size of any effects identified in relation to organic solvent exposure could be assessed in relation to those identified in relation to other exposures or factors known to impair performance, thus providing a reference point for interpretation of the data. The

Health and Safety Executive (HSE) therefore undertook an initial review (i) to investigate the validity of neuropsychological assessment in identifying neurobehavioural effects of organic solvent exposure, and (ii) to explore the feasibility of benchmarking, i.e. comparing neurobehavioural test data from organic solvent studies against neurobehavioural test data recorded in relation to other agents or factors known to affect nervous system functioning. This review (Kelly 1999) contained the following elements:

(i) a literature search on defined relevant databases using defined search terms

(ii) the selection of studies, according to defined criteria, for inclusion in a data evaluation exercise

(iii) the recording of the frequency of testing of particular functional domains and use of particular tests in selected studies

(iv) the recording of the frequency of results showing statistically significant differences between test scores on the six most frequently used tests

(v) an evaluation of group scores of organic solvent exposed workers in these six tests against certain criteria for clinical significance

(vi) a preliminary evaluation of the feasibility of using data obtained in relation to carbon monoxide exposure and long-term alcohol abuse for benchmarking purposes

The current review aims to build on this work and will incorporate some of the above information into a further exploration of the issues. Specifically the current review is intended to address the following questions:

1. Is there good evidence that occupational exposure to organic solvents can result in adverse effects on the nervous system, which are less severe than those included in the EU definition of CTE, but which are nevertheless discernible from the results of neurobehavioural tests?

2. Can existing neurobehavioural data relating to other agents or factors known to affect nervous system functioning provide reference points for the interpretation of the size of any effects observed in relation to organic solvent exposure?

The report is therefore divided into two sections. Section I contains a systematic review of the evidence for neurobehavioural effects, in terms of the methodological quality of the data, the consistency of observed effects and their size relative to current criteria for clinical (non-normal) significance. The focus is on studies of mixed solvent exposure rather than those which investigate the effects of a single solvent. This section of the report draws on the initial literature search conducted by HSE in preparing the previous report (Kelly 1999), and also includes reference to some of their descriptive data and analysis relating to the clinical significance of test scores. Where information from the previous report is included it is acknowledged accordingly.

Section II contains an evaluation of the feasibility of a benchmarking approach using data relating to the following:

- long-term sequelae of high level carbon monoxide exposure
- effects of long-term alcohol abuse
- acute effects of alcohol consumption
- effects of sleep loss
- normal ageing
- head injury

Information and conclusions relating to carbon monoxide exposure and long-term alcohol abuse is derived from the previous HSE report (Kelly 1999) and has not been specifically investigated for the current report. Its source is acknowledged accordingly.

2. Section I - The evidence for neurobehavioural effects

2.1 Methods

2.1.1 Literature search strategy❖

Terms used for the searches were as follows:

Neurobehavioural or behavioural or neuropsychological❖❖ or neuropsychiatric or cognitive or mental or memory/memories or thought disorder or amnesia or dementia
and
paint or organic solvents

The databases searched were as follows:

- HSELINE
- CISDOC
- NIOSHTIC
- MEDLINE
- PSYCHOINFO
- EXCERPTA MEDICA

A panel of experts was assembled by HSE to develop a set of exclusion and inclusion criteria to apply to the abstracts identified (Tables 1-2).

Because the initial search (carried out by HSE) only included papers published up to 1997 a further search covering the period 1997 - 2000 was carried out using the same databases and search terms but excluding, the terms “thought disorder, amnesia or dementia” which were not considered relevant by the author.

❖ initial search conducted by P Kelly, HSE; subsequent search carried out by the author

❖❖ the terms neurobehavioural, behavioural and neuropsychological are regarded as synonymous throughout the report

Table 1. Inclusion Criteria[❖]
Long-term occupational exposure to solvent mixtures investigated
Cross-sectional or longitudinal design
Differences between groups investigated
Estimated exposure levels included
Neuropsychological tests used (automated or manual)

Table 2. Exclusion Criteria[❖]
Short-term rather than long-term exposure (acute effects)
Solvent abuse rather than occupational exposure
Comparison of two solvents with no control group
Investigation of a single solvent
Investigation of toxins other than organic solvents
Comparison of solvents and other toxins
Controlled laboratory studies rather than field studies
Intervention studies
Investigations of recovery from solvent-related effects
Neuropsychiatric and affect effects rather than neurobehavioural
Neurophysiological effects rather than neurobehavioural
Neuropathological effects rather than neurobehavioural
Investigations of hearing loss
Investigations of symptoms rather than neurobehavioural effects
Studies using non-neurobehavioural tests
Studies failing to name the tests used
Studies failing to publish actual test scores
Studies reporting correlations rather than differences between groups
Case studies
Animal studies
Review papers rather than original studies
Duplicate publications of the same data
Evaluations of neuropsychological tests
Comparisons of computerised and manual tests
Non-English language[❖❖]

Studies were selected from both searches for inclusion in the review by application of these exclusion and inclusion criteria.

❖ adapted from those developed by HSE

❖❖ use of English language papers only was considered adequate since only 4 non-English language papers were identified in a search using the same terms and databases

2.1.2 Content of Data

Following the selection of studies a set of variables were specified to represent the information to be recorded from the selected papers (Table 3). Tables were constructed using these variables to summarise the content of the selected papers.

Table 3. Variables Included
Principle author
Year of publication
Country of study
Method of subject selection
Response rates
Number in exposed group(s)
Number in control group(s)
Gender
Age of exposed group(s) (mean and/or range)
Age of control group(s) (mean and/or range)
Highest exposure level of exposed group(s) (mean and/or range)
Exposure time of exposed group(s) (mean and/or range)
Solvents identified
Method(s) of assessing exposure
Control for last exposure stated
Confounders/modifiers accounted for
Control for test conditions stated
Method of test administration
Significant performance difference between groups observed in at least one test
Exposure-effect relationship observed in relation to at least one test

Variables in Table 3 included only basic information on (i) whether a statistically significant difference between exposed and control groups was observed in relation to any test outcome measure, and (ii) whether an exposure-effect relationship was observed in relation to any test outcome measure.

A further analysis was therefore carried out to describe:

(i) the number and identity of tests used in each selected study

(ii) the results in relation to each test (statistically significant differences between exposed and control groups and/or exposure-effect relationship)

This information was also summarised in tabular form.

2.1.3 Quality of Data

A set of criteria were constructed in order to rank selected studies in terms of quality. These were derived from the EU criteria document (Institute of Occupational Health 1997). This allowed an assessment of the data in terms of the number of positive results recorded in high or medium quality studies only.

2.1.4 Exposure

Studies were classified in terms of reported exposure (> or ≤TLV/OES and <10 or 10+ years) and an assessment was made of the number of positive results in each case. The assessment of > or ≤ the TLV/OES was made in relation to current occupational exposure levels in the UK rather than the researchers own assessment since (i) exposure standards in the country in question might differ from those in the UK, and (ii) may have changed since the study was carried out.

2.2 Results

2.2.1 Literature search

A total of 688 abstracts were identified from the initial search. After application of inclusion/exclusion criteria this was reduced to 39 relevant papers containing 40 relevant studies. The subsequent search (1997 - 2000) identified a further 5 relevant papers, (5 studies)

2.2.2 Descriptive data

Data from the 45 selected studies were summarised in tabular form according to the variables specified in Table 3. This summary is shown in Annexe 1.

The majority of studies were carried out either in Western Europe (18 studies) or the USA (15 studies). Of the remainder 6 were carried out in the Far East, 2 in India, 2 in South Africa, 1 in South America and 1 in the former Yugoslavia. On the basis of the frequency of reporting relatively large scale studies, there appears to be no obvious lessening of interest in this subject in the USA or Western Europe as a whole. Since 1995 for example there have been 6 studies emanating from the USA and 3 from Europe. However, no large scale study has been carried out in the UK since 1994.

Taking the studies as a whole 32 report a statistically significant difference between the performance of the exposed and control groups on at least one psychological test. A further seven studies were designed in such a way that they did not include a control group and therefore this assessment was not applicable. Of these, three studies demonstrated significant differences between the performance of the lowest and highest exposed groups. On this basis therefore approximately 78% of studies report an effect of solvent exposure on test performance. However, a number of points should be made:

(i) there is a well-accepted bias towards the reporting and publication of positive studies which may well be operating here.

(ii) only 54% of the positive studies who also investigated exposure-effect relationships were able to demonstrate such a relationship between a measure of long-term, cumulative "solvent dose" and test scores, either in the exposed group or in the difference scores between exposed and controls. Clearly the demonstration of an

exposure-effect relationship considerably strengthens the findings, particularly in cross-sectional studies.

It should be noted, however, that few of the early studies attempted to investigate exposure-effect associations, probably reflecting a lack of appropriate methodology to carry out retrospective exposure assessment. In recent years such assessments have become more sophisticated and more data derived from regular workplace monitoring are available, particularly in large organisations which are usually involved in these types of studies. Of the 13 studies reported since 1995 seven report an exposure-effect relationship and only three studies which find a significant difference between exposed and control group performance are unable to demonstrate such an association.

(iii) the quality of the studies is variable and equal weight cannot therefore be attached to all. This point is addressed further below.

(iv) A wide variety of tests have been used and consistency of positive results is not necessarily reflected in consistency of effects on particular tests. This is also discussed in more detail below.

2.2.3 Study quality

Drawing on the EU document[1] which describes criteria for evaluating neurobehavioural studies, a set of criteria were established to permit classification of studies as I (superior), II (good), III (fair) and IV (poor).

Criteria are shown in Table 4.

Table 4. Study Classification Criteria
1. Population of adequate size relative to the number of tests used. Usually >40 per group❖
2. Adoption of a subject selection method which avoids bias for the exposed group
3. Adoption of a subject selection method which avoids bias for the control group
4. Pre-stated exclusion/inclusion criteria for study participants
5. High response rate for the exposed group. Usually >60%
6. High response rate for the control group. Usually >60%
7. Control or adjustment for important confounders/modifiers of performance, notably age, gender, social class or job type, educational level or initial intelligence and alcohol consumption
8. Inclusion of quantitative or semi-quantitative assessment of long-term exposure
9. Control for recent exposure
10. An indication of standardisation of testing conditions

For a study to be classified as I it was required to fulfil all the above criteria. Studies were classified as II if they fulfilled at least 7 of the criteria and as III if they fulfilled at least 5 of the criteria. Studies failing to fulfil at least 5 of the criteria were classified as IV.

Classification of studies according to this system is shown in the first column of Annexe 1.

I	****	(score 10)
II	***	(score 7+)
III	**	(score 5+)
IV	*	(score <5)

On the basis of this classification system only one study, (that of Daniell *et al*, 1999) was classified as I Superior. A further 28 studies were classified as II Good. It is noteworthy that eleven of the fourteen studies carried out since 1995 were rated as I or II, reflecting a general improvement in methodology. Of those studies rated I or II a total of 21 (72%) reported a statistically significant difference between the performance of exposed and control groups or

❖ this definition of an "adequate size" is somewhat arbitrary however since it should be noted that only one study reports power calculations

between the performance of highest and lowest exposed groups on at least one psychological test. Fourteen of the 25 higher rated studies who investigated exposure-effect relationships, demonstrated such a relationship. Where only better studies are considered therefore the percentage of positive results (72%) and exposure-response relationships demonstrated (56%) is very similar to the percentage for all studies (78% and 54% respectively).

2.2.4 Exposure

Current opinion tends to the view that neurobehavioural effects are unlikely to occur until exposure has exceeded 10 years duration. Other authors have suggested that chronic effects on the nervous system can be prevented if exposures remain below levels intended to protect workers from acute narcotic or irritancy effects (Spurgeon *et al* 1994). An attempt was made therefore to determine whether positive results tended to occur more often in studies where exposures had, at least some of the time, exceeded currently accepted exposure standards and/or exposure had continued for more than 10 years for at least some of the subjects.

The majority of studies included some form of quantitative exposure assessment. Approximately one third based their assessment on current air and/or biological monitoring. Approximately two thirds attempted an assessment based on job history, producing either a purely descriptive or a quantitative assessment. Some of these were expressed in a form which made it impossible to identify highest exposure levels or longest exposure time. However of the 22 studies which provided some information on both of these, the results relating to exposure levels (expressed in terms of current UK occupational exposure standards) and duration are shown in Table 5.

Table 5. Results in relation to exposure levels and duration (22 studies)

Highest Exposure	Longest Exposure < 10 years		Longest Exposure 10 + years	
>OES*	number of studies	5	number of studies	9
	positive effect	3	positive effect	9
	dose-effect	1	dose-effect	6
≤OES*	number of studies	2	number of studies	6
	positive effect	2	positive effect	4
	dose-effect	2	dose-effect	1

*current UK

These data show that as expected effects, including exposure-effect relationships, are reported more consistently where exposure levels are highest and more than 10 years. However, the trend for the other cells in the table does not conform clearly to the expected pattern. For example, both studies involving exposures below the current TLV for less than 10 years report effects and exposure-effect relationships.

It should be emphasised however that in all these studies exposure was to a solvent mixture reducing the likelihood that consistent exposure-effect relationships across studies will be demonstrated. Further, exposure assessments in studies of this type rarely, if ever, take account of patterns of exposure in terms of peaks, steady state or periods of zero exposure and thus at best represent very crude estimates.

2.2.5 Test data

A list of the tests used in different studies and the results relating to these are summarised in Annexe 2, Tables 1-9. For convenience these have been grouped loosely into tests of different functional domains. However, it should be noted that there is considerable overlap between tests in this respect and no tests can be considered a “pure” measure of one function. The most striking feature is the wide variety of tests used, many of which are employed only

in a single study. In addition, where tests have been used in several studies the results are inconsistent. This is clearly illustrated by reference to the most commonly used test, the Digit Symbol Substitution Test from the Wechsler Adult Intelligence Scale – Revised (WAIS-R). This appears in a total of 25 studies, 15 of which demonstrate a significant difference between exposed and control group performance and only two of which demonstrate an exposure-effect relationship.

All studies use at least four tests and most use more than this. The average number of tests used was eight, although a number of studies derive more than one outcome from each test. The largest number of tests used in a single study was 21 and the largest number of test outcomes 31. Despite the inevitability of multiple comparisons (and hence increased risk of chance significance) only three studies report the use of Bonferroni correction. One further author made reference to the problem of multiple comparisons but argued that interpretation of the data in terms of functional areas (patterns of scores) rather than individual test results, countered this criticism. Inspection of the data as a whole, however, does not support the view that it can be meaningfully interpreted in this way.

An attempt to group results by areas of function yielded the following results.

Table 6

Functional area	Frequency of tests employed (studies)	% showing positive effect
General	43	58
Memory (WAIS-R, WMS)	23	65
Memory (other)	63	40
Learning	28	39
Motor speed/control	50	30
Visuo-spatial organisation	41	56
Vigilance/attention	24	38
Reasoning	11	18
Reaction time	24	38

It would appear that there is very little consistency and no discernible pattern in the test results. In addition to problems of inconsistency there are also concerns about the interpretation of the results of tests from the WAIS-R and the Wechsler Memory Scale (WMS) which are the most frequently used tests. These concerns were discussed in the HSE study/report and are summarised below.

2.2.6 Presentation of scores from WAIS-R❖

The HSE study/report carried out on analysis of the scores which were reported in a number of studies which used tests from the WAIS-R. This is a neuropsychological test battery used widely in clinical settings to assess intellectual functioning.

It is usual when scoring this test to convert an individual's raw score into a "scaled score" which is essentially a score which has been weighted to allow for effects of age and gender. Thus each individual obtains a scaled score on each test which can be related to an "average" scaled score of 10. Since the WAIS-R contains ten tests, an individual with an "average IQ"

❖ detailed data relating to this section are presented in the HSE report, pages 8-13.

would score 10 on every test, thus obtaining a total scaled score of 100. In fact any scaled score within the range 90 - 110 is considered to fall within the average IQ range. Clinically, psychologists tend to regard test scores which fall beyond one standard deviation of the mean to be indicative of evidence of neuropsychological impairment, either over all scores or for individual test scores.

In many cases, where WAIS-R tests have been used in solvent studies, raw scores rather than scaled scores have been reported. Thus questions arise about the validity of assessing effect sizes on this basis. It can be shown, for example, that where raw scores are converted to scaled scores they frequently fall within one standard deviation of the mean, with a minority of scores falling outside this. It could be argued therefore that few results are truly indicative of an effect as assessed by clinical standards. The lack of scaled score conversion is of concern, particularly since it raises questions about the familiarity of the investigators with the psychometric tools they are using. However, certain other factors should perhaps be considered in relation to this:

(i) Scores reported in solvent studies are mean scores relating to groups and therefore conversion of this mean to its corresponding scale score will not adequately reflect the performance of the individuals within the group. To address this problem properly would require conversion of each individual's score to a scaled score before summation and subsequent analysis. Clearly it is impossible to carry out this exercise retrospectively without access to individually-based raw data.

(ii) Investigators carrying out solvent studies might have some justification for reporting only raw scores where they are dealing with large groups of people in an epidemiological context, and have been careful to carry out age and gender matching of groups. It might be argued that such matching, especially if it also includes

educational level, is superior to the use of scaled scores which may have been obtained by reference to a culturally dissimilar population and also obtained many years before.

This said, however, the use of non-standard scoring procedures appears to be a general problem in this field of work. At the very least this makes it difficult to compare the results of different studies, even where they purport to use the same tests and where these tests have established procedures for administration and scoring.

3. Summary Statements - Section I

1. Examination of the results of 45 cross-sectional studies shows a majority (35 studies or 78%) found significant neurobehavioural effects of solvent exposure, in at least one test.

2. Where studies rated either superior or good (according to pre-determined criteria) are considered separately, the evidence is consistent with (1) above (21 out of 29 studies or 72% demonstrated an effect).

3. For all studies 54% (19 out of 35 studies) report an exposure-effect association. For higher rated studies this figure rises to 56% (14 out of 25 studies)

4. The high number of studies reporting an effect may reflect a reporting bias in favour of positive studies.

5. There was an indication that studies reporting the longest exposure (> 10 years) and the highest exposures (>TLV at least some of the time) reported effects more frequently. There was no consistent pattern in relation to shorter or lower exposures. However methods for retrospective assessment vary in quality and the use of relatively crude exposure assessments reduces the likelihood of such a pattern emerging.

6. There was no observable consistent pattern in terms of effects on particular tests. However, the wide variety of tests, test outcomes and scoring methods employed make it difficult to compare results between studies.

7. There are concerns about the appropriate scoring and data interpretation of tests normally employed in a clinical setting. In some cases this may invalidate the conclusion of a “significant” difference between the scores of exposed and control groups.

8. There are concerns about the (low) sample sizes in many studies and, alongside this, the use of large numbers of tests. Only one study reports power calculations and very few include correction for multiple comparisons of outcome measures.

3.1 General Conclusions

Notwithstanding a number of methodological concerns the balance of the evidence does suggest that neurobehavioural effects may occur following long-term occupational exposure to solvent mixtures. However the clinical significance of these effects remains to be defined. More recent studies (since 1995) show evidence of improved methodology compared with earlier investigations. However, this is largely in respect of epidemiological concerns, such as subject selection, response rates and control of confounders, and in relation to retrospective exposure assessment which is also becoming more sophisticated. There is little evidence of any parallel improvements or new developments in test outcome measures to assist evaluation and interpretation of the data. One possible approach to this is discussed in Section II.

4. Section II - The feasibility of a benchmarking approach

4.1 Benchmarking

The previous section has been concerned with the quality and consistency of the data purporting to demonstrate neurobehavioural effects occurring in response to solvent exposure. A separate but related question concerns the interpretation of the data in terms of the biological or psychological significance of statistically significant effects. This is also related to what may be termed "social" significance, i.e. the importance of any demonstrated effects for the individual in terms of everyday functioning, and for society in terms of acceptable risk.

A possible approach to this question lies in the use of benchmarking. In this context benchmarking refers to the comparison, in terms of size and perhaps nature, of effects associated with solvent exposure with those which have been demonstrated in relation to other factors known to impair nervous system functioning. This approach currently appears to be generating some interest particularly in relation to the application of widely accepted standards for alcohol consumption and driving. For example in relation to sleep deprivation Williamson and Feyer (2000) argue "…by comparing the change in performance due to alcohol consumption at concentrations widely agreed to be hazardous, (0.5% blood alcohol concentration), with the same behaviour after sleep deprivation, it should be possible to assess the amount of sleep deprivation at which equivalent deficits occur."

Application of this approach in relation to solvent exposure has not so far been explored in any detail. An immediate question arises as to which particular agents would be useful markers with which to compare solvent-related effects. In this and the previous HSE report

the choice of agents has been largely determined on an intuitive basis. This has been in terms of agents which are known to affect nervous system functioning, which may have legal or socially accepted standards attached to them (as in the case of alcohol consumption and driving) and which may have scientific data associated with them which provides consistent evidence of effects. On these bases the following agents were initially selected:

- acute or chronic effects of carbon monoxide exposure
- chronic effects of alcohol consumption
- acute effects of alcohol consumption
- acute effects of sleep deprivation
- effects of normal ageing

In addition, a further factor, that of head injury was added during workshop discussions.

The assessment of each of these agents as possible candidates for benchmarkers in relation to solvent exposure are described below.

4.2 Carbon monoxide exposure

The data relating to carbon monoxide exposure was discussed as part of the HSE study/report (Kelly, 1999). Much of the data has arisen as a consequences of self- inflicted life-threatening intoxication. As such, information is frequently presented in the form of individual case-studies or case-series and often relates to individuals with severe psychological problems. Although neurological impairment frequently occurs following intoxication the pattern, severity and prognosis associated with this appears to be very variable depending on individual differences as well as dose. In addition to case histories there is a moderate literature on small-scale volunteer studies involving administration of neuropsychological tests following controlled exposures. This literature has been reviewed

by Green *et al* (1998) who noted that many of the studies were of questionable value due to low subject numbers, non-standard experimental designs and lack of detail relating to statistical analysis. While many of these studies appear to use similar tests to those used in solvent studies therefore there are concerns about the quality of the data and it seems unlikely that they would be sufficiently reliable to be used for benchmarking purposes.

4.3 Chronic effects of alcohol consumption

Information relating to impairment associated with chronic alcoholism was also reviewed as part of the HSE study/report (Kelly, 1999). A total of 95 papers were selected for examination on the basis that they employed at least one of the six tests most commonly used in solvent studies. A sample of these papers are described in the HSE report.

Initial inspection of the data suggested that it was inconsistent in terms of demonstrating effects on specific tests. Further, when the results of individual tests were examined and scores were scaled (as opposed to presentation of raw scores - see Section I) cognitive impairment in many cases appeared to be very minor or even non-existent. Since these data were derived largely from individuals with a history of long-term alcohol abuse it would seem implausible that some effects would not occur. Moreover, alcohol consumption has sometimes been suggested as the most suitable model for predicting the effects of long-term solvent exposure. If no effects occur following heavy prolonged use of alcohol it seems less likely that relatively low exposure to solvents in an occupational setting would produce such effects.

A possible explanation for the negative results in relation to alcoholism however lies in the use of cross-sectional data. Since the individuals concerned were assessed at one point in

time, rather than on a longitudinal basis, it is quite possible that impairment relative to earlier functioning had occurred. Thus in individuals of higher or superior initial ability, 'normal' scores represent a considerable drop in functional level. For this reason assessment of the data on a test by test basis across studies may mask the presence of a consistent effect. In the present context therefore it may be more appropriate to examine the pattern of test results in each individual or group of individuals. In particular, where this is available, the results of a 'hold' test (i.e. test less susceptible to nervous system insult) will yield important information about earlier levels of functioning which can be set against current performance on other tests. Studies which have used the WAIS are likely to provide this information, for example in the form of vocabulary and other verbal tests.

While, therefore, it is clear that there are likely to be considerable psychosocial differences between occupational populations and those included in studies of alcohol abuse it would nevertheless seem prudent to re-examine the data relating to the latter group to determine (i) whether there is good evidence of impairment, and (ii) whether a consistent pattern of impairment can be identified which might represent a more severe form of that associated with long-term occupational solvent exposure.

4.4 Other agents

Examination of the data in relation to the other agents (acute effects of alcohol, sleep deprivation and normal ageing) was carried out as part of the current project and is reported below. The additional suggestion, relating to the use of data on head injury is also discussed briefly.

4.5 Methods

4.5.1 Literature search strategy

Three separate searches were carried out in respect of data relating to:

(i) acute effects on cognitive functioning of alcohol consumption

(ii) acute effects on cognitive functioning of sleep loss

(iii) effects of normal ageing on cognitive functioning

Search terms employed were as follows:

(i) alcohol, cognitive function, performance tests

(ii) fatigue or sleep deprivation, cognitive function, performance tests

(iii) ageing, cognitive function, performance tests

The following databases were searched:

- HSELINE
- CISDOC
- NIOSHTIC
- MEDLINE
- PSYCHLIT
- MBASE

The search covered the period 1990 - 2000 and included English language papers only.

Papers were initially selected from abstracts if they reported or implied the use of at least one neurobehavioural test which had been used in at least one solvent study. These papers were acquired and were included if they fulfilled the following criteria:

(i) actual test scores were recorded

(ii) a significant effect was noted as a result of the exposure or factor in question in at least one test

In addition some older papers were identified from the reference lists of selected papers. These were chosen on the basis of apparent relevance according to their titles. One data set from an unpublished report (Hooisma *et al*, 1988) was obtained directly from the author.

Data recorded in the selected papers in relation to the three factors (alcohol, sleep loss and ageing) were examined to determine:

(i) the levels of 'exposure' at which effects occurred

(ii) the consistency of effect sizes across studies

A detailed analysis of the quality of studies was not conducted. Since data were limited it was decided to include all that was available. However, some relevant comments regarding methodological difficulties are included below.

4.6 Results

4.6.1 Acute effects of alcohol consumption

Of a total of 25 papers identified from abstracts 13 were selected which fulfilled the stated criteria. Although many studies tested similar functions to those tested in the solvent studies very few employed exactly the same test and the data in many papers were not therefore directly comparable. In addition many papers did not record details of test scores preferring to report the results of statistical tests.

Selected studies together with the tests used and results are summarised in Annexe 3.

Overall, the data were difficult to interpret. The most frequently used test was the Digit Symbol Substitution Test (DSST), used in six of the thirteen studies. In five of these studies a significant acute effect of alcohol consumption was noted, in doses ranging from 0.15 - 1.0 g/kg. Although no blood alcohol level was presented in relation to the lowest dose of 0.15 g/kg another study reported blood alcohol levels around 80 mg/dl (the current legal limit for driving) in relation to a dose of 0.56 g/kg. This suggests that a dose of 0.15 g/kg which was associated with effects on DSST, is likely to produce blood alcohol levels below the current legal driving limit.

By contrast another study reported blood alcohol levels around 95 mg/dl and reported no effects on DSST.

Two studies tested immediate memory using the Digit Span test and found no acute effects at blood alcohol concentrations up to 0.06%. Two further studies used a syntactic reasoning test and both found acute effects following alcohol consumption. One study reported alcohol dose (0.15 g/kg) and the other blood alcohol concentration (0.025 - .1%). No effects were found on Sternberg's Memory Scanning test at an alcohol dose of .25 - 1.0 g/kg or at blood alcohol concentrations up to 0.08%. Similarly no effects were found on the Block Design test of the WAIS-R at a blood alcohol concentration of 0.05%.

Since alcohol has a depressant effect on the nervous system effects on tests primarily concerned with attentional control might reasonably be expected. One study did show that performance on the Stroop Colour - Word test was affected at blood alcohol concentrations from .062 - .073% but no other studies used this or similar tests.

4.6.1.1 General conclusions in relation to alcohol consumption

The investigation of the acute effects of alcohol on performance is complex and requires attention to a number of methodological factors notably individual and situation-based variables, such as gender, age, time since dose and expectancy. However, this subject has been studied for many years and a large amount of data has been amassed. Unfortunately in the present context, these data have limited use, since tests currently employed in solvent studies rarely appear in studies of alcohol effects. Data relating to these tests is therefore patchy and inconclusive and is unlikely to be helpful in terms of interpreting existing data.

It is clear however that much is known about the pattern of cognitive impairment associated with alcohol consumption and that certain tests, addressing particular aspects of functioning, are more responsive to small doses of alcohol than others. Further, there are well-established social and legal norms regarding acceptable levels and performance decrements associated with these. It is possible therefore that the future acquisition of data relating to the effects of alcohol on tests used in solvent studies could help in the interpretation of solvent-related effects.

4.7 Effects of sleep loss

Only 12 papers were identified from abstracts as being potentially relevant. Of these only five fulfilled the stated criteria for inclusion. Their results are summarised in Annexe 4. All these studies used the Syntactic Reasoning Test and one used Digit Span. No other tests were employed which were relevant to solvent investigations.

The study using the Digit Span test investigated the consequences of one nights' sleep loss. No effects on test performance were observed. Only one study, which investigated the effects

of the longest period of sleep deprivation (64 hours), found effects on syntactic reasoning, although most studies found effects on other tests. These limited data cannot therefore be considered useful for benchmarking purposes in relation to solvent associated effects. While there is an abundance of data concerned with the effects of sleep loss, the problem in the present context is again one of variability in test methods and thus non-comparable data.

4.8 Effects of normal ageing

Of the 51 papers identified from abstracts 28 were selected which fulfilled the stated criteria. Results reported in these studies in relation to specific tests are shown in Annexe 5, Tables 1-11.

There was some variation in the methods for reporting test scores. For example, some studies record the total time to complete a test, others record the time per test item and others the number or proportion of items correct. Where possible, data have been converted or inferred from that explicitly reported, for example in the case of graphical presentation, in order to facilitate comparison with those in solvent studies.

The data relating to age-effects on tests is more extensive than that relating to either alcohol consumption or sleep deprivation. Some data are available on the following tests which have also been used in solvent studies:

Digit Symbol Substitution Test (WAIS-R)(WHO-NCTB, MANS)
Tests from the Neurobehavioural Evaluation System
Stroop Colour-Word test
Wisconsin card sorting
Sternberg Test
Syntactic Reasoning
Digit Span (WMS) (WAIS-R)(NES)(WHO-NCTB)

Block Design (WAIS-R)
Rey's Auditory Verbal Learning Test
Santa Ana Test
Trails Test
Benton Visual Retention Test
Pursuit Aiming (WHO.National Core Test Battery)
Corsi Block Test

Although in the case of some tests data on age effects is derived from a single study it would seem reasonable to use this where the study is methodologically sound. For example, a detailed set of data is available for Rey's Auditory Verbal Learning Test (Vakil 1997). In the case of more frequently used tests such as the Digit Symbol Substitution Test (DSST), the data is in fact more difficult to interpret since a number of studies have investigated age effects and the age-groupings employed overlap to some extent. In the case of other tests, notably those which comprise the Neurobehavioural Evaluation System (NES), existing data covers only two age groups (26-47 and 60-73) and this provides rather limited information in terms of assessing effects relative to ageing throughout the working life. The relationship between ageing data on the various tests and solvent study data is discussed below. This section employs only solvent study data derived from studies rated as I or II in Section I.

4.8.1 Digit Symbol Substitution Test (DSST)

This is one of the most frequently used test and exists in two main forms, that contained in the original WAIS-R battery and that contained in the WHO National Core Test Battery (WHO-NCTB) developed for use in neurobehavioural studies. The two tests are very similar and administration is non-automated. A computer-administered version contained in the WHO-Milan Automated Neurobehavioural System (MANS) was subsequently developed and has been used in one study. Data from those studies (raw scores) using the DSST are shown in Table 7.

Table 7. DSST in 7 Solvent Studies				
Test	**Age**		$\bar{x}$ **Scores**	
	EG	**CG**	**EG**	**CG**
WAIS-R	31.5 (a) 33.0 (b)	34.5	33.8 (a) 27.9 (b)	40.5
WAIS-R	33.3	32.3	52.7	58.6
WAIS-R	43.4	43.5	4.8 (md)	5.8 (md)
WHO-NCTB	43.0	48.0	21.5	23.8
WHO-MANS	32-55	32-55	32.9	34.9
WHO-NCTB	33.0	30.0	28.3	43.9
WAIS-R	NS	NS	46.2 (c) 41.2 (d) 39.7 (e)	42.3

EG = exposed group
CG = control group
a = occasionally exposed
b = continuously exposed
c = low exposed
d = intermediate exposed
e = highly exposed
md = median score
WAIS-R = original Wichsler version
WHO-NCTB = version contained in WHO-National Core Test Battery
WHO-MANS = computer-administered version of WHO-NCTB (Milan Automated Neurobehavioural System)
NS = not stated

It can be seen that most subjects in these studies are aged between 30 and 50 years. Examination of the scores of control group subjects indicates a lack of consistency in scores in relation to age. For example, three groups with a similar mean age, 34.5, 32.3 and 30.0 years have mean DSST scores of 40.5, 58.6 and 43.9 respectively. Moreover one group with a mean age of 48 years has a very low mean score of 23.8. Given the lack of consistency in control group data it is difficult to relate this to the ageing data (provided in Annexe 5, Table 1) which is itself inconsistent. Although quite large differences are often demonstrated between exposed and control group performance, therefore, it is difficult to see how these can be interpreted by reference to age-related changes.

The wide variation in scores in Annexe 5 raises some concerns about the validity of the data. Based on the information given in the papers, six of the studies cited in Annexe 5, Table 1 used the WAIS-R version of the test and presumably therefore administered it according to standard instructions. One study (Van Boxtel, 1997) used a modified version in that letters were substituted for symbols and one study (Hooisma, 1988) used the WHO-NCTB version. A number of possible explanation exist for the variation in scores:

- poor inter or intra tester reliability
- lack of standardisation of test conditions
- variation in ability or educational level of subjects (most studies report the use of community-based volunteer samples)

Regardless of the reason for this variation, however, the fact remains that the data do not appear to provide a strong basis for interpreting differences between scores of exposed and control groups.

4.8.2 Tests from the Neurobehavioural Evaluation System (NES)

This test battery was developed by Baker *et al* in 1985 specifically for use in occupational and environmental studies. It comprises a range of tests for automated administration. The original intention was that control group data from those studies using the NES could be progressively pooled to provide age-related norms. Unfortunately this objective was not achieved and currently very few data are available relating to age effects on performance. These are entirely provided from one study by Hooisma in 1988 which reported on only two age ranges, 26-47 years and 60-73 years, (Annexe 5 Table 2).

Examination of the data in those studies using the NES reveal a considerable amount of variation in the form of reporting results. This is because numerous options are available in

the computer programme, both in terms of test parameters and data records. Much of the data recorded in the solvent studies was found to be in a form non-comparable to that used by Hooisma. The only exceptions were in respect of two tests, the Symbol Digit Substitution Test (SDS) and the Continuous Performance Test (CPT).

The SDS is similar to the DSST in that it is a coding test (matching symbols and numbers) but is carried out in separate trials of nine items each and records response time in terms of seconds per item. (The DSST records the number of items completed in 90 seconds). The range of control group scores recorded in the solvent studies was between 2.5 and 2.9 seconds per item. The range of exposed group scores was between 2.6 and 3.2 seconds per item. Hooisma records mean scores of 2.4 for males and 2.2 for females in the younger age-group and 3.1 for males and 3.3 for females in the older age-group.

The CPT is a vigilance test requiring the subject to respond each time a certain signal appears on the screen. Mean response times in milliseconds are recorded. In the solvent studies control group times ranged from 385.0 - 479.1 milliseconds and exposed group times from 379.7 - 498.6 milliseconds. Hooisma's data records 387.3 (males) and 386.2 (females) for younger subjects and 378.2 (males) and 381.4 (females) for older subjects.

In both these cases (SDS and CPT) the data are somewhat inadequate for interpretative purposes as they currently stand but suggest the possibility of developing more useful data in relation to NES tests. This is particularly the case since computer-administration considerably enhances standardisation of testing conditions.

4.8.3 Other frequently used tests

4.8.3.1 Stroop Colour-Word test

This is a test requiring attentional control. Unfortunately, although it was used in two studies, neither presented data in a form comparable to that in Annexe 5, Table 3. In fact the data presented in Annexe 5 appear to show a reasonably consistent pattern in relation to ageing which should provide a sound basis for interpretation of future data if this is recorded in a comparable form.

4.8.3.2 Wisconsin card-sorting test

This test involves concept formation. The data provided by Haaland (Annexe 5, Table 4) provides a limited basis on which to interpret ageing effects but is confined largely to older age groups. That provided by Compton in the same table seems incompatible and certain statements in the paper suggest the test may have been slightly modified from its original form. Two scores are usually recorded, categories and errors. In the two solvent studies reporting results on this test however both record category scores which appear to be rather high in terms of Haaland's data, and error scores which are rather low. Added to this Haaland includes only one age category which could be said to be representative of any section of the working population (17-25). Again therefore information appears inadequate for present purposes.

4.8.3.3 Sternberg's Memory Scanning Test

This test requires the subject to scan a set of numbers in order to identify the presence or absence of numbers presented previously. It exists in a manual or computer-administered (NES) version. A considerable amount of data has been published in relation to this test and it has a well-developed theoretical basis.

Examination of data published during the last ten years, however, reveals considerable variation in scores such that it is difficult to accept that the test was administered in a consistent manner (Annexe 5, Table 5). It was used only once in the solvent studies. Since the NES computer-administered version was used these data cannot be compared with that deriving from the non-automated form of the test.

4.8.3.4 Syntactic Reasoning

This is a test of higher level reasoning ability. Some limited data are available (Annexe 5, Table 6) but do not discriminate between separate age ranges. Moreover this test appears to be rarely used in solvent studies.

4.8.3.5 Digit Span

This test measures memory span, either forwards or backwards. A body of opinion supports the view that while forward span is relatively fixed throughout life (normally regarded as 7 ± 2), backward span places considerably more demands on the information processing system and is more susceptible to nervous system insult. The summation of forward and backward spans is therefore considered inappropriate. In the context of solvent-related effects, forward span has often been designated a “hold” test (i.e. resistant to change) and has been used as a measure of pre-morbid ability by some investigators. The measure of interest has thus tended to be backward span, or alternatively the difference between the forward and backward spans. A large number of solvent studies use this test and the results are shown in Table 8. Although a number of versions exist, the test is relatively straightforward and for convenience therefore the results have been presented together.

Digit Span Forward (Backward) in 14 Solvent Studies

Test	Age		x score	
	Exposed	Controls	Exposed	Controls
WAIS-R	NS	NS	6.4 (5.5)	6.3 (5.8)
WAIS-R	39.0	39.0	6.2 (4.5)	6.9 (5.1)
WHO-NCTB	48.0	48.2	5.6 (3.6)	6.9 (3.8)
NES	40.8	NA	6.2 (5.2)	NA
NES	35.8	36.9	5.5 (5.1)	5.6 (5.4)
WHO-NCTB	35.8	36.9	5.9 (5.4)	5.9 (5.9)
NES	63.2	62.7	5.6 (5.2)	5.3 (5.3)
WHO-NCTB	63.2	62.7	5.9 (5.7)	5.5 (5.2)
WAIS-R	38.0 52.0	34.0	5.7 5.8	5.6
WHO-NCTB	48.0	43.5	5.1 (3.3)	5.3 (3.3)
NES	35.5	37.6	5.3	5.8
NES	40.9	56.0	5.8 (4.7)	5.7 (4.8)
NES	38.5 38.9	37.9	7.4 (6.6) 7.2 (6.6)	6.9 (6.1)
NES	42.8	NA	6.3 (5.2)	

In fact many of the forward spans reported in these studies are rather low, especially in the context of the data presented in Annexe 5, Table 7. It is perhaps unfortunate that so little age-related data is presented which relates to backward span since this would appear to be the more important measure where neurotoxicant exposure is concerned.

4.8.3.6 Block Design Test

This test is part of the original WAIS-R and is related to visuo-spatial organisation. It is of concern that in those solvent studies reporting scores on this test, those of both the exposed and control groups (age 33 and 32 years respectively) appear to fall in the 50+ age range according to the data presented in Annexe 5, Table 8. Again, however, normative data is very limited and requires further investigation before drawing any firm conclusions from this observation.

4.8.3.7 Rey's Auditory Verbal Learning Test (RVLT)

This test is carried out over five trials and reports incremental learning. Vakil (1997) has provided a considerable amount of data related to age decades between 20 and 90, for both males and females. One study reports data in a comparable form except that total scores over five trials are reported as opposed to individual trial scores. In this (solvent) study the exposed group (mean age 42 years) achieved a total score over five trials of 49.1 items and the control group (mean age 45 years) 50.0 items. The comparable score for males aged 40-49 from Vakil's data is 54.3 and for males aged 50-59 years is 50.8. According to these age-related data therefore both groups were slightly under-performing for their age which may be accounted for by pre-morbid ability or educational level. Vakil's data demonstrate the possibility of useful age-related norms, but points to the possible importance of other influences on performance, which are likely to be test specific.

4.8.3.8 Santa-Ana Test

This is a peg-board type of test and is thus concerned with motor control. It is included in the WHO-NCTB. The only age-related data available is from the two age-groups of Hooisma (Annexe 1, Table 10). Of the five studies employing this test two report non-comparable data which appears to be the summation of scores over several trials. The remaining three studies report data which is compatible with Hooisma's age-group data. However the age-groups are not entirely consistent with his (for example 40+ years) and this makes it difficult to interpret in terms of age norms. The data do suggest however, that the preparation of such norms should be possible.

4.8.4 Other Selected Tests

The following tests have been groups together (Annexe 5, Table 11) either because age-related data is particularly limited, or because they have been used so infrequently in solvent studies.

4.8.4.1 Trails (Trailmaking) Test

This test is frequently used in clinical settings to assess general impairment. Few normative data appear to have been published in recent years. However, some is available from Compton (2000). The scores achieved by both subjects and controls in solvent studies appear to be remarkably low by comparison with Compton's data. It should be noted however, that Compton's subjects were described as 'highly educated', again underlining the need to create norms which reflect educational level as well as age.

4.8.4.2 Benton Visual Retention Test

Three solvent studies report data on this test and norms are provided for two core groups by Hooisma (1988). Again both subjects and controls in solvent studies appear to produce rather low scores, range 5.6-8.9 as opposed to 8.3-8.7 in Hooisma's data, raising similar concerns about educational factors.

4.8.4.3 Pursuit Aiming

This test, involving visual-motor control is part of the WHO-NCTB. Two studies report data in a form which is comparable to that of Hooisma (1988). Broadly the data are compatible in that mean scores of both the exposed and control groups in the two solvent studies fall within a few points of those reported by Hooisma for similar ages. However, as noted earlier,

Hooisma's data are limited to two broad age ranges, one of which covers virtually all the employed population. As such they cannot provide detailed age-related norms.

4.8.4.4 Corsi Block Test

Data have been provided for the upper age ranges for this test with only one group relating to the bulk of the employed population. One solvent study used this test and mean scores for both exposed and control groups were compatible with scores for subjects below the age of 55 years. Little more information can be derived from these data.

4.8.5 General conclusions in relation to ageing

Data relating to normal ageing are more extensive than those relating to the other agents studied. Most appear to be of limited relevance however, because they relate largely to older age-groups not present in the working population. In addition many of the test parameters reported are different from those reported in solvent studies. It is possible that some data sets exist which relate to early test development and therefore would not be identified in a literature search covering the period 1990-2000. They might for example be available from the test developers. However this would not include more recently developed computer-administered batteries such as the NES and WHO-MANS. Authors of these confirm the absence of age-related data (personal communication).

Those data which have been identified for a limited number of tests suggest that the use of normal ageing for benchmarking purposes is feasible in principle. However its use in the future will probably require the development of new data sets.

4.9 Effects of head injury

In addition to the five agents originally considered a further suggestion as a benchmarking candidate is that of head injury classification. There are clear clinical parameters which are currently employed when patients present with head injuries. Such injuries are categorised as 'mild', 'moderate' or 'severe' according to an impairment scale relating to functional capabilities. Neuroradiological and other data may also be available. Patients with 'mild' head injury, for example, might experience some degree of initial impairment but would be expected to recover fully within 3-6 months. It is possible that some data are available in clinical records which relate to neuropsychological test performance in some of these patients. This could therefore provide a means of relating test scores to categories of injury which (i) are based on the actual functional state of the individual and (ii) are widely accepted to have a degree of validity in terms of predicting future functioning. It is important to emphasise, however, that any 'read across' from head injury data to solvent exposure data would be simply in terms of evaluating the size or severity of any effect and would not imply any similarity in prognosis.

5. Summary Statements - Section II

1. Data relating to the effects of carbon monoxide exposure would appear to be inappropriate for benchmarking purposes since they are inconsistent and inconclusive and frequently relate to individual cases of patients experiencing life-threatening intoxication as a result of self-harm.

2. Data relating to the chronic effects of alcohol consumption similarly relates largely to small numbers of cases of individuals suffering from psychological distress. Initial inspection of the data suggests minimal impairment even in more severe cases. However, this may result from reliance on a single assessment which does not take into account change from earlier (possibly higher) levels of functioning. Re-examination of the data on this basis may provide a useful model of more severe effects.

3. Data relating to the acute effects of alcohol consumption are limited for present purposes in that few studies use the same tests as those used in solvent studies. This is a complex area of study in that effects appear to be influenced by a range of psychological factors (for example, expectancy of effect) and display large individual differences and time course patterns. However a large amount of data has been amassed and these show some consistency in suggesting that the acute effects of alcohol are function specific (and therefore test-specific). Therefore, while existing data cannot be used for interpretation of existing solvent data, future acquisition of data which describes alcohol effects in relation to tests used in solvent studies could prove useful.

4. Data relating to sleep deprivation are very limited in terms of studies using the same tests as those used in solvent studies. The data are inconsistent and show the influence of various individual differences and environmental factors. They are thus of limited relevance for the present purpose.

5. Data relating to normal ageing appear to be concentrated in the older age-groups and are thus of limited relevance to the working population. It is possible that a search of earlier literature might yield more data collected at the time some tests were originally developed, but only for a few tests. Therefore although normal ageing could provide a useful benchmark for the effects of solvent exposure this will probably require the development of new data sets.

6. Data may be available linking psychological test performance to an accepted classification system for severity of head injury, which could serve as a benchmark for the severity of effects of solvent exposure. Alternatively it may be possible to collect such data in the future.

6. Overall Conclusions

A large number of studies of neurobehavioural effects of solvents have been carried out during the last 20 years. Despite an apparent consensus among many researchers that such effects are not in doubt there is no sign that interest in researching this subject has abated in recent years. The studies are variable in quality with most of the better ones being carried out during the last five years. In particular more recent studies have more frequently attempted quantitative retrospective exposure assessment and have attempted to investigate exposure-effect relationships.

The overwhelming majority of studies, (78%), demonstrate a statistically significant difference between the performance of exposed and control groups in at least one neurobehavioural test. Fifty-four percent of those studies investigating exposure-effect associations report such a relationship. When only highly rated studies, according to pre-determined methodological criteria, are considered, a similar percentage are positive (72%) and demonstrate an exposure-effect relationship (56%). On this basis it is tempting to conclude that there is convincing evidence for the existence of neurobehavioural effects following long-term low-level exposure to solvent mixtures. However, a number of factors while not entirely negating the evidence, suggest it should be viewed with caution.

1. Reporting bias tends to result in only the publication of studies finding a positive effect.
2. Examination of the data on exposure did not indicate a consistent pattern. Although effects occurred most often where exposure was both high and of longer duration,

some studies where exposure was relatively low and of short duration found effects equal to those of studies where exposure was higher and longer.

3. Exposure was to solvent mixtures in these studies. Although this is unavoidable in this kind of work it nevertheless raises questions about the probability of identifying specific effects on the nervous system which are consistent across studies.
4. A wide range of tests was used and frequently, even where tests were similar, different measures were reported, making it difficult to compare data across studies.
5. There was no consistent pattern in test results, even in terms of a very broad classification of functional areas.
6. Where traditional clinical tests were employed there was a tendency to report raw rather than scaled scores. When raw scores were converted to scaled scores (i.e. age adjusted) there were indications that the scores of both exposed and control groups frequently fell within what would usually be considered a 'normal' range.
7. Studies frequently used large numbers of tests without carrying out correction for multiple comparisons. Hence many of the observed significant differences may have occurred by chance.
8. Sample sizes were frequently small (only one third of studies tested more than 100 exposed workers) and only one study reported power calculations

Given this range of difficulties the data relating to neurobehavioural effects should perhaps be viewed as suggestive but equivocal. The questions raised by the current state of the evidence are however unlikely to be answered by more neurobehavioural studies of the type which continue to be reported. In terms of test outcome measures in particular, recent studies are very similar to those which were carried out in the past and only serve to add to the confusion

surrounding this issue. A specific area of controversy relates to the size and biological or social significance of any identified effects. The question of whether statistical significance relates to psychological significance remains, even amongst those who accept the validity of the former. The proposal that a form of benchmarking against the known effects of other agents might address this question was explored in Section II of the current report. A number of agents were examined in terms of whether their effects could be used in this way. It became clear that for two of the agents examined (carbon monoxide exposure and sleep deprivation) this was unlikely to be feasible. Firstly few data were available which related to the specific tests used most often in solvent studies, and in addition much of the data was inconsistent. This was probably because each of these areas is itself a difficult field of research with numerous methodological problems.

Initial examination of the data relating to a third factor, the chronic effects of alcohol abuse, indicated that effects on cognitive functioning were likely to be relatively mild. However, consideration of cross-sectional data alone may mask progressive deterioration in some individuals who, while currently functioning within 'normal' limits, may have functioned at a higher level prior to the development of alcoholism. Further examination of the data, taking into account performance on 'hold' tests measuring pre-morbid functioning might therefore yield useful information about patterns of cognitive impairment.

Existing data on the acute effects of alcohol consumption like those on carbon monoxide and sleep deprivation appears not to be useful for current purposes because of the use of different tests. However, although this field is complex, it is well-developed methodologically and

might provide a means whereby benchmarking data can be acquired in the future in relation to the tests used in solvent studies.

A further possible candidate for benchmarking purposes is the existing classification system for head injuries. Currently there are clear clinical parameters for definition of 'mild', 'moderate' and 'severe' effects, measured on an impairment scale and sometimes accompanied by neuroradiological data. The association of neurobehavioural test data with these categories, either derived from the existing literature or constructed in the future, could provide a means of interpreting the functional significance of effects which have been identified in relation to solvent exposure.

Consideration of the remaining alternative for the benchmarking process, that of normal ageing effects resulted in the following conclusions:

1. There are some data available on the tests which are frequently used in solvent studies.
2. The data are variable in terms of quantity and quality. For a small number of tests a set of reference data exists which could be used for reference purposes.
3. For most tests the current data are too limited to be used in their present form. Most investigators concerned with the ageing process are interested in the later decades and therefore data tend to be concentrated in these age-groups (50+ years).
4. There is considerable variation in the test parameters reported on and these frequently do not correspond to those reported in the solvent studies.

5. Many solvent studies do not report mean scores on tests but prefer to report the results of statistical tests of significance, making it difficult to assess psychological significance.

6. There are indications in the ageing data that both gender and educational level should be taken account of separately in most tests.

7. It is possible that some further data may be identified from a search of earlier literature, relating to original test development. This would not, however, cover computer administered test batteries.

8. The existence of some data in relation to some tests suggests that this may in principle provide a way forward in test interpretation. However, considerable effort would need to be invested in the creation of age and gender - related normative data.

It would appear therefore that most existing data are not currently in a form which would allow benchmarking to be carried out, at least in the immediate term. However, a number of courses of action can be suggested which may make such a process possible in the future. Some of these relate to existing information and others will require collection of new data.

7. Recommendations

7.1 Existing data

1. Data on alcohol abusers could be re-examined, taking into account performance on "hold" tests relative to those on other tests. The objectives would be to investigate (i) evidence of intellectual deterioration (ii) patterns of deterioration. These data may be

available in the published literature and may also exist in unpublished form having been collected by those concerned with the assessment and treatment of this group.

2. Further exploration of data on ageing could be carried out which includes (i) earlier literature published at the time of original test development or early data collected by test developers, and (ii) earlier literature on the ageing process in general which may be more likely to use tests used in solvent studies. It should be noted however, that the latter is still unlikely to cover age-groups relevant to working populations since the study of "ageing" is largely the study of those over the age of 50.

3. An initial exploration could be carried out of the data held by clinicians on those assessed following head injury, together with any published data. It may be possible to use the data to model the severity of effects using the current classification system employed by clinicians (mild, moderate, severe).

4. The current report has been concerned exclusively with mixed solvent exposure. A number of studies have been carried out which investigate the effects of single solvent exposure where it has been possible to identify populations with such exposures. In the interests of attempting to identify patterns of effect it may be useful to carry out a similar exercise to that conducted here for a number of single solvents where exposure is less complex. Those identified as possible candidates are trichloroethylene, perchloroethylene, toluene and styrene.

7.2 Future data

1. Consideration should be given to the possibility of conducting studies targeted specifically at investigating the acute effects of alcohol consumption on test

performance, to provide purpose-developed benchmarking data. Tests would be selected on their frequency of use in solvent studies.

2. Consideration should be given to creating data sets which identify ageing effects on tests frequently used in solvent studies. Methodologically this is relatively straightforward and would provide norms similar to those available in certain neurophysiological tests, for example lung function.

3. Depending on the results of an initial exploration of head injury data, it may be useful to construct data sets in the future which relate test performance to the existing classification system for functional impairment following head injury.

The above recommendations are not necessarily discrete alternatives. It may emerge that the use of a number of different forms of benchmark are required to develop a valid and reliable approach to the interpretation of test data.

7.3 Neurobehavioural testing - future directions

The above report has essentially been concerned with the interpretation of existing neurobehavioural data in order to address current issues surrounding the effects of solvent exposure. This is necessary because a large body of literature already exists and also because new studies which continue to be carried out around the world invariably employ the same methods as those used in the past. If, however current difficulties are to be avoided in the future more fundamental issues of test development, will need to be addressed incorporating a more theoretically driven approach, in line with current thinking in cognitive psychology. While this may not have direct relevance to the immediate questions in the field of solvent exposure discussed here, it should be noted that other health concerns have already arisen and

are likely to arise in the future where neurobehavioural data plays a central role. Without the development of a theoretically-based approach to test selection, development and the recording and analysis of data therefore the current confusion surrounding the interpretation of neurobehavioural data is likely to become a recurrent problem.

References

Anstey, K. J., and Smith, G. A. (1999). Interrelationships among biological markers of ageing, health, activity, acculturation, and cognitive performance in late adulthood. *Psychological Ageing* **14**(4), 605-618.

Baker, E. L., Letz, R. E., Eisen, E. A., Pothier, L. J., Plantamura, D. L., Larson, M., and Wolford, R. (1988). Neurobehavioural effects of solvents in construction painters. *Journal of Occupational Medicine* **30**(2), 116-123.

Baker, S. J., Chrzan, G. J., Park, C. N., and Saunders, J. H. (1986). Behavioural effects of 0 and 0.05% blood alcohol in male volunteers. *Neurobehavioural Toxicology and Teratology* **8**, 77-81.

Becker, J. T., and Milke, R. M. (1998). Cognition and ageing in a complex work environment: relationships with performance among air traffic control specialists. *Aviation, Space and Environmental Medicine* **69**(10), 944-951.

Bieliauskas, L. A., Roper, B. R., Trobe, J., Green, P., and Lacy, M. (1998). Cognitive measures, driving safety, and Alzheimer's disease. *The Clinical Neuropsychologist* **12**(2), 206-212.

Bleecker, M. L., Bolla, K. I., Agnew, J., Schwartz, B. S., and Ford, D. P. (1991). Dose-related subclinical neurobehavioural effects of chronic exposure to low levels of organic solvents. *American Journal of Industrial Medicine* **19**(6), 715-728.

Bolla, K. I., Schwartz,, BS, Stewart, W., Rignani, J., Agnew, J., and Ford, D. P. (1995). Comparison of neurobehavioural function in workers exposed to a mixture of organic and inorganic lead and in workers exposed to solvents. *American Journal of Industrial Medicine* **27**(2), 231-246.

Boone, K. B., Berman, N. G., and D'Elia, L. F. (1993). Rey-Osterrieth Complex Figure performance in healthy, older adults: relationship to age, education, sex and IQ. *The Clinical Neuropsychologist* **7**(1), 22-28.

Bowler, R. M., Mergler, D., Huel, G., Harrison, R., and Cone, J. (1991). Neuropsychological impairment among former microelectronics workers. *Neurotoxicology* **12**, 87-104.

Broadwell, D. K., Darcey, D. J., Hudnell, H. K., Otto, D. A., and Boyes, W. K. (1995). Work-site clinical and neurobehavioural assessment of solvent-exposed microelectronics workers. *American Journal of Industrial Medicine* **27**, 677-698.

Chaves, M., Ilha, D., Maia, A. L., Motta, E., Lehmen, R., and Oliveira, L. M. (1999). Diagnosing dementia and normal ageing: clinical relevance of brain ratios and cognitive performance in a Brazilian sample. *Braz J Med Biol Res* **32**(9), 1133-1143.

Colvin, M., Myers, J., Nell, V., Rees, D., and Cronje, R. (1993). A cross-sectional survey of neurobehavioural effects of chronic solvent exposure on workers in a paint manufacturing plant. *Environmental Research* **63**(1), 122-132.

Compton, D. M., Bachman, L. D., Brand, D., and Avet, T. L. (2000). Age-associated changes in cognitive function in highly educated adults: emerging myths and realities. *International Journal of Geriatric Psychiatry* **15**(1), 75-85.

Corbett, R., Meagher, J. B. A., and Leonard, B. E. (1991). The effect of acute alcohol intoxication on psychometric testing and 5-HT-induced platelet aggregation in normal subjects; Modulatory role of Evening Primrose oil. *Human Psychopharmacology* **6**, 253-256.

Crossen, J. R., and Wiens, A. N. (1988). Wechsler Memory Scale-Revised: Deficits in performance associated with neurotoxic solvent exposure. *The Clinical Neuropsychologist* **2**(2), 181-187.

Czigler, I. (1996). Age, colour processing and meaningfulness: an event-related potential study. *International Journal of Psychophysiology* **22**(1-2), 25-34.

Daniell, W. E., Claypoole, K. H., Checkoway, H., Smith-Weller, T., Dager, S. R., and Townes, B. D. (1999). Neuropsychological function in retired workers with previous long-term occupational exposure to solvents. *Occupational and Environmental Medicine* **56**(2), 93-105.

Elofsson, S. A., Gamberale, F., Hindmarsh, T., Iregren, A., Isaksson, A., Johnsson, I., Knave, B., Lydahl, E., Mindus, P., Persson, H. E., Philipson, B., Steby, M., Struwe, G., Soderman, E., Wennberg, A., and Widen, L. (1980). Exposure to organic solvents. A cross-sectional epidemiologic investigation on occupationally exposed car and industrial spray painters with special reference to the nervous system. *Scandinavian Journal of Work, Environment and Health* **6**(4), 239-273.

Emmen, H. H., Hoogendijk, E. M. G., Hooisma, J., Orlebeke, J. F., and Uijtdehaage, S. H. J. (1988) Adaptation of two standardized international test batteries for use in the Netherlands for detection of exposure to neurotoxic compounds.

Escalona, E., Yanes, L., Feo, O., and Maizlish, N. (1995). Neurobehavioural evaluation of Venezuelan workers exposed to organic solvent mixtures. *American Journal of Industrial Medicine* **27**, 15-27.

Fidler, A. T., Baker, E. L., and Letz, R. E. (1987). Neurobehavioural effects of occupational exposure to organic solvents among construction painters. *British Journal of Industrial Medicine* **44**(5), 292-308.

Foo, S. C., Lwin, S., Chia, S. E., and Jeyaratnam, J. (1994). Chronic neurobehavioral effects in paint formulators exposed to solvents and noise. *Annals Academy of Medicine* **23**(5), 650-654.

Green E and Short S (1998). Indoor Air Quality in the Home (2): Carbon Monoxide. MRC Institue for Environment and Health, University of Leicester.

Grosch, J. W., Neale, A. V., and Demers, R. Y. (1996). Neurobehavioural and health-related deficits in solvent-exposed painters. *American Journal of Industrial Medicine* **30**(5), 623-632.

Gupta, B. N., Kumar, P., and Srivastava, A. K. (1990). An investigation of the neurobehavioral effects on workers exposed to organic solvents. *Journal of the Society of Occupational Medicine* **40**, 94-96.

Guralnik, J. M., Ferrucci, L., Penninx, B. W., Kasper, J. D., Leveille, S. G., Bandeen-Roche, K., and Fried, L. P. (1999). New and worsening conditions and change in physical and cognitive performance during weekly evaluations over 6 months: the Women's Health and Ageing Study. *Journal of Gerontology* **54**(8), M410-422.

Gustafson, R., and Källmén, H. (1990). Alcohol and the compensation hypothesis: A test with cognitive and psychomotor tasks. *Perceptual and Motor Skills* **71**, 1367-1374.

Haaland, K. Y., Vranes, L. F., Goodwin, J. S., and Garry, P. J. (1987). Wisconsin Card Sort Test performance in a healthy elderly population. *Journal of Gerontology* **42**(3), 345-346.

Hänninen, H., Eskelinen, L. A., Husman, K. A. J., and Nurminen, M. (1976). Behavioral effects of long-term exposure to a mixture of organic solvents. *Scandinavian Journal of Work, Environment and Health* **4**, 240-255.

Hanninen, H., Antti-Poika, M., Juntunen, J., and Koskenvuo, M. (1991). Exposure to organic solvents and neuropsychological dysfunction: a study on monozygotic twins. *British Journal of Industrial Medicine* **48**(1), 18-25.

Heishman, S. J., Arasteh, K., and Stitzer, M. L. (1997). Comparative effects of alcohol and marijuana on mood, memory and performance. *Pharmacology Biochemistry and Behaviour* **58**(1), 93-101.

Hertzog, C., Cooper, B. P., and Frisk, A. D. (1996). Ageing and individual differences in the development of skilled memory search performance. *Psychological Ageing* **11**(3), 497-520.

Hill, R. D., and Bäckman, L. (1995). The relationship between the Mini-Mental State examination and cognitive functioning in normal elderly adults: A componential analysis. *Age and Ageing* **24**, 440-446.

Hindmarch, I., Bhatti, J. Z., Starmer, G. A., Mascord, D. J., Kerr, J. S., and Sherwood, N. (1992). The effects of alcohol on the cognitive function of males and females and on skills relating to car driving. *Human Psychopharmacology* **7**, 105-114.

Hindmarch, I., Kerr, J. S., and Sherwood, N. (1991). The effects of alcohol and other drugs on psychomotor performance and cognitive function. *Alcohol and Alcoholism* **26**(1), 71-79.

Hooisma, J., Hänninen, H., Emmen, H. H., and Kulig, B. M. (1993). Behavioural effects of exposure to organic solvents in dutch painters. *Neurotoxicology and Teratology* **15**, 397-406.

Houx, P. J., Jolles, J., and Vreeling, F. W. (1993). Stroop interference: ageing effects assessed with the Stroop Colour-Word Test. *Experimental Ageing Research* **19**(3), 209-224.

Institute of Occupational Health, University of Birmingham. (1997) Occupational Exposure Limits. Key document. Criteria for the qualitative evaluation of human neurobehavioural studies of neurotoxicity. European Commission. EUR 17390 EN. Employment and Social Affairs.

Iregren, A. (1986). Subjective and objective signs of organic solvent toxicity among occupationally exposed workers. An experimental evaluation. *Scandinavian Journal of Work, Environment and Health* **12**(5), 469-475.

Kauranen, K., and Vanharanta, H. (1996). Influences of ageing, gender and handedness on motor performance of upper and lower extremities. *Perceptual and Motor Skills* **82**(2), 515-525.

Kelly, P. (1999) The validity of neuropsychological assessment. Phase One. Phases Two to Four. Report to HSE.

Kelly, T.L., Ryman,D.H., Schlangen, K., Gomez, S.A. and Elsmore, T.F. (1997) The effects of a single dose of pemoline on performance and mood during sleep deprivation. Military Psychology, 9(3), 213-225.

Kennedy, R. S., Dunlap, W. P., Turnage, J. J., and Fowlkes, J. E. (1993). Relating alcohol-induced performance deficits to mental capacity: A suggested methodology. *Aviation, Space and Environmental Medicine* **64**(12), 1077-1085.

Kishi, R., Harabuchi, I., Katakura, Y., Ikeda, T., and Miyake, H. (1993). Neurobehavioural effects of chronic occupational exposure to organic solvents among Japanese industrial painters. *Environmental Research* **62**(2), 303-313.

Kramer, A. F., Hahn, S., Irwin, D. E., and Theeuwes, J. (1999). Attentional capture and ageing: implications for visual search performance and oculomotor control. *Psychological Ageing* **14**(1), 135-154.

Kuitunen, T. (1994). Drug and ethanol effects on the clinical test for drunkenness: Single doses of ethanol, hypnotic drugs and antidepressant drugs. *Pharmacology and Toxicology* **75**, 91-98.

Lagarde, D., and Batejat, D. (1995). Disrupted sleep-wake rhythm and performance: Advantages of modafinil. *Military Psychology* **7**(3), 165-191.

Lee, D. H., Park, I. G., Kim, J. H., Lee, Y. H., Kim, D., and Kang, S. K. (1998). Neurobehavioural Changes in Shoe Manufacturing Workers. *Neurotoxicology and Teratology* **20**(3), 259-263.

Lee, S. H., and Lee, S. H. (1993). A study on the neurobehavioural effects of occupational exposure to organic solvents in Korean workers. *Environmental Research* **60**, 227-232.

Linde, L., and Bergström, M. (1992). The effect of one night without sleep on problem-solving and immediate recall. *Psychological Research* **54**, 127-136.

Lindstrom, K., and Wickstrom, G. (1983). Psychological function changes among maintenance house painters exposed to low levels of organic solvent mixtures. *Acta Psychiatrica Scandinavica Suppl* **303**, 81-91.

Lukas, S. E., Lex, B. W., Slater, J. P., Greenwald, N. E., and Mendelson, J. H. (1989). A microanalysis of ethanol-induced disruption of body sway and psychomotor performance in women. *Psychopharmacology* **98**, 169-175.

Lundberg, I., Michelsen, H., Nise, G., Hogstedt, C., Hogberg, M., Alfredsson, L., Almkvist, O., Gustavsson, A., Hagman, M., Herlofson, J., et al. (1995). Neuropsychiatric function of housepainters with previous long-term heavy exposure to organic solvents. *Scandinavian Journal of Work, Environment and Health* **21**(1), 1-44.

Maizlish, N. A., Langolf, G. D., Whitehead, L. W., Fine, L. J., Albers, J. W., Goldberg, J., and Smith, P. (1985). Behavioural evaluation of workers exposed to mixtures of organic solvents. *British Journal of Industrial Medicine* **42**, 579-590.

Mazaux, J. M., Dartigues, J. F., Letenneur, L., Darriet, D., Wiart, L., Gagnon, M., Commenges, D., and Boller, F. (1995). Visuo-spatial attention and psychomotor performance in elderly community Residents: Effects of age, gender and education. *Journal of Clinical and Experimental Neuropsychology* **17**(1), 71-81.

McCann, U. D., Penetar, D. M., Shaham, Y., Thorne, D. R., Gillin, J. C., Sing, H. C., Thomas, M. A., and Belenky, G. (1992). Sleep deprivation and impaired cognition. Possible role of brain catecholamines. *Biological Psychiatry* **31**, 1082-1097.

Mikkelsen, S., Jørgensen, M., Brown, E., and Gyldensted, C. (1988). Mixed solvent exposure and organic brain damage: A study of painters. Scandinavian Journal of Work, Environment and Health.*Suppl.* **78**(118).

Milanovic, L., Spilch, G., Vucinic, G., Knezevic, S., Ribaric, B., and Mubrin, Z. (1990). Effects of occupational exposure to organic solvents upon cognitive performance. *Neurotoxicology and Teratology* **12**, 657-660.

Morrow, L. A., Robin, N., Hodgson, M. J., and Kamis, H. (1992). Assessment of attention and memory efficiency in persons with solvent neurotoxicity. *Neuropsychologia* **30**(10), 911-922.

Morrow, L. A., Ryan, C. M., Hodgson, M. J., and Robin, N. (1990). Alterations in cognitive and psychological functioning after organic solvent exposure. *Journal of Occupational Medicine* **32**(5), 444-450.

Morrow, L. A., Steinhauer, S. R., Condray, R., and Hodgson, M. (1997). Neuropsychological performance of journeymen painters under acute solvent exposure and exposure-free conditions. *Journal of the International Neuropsychological Society* **3**, 269-275.

Muijser, H., Geuskens, R. B. M., Hooisma, J., Emmen, H. H., and Kulig, B. M. (1996). Behavioral effects of exposure to organic solvents in carpet layers. *Neurotoxicology and Teratology* **18**(4), 455-462.

Myers, J. E., Nell, V., Colvin, M., Rees, D., and Thompson, M. L. (1999). Neuropsychological function in solvent-exposed South African paint makers. *Journal of Occupational and Environmental Medicine* **41**(11), 1011-1018.

Nasterlack, M., Dietz, M. C., Frank, K. H., Hacke, W., Scherg, H., Schmittner, H., Stelzer, O., Zimber, A., and Triebig, G. (1999). A multidisciplinary cross-sectional study on solvent-related health effects in painters compared with construction workers. *International Archives of Occupational and Environmental Health* **72**(4), 205-214.

Newman, D., Speake, D. J., Armstrong, P. J., and Tiplady, B. (1997). Effects of ethanol on control of attention. *Human Psychopharmacology* **12**, 235-241.

Ng, T. P., Ong, S. G., Lam, W. K., and Jones, G. M. (1990). Neurobehavioural effects of industrial mixed solvent exposure in Chinese printing and paint workers. *Neurotoxicology and Teratology* **12**, 661-664.

Petersen, R. C., Smith, G., Kokmen, E., Ivnik, R. J., and Tangalos, E. G. (1992). Memory function in normal ageing. *Neurology* **42**, 396-401.

Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., Lawrence, A. D., McInnes, L., and Rabbitt, P. M. (1998). A study of performance on tests from the CANTAB battery sensitive to frontal lobe dysfunction in a large sample of normal volunteers: implications for theories of executive functioning and cognitive ageing. Cambridge Automated Neuropsychological Test Battery. *Journal of the International Neuropsychological Society* **4**(5), 474-490.

Rouleau, N., and Belleville, S. (1996). Irrelevant speech effect in ageing: An assessment of inhibitory processes in working memory. *Journal of Gerontology* **51B**(6), 356-363.

Ruijten, M. W. M. M., Hooisma, J., Brons, J. T., Habets, C. E. P., Emmen, H. H., and Muijser, H. (1994). Neurobehavioural effects of long-term exposure to xylene and mixed organic solvents in shipyard spray painters. *Neurotoxicology* **15**(3), 613-620.

Rush, C. R., Higgins, S. T., Hughes, J. R., Bickel, W. K., and Wiegner, M. S. (1993). Acute behavioural and cardiac effects of alcohol and caffeine, alone and in combination, in humans. *Behavioural Pharmacology* **4**, 562-572.

Salthouse, T. A., and Meinz, E. J. (1995). Ageing, inhibition, working memory, and speed. *Journal of Gerontology* **50B**(6), 297-306.

Sarretto, G., Oliveira, F. C., Kato, M., Reis, E. J. F. B., Castro, R. C., Dorigatti, F., Camerino, D., and Gilioli, R. (1997). Exposure to organic solvents and its effects on the central nervous system in workers of the Camacari petrochemical complex in Brazil. *G Ital Med Lav Ergon* **19**(3), 66-71.

Schumacher, J. M. (1981) "The effects of ageing on four behavioural tests," Master's Degree in Industrial Engineering Thesis, The University of Michigan, Ann Arbor, Michigan.

Seppäläinen, A., M, Lindström, K., and Martelin, T. (1980). Neurophysiological and psychological picture of solvent poisoning. *American Journal of Industrial Medicine* **1**, 31-42.

Smith, M. W., Sharit, J., and Czaja, S. J. (1999). Ageing, motor control, and the performance of computer mouse tasks. *Human Factors* **41**(3), 389-396.

Spurgeon, A., Glass, D., C, Calvert, I., A, Cunningham-Hill, M., and Harrington, J., M (1994). Investigation of dose related neurobehavioural effects in paintmakers exposed to low levels of solvents. *Occupational and Environmental Medicine* **51**, 626-630.

Spurgeon, A., Gray, C. N., Sims, J., Calvert, I., Levy, L. S., Harvey, P. G., and Harrington, J. M. (1992). Neurobehavioural effects of long-term occupational exposure to organic solvents: two comparable studies. *American Journal of Industrial Medicine* **22**(3), 325-335.

Streufert, S., Pogash, R. M., Roache, J., Gingrich, D., Landis, R., Severs, W., Lonardi, L., and Kantner, A. (1992). Effects of alcohol intoxication on risk taking, strategy, and error rate in visuomotor performance. *Journal of Applied Psychology* **77**(4), 515-524.

Triebig, G., Barocka, A., Erbguth, F., Holl, R., Lang, C., Lehrl, S., Rechlin, T., Weidenhammer, W., and Weltle, D. (1992). Neurotoxicity of solvent mixtures in spray painters. II. Neurologic, psychiatric, psychological, and neuroradiologic findings. *International Archives of Occupational and Environmental Health* **64**(5), 361-372.

Tripathi, S. R., Bhattacharya, S. K., Chattopadhya, P., and Kashyap, S. K. (1989). Neurobehavioural disturbances in workers engaged in high-pressure spray painting. *Journal of Human Ergology* **18**, 191-198.

Troyer, A. K., Cullum, C. M., Smernoff, E. N., and Kozora, E. (1994). Age effects on Block Design: Qualitative performance features and extended-time effects. *Neuropsychology* **8**(1), 95-99.

Tsai, S. Y., Chen, J. D., Chao, W. Y., and Wang, J. D. (1997). Neurobehavioural effects of occupational exposure to low-level organic solvent among Taiwanese workers in paint factories. *Environmental Research* **73**(1-2), 146-155.

Tzambazis, K., and Stough, C. (2000). Alcohol impairs speed of information processing and simple and choice reaction time and differentially impairs higher-order cognitive abilities. *Alcohol and Alcoholism* **35**(2), 197-201.

Vakil, E., and Blachstein, H. (1997). Rey AVLT: Developmental norms for adults and the sensitivity of different memory measures to Age. *The Clinical Neuropsychologist* **11**(4), 356-369.m

Valciukas, J. A., Lilis, R., Slinger, R. M., Glickman, L., and Nicholson, W. J. (1985). Neurobehavioural changes among shipyard painters exposed to solvents. *Archives of Environmnental Health* **40**(1), 47-52.

Van Boxtel, M. P., Paas, F. G., Houx, P. J., Adam, J. J., Teeken, J. C., and Jolles, J. (1997). Aerobic capacity and cognitive performance in a cross-sectional ageing study. *Medical Science and Sports Exercise* **29**(10), 1357-1365.

Verhaeghen, P., Vandenbroucke, A., and Dierckx, V. (1998). Growing slower and less accurate: adult age differences in time-accuracy functions for recall and recognition from episodic memory. *Experimental Ageing Research* **24**(1), 3-19.

White, R. F., Proctor, S. P., Echeverria, D., Schweikert, J., and Feldman, R. G. (1995). Neurobehavioural effects of acute and chronic mixed-solvent exposure in the screen printing industry. *American Journal of Industrial Medicine* **28**, 221-231.

Wilkinson, C. J. (1995). The acute effects of zolpidem, administered alone and with alcohol, on cognitive and psychomotor function. *Journal of Clinical Psychiatry* **56**, 309-318.

Williamson, A. M., and Feyer, A. M. (2000). Moderate sleep deprivation produces impairments in cognitive and motor performance equivalent to legally prescribed levels of alcohol intoxication. *Occupational and Environmental Medicine* **57**, 649-655.

Woodard, J. L., Dunlosky, J. A., and Salthouse, T. A. (1999). Task decomposition analysis of intertrial free recall performance on the Rey Auditory Verbal Learning Test in normal ageing and Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology* **21**(5), 666-676.

Annexe 1

Study details of 45 selected studies

KEY TO TABLE HEADINGS				
PA (C)	Principle Author (Country)		**ET years**	Exposure Time - Exposed Group (years)
YOP	Year of Publication		**HE**	Highest Exposure - Exposed Group
SSM	Subject Selection Method		**S**	Solvents
RR	Response Rate		**MEA**	Method of Exposure Assessment
I/E	Inclusion/Exclusion Criteria Stated		**CLE**	Control for Last Exposure
No: EG	Number Exposed Group		**C/M**	Confounders/Modifiers Controlled
No: CG	Number Control Group		**STC**	Standard Test Conditions
G	Gender		**TA**	Test Administration Type
Age EG	Age Exposed Group		**SD**	Significant Difference(s) between EG and CG
Age CG	Age Control Group		**EE**	Exposure-Effect relationship

KEY				
S	stratified		**TD**	time of day
A	all employees included		**M**	medication
V	volunteers		**I**	injury
R	random		**SL**	sleep loss
P	patients		**TE**	test effort
M	male		**OE**	other exposure
F	female		**VT**	vocational training
M/F	both genders		**H**	handedness
TLV	threshold limit value		**SES**	socio-economic status
UM	unidentified mixed		**FS**	family size
Air	air monitoring		**VI**	viral infection
Bio	biological monitoring		**CE**	computer experience
JH(Q)	job history, quantitative estimate		**BMI**	body mass index
JH(D)	job history, description only		**J Sat.**	job satisfaction
E	exposed group		**Geog.**	geographical area
C	control group		**Man**	manual
II	initial intelligence		**Comp**	computer-administered
JT	job type		**NS**	not stated
ES	employment status		**NA**	not applicable
ED	employment duration			

STUDY Classification

****	= I	superior
***	= II	good
**	= III	fair
*	= IV	poor

PA (C)	YOP	SSM	RR	I/E	No: EG	No: CG	G	Age EG	Age CG	ET years	HE	S	MEA	CLE	C/M	STC	TA	SD	EE
Hanninen (Finland) **	1976	S(E) NS(C)	98.5%(E)	Yes	100	101	NS	20-65 35($\bar{x}$)	matched within 1-2 years	1-40 14.8($\bar{x}$)	<TLV	IM	Air	NS	Age JT II	NS	Man	Yes	No
Elofsson (Sweden) **	1980	S(E) NS(C)	83.7%(E) 61.0%(C1) 70.0%(C2)	Yes	80	(1)40 (2)40	NS	>25-65	matched	NS	<TLV	UM	JH(Q)	NS	Age Education JT ES	NS	Man	Yes	No
Seppalainen (Finland) **	1980	P	NS	Yes	48(M) 59(F)	NA	M/F	(M)35.8($\bar{x}$) (F)42.0($\bar{x}$)	NA	(M)9.6($\bar{x}$) (F)7.6($\bar{x}$)	≤TLV	IM	JH(Q)	NA	Age Gender II	NS	Man	NA	Yes
Lindstrom (Finland) **	1983	A(E) NS(C)	66.8(E) NS(C)	NS	219	229	NS	42.4	41.9	NS	NS	UM	JH(Q)	Yes	Age Education II JT alcohol M Geog	NS	Man	Yes	Yes
Cherry (UK) ***	1985	R(E) NS(C)	100%	Yes	44	44	M	41.0($\bar{x}$)	40.8($\bar{x}$)	11.7($\bar{x}$)	>TLV	IM	JH(D) Air	NS	Age II Education Alcohol ED	Yes	Man	Yes	No
Valciukas (USA) *	1985	NS	NS	NS	55	55	M/F	58.2($\bar{x}$)	58.3($\bar{x}$)	<10 - >40	NS	IM	JH(D)	NS	Age Race Gender Education	NS	Man	Yes	No
Maizlish (USA) ***	1985	A(E) NS(C)	50%(E) 38%(C)	Yes	124	116	M/F	35($\bar{x}$) Total Group		up to 7	>TLV	IM	Air	Yes	Age Gender Education JT TD Alcohol	Yes	Man	No	No

PA (C)	YOP	SSM	RR	I/E	No: EG	No: CG	G	Age EG	Age CG	ET years	HE	S	MEA	CLE	C/M	STC	TA	SD	EE
Iregren (Sweden) ***	1986	S(E) NS(C)	81%(E) 81%(C)	Yes	80	80	M	NS	NS	NS	<TLV	UM	Years	Yes	Age II Education JT ES	NS	Man/ Comp	NS	No
Fidler (USA) ***	1987	V	26.5%	Yes	101	NA	M	(19-66) 42.8($\bar{x}$)	NA	17.99($\bar{x}$)	NS	UM	JH(Q)	Yes	Age Education II SES Alcohol Caffeine Blood lead M I SL	Yes	Comp	NA	No
Mikkelsen (Denmark) ***	1988	R(E) R(C)	89% 82%	Yes	85	85	M	53.9($\bar{x}$)	53.5($\bar{x}$)	32.5($\bar{x}$)	>TLV	UM	JH(Q)	Yes	Age Education II Job type Alcohol	NS	Man	Yes	Yes
Baker (USA) ***	1988	V	37.1%	Yes	186	NA	M	(20-67) 40.0($\bar{x}$)	NA	.3-47 16.2($\bar{x}$)	NS	UM	JH(Q)	Yes	Age Education II SES Race Alcohol TE	Yes	Comp	NA	Yes
Crossen (USA) ***	1988	P	90%	Yes	20	NA	M	46.9($\bar{x}$)	NA	17.6($\bar{x}$)	NS	NS	NA	NA	Age Education ES	Yes	Man	NA	No

PA (C)	YOP	SSM	RR	I/E	No: EG	No: CG	G	Age EG	Age CG	ET years	HE	S	MEA	CLE	C/M	STC	TA	SD	EE
Tripathi (India) ***	1989	NS	NS	NS	45	25(1) 25(2)	M	25-40 $\bar{x}$ 32.4	25-40 $\bar{x}$ 33.1(1) 25-40 $\bar{x}$ 32.8(2)	5-6 $\bar{x}$ 10.6	>OEL	UM	JH(D)	Yes	Age II SES Alcohol Smoking ED	Yes	Man	Yes	Yes
Milanovic (Yugoslavia) **	1990	NS	NS	Yes	23	23	M/F	$\bar{x}$ 41.4	$\bar{x}$ 41.0	4	>TLV	UM	Air	Yes	Education JT Alcohol M	NS	Man	Yes	No
Morrow (USA) **	1990	P(E) R(C)	NS	Yes	32	32	NS	$\bar{x}$ 39.7	$\bar{x}$ 40.1	1-19 $\bar{x}$ 9	NS	UM	NS	Yes	Age Education II JT	NS	Man	Yes	No
Ng (Hong Kong) ***	1990	A(E) NS(C)	86%(E) 92%(C)	Yes	78	145	M	16-76 $\bar{x}$ 33.3	19-67 $\bar{x}$ 32.3	NS	>TLV	IM	JH(Q)	Yes	Age Education SES Alcohol TD	Yes	Man	Yes	No
Gupta (India) ***	1990	A(E) NS(C)	100%	Yes	30(1) 15(1)	25	NS	$\bar{x}$ 33(1) $\bar{x}$ 31.5(2)	$\bar{x}$ 34.5	$\bar{x}$ 10.2(1) $\bar{x}$ 4.9(2)	>TLV	IM	Air	Yes	Age Education JT ET TD	NS	Man	Yes	No
Bleeker (USA) ***	1991	A	91% 64% (2 sites)	NS	187	NA	M	28-63 $\bar{x}$ 42	NA	NS	<TLV	IM	JH(Q)	Yes	Age Race II Alcohol Smoking TD	Yes	Man	NA	Yes

PA (C)	YOP	SSM	RR	I/E	No: EG	No: CG	G	Age EG	Age CG	ET years	HE	S	MEA	CLE	C/M	STC	TA	SD	EE
Bowler (USA) ***	1991	P(E) V(C)	NS	Yes	67	67	M/F	43.4($\bar{x}$)	43.5($\bar{x}$)	6.7($\bar{x}$)	>TLV	UM	JH(Q)	Yes	Age Education Race Gender FS ED	Yes	Man	Yes	No
Hanninen (Finland) ***	1991	A(E) V(C)	100%(E) NS(C)	Yes	21	21(1) 28(2) 28(3)	M/F	28-55	28-55 29-64 29-64	5-30	≥TLV	IM	JH(Q)	Yes	Age Genes (Twins)	Yes	Man	Yes	No
Morrow (USA) ***	1992	NS(E) V(C)	NS	Yes	40	40	M/F	39.0	37.2	6.43($\bar{x}$)	>TLV	IM	JH(D)	Yes	Age Education II	NS	Man	Yes	No
Spurgeon (a) (UK) ***	1992	A(E) R(C)	67%(E) NS(C)	Yes	90	90	M	21-65	21-65	1-30	>OEL	UM	JH(D)	Yes	Age Education Alcohol II M VI SL CE TD	YES	Comp	Yes	Yes
Spurgeon (b) (UK) ***	1992	A(E) R(C)	95%(E) NS(C)	Yes	144	144	M	21-65	21-65	1-30	>OEL	UM	JH(D)	Yes	Age Education Alcohol II JT M VI SL CE TD	Yes	Comp	Yes	Yes
Triebig (Germany) **	1992	NS	NS	Yes	83	42	M	44.2($\bar{x}$)	40.5($\bar{x}$)	10-44	NS	NS	Air Bio	NS	Age II JT SES	NS	Man	No	No

PA (C)	YOP	SSM	RR	I/E	No: EG	No: CG	G	Age EG	Age CG	ET years	HE	S	MEA	CLE	C/M	STC	TA	SD	EE
Hooisma (Netherlands) ***	1993	V(E) V(C)	51%(E)(1) 59.5%(C)(2) 29.8%(E)(1) 33.8%(C)(2)	Yes	47(1) 45(2)	53(1) 43(2)	M	31-40 35.8($\bar{x}$)(1) 58-71 63.2($\bar{x}$)(2)	32-40 36.9($\bar{x}$)(1) 56-72 62.7($\bar{x}$)(2)	NS	>TLV	UM	JH(Q)	NS	Age Gender II Alcohol JT	Yes	Comp/ Man	No	No
Colvin (South Africa) ***	1993	A(E) NS(C)	91%	Yes	43	24	M	48.0($\bar{x}$)	43.5($\bar{x}$)	>5	<TLV	IM	Air JH(Q)	NS	Age Education Alcohol Race JT	Yes	Comp/ Man	Yes	Yes
Lee (Korea) **	1993	V(E) V(C)	NS NS	NS	113	81	M	33.3($\bar{x}$)	34.7($\bar{x}$)	NS	>TLV	IM	Air	NS	Age Education Gender Alcohol Smoking JT	NS	Man	Yes	No
Kishi (Japan) *	1993	NS(E) R(C)	NS	NS	20	20	M	25-60 39.5($\bar{x}$)	25-60 39.5($\bar{x}$)	NS	>TLV	IM	Air Bio JH(Q)	NS	Age Gender Education II Alcohol Smoking	NS	Man	Yes	Yes
Ruijten (Netherlands) *	1994	NS	NS	Yes	28	25	NS	38.6($\bar{x}$)	38.7($\bar{x}$)	16.9($\bar{x}$)	<TLV	IM	Air Bio JH(Q)	NS	Age Education Gender Alcohol JT	NS	Comp	Yes	Yes

PA (C)	YOP	SSM	RR	I/E	No: EG	No: CG	G	Age EG	Age CG	ET years	HE	S	MEA	CLE	C/M	STC	TA	SD	EE
Spurgeon (UK) ***	1994	R(E) R(C)	42%(E) NS(C)	Yes	110	110	M	NS	NS	>11 - >30	<OES	IM	JH(Q)	Yes	Age Gender Education II JT Alcohol Smoking J Sat N H VI SL CE	Yes	Comp	No	No
Foo (Singapore) **	1994	NS	NS	Yes	21	21	M	27-53 41.3($\bar{x}$)	25-53 40.8($\bar{x}$)	7-39 20.2($\bar{x}$)	<TLV	IM	Air	NS	Age Gender Education Race Alcohol	NS	Man	Yes	Yes
Bolla (USA) ***	1995	A(E) V(C)	76%(E) NS(C)	Yes	144	52	M	31-63 42($\bar{x}$)	31-63 45($\bar{x}$)	15($\bar{x}$)	NS	IM	Air JH(Q)	Yes	Age Education II Race JT Alcohol Smoking	Yes	Man	Yes	No

PA (C)	YOP	SSM	RR	I/E	No: EG	No: CG	G	Age EG	Age CG	ET years	HE	S	MEA	CLE	C/M	STC	TA	SD	EE
Lundberg (Sweden) ***	1995	A(E) NS(C)	91%(E) 84%(C)	Yes	135	71	M	NS	NS	NS	>OEL	IM	JH(Q)	Yes	Age Education II JT Alcohol I M Smoking OE VT H	Yes	Man	Yes	Yes
White (USA) ***	1995	A	89% (year 1) 84% (year 2)	NS	30	NA	M/F	21-62 34($\bar{x}$)	NA	NS	<TLV	IM	Air JH(Q)	No	Age Gender Education	Yes	Man	NA	No
Broadwell (USA) **	1995	A(E) S(C)	71%(E) NS(C)	Yes	25	32	M/F	47.0($\bar{x}$)	47.6($\bar{x}$)	3-18 9.2($\bar{x}$)	>TLV	IM	Air JH(Q)	NS	Age Gender Education Race Alcohol Smoking M	NS	Man / Comp	Yes	No
Escalona (Venezuela) ***	1995	NS	70%(E) 54%(C)	Yes	67	82	M/F	16-45 33($\bar{x}$)	16-45 30($\bar{x}$)	<5 - >10 7($\bar{x}$)	>TLV	IM	Air	NS	Age Gender Education JT	Yes	Man	Yes	Yes

PA (C)	YOP	SSM	RR	I/E	No: EG	No: CG	G	Age EG	Age CG	ET years	HE	S	MEA	CLE	C/M	STC	TA	SD	EE
Grosch (USA) ***	1996	V(E) V(C)	25%(E) NS(C)	NS	133	51	M/F	40.9($\bar{x}$)	56.0($\bar{x}$)	NS	NS	IM	JH(Q)	Yes	Age II Alcohol Smoking JT	NS	Comp	Yes	Yes
Muijser (Netherlands) ***	1996	V(E) NS(C)	NS	Yes	77	71	M	35.5($\bar{x}$)	37.6($\bar{x}$)	NS	>TLV	IM	Air	Yes	Age Education JT	Yes	Comp	No	No
Tsai (Taiwan) ***	1997	A(E) NS(C)	68%(E) NS(C)	Yes	47(1) 34(2) 88(3)	72(1) 57(2)	M/F	37.9($\bar{x}$)(1) 38.5($\bar{x}$)(2) 38.9($\bar{x}$)(3)	33.2($\bar{x}$)(1) 30.6($\bar{x}$)(2)	NS	>TLV	IM	Air	Yes	Age Gender Education Alcohol Smoking Caffeine SES H M	NS	Comp	Yes	Yes
Morrow (USA) **	1997	V(E) V(C)	NS	No	38	36	M/F	38($\bar{x}$)	35($\bar{x}$)	NS	NS	UM	JH(Q)	Yes	Age JT Alcohol	NS	Man	Yes	No
Saretto (Brazil) *	1997	NS	NS	NS	188	188	M	37.8($\bar{x}$)	38.6($\bar{x}$)	11-20	NS	UM	NS	NS	Age Gender Education JT Alcohol	NS	Comp	Yes	No
Lee (Korea) ***	1998	A(E) NS(C)	NS	Yes	40	28	F	30-59 $\bar{x}$ 48.0	30-59 $\bar{x}$ 48.2	>5 years	>TLV	IM	Air	Yes	Age Education Alcohol Smoking	Yes	Man	Yes	Yes
Daniell (USA) ****	1999	A(E) R(C)	47-52% (over 3 groups)	Yes	67(1) 22(2)	126	M	62-74 (retired)	62-74 (retired)	37(1) 31(2)	>TLV	UM	JH(Q)	Yes	Age Education Alcohol JT IE	Yes	Man	Yes	Yes

PA (C)	YOP	SSM	RR	I/E	No: EG	No: CG	G	Age EG	Age CG	ET years	HE	S	MEA	CLE	C/M	STC	TA	SD	EE
Myers (South Africa) ***	1999	A	NS	Yes	228	NA	M	46.0 $\bar{x}$	NA	13.5($\bar{x}$)	<TLV	IM	Air JH(Q)	Yes	Age Education JT Alcohol SES Race	Yes	Man	NA	No
Nasterlack (Germany) ***	1999	V(E) NS(C)	NS	Yes	366	193	M	41.7($\bar{x}$)	44.6($\bar{x}$)	>10	NS	UM	JH(Q)	NS	Age Education JT Alcohol	NS	Man	Yes	Yes

Annexe 2

Tests and results in 44 studies

N.B. Results from one study (Triebig *et al*) have been omitted due to an idiosyncratic test battery which cannot readily be understood by non-German speakers

Table 1

Tests of general intellectual functioning

Test	Total Studies	Total SD	Total EE
WAIS Digit Symbol	25	15	2
NES Symbol digit	12	7	6
Trails test	6	3	1

WAIS	Wechsler Adult Intelligence Scale
NES	Neurobehavioural Evaluation System
SD	Significant difference between scores demonstrated where exposed and control groups were compared
EE	Significant association between test scores and long-term exposure (exposure-effect)

Table 2

Tests for memory using WMS and WAIS-R

Test	Total Studies	Total SD	Total EE
Visual reproduction	8	6	1
Logical memory	5	3	0
Visual memory	3	1	2
Digit span	3	2	0
Information/orientation	1	0	0
General memory	1	1	0
Delayed recall	1	1	0
Mental control	1	1	0

WMS	Wechsler Memory Scale
WAIS	Wechsler Adult Intelligence Scale-Revised
SD	Significant difference between scores demonstrated where exposed and control groups were compared
EE	Significant association between test scores and long-term exposure (exposure-effect)

Table 3

Other tests for memory

Test	Total Studies	Total SD	Total EE
WAIS Digit span	16	10	2
NES pattern memory	11	4	2
Benton visual retention	12	4	1
NES digit span	5	0	1
Rey-Osterrieth test	2	1	0
Memory reproduction	1	0	0
Brown-Peterson/4 word	3	1	0
Memory scanning	2	0	0
Visual digit span	2	1	1
CETM	1	1	0
Paragraph memory	2	0	0
Recurring words	1	1	0
Memory span	1	0	0
Continuous recognition memory	1	0	0
Memory test	1	1	0
Incidental recall	1	1	0
Corsi block test	1	0	0

WAIS	Wechsler Adult Intelligence Scale
NES	Neurobehavioural Evaluation System
CETM	Contextual Effects on Textual Memory test
SD	Significant difference between scores demonstrated where exposed and control groups were compared
EE	Significant association between test scores and long-term exposure (exposure-effect)

Table 4

Tests for learning

Test	Total Studies	Total SD	Total EE
WMS Verbal PAL	7	4	0
NES AL	5	1	0
NES SDLT	4	0	0
NES AR	2	1	0
Rey AVLT	2	0	0
Buschke LTS	1	0	0
WAIS DS recall	2	1	0
LR	1	1	1
LLR	1	1	1
Benton SDLT	1	1	1
WMS Visual PAL	1	1	0
Claeson-Dahl test	1	0	0

WMS	Wechsler Memory Scale
PAL	Paired Associate learning
NES	Neurobehavioural Evaluation System
AL	Associate learning
AR	Associate recall
SDLT	Serial digit learning test
AVLT	Auditory verbal learning test
LTS	Long term store
DS	Digit symbol
LR	Learning and recall
LLR	Logical learning and recall
SD	Significant difference between scores demonstrated where exposed and control groups were compared
EE	Significant association between test scores and long-term exposure (exposure-effect)

Table 5

Tests for motor speed/control

Test	Total Studies	Total SD	Total EE
Finger tapping	11	0	0
Santa-Ana peg-board	11	5	3
Hand-Eye Co-ord. (NES)	5	0	0
Grooved peg-board	4	2	2
Purdue peg-board	4	1	0
Dotting test	3	2	0
Pursuit aiming	5	1	1
Mira psychomotor test	3	2	0
Tweezer dexterity	1	1	1
Manual dexterity	1	1	0
Pauli test	1	0	0
Fitts Law test	1	0	0

NES Neurobehavioural Evaluation System
SD Significant difference between scores demonstrated where exposed and control groups were compared
EE Significant association between test scores and long-term exposure (exposure-effect)

Table 6

Tests for Visuo-spatial organisation

Test	Total Studies	Total SD	Total EE
WAIS Block design	17	13	6
Picture completion	2	1	0
Embedded figures	2	3	0
Symmetry drawing	3	1	1
NES Pattern Comparison	5	3	2
Figure classification	3	0	0
Spatial transposition	2	0	0
Visual gestalt ability	2	0	0
Mirror drawing test	1	1	0
Geometric shape	2	0	0
Mental rotation	1	0	0
Koh's Blocks	1	1	0

WAIS	Wechsler Adult Intelligence Scale
NES	Neurobehavioural Evaluation System
SD	Significant difference between scores demonstrated where exposed and control groups were compared
EE	Significant association between test scores and long-term exposure (exposure-effect)

Table 7

Tests for vigilance/attention

Test	Total Studies	Total SD	Total EE
NES CPT	6	2	1
NES CWV	4	2	2
Stroop	4	2	1
NES SA	3	0	1
CP	2	0	0
Number search	1	0	0
Perceptual speed	1	1	0
Shape comparison	1	0	0
WMS A/C	1	1	0
Figure Identification	1	1	0

NES	Neurobehavioural Evaluation System
WMS	Wechsler Memory Scale
CPT	Continuous Performance Test
CWV	Colour Word Vigilance
SA	Switching Attention
CP	Continuous Performance
A/C	Attention Concentration
SD	Significant difference between scores demonstrated where exposed and control groups were compared
EE	Significant association between test scores and long-term exposure (exposure-effect)

Table 8

Tests for reasoning ability

Test	Total Studies	Total SD	Total EE
WAIS Similarities	5	1	0
Wisconsin Test	2	1	0
mental arithmetic	2	0	0
Syntactic reasoning	1	0	0
Semantic reasoning	1	0	0

WAIS	Wechsler Adult Intelligence Scale
SD	Significant difference between scores demonstrated where exposed and control groups were compared
EE	Significant association between test scores and long-term exposure (exposure-effect)

Table 9

Tests of Reaction Time*

Test	Total Studies	Total SD	Total EE
SRT	22	7	4
CRT	2	2	2

SRT	Simple Reaction time
CRT	Complex or Choice Reaction Time
SD	Significant difference between scores demonstrated where exposed and control groups were compared
EE	Significant association between test scores and long-term exposure (exposure-effect)

*There is considerable variation in the specific procedures employed to conduct a reaction time test. Few studies describe these in any detail. Therefore the tests have been grouped together under the general categories of SRT and CRT

Annexe 3

Performance on specific tests in relation to acute effects of alcohol consumption

Studies of alcohol consumption and acute effects on selected performance tests							
Principle Author (year)	**No: of subjects (gender)**	**No: of controls (gender)**	**Alcohol Dose**	**Alcohol Level (range or highest)**	**Test Used in Solvent Studies**	**Effect**	**Effects in Other Tests**
Baker (1986)	21(M)	NA	NS	0.05-0.06%	Digit Span	No	Yes
Lukas (1989)	20(F)	NA	0.56 g/kg	up to 80 mg/dl	DSST	Yes	No
Gustafson (1990)	18(M) 18(F)	9(M) 9(F)	1 ml/kg	.062-0.073%	Stroop	Yes	Yes
Hindmarch (1991)	9(M) 9(F)	NA	0.25-1.0 g/kg	NS	Sternberg (Memory scanning)	No	Yes
Corbett (1991)	30(M/F)	15(M/F)	NS	95 mg/dl	DSST	No	No
Kennedy (1993)	20(M)	NA	0.15 g/kg	NS	DSST Syntactic Reasoning	Yes Yes	Yes
Rush (1993)	8(M/F)	NA	0-1.0 g/kg	NS	DSST	Yes	Yes
Kuitunen (1994)	12(M/F)	12(M/F)	.8 g/kg	0.79 g/l	DSST	Yes	Yes
Wilkinson (1995)	24(M)	NA	.75 g/kg	0.08%	Sternberg (Memory scanning)	No	Yes
Newman (1997)	18(M/F)	NA	.88g/kg	60-80 mg/ 100ml	DSST	Yes	Yes

Principle Author (year)	No: of subjects (gender)	No: of controls (gender)	Alcohol Dose	Alcohol Level (range or highest)	Test Used in Solvent Studies	Effect	Effects in Other Tests
Heishman (1997)	5(M)	NA	0-1.0g/kg	90mg/dl	DSST	Yes	Yes
Williamson (2000)	39(M/F)	NA	NS	0.025 - .1%	Syntactic reasoning	Yes	Yes
Tzambazis (2000)	16(M/F)	NA	40ml	0.05%	Block Design Digit Symbol Digit Span	No No No	Yes

DSST = Digit Symbol Substitution Test
NA = Not Applicable
NS = Not Stated

Annexe 4

Performance on specific tests in relation to sleep loss

Studies of sleep deprivation and acute effects on selected performance tests						
Principle Author (year)	**No: of subjects (gender)**	**No: of controls (gender)**	**Time without sleep**	**Tests used in solvent studies**	**Effect**	**Effects in Other Tests**
Linde (1992)	8(M/F)	8(M/F)	1 day + 1 night	Syntactic Reasoning Digit Span	No No	Yes
McCann (1992)	20(M)	20(F)	1 day + 1 night + 1 day	Syntactic Reasoning	No	Yes
Lagarde (1995)	8(M)	NA	60 hours	Syntactic Reasoning	No	No
Kelly (1997)	7(M)	7(F)	64 hours	Syntactic Reasoning	Yes	Yes
Williamson (2000)	39(M/F)	NA	17-19 hours	Syntactic Reasoning	No	Yes

NA = Not Applicable

Annexe 5

Performance on specific tests in relation to ageing

Table 1
Performance on Digit Symbol Substitution Test - WAIS-R & derivatives

Principal Author (year)	**Sample Size (gender)**	**Age-group(s)**	$\bar{x}$ **Scores**
Hooisma (1988)	24(M) 27(F) 24(M) 27(F)	26-47 26-47 60-73 60-73	60.3 67.9 48.3 46.9
Streufert (1992)	44(M)	21-45	70.0
Kuitenen (1994)	24(M/F)	20-28	71.0
Mazaux (1995)	796(M) 1003(F)	65-69 70-74 75-80 >80	39.6 MHE 28.8 MLE 38.9 FHE 29.6 FLE 38.3 MHE 26.5 MLE 35.7 FHE 26.2 FLE 32.2 MHE 24.7 MLE 31.9 FHE 22.8 FLE 28.6 MHE 21.5 MLE 27.8 FHE 21.3 FLE

Principal Author (year)	**Sample Size (gender)**	**Age-group(s)**	**($\bar{x}$) Scores**
Hertzog (1996)	97(M/F) 104(M/F)	18-32 65-81	76.58 48.40
Van Boxtel (1997)	132(M/F)	25-34 35-44 45-54 55-64 65-74 75+	65.5 61.4 61.5 54.4 52.3 48.0
Newman (1997)	18(M/F)	20-50	73.2
Compton (2000)	102(M/F)	30-39 40-49 50-59 60+	(scaled) 11.84 12.03 10.75 10.77

HE = High educational level
LE = Low educational level

Table 2
Performance on tests from the Neurobehavioural Evaluation System
Data from Hooisma 1988

Age-groups (gender)	**n**	**Pattern Comparison**	**Pattern Memory**	**Hand-Eye Co-ordination**	**Finger Tapping**	**Finger Tapping**	**Finger Tapping**
		RT/correct item secs.	**RT/correct item secs.**	$\bar{x}$ **errors**	$\bar{x}$ **taps preferred hand**	$\bar{x}$ **taps non-preferred hand**	$\bar{x}$ **taps alternating hands**
26-47 (M)	24	4.4	6.7	4.5	63.0	60.3	53.0
26-47 (F)	27	4.2	7.6	5.5	56.4	52.9	48.4
60-73 (M)	24	5.4	7.0	5.7	60.0	56.5	47.5
60-73 (F)	27	5.6	7.0	9.1	46.9	46.9	41.1

Performance on tests from the Neurobehavioural Evaluation System Data from Hooisma 1988							
Age-groups (gender)	**n**	**Associate Learning**	**Associate Recall**	**Serial Digit Learning**	**Colour Word Vigilance**	**Continuous Performance Test**	**Symbol Digit Substitution**
		Total 3 trials	**No. Correct**	**Total Correct**	$\bar{x}$ **RT m.secs.**	$\bar{x}$ **RT m.secs.**	$\bar{x}$ **time/item (secs.)**
26-47 (M)	24	17.1	6.4	68.4	621.7	387.3	2.4
26-47 (F)	27	19.0	6.9	67.2	606.8	386.2	2.2
60-73 (M)	24	13.1	5.3	60.0	642.7	378.2	3.1
60-73 (F)	27	12.7	4.8	53.8	632.7	381.4	3.3

Table 3 Performance on Stroop Test			
Principal Author (year)	**Sample Size (gender)**	**Age-group(s)**	$\bar{x}$ **Scores**
Comalli reported in Schumacher (1981)			Interference effect Card C-B
	NS	17-19	47 secs.
	NS	25-34	45 secs.
	NS	35-44	52 secs.
	NS	65-80	96 secs.
	15(M)	80+	160 secs.
Gustafson (1990)	52(M/F)	$\bar{x}$ 25.9	Time/item (100 items) 1.0 secs.
Hindmarch (1992)	18(M/F)	19-26	Time/item 2.8 secs.
Houx (1993)	247(M/F)		Time/item
		$\bar{x}$ 20.1	.8 secs.
		30.4	.8 secs.
		39.9	.9 secs.
		49.8	.8 secs.
		59.8	.9 secs.
		69.7	1.0 secs.
		79.1	1.1 secs.
Salthouse (1995)	112(M/F)		Time/item (20 items)
		20-29	0.7 secs.
		30-39	0.8 secs.
		40-49	0.9 secs.
		50-59	0.9 secs.
		60-69	1.0 secs.
		70+	1.2 secs.

Principal Author (year)	**Sample Size (gender)**	**Age-group(s)**	$\bar{x}$ **Scores**
Van Boxtel (1997)	132 (M/F)		Time/item
		25 $\pm$ 1	.8 secs.
		35 $\pm$ 1	.8 secs.
		45 $\pm$ 1	.9 secs.
		55 $\pm$ 1	1.1 secs.
		65 $\pm$ 1	1.1 secs.
		75 $\pm$ 1	1.3 secs.

Table 4 Performance on Wisconsin Test				
Principal Author (year)	**Sample Size (gender)**	**Age-group(s)**	$\bar{x}$ **Scores**	
			Categories	Errors
Haaland (1987)	95(M/F)	17-25	2.4	19.1
		64-69	2.6	12.2
		70-74	2.2	18.6
		75-79	2.0	21.8
		80-87	1.5	26.1
			Categories	Errors
Bielianskas (1998)	9(M/F)	$\bar{x}$ 71.7	3.1	11.3
			Categories	Errors
Compton* (2000)	102(M/F)	30-39	5.0	8.0
		40-49	4.9	8.4
		50-59	4.0	11.2
		60+	3.3	6.4

*possibly different test version

Table 5
Performance on Sternberg's Memory Scanning Test

Principal Author (year)	**Sample Size (gender)**	**Age-group(s)**	$\bar{x}$ **Scores**	
			Intercept (m.sec.)	**Slope (m.sec./digit)**
Anders (1972) reported in Schumacher (1981)	30(M/F)	$\bar{x}$ 20.0 $\bar{x}$ 37.5 $\bar{x}$ 68.1	623 619 816	39 63 71
Anders (1973) reported in Schumacher (1981)	16(M/F)	$\bar{x}$ 21.2 $\bar{x}$ 55.5	680 892	23 45
Eriksen (1973) reported in Schumacher (1981)	18(M/F)	20-25 35-40 50-55	394 383 459	27 28 43
Hindmarch (1991)	9(M) 9(F)	19-26	439(M) 525(F)	
Hindmarch (1992)	9(M) 9(F)	19-22	438.5(M) 551.5(F)	
Wilkinson (1995)	24(M)	21-40	$\bar{x}$ No. correct 29.0	

Table 6 **Performance on Syntactic Reasoning Test**			
Principal Author (year)	**Sample Size (gender)**	**Age-group(s)**	$\bar{x}$ **Scores**
Linde (1992)	8(M/F)	$\bar{x}$ 23.3	No. correct 26.4
Kennedy (1993)	20(M)	21-42	No. correct 22.0
Lagarde (1995)	8(M)	22-31	No. correct 25.7
Williamson (2000)	39(M/F)	30-50+	No. correct 23.19

Table 7 Performance on Digit Span Test			
Principal Author (year)	**Sample Size (gender)**	**Age-group(s)**	$\bar{x}$ **Scores**
Goldberg In: Schumacher (1981)	7(M/F) 7(M/F) 9(M/F)	20-24 32-51 58-91	7.3 (f) 7.6 (f) 6.1 (f)
Hooisma (1988)	24(M) 27(F) 24(M) 27(F)	26-47 26-47 60-73 60-73	12.8 (f+b) 12.3 (f+b) 10.8 (f+b) 11.6 (f+b)
Rouleau (1996)	16(M/F) 16(M/F)	18-24 65-91	6.5 (f) 6.0 (f)
Hertzog (1996)	97(M/F) 104(M/F)	18-32 65-81	10.6 (b) 8.27 (b)
Tzambazis (2000)	18(M/F)	18-29	8.6 (f) 7.2 (b)

(f) = digits forward
(b) = digits backward

Table 8 Performance on Block Design Test - WAIS-R			
Principal Author (year)	**Sample Size (gender)**	**Age-group(s)**	$\bar{x}$ **Scores**
Troyer (1994)	125(M/F)	50-59 60-69 70-79 80-89 90-95	(raw) 30 30 25 23 20
Hill (1995)	251(M/F)	>75	(scaled) 13.5
Tzambazis (2000)	16(M/F)	18-29	(raw) 70.3

Table 9
Performance on Rey's Auditory Verbal Learning Test

Principal Author (year)	Sample Size (gender)	Age-group(s)	$\bar{x}$ Scores				
			No. Correct				
Vakil			T_1	T_2	T_3	T_4	T_5
(1997)	257(M)	20-29	7.7	10.4	12.1	13.2	13.4
		30-39	7.3	9.9	11.7	12.6	13.1
		40-49	7.4	9.9	11.7	12.6	12.7
		50-59	6.5	9.2	10.7	12.0	12.4
		60-69	6.1	8.3	10.0	11.0	11.4
		70-91	5.3	7.0	8.5	9.8	10.8
	271(F)	20-29	7.9	11.1	12.8	13.6	14.2
		30-39	7.6	10.8	12.4	13.3	13.9
		40-49	7.6	10.7	11.5	12.6	13.1
		50-59	7.2	10.3	12.2	12.9	13.3
		60-69	6.5	9.1	10.5	11.8	12.3
		70-91	5.3	7.5	9.1	9.8	10.7
			No. Correct				
Woodard			T_1	T_2	T_3	T_4	T_5
(1999)	11(M/F)	64-78	6.0	7.5	9.0	9.8	10.5
Petersen		60-69	40.8			5 trials	
(1992)	51(M)	70-79	37.6			5 trials	
	110(F)	80-89	35.2			5 trials	
		90-99	31.9			5 trials	

Table 10 Performance on Finger Tapping Test			
Principal Author (year)	**Sample Size (gender)**	**Age-group(s)**	$\bar{x}$ **Scores**
Hooisma			(preferred hand)
(1988)	24(M)	26-47	22.2
	27(F)	26-47	21.4
	24(M)	60-73	18.8
	27(F)	60-73	17.0
			(non-preferred hand)
	24(M)	26-47	21.4
	27(F)	26-47	20.5
	24(M)	60-73	18.2
	27(F)	60-73	16.3

Table 10 Performance on the Santa Ana Test				
Principal Author (year)	**Test**	**Sample Size (gender)**	**Age-group(s)**	$\bar{x}$ **Scores**
Hooisma (1988)	WHO-NCTB SRT	 24(M) 27(F) 24(M) 27(F)	 26-47 26-47 60-73 60-73	(secs.) 236.2 242.0 220.5 229.1
Hooisma (1988)	WHO-NCTB Benton Visual Retention	 24(M) 27(F) 24(M) 27(F)	 26-47 26-47 60-73 60-73	(no. correct) 8.7 8.3 8.5 8.3
Hooisma (1988)	WHO-NCTB Pursuit Aiming	 24(M) 27(F) 24(M) 27(F)	 26-47 26-47 60-73 60-73	(no. correct) 149.7 166.1 114.6 107.8
Boone (1993)	Rey Osterrieth Test	91(M/F)	45-59 60-69 70-83	Copy 34.2 Recall 18.9 % retention 55.0 Copy 33.8 Recall 17.3 % retention 51.2 Copy 31.3 Recall 13.8 % retention 43.8

Principal Author (year)	**Test**	**Sample Size (gender)**	**Age-group(s)**	$\bar{x}$ **Scores**	
Robbins (1998)	Corsi Block Test	341(M/F)		(no. correct)	
			>55	5.3	
			55-59	5.4	
			60-64	5.3	
			65-69	4.9	
			70-74	4.7	
			75-79	4.7	
Compton (2000)	Trails Test	102(M/F)		(secs.)	
				A	B
			30-39	27.4	55.6
			40-49	30.7	57.8
			50-59	30.1	62.0
			60+	36.9	82.6

Printed and published by the Health and Safety Executive
C1 08/01

MAIL ORDER
HSE priced and free
publications are
available from:
HSE Books
PO Box 1999
Sudbury
Suffolk CO10 2WA
Tel: 01787 881165
Fax: 01787 313995
Website: www.hsebooks.co.uk

RETAIL
HSE priced publications
are available from booksellers

HEALTH AND SAFETY INFORMATION
HSE InfoLine
Tel: 08701 545500
Fax: 02920 859260
e-mail: hseinformationservices@natbrit.com
or write to:
HSE Information Services
Caerphilly Business Park
Caerphilly CF83 3GG

HSE website: www.hse.gov.uk

CRR 355

£20.00

ISBN 0-7176-2060-3