Available online at www.sciencedirect.com
SCIENCE DIRECT®
NeuroToxicology 27 (2006) 147–152
NeuroToxicology

Evaluating and interpreting exposure–response relationships for manganese and neurobehavioral outcomes

Mary Lou Thompson [a,b], J.E. Myers [a,*]

[a] Occupational and Environmental Health Research Unit, School of Public Health and Family Medicine, University of Cape Town, Anzio Road, Observatory, Cape Town, Western Cape 7925, South Africa
[b] Biostatistics Unit, South African Medical Research Council, South Africa

Received 30 May 2005; accepted 2 August 2005
Available online 12 September 2005

Abstract

Exposure–response relationships (ERRs) are of interest in many areas of epidemiology. In occupational and environmental epidemiology in particular, the nature of such associations may have practical implications for policy and the setting of standards. We use a study undertaken to estimate the association between neurobehavioral test scores and manganese exposure to illustrate that tests for linear trend may be significant in the presence of highly non-linear ERRs. We illustrate this point further with simulated data where the form of the ERR is known. We provide guidelines for exploring the nature of ERRs, in the absence of a priori knowledge of the response of the outcome of interest to exposure.

Keywords: Manganese; Occupational exposure; Neurobehavioral; Exposure–response relationship; Linear trend; Significance testing

1. Introduction

The practice of epidemiology is sometimes persistently at variance with theoretical knowledge. We will focus here on an issue that falls under the concept of analysis deviation—a term used by Steineck and Ahlbom (1992) to identify an important source of bias in epidemiologic studies over and above confounding, selection and information bias. Analysis deviation arises from erroneous calculations or from application of inappropriate statistical models. This relatively neglected aspect of bias has received little attention.

An example of analysis deviation arises when a modeled exposure–response relationship (ERR) is statistically significant, but is based on an inappropriate analytic model. The term ''exposure–response relationship'' will be used here to refer to the form of the association between an outcome of interest and some measure of exposure which is assumed to be recorded on a continuous scale. In order to develop a deeper understanding of the mechanisms by which an exposure impacts an outcome of interest, it is necessary to move beyond the arena of hypothesis testing (assessing the presence or absence of an association between exposure and outcome) to an exploration of the *nature* of the response of the outcome to increasing doses of exposure. Improved knowledge can be valuable in guiding policy development and planning preventive interventions. For example, the setting of occupational exposure limits should be guided by demonstrable exposure–response relations. Knowledge of relationships exhibiting a threshold, below which exposure does not have an adverse affect, can be used to protect exposed workers. Identification of relationships without such a threshold (e.g. carcinogens) could have severe policy and practical implications for implementation constituting pressure for banning the exposure as opposed to setting a limit.

Although the issue has been addressed in the epidemiologic literature (e.g. Maclure and Greenland, 1992; Witte and Greenland, 1997), there remains a tendency, when linear regression models have been fitted, to interpret a significant linear ERR without assessing the validity of the linearity assumption for the data (e.g. Lucchini et al., 1999; Bast-Pettersen et al., 2004). There are many non-linear ERRs which would yield linear regression coefficients that are significantly different from zero. In the absence of a priori theoretical grounds for expecting exposure–response to have a linear form on the appropriate (e.g. biologically plausible) scale, adequate

* Corresponding author. Tel.: +27 214066898; fax: +27 216836443.
E-mail address: myers@cormack.uct.ac.za (J.E. Myers).

0161-813X/$ – see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuro.2005.08.001

model validation is essential. It is this form of analysis deviation as an important source of bias that will be further explored here. Our particular focus is occupational epidemiology, where the problem may be exacerbated by the skew distribution of exposure and where results may have serious policy implications. However, the issue also arises in epidemiology more generally.

Our interest in this question arose during the analysis of a study of the nervous system effects of manganese (Mn) in smelter workers (Myers et al., 2003). Several statistically significant linear and logistic (linear) regression coefficients were obtained for Mn effects on a range of neurobehavioral tests. However, graphical inspection of the data showed the observed ERRs to be at odds with the models imposed, and in some cases, further showed ERRs that were biologically implausible. The peculiarity of our manganese study was that an unusually large number of subjects was studied, which allowed meaningful graphical and non-parametric inspection of the nature of the ERR without our being forced to impose an analytic model. Occupational epidemiology is typically characterized by small subject numbers and hence it can be argued that such studies are particularly susceptible to this form of analysis deviation. For example, of 13 studies reviewed by Iregren (1999) which used psychological tests to document neurotoxic effects from manganese exposure in active workers, subject numbers (including exposed and unexposed subjects) ranged from 21 to 240 with 10 of 13 studies having fewer than 100 subjects and 6 fewer than 50 subjects. In a recent commentary on possible neurological effects of manganese on welders, Finley and Santamaria (2005) emphasized the need for careful evaluation of ERRs in this industry.

In this article, we illustrate in a variety of settings how one may be misled by interpreting statistical significance of tests for linear trends as validation of the existence of such trends. We are not intending to discuss or evaluate alternative *statistical* approaches to fitting ERRs—this topic has been adequately addressed in the above-mentioned statistical literature. Rather, we want to emphasize the need for careful model assessment and interpretation. Uncritical estimates of ERRs may constitute a source of bias, which can lead to inappropriate interpretation with consequences for health policy.

We commence in Section 2 with a practical example from our study in which neurobehavioral test outcomes were related to exposure in a manganese smelter. This is followed in Section 3 by a series of simulated examples, where the true non-linear form of the ERR is known and so we are able to illustrate precisely how one might be misled by fitting a linear trend. In Section 4, we provide some practical guidelines for exploring the nature of the ERR, and Section 5 concludes with a discussion.

2. An example of neurobehavioral response to manganese exposure

A cross-sectional study (Myers et al., 2003) was conducted in South Africa on 509 subjects drawn from eight production environments in a Mn smelting works and 67 unexposed workers in a control workplace. Results of the Digit Symbol (measuring higher cognitive function for collating familiar numbers with obscure symbols from a lookup table) and Digit Span (measuring short term memory for series of spoken numbers) tests from the World Health Organisation Neurobehavioral Core Test Battery (WHO NCTB) are used here in relation to a Mn cumulative exposure index measured in mg years/m^3, calculated for each subject by summing the product of the average exposure intensity for each job held by the subject by the number of years this activity was performed. Given that Mn is an essential element in human metabolism, a threshold would be expected at low doses, above which there would be an increasingly poor test performance (here lower scores on both tests), and below which, given that there is no known condition in humans for Mn deficiency, one would expect to observe no impairment.

Table 1 summarises the results of estimating the (unadjusted) mean test response by exposure category, relative to an unexposed reference group, and of fitting linear trend models associating these outcomes to cumulative exposure. All categorical coefficient estimates are relative to the unexposed referents. So as not to obscure our general argument, we do not include any adjustments for possible confounders. Essentially the same observations would be made, however, for the corresponding adjusted analyses. It would not be uncommon in

Table 1
Exposure–response estimates for selected WHO NCTB tests

WHO NCTB test		Digit Span		Digit Symbol	
Unexposed referents	Mean score	15.6		33.8	
		$\hat{\beta}$	p	$\hat{\beta}$	p
Dichotomous comparisons	All exposed vs. external referents	−4.5	<0.0005	−10.9	<0.0005
	Rest of exposed vs. internal referents	−2.1	<0.0005	−6.2	<0.0005
Categorical Exposure (relative to external referents)	$0 < \text{Exp} \leq 1.3$	−2.1	<0.0005	−3.1	0.08
	$1.3 < \text{Exp} \leq 5.4$	−4.1	<0.0005	−8.5	<0.0005
	$5.4 < \text{Exp} \leq 10.6$	−5.1	<0.0005	−11.3	<0.0005
	$10.6 < \text{Exp} \leq 22.4$	−4.9	<0.0005	−13.2	<0.0005
	$\text{Exp} > 22.4$	−6.1	<0.0005	−18.5	<0.0005
Linear trend	mg years/m^3	−0.05	<0.0005	−0.20	<0.0005

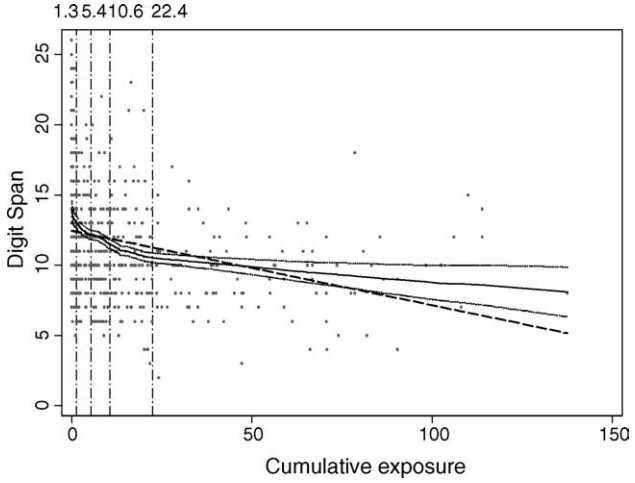

Fig. 1. Exposure–response relationship (ERR) between Digit Span test and cumulative manganese exposure (non-parametric ERR (solid); pointwise 95% confidence limits (dotted); fitted linear trend (dashed)).

the literature to see a table of this form, accompanied by the conclusion that a significant linear trend exists, followed perhaps by some interpretation of this trend. For instance, that a 10 mg years/m^3 increase in cumulative exposure would be expected to be associated with a decrease in Digit Span score of 0.5 and a decrease in Digit Symbol score of 2. Note that, in this study, exposure ranged from unexposed to 137.6 mg years/m^3 and 75% of exposures were below 17.5 mg years/m^3. The sparsity of data at high exposures is also typical of occupational epidemiology. For the average duration of exposure (18 years) of smelter workers in this study, 3.6 mg years/m^3 is the average cumulative exposure corresponding to the American Conference of Governmental Industrial Hygienists Threshold Limit Value (ACGIH TLV) of 0.2 mg years/m^3.

Figs. 1 and 2 show plots of these two test scores from the study versus Mn exposure, with the fitted linear regression lines (dashed) superimposed. Also shown are smooth non-

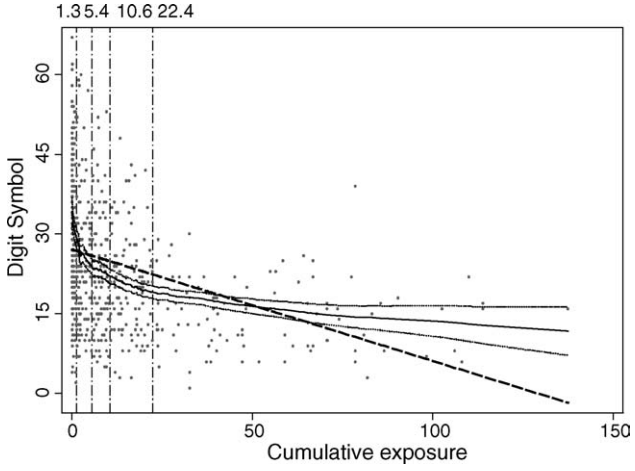

Fig. 2. Exposure–response relationship (ERR) between Digit Symbol test and cumulative manganese exposure (non-parametric ERR (solid); pointwise 95% confidence limits (dotted); fitted linear trend (dashed)).

parametric estimates of the ERR (solid) (using locally weighted regression, see, e.g. Cleveland, 1979; Sasieni and Royston, 1998), with associated pointwise 95% confidence bands (dotted). The vertical lines on the figures are drawn at the boundaries of the exposure categories considered in Table 1. It is important to note the varying ranges of these exposure categories, which were chosen, as is commonly the case, so as to have roughly equal numbers of subjects in each category. The first (exposed) category contains individuals with exposures below 1.3 mg years/m^3, whereas the highest exposure category includes individuals with exposures ranging from 22.4 to 137.6 mg years/m^3. These differences in ranges of exposure must be considered when attempting to assess linearity of change in outcome across categories.

Taken in conjunction with Table 1, it can be seen that the ERRs for both Digit Symbol and Digit Span exhibit a rapid decrease in response at very low levels of exposure. For both neurobehavioral tests considered here, the decline in response with increasing cumulative exposure is more rapid below this exposure threshold than above it. Note further that the fitted linear regression lines lie outside the confidence bands for the smoothed ERR over most of the exposure range in both cases. Simply carrying out a linear trend analysis without accompanying model validation would have obscured these features and led one to assume a continuity of response of the test result to increasing levels of exposure which is not supported by the data.

The observed ERR in both these examples is surprising and somewhat implausible from a biologic perspective. In both cases, the association of test score with *logged* exposure *is* approximately linear, but there is no scientific basis to support an association on this scale; namely that at very low exposure, increasing exposure has the greatest effect on outcome and that, thereafter, increasingly large increments of exposure can be accommodated before similar declines in performance are observed. It should be noted of course that understandings of biological plausibility may evolve over time.

Our primary focus is addressing analysis deviation which may arise when imposing a linear model on the data. We wish to make two main points here. Firstly, that the nature (shape) of the ERR is of interest in its own right. Secondly, that the common practice of basing inference on significance of a linear trend (in large part because of the ready availability of linear regression software) may be misleading and may obscure features of interest.

While Figs. 1 and 2 show quite clearly that a linear trend may not be appropriate to represent these associations, the picture may not be as clear with a smaller data set, particularly one which does not cover as wide a range of exposures as in the Mn smelter study. Consider Fig. 3, which shows the association between exposure and the test outcome Digit Span based on a random sub-sample of 50 observations from the study. Here a linear trend might appear to be a reasonable fit as the fitted line lies completely within the confidence bands (which are however very wide) for the non-parametric ERR. The linear trend was statistically significant with $p = 0.04$.

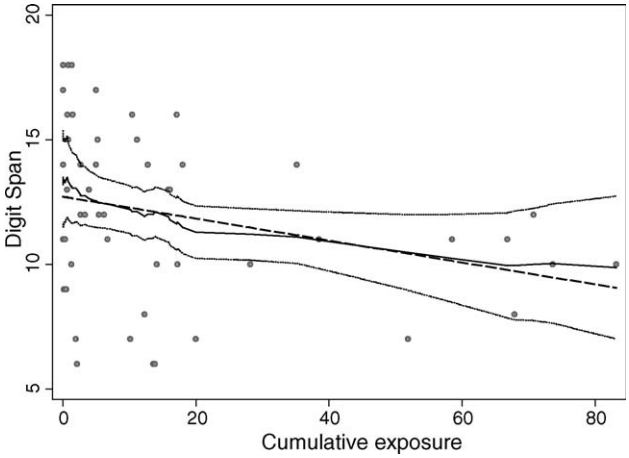

Fig. 3. Exposure–response relationship (ERR) between Digit Span test and cumulative manganese exposure, based on random sub-sample of 50 observations (non-parametric ERR (solid); pointwise 95% confidence limits (dotted); fitted linear trend (dashed)).

Exposure distributions in occupational epidemiology are typically skew, with the bulk of exposures clustered around the low end of the range. With a small dataset, the relatively few data points at high exposure are likely to have undue influence over the shape of the ERR, making the estimate unstable. Large datasets with adequate exposure range allow more flexible non-parametric evaluation of the nature of the ERR. Narrower confidence bands associated with increased sample size will have greater power to indicate an inappropriate assumption of linearity. Consequently, estimated ERRs based on studies with small numbers of subjects should be interpreted cautiously.

3. Some simulated examples

We simulated datasets of 50 subjects with cumulative exposure distribution generated to mimic that which we observed in the manganese smelter study. Let Z denote throughout a standard normal random variable. Then cumulative exposure, E, for each "subject" was generated from the expression:

$$E_i = \exp(1 + 1.4Z_i) \qquad i = 1, 2, \ldots, 50.$$

The median cumulative exposure was 2.2, range (0.2, 106.1), IQR (0.7,7.5).

We considered three non-linear ERRs for a hypothetical neurobehavioral test score, T_n, relative to this exposure, E:

(i) a step function:
$T_1 = 18 + Z$ for $E < 5$
$= 15 + Z$ for $E \geq 5$
i.e. no effect of exposure up to a certain threshold (5 units) and then a constant effect for exposure levels above the threshold;

(ii) single threshold model:
$T_2 = 20 + 1.5Z$ for $E < 25$
$= 20 + 0.5(25 - E) + 1.5Z$ for $E \geq 25$
i.e. no effect of exposure up to a threshold (25 units) and linear trend thereafter;

(iii) non-linear ERR:
$T_3 = 18 + 2\exp(-.2E) + Z$
i.e. response declining to a threshold beyond which further exposure has essentially no additional impact, and where the region of decline in outcome is predominantly at low exposure levels (as in our manganese smelter study).

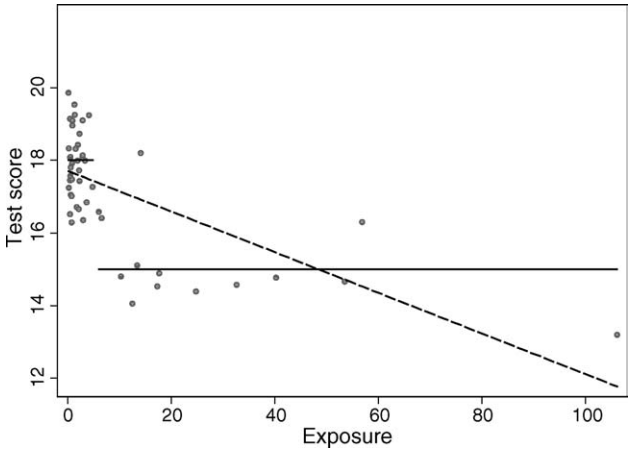

Fig. 4. Step function true ERR and fitted linear trend, simulated sample with $n = 50$ (true ERR (solid); fitted linear trend (dashed)).

Simulated data following each of these models are shown in Figs. 4–6. Each figure also shows the true ERR (solid) and the fitted linear trend (dashed), which was statistically significant ($p < 0.01$) in each instance. While there is, in each case, a real association between exposure and outcome, the associations are not linear and the interpretation of the estimated linear regression coefficient would not only be misleading, it would also cause one to overlook important insights into the exposure–response structure.

The true ERR shown in Fig. 4 might be due to a systematic difference between external controls and the exposed population. In the terminology of Steineck and Ahlbom (1992), this could be misrepresentation bias (or selection bias) and the

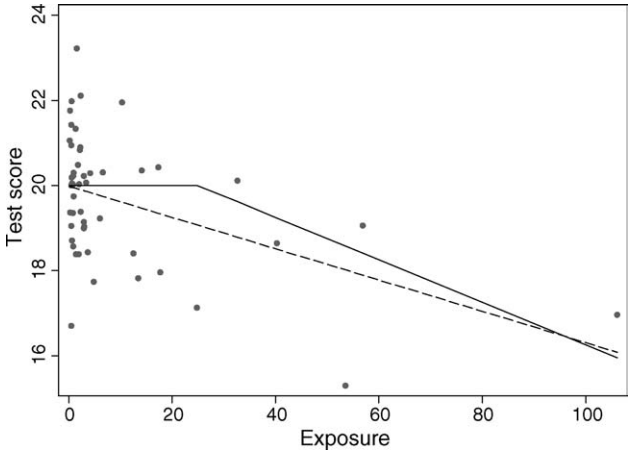

Fig. 5. Threshold function true ERR and fitted linear trend, simulated sample with $n = 50$ (true ERR (solid); fitted linear trend (dashed)).

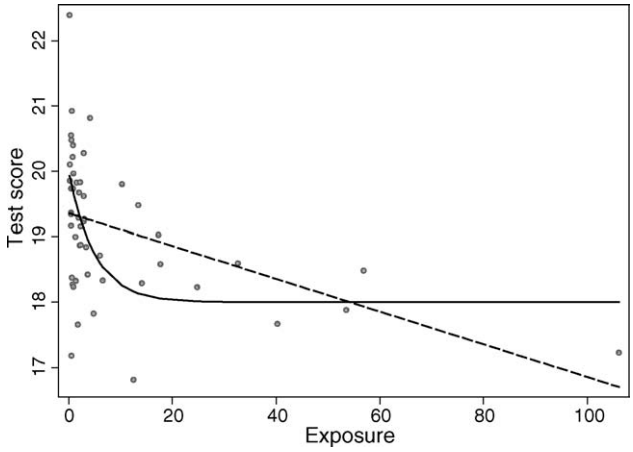

Fig. 6. Non-linear true ERR and fitted linear trend, simulated sample with $n = 50$ (true ERR (solid); fitted linear trend (dashed)).

''effect'' is due to inherent non-comparability between the two groups, and not due to exposure.

Fig. 5 shows an ERR which might have important policy implications in terms of identifying exposure thresholds which would be missed were a linear trend, suggesting a continuity of exposure–response, fitted to the data. The fitted linear trend additionally underestimates the slope of the true decline after the threshold. Subjects exposed above the threshold will be experiencing a more rapid decline than that estimated by the fitted line.

Fig. 6 is an example of the kind of ERR found in the manganese smelter study for the Digit Symbol test. While the estimate of linear trend is highly statistically significant, the true ERR seems implausible and difficult to interpret.

These examples show once again that while non-linear ERRs may lead to statistical significance in tests for linear trend, the results of such hypothesis testing should be treated with caution when sample sizes are not adequate to allow model validation.

4. Practical guidelines

Ideally, a cohort study design would be preferable; however, these are impracticable in many occupational settings. What follows is applicable to the more common cross-sectional design with retrospective cumulative exposure estimation.

Where possible, design of an exposure assessment study should be made with a view to adequately estimating the ERR. Two aspects of study design deserve particular consideration. Given that exposure distributions in occupational epidemiology are frequently skewed, it would, for instance, be desirable to plan for stratified sampling, with over-sampling of those with high exposure. Random sampling of the workforce will lead to high exposures being relatively underrepresented and hence estimates of the ERR in regions of high exposure being poorly estimated and unstable. Secondly, to identify non-linear ERRs, the study should ideally have sufficient subject numbers in all the strata to have adequate power to compare mean outcomes between adjacent strata (see (b) below). Consequent sample size requirements are greater than those required for fitting a linear trend. The manganese study considered here is possibly unusual in occupational settings in that there was a large workforce. In many contexts, a multicentre study will be required to generate sufficient subject numbers, but this may introduce additional variability due to different settings from which subjects are drawn.

If sample size permits, the analysis plan for investigating ERRs should include the following features (adjusted for all potential confounders):

(a) Comparing:
  (i) external unexposed referents with all other subjects, and
  (ii) internal low exposed referents with the rest of the exposed.
(b) Examining the association between outcome and exposure with multiple categories.
  At a minimum, there must be three exposure categories, but the number of categories across which meaningful comparisons of the outcome could be made will depend on the sample size and the variability of the outcome measure. Typically at the start of a study, one only has a broad sense of exposure distributions in different parts of the workplace. Stratifying into notional initial categories of low, medium and high is a commonly used occupational hygiene approach.
(c) Graphical exploration of the association between outcome and continuous exposure via non-parametric ERR estimation (e.g Cleveland, 1979).
(d) Estimating this association with a linear trend, where appropriate, or by more sophisticated modeling, e.g. smoothing splines or fractional polynomials (e.g. Greenland, 1995).
(e) Forming a judgement based on careful examination of the panel of all these results as to the presence, character and significance of the ERR. (Note, as commented above, that, with small data sets ($n \ll 100$), this may be hard or impossible to implement and hence results of such studies should be interpreted with particular caution.)

Ideally, scientific understanding of the nature of the ERR would a priori suggest the functional form that should be included in the statistical modeling. In the absence of such an a priori model, it should be acknowledged that any *data-driven* estimation of the ERR is exploratory (rather than confirmatory) and will need subsequent validation in an appropriately designed separate study. Any fitted exposure response relationship should of course be examined in terms of its biological plausibility while noting that understandings of plausibility may change over time.

5. Discussion

In summary, we have illustrated that:

(i) a test for linear trend may be statistically significant when the actual ERR is non-linear;

(ii) interpreting statistical significance as confirming the presence of a linear trend and then proceeding to interpret the estimated trend itself, may result in erroneous conclusions as to the nature of the ERR—the ERR may be non-linear or, if linear over part of the exposure range, this trend may be under or over estimated by an overall linear trend.

We note further that small studies may have inadequate power to identify deviations from linearity.

In our manganese smelter study, 20 of 110 neurobehavioral tests were associated with statistically significant linear trends with Mn exposure. Uncritical interpretation of these results would have led to conclusions supporting the existence of an ERR. However, more careful examination along the lines described in Section 4 above revealed plausible ERRs for only 2 of these 20 significant tests (constituting under 5% of comparisons made). Neither of these two tests was the Digit Symbol or the Digit Span test (Myers et al., 2003).

In the absence of a priori knowledge of the nature of the ERR, further appropriately designed study (adequate data points with over-sampling of high exposures) will always be needed to validate the form of the ERR observed in a preliminary or exploratory study. It is only when there is a priori knowledge that a linear trend is the appropriate form of the ERR, and the data from a study with sufficiently large subject numbers are compatible with this assumption on exploratory examination, that inferences based on a linear ERR may be assumed to be valid.

These issues have ramifications for occupational health at the level of (1) policy, (2) implementation of preventive interventions and (3) understanding mechanisms of injury or disease. Perhaps more so than in other areas of epidemiology, inferences about the nature of the ERR may have important sequelae in the domains of policy and regulation. When new studies show exposure–response below the level of the accepted occupational exposure limit (OEL), especially when all the measured exposures are below this level, pressure is generated for downwards revision of the OEL. Rational preventive interventions may be undermined by the application of inappropriate linear exposure–response estimates which imply a continuum of response as exposures decrease and hence no safe OEL. Where there is a true level below which there is no adverse effect, such linear modeling underestimates the slope of the ERR above that level. Primary preventive measures may consequently be misspecified, while the initiation of primary and secondary preventive measures (such as medical surveillance) may take place too late.

Where causal mechanisms are understood, they may serve as a validity check for patterns in the data. Where these are not clear, analysis deviation can add to the confusion about empirical indicators for effective prevention at primary and secondary levels.

References

Bast-Pettersen R, Ellingsen DG, Hetland SM, Thomassen Y. Neuropsychological function in manganese alloy plant workers. Int Arch Occup Environ Heath 2004;77(4):277–87.

Cleveland WS. Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc 1979;74:829–36.

Finley BL, Santamaria AB. Current evidence and research needs regarding the risk of manganese-induced neurological effects in welders. Neurotoxicology 2005;26:285–9.

Greenland S. Dose–response and trend analysis in epidemiology: alternatives to categorical analysis. Epidemiology 1995;6(4):356–65.

Iregren A. Manganese neurotoxicity in industrial exposures: proof of effects, critical exposure level, and sensitive tests. Neurotoxicology 1999;20:315–26.

Lucchini R, Apostoli P, Perrone C, et al. Long-term exposure to ''low levels'' of manganese oxides and neurofunctional changes in ferroalloy workers. Neurotoxicology 1999;20(2–3):287–97.

Maclure M, Greenland S. Tests for trend and exposure–response: misinterpretations and alternatives. Am J Epidemiol 1992;135(1):96–104.

Myers JE, Thompson ML, Ramushu S, et al. The nervous system effects of occupational exposure on workers in a South African manganese smelter. Neurotoxicology 2003;24(6):885–94.

Sasieni P, Royston P. Pointwise confidence intervals for running. College Station, TX: Stata Corp; 1998 (STB-41).

Steineck G, Ahlbom A. A definition of bias founded on the concept of the study base. Epidemiology 1992;3(6):477–82.

Witte JS, Greenland S. A nested approach to evaluating dose-response and trend. Ann Epidemiol 1997;7(3):188–93.