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.

The major targets of lead toxicity are the peripheral and central nervous systems (PNS and CNS). In the CNS, symptoms of lead poisoning include dullness, forgetfulness, irritability, poor attention span, headache, fatigue, impotence, dizziness, and depression. Lead encephalopathy, a progressive and potentially fatal degeneration of the brain, is the most severe neurological effect of lead poisoning. The concentration at which lead begins to exert adverse health effects is not known. Several studies have suggested the existence of subclinical abnormalities in the absence of overt signs and symptoms of clinical lead poisoning among people with moderately (≤70 µg/dl) increased blood lead concentrations. By contrast, Parkin-son 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.” 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. 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. 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 occupational”. 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).

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

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

Accepted 17 October 2001


www.occenvmed.com
on each neurobehavioural test was reviewed from standard
textbooks.

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. 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

Figure 1  Plan of statistical analysis.
the test may be insufficient to detect significant heterogeneity; (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.

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.

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

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

(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 $T_i$ are given by

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

(2)

where $\nu_i$ is the conditional variance in each study. The average effect size $\bar{T}$ has a conditional variance $\nu$, given by

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

(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. The statistic $d_i$ is calculated as follows:

$$d_i = \frac{(X_{exp} - X_{c})}{s} \left(1 - \frac{3}{4N_i - 9}\right) \left(\frac{1}{r_i}\right)$$

(4)

where $X_{exp}$ is the mean of the exposed group, $X_{c}$ is the mean of the control group, $s$ 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:

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

(5)

where $n_i^{exp}$ and $n_i^c$ are the sample sizes of the exposed and control groups respectively.

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

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

(6)

When the assumption holds, $Q$ has a $\chi^2$ 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:

$$V_i^* = \sigma^2 + \nu_i$$

(7)

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

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

(8)

The random variance is then estimated by:

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

(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:

$$\sigma^2 = \frac{Q - (k - 1)}{\sum_{i=1}^{k} w_i - \sum_{i=1}^{k} w_i / \nu_i}$$

(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 $\bar{T}$ and its variance $\nu$ are calculated from equations 1–3; however, $\nu^*$ is substituted for $\nu$. 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

<table>
<thead>
<tr>
<th>First author</th>
<th>Year</th>
<th>Occupational group strata</th>
<th>Exposed (n)</th>
<th>Unexposed (n)</th>
<th>Exposed (mean BPb)</th>
<th>Unexposed (mean BPb)</th>
<th>BPb range &lt;70µg/dl?</th>
<th>Observers blinded?</th>
<th>Adjusted?</th>
<th>Test results before exposure?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Age</td>
</tr>
<tr>
<td>Araki15</td>
<td>1986</td>
<td>Gun metal foundry workers</td>
<td>19</td>
<td>12</td>
<td>42</td>
<td>13</td>
<td>Y (8–64)</td>
<td>Y</td>
<td>N*</td>
<td>N*</td>
</tr>
<tr>
<td>Baker4</td>
<td>1984</td>
<td>Foundry workers</td>
<td>160</td>
<td>0</td>
<td>32.8</td>
<td>not given</td>
<td>N [10–80]</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BPb 0–20</td>
<td>26</td>
<td>†</td>
<td>not given</td>
<td>Not given</td>
<td>0–20</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BPb 21–40</td>
<td>97</td>
<td>†</td>
<td>Not given</td>
<td>Not given</td>
<td>21–40</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BPb 41–60</td>
<td>28</td>
<td>†</td>
<td>Not given</td>
<td>Not given</td>
<td>41–60</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Bolla14</td>
<td>1995</td>
<td>Organolead manufacturing</td>
<td>190</td>
<td>52</td>
<td>24</td>
<td>Not given</td>
<td>†</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Braun45</td>
<td>1991</td>
<td>Secondary lead smelter workers</td>
<td>41</td>
<td>37</td>
<td>52.75</td>
<td>Not given</td>
<td>Y [30–66]</td>
<td>N</td>
<td>N*</td>
<td>N*</td>
</tr>
<tr>
<td>Campara20</td>
<td>1984</td>
<td>Electric storage battery</td>
<td>40</td>
<td>20</td>
<td>20.4</td>
<td>Y [11.11–60]</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Low exposure</td>
<td>20</td>
<td>†</td>
<td>31.7</td>
<td>†</td>
<td>25–35</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High exposure</td>
<td>20</td>
<td>†</td>
<td>52.2</td>
<td>†</td>
<td>45–60</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Chia24</td>
<td>1997</td>
<td>Lead battery manufacturing</td>
<td>50</td>
<td>97</td>
<td>37.1</td>
<td>6.1</td>
<td>Y [2.4–64.6]</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Espinosa26</td>
<td>1990</td>
<td>Lead exposed workers; no other information</td>
<td>31</td>
<td>31</td>
<td>41</td>
<td>7</td>
<td>‡</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Haenninen28</td>
<td>1978</td>
<td>Storage battery or machine shop</td>
<td>49</td>
<td>24</td>
<td>32.3</td>
<td>11.9</td>
<td>‡</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Hogstedt26</td>
<td>1993</td>
<td>Lead smelter or battery factory</td>
<td>49</td>
<td>27</td>
<td>42</td>
<td>14.7</td>
<td>‡</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Jeyaratnam23</td>
<td>1986</td>
<td>Lead based stabiliser manufacturing</td>
<td>49</td>
<td>36</td>
<td>49.35</td>
<td>14.7</td>
<td>‡</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Johnson50</td>
<td>1980</td>
<td>Lead smelter (high exposure) or zinc plant (low exposure)</td>
<td>358 M, 45 F</td>
<td>108 M, 31 F</td>
<td>‡</td>
<td>‡</td>
<td>‡</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male controls</td>
<td>15</td>
<td>108</td>
<td>15.2</td>
<td>‡</td>
<td>‡</td>
<td>‡</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male zinc exposure</td>
<td>194</td>
<td>†</td>
<td>39.3</td>
<td>‡</td>
<td>‡</td>
<td>!</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male lead exposure</td>
<td>164</td>
<td>†</td>
<td>Not given</td>
<td>Not given</td>
<td>‡</td>
<td>!</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female controls</td>
<td>15</td>
<td>31</td>
<td>10.3</td>
<td>‡</td>
<td>‡</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female lead exposure</td>
<td>45</td>
<td>†</td>
<td>29.5</td>
<td>‡</td>
<td>‡</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Maizlish25</td>
<td>1995</td>
<td>Smelter workers</td>
<td>43</td>
<td>47</td>
<td>42.6</td>
<td>15.3</td>
<td>Y [9–67]</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Milburn31</td>
<td>1976</td>
<td>Lead acid battery manufacturing</td>
<td>15</td>
<td>16</td>
<td>61</td>
<td>28</td>
<td>N</td>
<td>‡</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Morgan24</td>
<td>1974</td>
<td>Automobile storage battery manufacturing</td>
<td>195</td>
<td>‡</td>
<td>Not given</td>
<td>Not given</td>
<td>N (&lt;29–80+)</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BPb ≤ 29</td>
<td>6</td>
<td>†</td>
<td>Not given</td>
<td>Not given</td>
<td>≤ 29</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BPb 30–39</td>
<td>14</td>
<td>†</td>
<td>Not given</td>
<td>Not given</td>
<td>30–39</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BPb 40–49</td>
<td>38</td>
<td>†</td>
<td>Not given</td>
<td>Not given</td>
<td>40–49</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BPb 50–59</td>
<td>43</td>
<td>†</td>
<td>Not given</td>
<td>Not given</td>
<td>50–59</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BPb 60–69</td>
<td>45</td>
<td>†</td>
<td>Not given</td>
<td>Not given</td>
<td>60–69</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Parkinson27</td>
<td>1986</td>
<td>Battery plant workers</td>
<td>288</td>
<td>179</td>
<td>40.01</td>
<td>Not given</td>
<td>‡</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Pasternak29</td>
<td>1989</td>
<td>Electrical components manufacturing</td>
<td>24</td>
<td>29</td>
<td>47.46</td>
<td>Not given</td>
<td>Y [4.0–67.67]</td>
<td>Y</td>
<td>N*</td>
<td>N*</td>
</tr>
<tr>
<td>Pfister25</td>
<td>1999</td>
<td>Non-ferrous metal smelter workers</td>
<td>26</td>
<td>48</td>
<td>37.7</td>
<td>Not given</td>
<td>‡</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Repko24</td>
<td>1978</td>
<td>Storage battery manufacturing</td>
<td>83–84</td>
<td>52–55</td>
<td>46</td>
<td>18</td>
<td>N [all &lt; 79]</td>
<td>‡</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Repko24</td>
<td>1978</td>
<td>Storage (lead acid) battery manufacturing</td>
<td>316</td>
<td>112</td>
<td>63.09</td>
<td>15.07</td>
<td>N [0–243]</td>
<td>‡</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Ryan45</td>
<td>1987</td>
<td>Auto and motorcycle storage battery manufacturing</td>
<td>288</td>
<td>181</td>
<td>40.1</td>
<td>7.2</td>
<td>‡</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Age 21–30</td>
<td>121</td>
<td>54</td>
<td>40.6</td>
<td>7.5</td>
<td>‡</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Age 31–40</td>
<td>90</td>
<td>45</td>
<td>40.3</td>
<td>7.1</td>
<td>‡</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Age 41–50</td>
<td>43</td>
<td>44</td>
<td>39.6</td>
<td>7.3</td>
<td>‡</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Age 51–60</td>
<td>34</td>
<td>38</td>
<td>38.6</td>
<td>7.9</td>
<td>‡</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Valciukas23</td>
<td>1978</td>
<td>Secondary lead smelter workers</td>
<td>86–90</td>
<td>25</td>
<td>Not given</td>
<td>Not given</td>
<td>N [&lt;40–50]</td>
<td>!</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Williamson23</td>
<td>1986</td>
<td>Battery manufacturing and lead smelter</td>
<td>59</td>
<td>59</td>
<td>49.8</td>
<td>Not given</td>
<td>N [23.0–82.3]</td>
<td>!</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

*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.
method of combining p values. 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 concentrations.

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 as a single estimate (after combining the results of the four strata) produced a meaningful random effects model result. Six of the studies (Bolla et al., Campara et al., Chia et al., Johnson et al., Pasternak et al., and Valciukas et al.) 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 speed) 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. 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. 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

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

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

*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); Kw=fail safe N.

www.occenvmed.com

Occup Environ Med: first published as 10.1136/oem.59.4.217 on 1 April 2002. Downloaded from http://oem.bmj.com/ on April 19, 2022 by guest. Protected by copyright.
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.) 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. reported the mean test scores and SDs before and after 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 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. 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, Haenninen et al., and Repko et al. 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., Baker et al., Bolla et al., Campara et al., Chia et al., Hogstedt et al., Jeyaratnam et al., Maizlish et al., and Valcinaks et al., 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., Pasternak et al., and Ryan et al., 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 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 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, 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., 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. reported score results for cumulative exposure categories only. Therefore, it is inappropriate to combine Lindgren et al. 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. and the pursuit aiming test from Maizlish et al. and Chia et al. However, eye-hand coordination measures time “between successive pulses,” whereas pursuit

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.
Neurobehavioural testing in workers occupationally exposed to lead

aiming measures the “number of dotted circles.” Thus, the study by Repko et al. 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., Haenninen et al., Valciukas et al., Campara et al., and Mantere et al. As already stated, the study by Lindgren should have been excluded, but papers by Baker et al. and Parkinson et al. that contain block design data should have been included.

Comparison of our results with those of Meyer-Baron and Seебer 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 Seебer, was significant. In conclusion, other studies imply that 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 LoVerdo, E D Foster, J Iannuzzi, Exponent Health Group, Alexandria, VA, USA
C Clarke, Johns Hopkins University School of Public Health, Baltimore, MD, USA
J Mondel, Exponent Health Group, Menlo Park, CA, USA

REFERENCES


