REVIEW

Neurobehavioural testing in workers occupationally exposed to lead: systematic review and meta-analysis of publications

M Goodman, N LaVerda, C Clarke, E D Foster, J Iannuzzi, J Mandel

....

Although the toxic effects of lead on the central nervous system have been well described, the blood concentration at which lead begins to exert adverse effects remains the focus of debate. A meta-analysis of occupational studies was conducted evaluating the association between neurobehavioural testing results and moderate blood lead concentrations.

....

The inclusion criteria for the meta-analysis were (*a*) central tendency for blood lead concentration less than 70 μg/dl; (*b*) numbers of exposed and unexposed workers were reported; and (*c*) test score arithmetic means and measures of dispersion were available for the exposed and the unexposed workers. The data were extracted only for those tests that were included in three or more studies. Analyses involved both the fixed and the random effects models, adjusted, whenever possible, for the test reliability. Publication bias was evaluated by calculating the fail safe N, defined as the number of studies with a non-significant result (p>0.05) that would bring a significant pooled analysis to non-significant levels. Additional analyses examined only those studies that implemented blinding procedures. 22 Studies provided information on 22 tests. Only two of the 22 tests showed an unequivocal significant difference between people with high blood concentrations and controls. The results were sensitive to stratification of studies, adjustment for reliability, and choice of statistical analysis, and were not consistent with the results of an earlier meta-analysis. The data available to date are inconsistent and are unable to provide adequate information on the neurobehavioural effects of exposure to moderate blood concentrations of lead. Lack of true measures of premorbid state, observer bias, and publication bias affect the results.

See end of article for authors' affiliations

....

Correspondence to:
Dr M Goodman, Exponent Health Group, 1800 Diagonal Road, Suite 355, Alexandria, Virginia 22314, USA;
mgoodman@exponent.com

....

The major targets of lead toxicity are the peripheral and central nervous systems (PNS and CNS).[1] In the CNS, symptoms of lead poisoning include dullness, forgetfulness, irritability, poor attention span, headache, fatigue, impotence, dizziness, and depression.[1] Lead encephalopathy, a progressive and potentially fatal degeneration of the brain, is the most severe neurological effect of lead poisoning.[1]

The concentration at which lead begins to exert adverse health effects is not known. Several stud-

ies have suggested the existence of subclinical abnormalities in the absence of overt signs and symptoms of clinical lead poisoning among people with moderately ($\leqslant$70 μg/dl) increased blood lead concentrations.[2-4] By contrast, Parkinson *et al* found few significant differences in neurobehavioural performance between workers exposed and not exposed to lead and concluded that concentrations of current exposure of lead have "no detectable impact on psychological functioning."[5]

A review of the literature on the neurobehavioural effects of cumulative exposure to lead concluded that the current scientific evidence is flawed because of inadequate estimation of cumulative exposure to or absorption of lead and inadequate adjustment for age and intellectual ability before exposure.[6] Another review by Ehle and McKee discussed the difficulties of applying psychological testing in field research and concluded that, for exposure to low to moderate concentrations (<70 μg/dl) of lead, the evidence of neurobehavioural effects is suggestive, but not conclusive.[7] To quantitatively summarise the available evidence, we conducted a meta-analysis of occupational studies evaluating the association between results of neurobehavioural testing and moderate blood lead concentrations.

METHODS
Literature review and study selection
Study selection involved Medline searches using the keywords: "lead AND neurotoxic AND occupational exposure"; "lead AND neurotoxic AND occupation"; "blood lead AND neurobehavioural effects"; "lead AND health effects AND occupation". A research librarian conducted additional searches of other relevant databases, including those referencing ongoing research. These searches were supplemented by a review and retrieval of references from the *Toxicological profile for lead* published by the Agency for Toxic Substances and Disease Registry (ATSDR).[1]

As well as studies published in English, potentially relevant reports were translated from Hebrew, Chinese, Japanese, German, and Danish. In situations where information was missing, attempts were made to contact the author(s) to obtain the missing data. Descriptive information

....

Abbreviations: PNS peripheral nervous system; CNS, central nervous system; ATSDR, Agency for Toxic Substances and Disease Registry
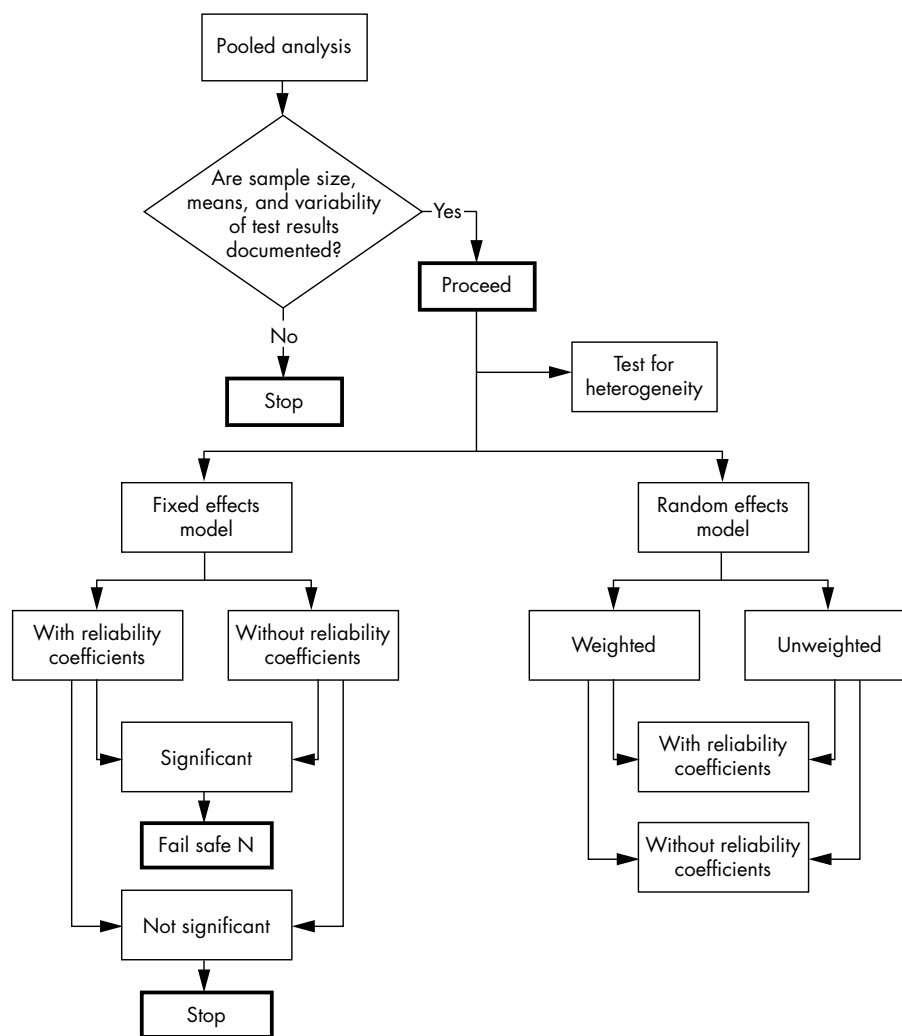
Figure 1 Plan of statistical analysis.

on each neurobehavioural test was reviewed from standard textbooks.[8–10]

The following inclusion criteria were used for the meta-analysis:

- Central tendency for lead exposure was less than 70 μg/dl
- Numbers of exposed and unexposed were reported
- Test score arithmetic means and measures of dispersion were reported for both the exposed and the unexposed workers.

Data extraction

Pertinent data from the studies were entered into an electronic database including central tendency measures for blood lead concentrations among the occupationally exposed and unexposed groups. The outcome information included mean test scores and their corresponding units and measures of dispersion for the exposed and unexposed (control) groups. Before data entry, blood lead concentrations given in μmol/l were converted to μg/100 ml and all measures of dispersion were converted to SDs. Some studies included test score infor-mation for subgroups of exposed workers with a mean blood lead concentration >70 μg/dl. For these studies, only information on those subgroups of exposed workers with an average blood lead concentration <70 μg/dl were extracted. Several studies provided test scores for certain strata within the exposed and unexposed groups. In these instances, infor-mation on each stratum was entered separately. Thus, some articles provided more than one study group.

Data were extracted only for neurobehavioural tests included in three or more studies. We omitted tests involving self rating of affect such as profile of mood states and multiple adjective affect checklist. Only studies reporting non-overlapping data were included.

Data analysis

The analysis of the data included two steps. The first step involved a review of the studies to assess their quality using the following criteria:

- Evaluation of the pre-exposure status
- Adjustment for age
- Adjustment for other occupational exposures
- Adjustment for alcohol use
- Adjustment for socioeconomic confounding factors such as income level, education, etc
- Use of blinding procedures.

The second step included a quantitative meta-analysis of the pooled data. The statistical analysis protocol is illustrated in figure 1. There are two approaches for combining the data: a fixed effects model and a random effects model. The fixed effects method assumes no heterogeneity between studies and attributes all observed variation between results to sampling error alone.[11] The random effects model assumes that the study specific effect sizes come from a random distribution of effect sizes with a certain mean and variance.

Although several tests of heterogeneity have been proposed, their interpretation is problematic because: (*a*) the power of

the test may be insufficient to detect significant heterogeneity[12]; (*b*) the studies may be subject to similar design flaws leading to consistent bias and providing a false impression of homogeneity; and (*c*) the effect sizes may seem to be more consistent than they really are if studies with zero or negative effects are less likely to be published.[13]

Because of the wide variation in testing procedures and scoring practices among the studies, as well as variations in the selected study populations, assumptions of homogeneity, even in the presence of a non-significant test, were generally inappropriate. Therefore, we used both the fixed and the random effects models for each of our analyses. Where the variance between studies was negligible (high level of homogeneity), the random effects models reduced to fixed effects models. To correct for measurement error, whenever possible, all results were adjusted for test-retest reliability.[8 14-17]

Using the fixed effects assumption, the general formula for the weighted average effect size of k studies is

$$\bar{T}. = \frac{\sum_{i=1}^{k} w_i T_i}{\sum^{k} w_i} \tag{1}$$

where T_i is the effect size estimate of the *i*th study and w_i is the weight associated with it. The weights that minimise the variance of T. are given by

$$w_i = \frac{1}{\nu_i} \tag{2}$$

where ν_i is the conditional variance in each study. The average effect size T. has a conditional variance *v*. given by

$$\nu. = \frac{1}{\sum_{i=1}^{k} \left(1/\nu_i\right)} \tag{3}$$

The effect size of each study (T_i) was estimated from the standardised mean difference statistic with the small sample bias correction applied as proposed in Hedges and Olkin.[14] The statistic d_i is calculated as follows:

$$d_i = \frac{\left(\bar{X}_i^{\exp} - \bar{X}_i^c\right)}{s_i} \left(1 - \frac{3}{4N_i - 9}\right) \left(\frac{1}{r_i}\right) \tag{4}$$

where $X^{\exp}_i$ is the mean of the exposed group, X^c_i is the mean of the control group, s_i is the pooled SD of the two groups, N_i is the total sample size of the two groups, and r_i is the reliability of the psychological test used in the study. The conditional variance of d_i is calculated as follows:

$$\nu_i = \left(\frac{n_i^{\exp} + n_i^c}{n_i^{\exp} n_i^c} + \frac{d_i^2}{2\left(n_i^{\exp} + n_i^c\right)}\right) \left(\frac{1}{r_i}\right) \tag{5}$$

where $n^{\exp}_i$ and n^c_i are the sample sizes of the exposed and control groups respectively.

A test of whether the fixed effects assumption holds uses the following statistic:

$$Q = \sum_{i=1}^{k} \left[(T_i - \bar{T}.)^2/\nu_i\right] \tag{6}$$

When the assumption holds, Q has a χ^2 distribution with k-1 degrees of freedom.

Equations for variance change for the random effects assumption. The total variance of an effect size estimate is given by:

$$V_i^* = \sigma^2 + v_i \tag{7}$$

where σ^2 is the random effects variance and v_i is the conditional variance already given. There are two methods for estimating σ^2. The first uses the ordinary unweighted sample estimate for the variance of the effect sizes computed as:

$$S^2(T) = \sum_{i=1}^{k} \left[\frac{(T_i - \bar{T})^2}{(k - 1)}\right] \tag{8}$$

The random variance is then estimated by:

$$\hat{\sigma}^2 = S^2(T) - \left(\frac{1}{k}\right) \sum_{i=1}^{k} v_i \tag{9}$$

If this estimate is negative, it is assumed to be zero and the fixed effects model applies.

The second method for estimating the random variance uses Q, which is taken as an estimate of the weighted sample estimate of the unconditional variance of T_i. In this method, the random variance is estimated by:

$$\hat{\sigma}^2 = \frac{[Q - (k - 1)]}{\sum_{i=1}^{k} w_1 - \left[\frac{\sum_{i=1}^{k} w_i^2}{\sum_{i=1}^{k} w_i}\right]} \tag{10}$$

This second method will give a non-zero estimate only when Q is greater than its expected value. Otherwise, it is assumed to be zero and the fixed effects model applies.

When the random effects model applies, the average effect size T. and its variance *v*. are calculated from equations 1–3; however, v^*_i is substituted for v_i.

When a study provided data for several subgroups, to prevent its overweighting in the meta-analysis, we combined subgroups to calculate a single effect size and then included the result in the final meta-analysis. If the strata represented different control and exposure groups, the combined effect size was calculated from the fixed effects model where k=number of strata. If the strata represented the same exposure and control groups that were tested two or more times with different test versions—such as simple reaction time with different types of stimuli—the effect size was calculated with the fixed effects model where k=number of tests. If a study consisted of one control group and two or more exposure groups, then the exposure groups were combined to calculate a common mean (SD), provided the exposure level fit the criteria.

Publication bias was examined by calculating the fail safe N, defined as the number of studies with a non-significant result (p>0.05) that would bring a significant pooled analysis to non-significant levels. The calculations were based on Rosenthal's application of the Stouffer-Liptar inverse normal

Table 1 Summary of studies included in the meta-analysis

First author	Year	Occupational group strata	Exposed (n)	Unexposed (n)	Exposed (mean BPb)	Unexposed (mean BPb)	BPb range <70µg/dl?	Observers blinded?	Age	Education	Alcohol	Other	Test results before exposure?
Araki[25]	1986	Gun metal foundry workers	19	12	42	13	Y (8–64)	?	N*	N*	N*	N	N
Baker[3]	1984	Foundry workers	160	0	32.8	not given	N (10–80)	N	Y	Y	Y	Y	N
		BPb 0–20	26	‡	not given	Not given	0–20						
		BPb 21–40	97	‡	Not given	Not given	21–40						
		BPb 41–60	28	‡	Not given	Not given	41–60						
Bolla[19]	1995	Organolead manufacturing	190	52	24	Not given	?	Y	Y	N*	N*	Y	N†
Braun[26]	1991	Secondary lead smelter workers	41	37	52.75	Not given	Y (30–66)	N	N*	N*	N*	Y	N
Campara[20]	1984	Electric storage battery	40	20	‡	20.4	Y (11.11–60)	Y	Y	Y	Y	Y	N
		Low exposure	20	‡	31.7	‡	25.9–35						
		High exposure	20	‡	52.2	‡	45.9–60						
Chia[21]	1997	Lead battery manufacturing	50	97	37.1	6.1	Y (2.4–64.6)	Y	Y	Y	Y	Y	N
Espinosa[27]	1990	Lead exposed workers; no other information	31	31	41	7	?	?	Y	Y	N*	N	N
Haenninen[28]	1978	Storage battery or machine shop	49	24	32.3	11.9	Y	?	Y	Y	N	Y	N
Hogstedt[29]	1983	Lead smelter or battery factory	49	27	42	14.7	?	N	Y	Y	N*	Y	N
Jeyaratnam[2]	1986	Lead based stabiliser manufacturing	49	36	49.35	14.7	?	N	N	Y	N	Y	N
Johnson[30]	1980	Lead smelter (high exposure) or zinc plant (low exposure)	358 M; 45 F	108 M; 31 F	‡	‡	?	Y	Y	N	Y	Y	N
		Male controls	‡	108	‡	15.2	?						
		Male zinc exposure	194	‡	39.3	‡	?						
		Male lead exposure	164	‡	56.1	‡	?						
		Female controls	‡	31	‡	10.3	?						
		Female lead exposure	45	‡	29.5	‡	?						
Maizlish[22]	1995	Smelter workers	43	47	42.6	15.3	Y(9–67)	?	Y	Y	Y	Y	N
Milburn[31]	1976	Lead acid battery manufacturing	16	15	61	28	N	?	Y	N	N	N	N
Morgan[24]	1974	Automobile storage battery manufacturing	195	‡	Not given	Not given	N (<29–80+)	N	N	N	N	N	N
		BPb ⩽ 29	6	‡	Not given	Not given	⩽ 29						
		BPb 30–39	14	‡	Not given	Not given	30–39						
		BPb 40–49	38	‡	Not given	Not given	40–49						
		BPb 50–59	43	‡	Not given	Not given	50–59						
		BPb 60–69	45	‡	Not given	Not given	60–69						
Parkinson[5]	1986	Battery plant workers	288	179	40.01	Not given	?	?	Y	Y	N	Y	N
Pasternak[32]	1989	Electrical components manufacturing	24	29	47.46	Not given	Y (4.0–67.67)	Y	N*	N*	N*	Y	N
Pfister[33]	1999	Non-ferrous metal smelter workers	26	48	37.7	Not given	?	?	N	N	Y	Y	N
Repko[34]	1978	Storage battery manufacturing	83–84	52–55	46	18	N (all < 79)	?	Y	Y	N	Y	N
Repko[35]	1975	Storage (lead acid) battery manufacturing	316	112	63.09	15.07	N (0–243)	?	Y	Y	N	Y	N
Ryan[36]	1987	Auto and motorcycle storage battery manufacturing	288	181	40.1	7.2	?	?	Y	Y	N*	N*	N
		Age 21–30	121	54	40.6	7.5	?						
		Age 31–40	90	45	40.3	7.1	?						
		Age 41–50	43	44	39.6	7.3	?						
		Age 51–60	34	38	38.6	7.9	?						
Valciukas[4]	1978	Secondary lead smelter workers	86–90	25	Not given	Not given	N (<40–⩾80)	Y	Y	Y	N	N	N
Williamson[23]	1986	Battery manufacturing and lead smelter	59	59	49.8	Not given	N (25.0–82.3)	?	Y	Y	Y	Y	N

*Authors state that groups were comparable; †authors used vocabulary test as a measure of "pre-morbid state"; ‡data available for strata, but not for the entire exposure group.
N, no; Y, yes; M, male; F, female; BPb, blood lead concentration.

Table 2 Meta-analysis results for three models: fixed effects, weighted random effects, and unweighted random effects

Test	Fixed effects			Weighted random effects Effect size (95% CI)	Unweighted random effects Effect size (95% CI)
	K	Effect size (95% CI)	K_0		
Santa Ana preferred	8	−0.11 (−0.27 to 0.04)		−0.11 (−0.35 to 0.13)	−0.11 (−0.34 to 0.13)
Santa Ana non-preferred hand	6	−0.12 (−0.30 to 0.05)		† (†)	† (†)
Santa Ana both hands	5	0.01 (−0.20 to 0.22)		0.02 (−0.28 to 0.32)	0.02 (−0.28 to 0.32)
Block design test*	11	−0.10 (−0.23 to 0.03)		−0.12 (−0.26 to 0.03)	−0.14 (−0.32 to 0.04)
Logical memory test*	4	0.06 (−0.30 to 0.42)		0.05 (−0.34 to 0.44)	0.05 (−0.36 to 0.46)
Digit symbol*	13	**−0.16 (−0.28 to −0.04)**	45	**−0.32 (−0.59 to −0.05)**	**−0.33 (−0.64 to −0.02)**
D2-errors	6	**0.28 (0.07 to 0.50)**	7	**0.30 (0.06 to 0.55)**	**0.31 (0.05 to 0.56)**
D2-speed	5	**0.37 (0.14 to 0.60)**	8	**0.37 (0.02 to 0.72)**	0.37 (−0.03 to 0.77)
Visual interference	2	−0.03 (−0.20 to 0.14)		−0.21 (−0.80 to 0.38)	−0.21 (−0.80 to 0.38)
Similarities*	6	−0.09 (−0.25 to 0.07)		−0.18 (−0.49 to 0.13)	−0.17 (−0.46 to 0.12)
Bento n visual retention*	5	−0.08 (−0.25 to 0.10)		−0.08 (−0.26 to 0.10)	−0.08 (−0.26 to 0.10)
Paired associates*	6	−0.18 (−0.37 to 0.00)		−0.23 (−0.79 to 0.33)	−0.23 (−0.76 to 0.30)
Visual reproduction*	5	**−0.21 (−0.40 to −0.01)**		−0.17 (−0.44 to 0.11)	−0.19 (−0.42 to 0.05)
Flicker fusion	8	−0.09 (−0.26 to 0.09)		−0.02 (−0.27 to 0.24)	−0.01 (−0.27 to 0.24)
Arithmetic*	5	−0.03 (−0.21 to 0.15)		† (†)	† (†)
Information*	5	−0.12 (−0.28 to 0.04)		−0.19 (−0.46 to 0.09)	−0.18 (−0.43 to 0.07)
Vocabulary*	7	**−0.50 (−0.69 to −0.32)**	35	−0.45 (−0.95 to 0.05)	−0.45 (−0.94 to 0.05)
Symbol digit learning	6	−0.08 (−0.23 to 0.08)		† (†)	† (†)
Digit span backward*	4	−0.09 (−0.32 to 0.15)		† (†)	† (†)
Digit span forward	5	−0.09 (−0.29 to 0.10)		−0.13 (−0.42 to 0.15)	−0.11 (−0.34 to 0.12)
Digit span score*	6	−0.08 (−0.25 to 0.09)	0	−0.26 (−0.64 to 0.13)	−0.24 (−0.58 to 0.11)
Trail making A*	8	**0.23 (0.08 to 0.38)**	9	0.23 (0.00 to 0.45)	0.23 (0.00 to 0.45)
Trail making B*	8	**0.35 (0.19 to 0.52)**	32	0.40 (−0.24 to 1.04)	0.40 (−0.21 to 1.01)
Simple reaction time preferred hand	14	**0.11 (0.01 to 0.21)**	21	0.20 (−0.02 to 0.42)	0.20 (−0.02 to 0.41)
Simple reaction time non-preferred hand	3	**0.35 (0.11 to 0.59)**	3	† (†)	† (†)
Picture completion*	6	**−0.17 (−0.34 to −0.01)**	9	−0.37 (−0.78 to 0.04)	−0.35 (−0.70 to 0.00)
Grooved pegboard dominant hand*	3	**0.25 (0.07 to 0.42)**	2	0.21 (−0.44 to 0.85)	0.22 (−0.30 to 0.74)
Grooved pegboard non-dominant hand*	2	**0.38 (0.19 to 0.56)**	9	0.50 (−0.03 to 1.04)	0.50 (−0.03 to 1.04)
Eye-hand coordination	5	**−0.13 (−0.25 to −0.01)**		−0.13 (−0.26 to 0.00)	−0.12 (−0.29 to 0.06)
Tapping rate non-dominant hand	3	−0.21 (−0.44 to 0.02)		−0.22 (−0.53 to 0.09)	−0.21 (−0.47 to 0.05)
Tapping rate dominant hand	4	−0.21 (−0.42 to 0.01)		−0.21 (−0.42 to 0.01)	† (†)

*Results were adjusted for the reliability coefficient reported in the literature, otherwise reliability coefficients assumed to be 1.0; †Random-effect model reduced to a fixed effect model; **Bold** print indicates significant results; K=number of studies (strata); K_0=Fail safe N.

method of combining p values.[18] To evaluate observer bias, studies that implemented blinding procedures were included in a separate analysis.

RESULTS
Review of the literature
About 140 papers, reports, and books were reviewed. Twenty two studies met the inclusion criteria (table 1). Lead concentrations among study subjects ranged from 24 to 63 μg/dl for exposed and from 0 to 28 μg/dl for unexposed workers. However, none of the studies had an overlap in terms of mean blood lead concentration.

There was considerable variation in quality among the 22 studies that met the inclusion criteria. None of the studies directly compared the test scores in their study populations before and after exposure. However, Bolla et al,[19] recognising the importance of controlling for premorbid verbal intelligence, used the vocabulary subtest score of the WAIS-R in her regression model.

Five studies (Baker et al,[3] Campara et al,[20] Chia et al,[21] Maizlish et al,[22] and Williamson and Teo[23]) adjusted results for the confounders age, education, and alcohol use. One study did not control for any of these variables.[24] The rest of the studies showed varying and inconsistent attention to controls for these potential confounders (table 1).

There seemed to be no consistent pattern for adjustment for other potential confounders. For example, Chia et al[21] adjusted for smoking history and ethnicity, Campara et al[20] matched the study group and referents on "domestic status" (marital status), distance from work, and number of chronic illnesses, and Pasternak et al[32] eliminated subjects with evidence of current illness or injury.

Six of the studies (Bolla et al,[19] Campara et al,[20] Chia et al,[21] Johnson et al,[30] Pasternak et al,[32] and Valciukas et al[4]) indicated that they had used observer blinding procedures.

Meta-analysis
Twenty two neurobehavioural tests met the inclusion criteria. The results of the meta-analysis for each test are presented in table 2. Two tests (digit symbol and D-2 errors) showed a significant effect for all three models used. The result for one test (D-2 speed) was significant when using the fixed effects model and weighted random effects model, but was not significant when using the unweighted random effects model. Also, several tests (simple reaction time, grooved pegboard, trail making A and B, picture completion, visual reproduction, eye-hand coordination, and vocabulary) showed a significant effect, but for the fixed effects model only.

Correction for reliability coefficients did not seem to affect the results substantially. When the data from different subgroups in a single study were combined as opposed to included separately, the results of the meta-analysis changed in some instances, but only for the fixed effects model. The visual reproduction test result changed from statistically non-significant to statistically significant, whereas the result for paired associates remained the same.

Separate inclusion of strata led to overestimation of homogeneity. For example, the random effects model for the visual reproduction test with Ryan et al[36] included as four separate strata was reduced to a fixed effects model because the variance between studies was estimated to be zero. By contrast, inclusion of Ryan et al[36] as a single estimate (after combining the results of the four strata) produced a meaningful random effects model result.

DISCUSSION
Our evaluation of the literature indicates that none of the studies reviewed allow definitive conclusions about the presence or absence of adverse neurobehavioural effects at low (<70 μg/dl) current blood lead concentrations. Perhaps the

most difficult aspect of observational cross sectional studies involving neurobehavioural outcomes is their limited ability to control for wide interpersonal variability of test results.

Only one study (Mantere et al[37]) attempted to use testing results before employment as a background measure of neurobehavioural performance. The authors then re-evaluated the lead workers and controls for the number of people whose testing improved, declined, or stayed the same. They also performed a linear regression analysis of the relation between blood lead time weighted average and performance score. However, these results were reported selectively for two tests only: Santa Ana coordination and block design. Unfortunately, the study by Mantere et al[37] reported the mean test scores and SD only before the start of exposure and, therefore, was not included in our analysis.

Various attempts to control for factors such as age, education, and alcohol use, although useful, are probably insufficient because of the many confounding variables. Therefore, our review of the literature indicates a need for prospective studies that would directly compare the neurobehavioural function before exposure to test results after exposure.

Our meta-analysis shows that changes in study selection criteria or methods of analysis may greatly affect the results. For example, the results for the visual reproduction test changed several times depending on the model used. This finding indicates a need for caution in interpreting the results of meta-analyses in general, and meta-analyses of studies evaluating minor subclinical changes in particular.

In conducting meta-analyses, it is necessary to also consider the possibility of publication bias. Easterbrook et al[38] surveyed 487 research projects approved by the Central Oxford Research Ethics Committee between 1984 and 1987. By 1990, only 52% of these studies were published. Studies with significant positive findings were far more likely to be published. The observational studies were at particularly high risk of such publication bias.

We considered publication bias with two approaches. Firstly, we attempted to contact authors of published studies to identify unpublished and ongoing research. Although we became aware of several potentially relevant studies, we were unable to obtain the data. Secondly, we calculated a fail safe N. These calculations showed that in some instances the results might have been attributed to publication bias.

Our review of individual study results showed substantial variability, even in the absence of a significant homogeneity test. In some cases, the differences in mean values reported by different studies leads one to question whether these studies did, indeed, use the same tests. For example, for simple reaction time (non-preferred hand), Braun and Daigneault,[26] Haenninen et al,[28] and Repko et al[34] reported means for the exposed groups of 178.78, 1310, and 29.62, respectively. The source of such discrepancy is not apparent, although it is possible that different studies used different numbers of items.

In some cases, the source of discrepancy is more evident. For example, the articles by Araki et al,[25] Baker et al,[3] Bolla et al,[19] Campara et al,[20] Chia et al,[21] Hogstedt et al,[29] Jeyaratnam et al,[2] Maizlich et al,[22] and Valciukas et al,[4] with means for the digit symbol test ranging from 29.4 to 59.9, indicate that this test is taken from the WAIS, whereas the papers by Parkinson et al,[5] Pasternak et al,[32] and Ryan et al,[36] with means ranging from 8.3 to 9.83 for the digit symbol test, indicate that the source for the test was the WAIS-R.

We think that in instances when the tests are evaluating the same psychological construct, although perhaps using different procedures and scoring, they legitimately may be combined in a meta-analysis using a random effects model. One methodological feature of this paper is the fact that analyses were conducted and presented with both the fixed effects and random effects models so that the results of these two models can be directly compared. Our reluctance to rigidly

Main messages

- None of the individual studies is conclusive or adequate in providing information on the subclinical neurobehavioural effects of exposure to lead.
- The reviewed studies showed lack of uniform testing method.
- The meta-analysis results are extremely sensitive to changes in study inclusion criteria and use of statistical methods.
- Lack of true measures of premorbid state, observer bias, and publication bias, as well as the selection of individual studies, can all affect the results of the meta-analysis.
- There is a need for prospective studies that would take into consideration variability between people by comparing test scores before and after exposure.

Policy implications

- The data available to date are inconsistent and do not provide adequate information to draw firm conclusions about the biological effects of exposure to lead at current blood lead concentrations of less than 70 µg/dl.
- It is not clear whether regulating exposure to lead based on current blood lead concentration will provide adequate protection against potential neurotoxic effects of lead.
- In making occupational health and safety decisions, the quality of scientific data is more important than the results of pooled analyses based on many studies.

follow a certain analytical model is related to the fact that each of the proposed approaches has its limitations.

A recent meta-analysis by Meyer-Baron and Seeber[39] evaluated 12 tests from 22 studies using similar selection criteria, but a different analytical approach. For each test, a fixed effects model was assumed, the effect sizes calculated, and a test of homogeneity performed. If the test for homogeneity was significant, then a random effects model was used.

With this approach, the results for the block design, logical memory, and Santa Ana (preferred hand only) tests showed significant effects. By splitting the studies that included the digit symbol test into two groups and dropping one study,[5] two homogeneous groups were formed and the effect sizes of both groups were significant. The authors concluded that "The evidence of neurobehavioural deficits at a current blood lead concentration of ~40 µg/100 ml is obvious," however, several concerns make this conclusion seem far reaching. Their meta-analysis included studies with a mean current lead concentration of <70 µg/dl. Therefore, many of the observed effects could be attributed to the inclusion of people whose lead concentrations were much higher than those claimed to be the focus of the study. The authors proposed that their calculated effect sizes for lead exposure might correspond to the effects of aging. This statement is not supported by the data. Moreover, as noted by Balbus-Kornfield et al,[6] the effects of current lead exposure may reflect only transient and reversible changes, whereas the interpretation offered by Meyer-Baron and Seeber assumes permanent effects.

Also, concerns about study selection, inclusion and exclusion of tests, and choice of statistical analyses cast doubt on the conclusions of their meta-analysis. For example, Lindgren et al[40] reported score results for cumulative exposure categories only. Therefore, it is inappropriate to combine Lindgren et al[40] with other studies. Exclusion of the Lindgren study would exclude the digit span, trail making, and logical memory tests in accordance with their study criteria.

For the aiming test, the authors used the eye-hand coordination test from Repko et al[34] and the pursuit aiming test from Maizlish et al[22] and Chia et al[21] However, eye-hand coordination measures time "between successive pulses," whereas pursuit

aiming measures the "number of dotted circles." Thus, the study by Repko et al[34] should have been excluded, thereby eliminating the aiming test from their meta-analysis.

Their block design test meta-analysis included data from Lindgren et al,[40] Haenninen et al,[28] Valciukas et al,[4] Campara et al,[20] and Mantere et al[37] As already stated, the study by Lindgren should have been excluded, but papers by Baker et al[3] and Parkinson et al[5] that contain block design data should have been included.

Comparison of our results with those of Meyer-Baron and Seeber[39] indicates that all three tests (block design, Santa Ana and logical memory) reported as showing significant effects were no longer significant in our analysis. However, our analysis of the digit symbol test with a fixed effects model, unlike the result of Meyer-Baron and Seeber, was significant. In conducting analyses involving a long list of tests, it would be expected that at least some tests would show significant effects by chance. To interpret the results as reflecting an underlying biological process, consistency of findings across studies would be expected. A lack of consistency should caution against far reaching conclusions about the true biological effects of low to moderate lead concentrations.

· · · · · · · · · · · · · · · · · · · ·

Authors' affiliations

M Goodman, N LaVerda, E D Foster, J Iannuzzi, Exponent Health Group, Alexandria, VA, USA
C Clarke, Johns Hopkins University School of Public Health, Baltimore, MD, USA
J Mandel, Exponent Health Group, Menlo Park, CA, USA

REFERENCES

1 **Agency for Toxic Substances and Disease Registry (ATSDR).** *Toxicological profile for lead.* Atlanta, GA: US Department of Health and Human Services, Public Health Service, Agency for Toxic Substances and Disease Registry, 1999.
2 **Jeyaratnam J**, Boey K, Ong C, et al. Neuropsychological studies on lead workers in Singapore. *Br J Ind Med* 1986;**43**:626–9.
3 **Baker EL**, Feldman RG, White RA, et al. Occupational lead neurotoxicity: a behavioral and electrophysiological evaluation. Study design and year one results. *Br J Ind Med* 1984;**41**:352–361.
4 **Valciukas JA**, Lilis R, Eisinger J, et al. Behavioral indicators of lead neurotoxicity: results of a clinical field survey. *Int Arch Occup Environ Health* 1978;**41**:217–36.
5 **Parkinson D**, Ryan C, Bromet E, et al. A psychiatric epidemiologic study of occupational lead exposure. *Am J Epidemiol* 1986;**123**:261–9.
6 **Balbus-Kornfeld JM**, Stewart W, Bolla KI, et al. Cumulative exposure to inorganic lead and neurobehavioral test performance in adults: an epidemiological review. *Occup Environ Med* 1995;**52**:2–12.
7 **Ehle AL**, McKee DC. Neuropsychological effect of lead in occupationally exposed workers: a critical review. *Crit Rev Toxicol* 1990;**20**:237–55.
8 **Lezak MD**. *Neuropsychological assessment.* New York: Oxford University Press, 1995.
9 **Johnson BL**. *Prevention of neurotoxic illness in working populations.* Chichester: Wiley, 1987.
10 **Zenz C**. *Developments in occupational medicine.* Chicago: Year Book Medical Publishers, 1980.
11 **Sutton AJ**, Abrams KR, Jones DR, et al. Systematic reviews of trials and other studies. *Health Technol Assess* 1998;**2**.
12 **Fleiss JL**. Analysis of data from multiclinic trials. *Control Clin Trials* 1986;**7**:267–75.
13 **Matt GE**, Cook TD, Cooper H, et al. (Eds.). *Threats to the validity of research synthesis.* New York: Russell Sage Foundation, 1994:503–520.
14 **Hedges LV**, Olkin I. *Statistical methods for meta-analysis.* London: Academic Press, 1985.
15 **Reitan RM**. Validity of the trail making test as an indication of organic brain damage. *Percept Mot Skills* 1958;**8**:271–6.
16 **Spreen O**, Strauss E. *A compendium of neuropsychological tests.* New York: Oxford University Press, 1991.
17 **Wechsler DA**. *The measurement of adult intelligence, 3rd ed.* Baltimore: Williams and Wilkins, 1944.
18 **Rosenthal R**. The "File Drawer Problem" and tolerance for null results. *Psychological Bulletin* 1979;**86**:638–641.
19 **Bolla KI**, Schwartz BS, Stewart W, et al. Comparison of neurobehavioral function in workers exposed to a mixture of organic and inorganic lead and in workers exposed to solvents. *Am J Ind Med* 1995;**27**:231–46.
20 **Campara P**, D'Andrea F, Micciolo R, et al. Psychological performance of workers with blood-lead concentration below the current threshold limit value. *Int Arch Occup Environ Health* 1984;**53**:233–46.
21 **Chia S-E**, Chia H-P, Ong C-N, et al. Cumulative blood lead levels and neurobehavioral test performance. *Neurotoxicology* 1997;**18**:793–804.
22 **Maizlish N**, Parra G, Feo O. Neurobehavioral evaluation of Venezuelan workers exposed to inorganic lead. *Occup Environ Med* 1995;**52**:408–14.
23 **Williamson A**, Teo R. Neurobehavioral effects of occupational exposure to lead. *Br J Ind Med* 1986;**43**:374–80.
24 **Morgan BB**, Repko JD. Evaluation of behavioral functions on workers exposed to lead. In: Xintaras C, et al. *Behavioral toxicology. National Institute for Occupational Safety and Health Report.* Cincinnati: NIOSH, 1974:248–66.
25 **Araki S**, Yokoyama K, Aono H, et al. Psychological performance in relation to central and peripheral nerve conduction in workers exposed to lead, zinc, and copper. *Am J Ind Med* 1986;**9**:535–42.
26 **Braun CMJ**, Daigneault S. Sparing of cognitive executive functions and impairment of motor functions after industrial exposure to lead: a field study with control group. *Neuropsychology* 1991;**5**:179–93.
27 **Espinosa AS**, Canepa CD, Hermosilla PD, et al. Study of neurobehavioral disturbances from lead exposure using neuropsychological indicators. In: Johnson B, ed. *Advances in neurobehavioral toxicology.* Chelsea, Michigan: Lewis, 1990:339–52.
28 **Haenninen H**, Hernberg S, Mantere P, et al. Psychological performance of subjects with low exposure to lead. *J Occup Med* 1978;**20**:683–9.
29 **Hogstedt C**, Hane M, Agrell A, et al. Neuropsychological test results and symptoms among workers with well-defined long-term exposure to lead. *Br J Ind Med* 1983;**40**:99–105.
30 **Johnson BL**, Burg JAR, Xintaras C, et al. A neurobehavioral examination of workers from a primary nonferrous smelter. *Neurotoxicology* 1980;**1**:561–81.
31 **Milburn H**, Mitran E, Crockford GW. An investigation of lead workers for subclinical effects of lead using three performance tests. *Ann Occup Hyg* 1976;**19**:239–49.
32 **Pasternak G**, Becker CE, Lash A, et al. Cross-sectional neurotoxicology study of lead-exposed cohort. *Clin Toxicol* 1989;**27**:37–51.
33 **Pfister E**, Bockelmann I, Darius S, et al. Diagnosis of early toxic effects of lead or mixed organic solvents: inclusion of psychopathological methods. *Fortschr Neurol Psychiatr* 1999;**67**:435–40.
34 **Repko JD**, Corum CR, Jones PD, et al. *The effects of inorganic lead on neurobehavioral and neurologic function.* Cincinnati: US Department of Health, Education, and Welfare, National Institute for Occupational Safety and Health (DHEW, NIOSH), 1978.
35 **Repko JD**, Norman Jr BB, Nicholson J. *Behavioral effects of occupational exposure to lead.* Cincinnati: US Department of Health, Education, and Welfare, National Institute for Occupational Safety and Health (DHEW, NIOSH), 1975.
36 **Ryan CM**, Morrow L, Parkinson D, et al. Low level lead exposure and neuropsychological functioning in blue collar males. *Int J Neurosci* 1987;**36**:29–39.
37 **Mantere P**, Haenninen H, Hernberg S, et al. A prospective follow up study on psychological effects in workers exposed to low levels of lead. *Scand J Work Environ Health* 1984;**10**:43–50.
38 **Easterbrook PJ**, Berlin JA, Gopalan R, et al. Publication bias in clinical research. *Lancet* 1991;**337**:867–72.
39 **Meyer-Baron M**, Seeber A. A meta-analysis for neurobehavioral results due to occupational lead exposure with blood lead concentrations <70 µg/ml. *Arch Toxicol* 2000;**73**:510–18.
40 **Lindgren K**, Masten V, Ford D, et al. Relation of cumulative exposure to inorganic lead and neuropsychological test performance. *Occup Environ Med* 1996;**53**:472–7.