NeuroToxicology 45 (2014) 238–246

Contents lists available at ScienceDirect

NeuroToxicology

Toward better research practice—Shortcomings decreasing the significance of epidemiological studies in the toxicological field

Monika Meyer-Baron *, Michael Schäper, Christoph van Thriel

IfADo – Leibniz Research Centre for Working Environment and Human Factors, Neurotoxicology and Chemosensation (Working Group), Dortmund, Germany

ARTICLE INFO

Article history:
Received 25 July 2013
Received in revised form 4 February 2014
Accepted 6 March 2014
Available online 19 March 2014

Keywords:
Risk assessment
Neurobehavioral toxicology
Neuropsychological test
Confounding
Methodological shortcomings

ABSTRACT

Neurobehavioral studies do not always gain the impact they should have, neither in the scientific nor in the regulatory field of neurotoxicology. Among others, shortcomings and inconsistencies across epidemiological studies may contribute to this situation. Examples were compiled to increase awareness of obstacles for conclusions. Meta-analyses were exploited since they sometimes allow the detection of deficits that are not obvious from individual studies.

Exposure assessment, performance measures, and confounding were scrutinized among 98 primary studies included in meta-analyses on mercury, solvents, manganese and pesticides.

Inconsistent and hardly comparable markers of exposure were found; figures, units or sampling periods were not always provided. The contribution of test materials to differences in test outcomes across studies could sometimes not be evaluated due to the insufficient description of the employed tests. Hypotheses for the selection of performance variables often remained undisclosed. Matching procedures prevailed with respect to the confounder age; the comparability of groups with respect to intelligence and gender remained more elusive. 8% and 16% of the studies did not even mention confounding from intelligence and gender, respectively. Only one third of the studies provided adjusted means for group comparisons; the proportion was slightly larger for studies published 2000–2010. While 50% of the studies considered confounders for their dose–response assessment, only 29% reported results for the total of test variables.

The outlined deficits impede, among others, the assessment of exposure–effect relationships and confounding across studies; thereby they limit the use of the studies for toxicological risk assessment and future prevention. Some shortcomings also impede a deeper insight into the mechanisms of toxicity: tests like the Digit Symbol show that something is affected, but not what is affected. Thorough description of measures employed is among the first consequences from the data. The consideration of mechanistic insights from research on animals and neurobiology may further help to increase the significance of epidemiological studies.

1. Introduction

Human health risk assessment comprises hazard identification, dose–response assessment, exposure assessment, and risk characterization (EPA (Environmental Protection Agency), 2013). As a result of the "weight of evidence" approach (Weed, 2005) increased attention has been allocated to systematic reviews and meta-analyses summarizing the evidence across studies for parts of this process. The increasing importance of these reviews also in the toxicological field may be concluded from the attempt of the National Toxicology Program to explore systematic-review methodology as a means to enhance transparency and increase efficiency in summarizing and synthesizing findings (Birnbaum et al., 2013).

The usefulness of meta-analyses for hazard identification and estimates of dose–response curves was described already about 20 years ago (Blair et al., 1995). Since then meta-analyses have undergone a thorough and continuous quality assessment taking into account the most recent outcomes from research. For example, the examination of a feasible publication bias became mandatory since there was evidence that the probability of studies being published is influenced by the direction and strength of

* Corresponding author at: *IfADo* – Leibniz Research Centre for Working Environment and Human Factors, Ardeystr. 67, D-44139 Dortmund, Germany. Tel.: +49 231 1084 416; fax: +49 231 1084 308.

E-mail address: meyerbaron@ifado.de (M. Meyer-Baron).

http://dx.doi.org/10.1016/j.neuro.2014.03.002
0161-813X/

findings (Dickersin, 1997; Dickersin et al., 1992). While publication bias reflects a selective reporting across studies, the selective reporting within studies gained increasing attention during the last years. Its examination was included in guidelines for systematic reviews and meta-analyses, for example the PRISMA statement (Liberati et al., 2009). It was triggered by reviews showing incomplete and biased reporting of trial outcomes that were sometimes inconsistent with study protocols (Chan et al., 2004a,b).

If quality assessment and requirements for systematic reviews remain an issue solely for the group of researchers conducting meta-analyses, the effect will probably be that fewer studies are considered by systematic reviews because of methodological shortcomings and deficits in the reporting of outcomes; at least the studies will be considered with reservations. This situation will finally weaken conclusions. But, it is also conceivable that original studies take advantage of systematic reviews and thereby contribute to the strengthening of conclusions. This may happen when problems occurring during the review process are analyzed and exploited for future studies or the reporting of future studies.

We conducted several meta-analyses of cross-sectional epidemiological studies investigating the neurobehavioral impact of toxicants (Meyer-Baron et al., 2002, 2007, 2008, 2009) and would like to share observations we made and obstacles we met.

We address researchers in the field of neurotoxicology, because we think that several of the problems are specific for our field. Generalizations are probably possible at several points, but not the primary goal of the paper: the intention is to focus the situation we currently encounter when we review neurobehavioral studies.

2. Methods and materials

Four meta-analyses on neurobehavioral performances of workers exposed to either mercury, solvents, manganese or pesticides will be exploited (Meyer-Baron et al., 2002, 2008, 2009, in preparation); a total of 98 studies was analyzed herein. Details on methods and inclusion criteria were described by Meyer-Baron et al. (2008) and will only be briefly summarized.

The included studies compared exposed and reference persons by neuropsychological tests that were also employed in other studies (at least 3 results available for one test). For each of the tests in each of the studies effect sizes were calculated. The effect sizes are standardized mean differences between performances of exposed and unexposed participants. The overall effect for each neuropsychological test was calculated as a weighted mean of the individual results by the use of a random effects model.

3. Observations on obstacles

3.1. Exposure measurement

Table 1 compiles markers of exposure provided by epidemiological studies on pesticides. Internal markers are given where available; alternatively markers of external exposure are displayed.

Already the coloring in Table 1 shows that many different ways were chosen to characterize exposure. In addition, the lack of units, figures, or figures for sub-samples is striking even among studies using the same marker for the exposure characterization.

This heterogeneity impedes estimates of dose–response relations across studies and questions analyses on the effectiveness of biomarkers. Even if the endeavor would be to draw conclusions from a group of studies using the same marker, the lack of details often remains a severe obstacle. Every researcher probably had reasons (or limitations) for the choice of the exposure marker, nevertheless it might be considered whether there are exposure markers that could be provided in addition or in a modified way (e.g. transformation into a common exposure metric) to increase comparability across studies.

Similar problems may occur when measurements of external exposure are provided by individual studies and need to be compared by a systematic review. Table 2 displays measurements of airborne concentrations of manganese. Again, similar markers or measurements are labeled in the same color, this time columnwise.

The exposure metric "airborne concentration" was operationalized in different ways: as a measurement of (total) dust, the inhalable or the respirable fraction. Differences in the sampling strategy with respect to place and time of sampling contribute to differences in representativeness and reliability of the measurements. As in the example on pesticides, even comparisons across studies using the same exposure marker are complicated hereby.

Differences in uptake and transport of manganese are related to the particle size (Andersen et al., 1999; Roels et al., 1997). Subsequently differing neurobehavioral effects might be elicited. Small particles also enter the brain via the olfactory system and thereby increase the manganese content in the brain. Information on the fraction measured (inhalable, respirable) is therefore crucial for the assessment of dose–response relationships, but the majority of studies that could be considered for this purpose did not provide these details. Due to a lack of information on the pore size of filters it is also impossible to identify the overlap between some of the measured fractions (total dust, dust, inhalable). Workplaces may exist where smaller particles are not present or were shown to be negligible, but also studies on welders, who are exposed to ultra-fine particles, did not always provide information on the respirable fraction. It might be argued that the distinctive importance of the particle size was not fully recognized when the first studies were conducted; however, the lack of an alteration among the studies published at distinctly different times suggests that this is not the only explanation for the lack of details.

3.1.1. Conclusions

Conclusions that might be derived from the systematic review would be the following: (1) We still need to spend considerable effort on the thorough description of exposure markers. (2) The exposure markers employed have to to take account of the knowledge we have about routes of exposure. (3) Only the use of the same exposure marker across studies enables the assessment of dose–response relationships across studies and allows to investigate the potential of the biomarker for the explanation of neurobehavioral effects.

3.2. Performance measurement

3.2.1. Comparability of test versions

A problem related to the testing of performances may be highlighted by outcomes on the test Digit Symbol. Table 3 compiles outcomes obtained by different studies investigating occupational exposure to pesticides. Distinctly different means were provided by the studies.

While the differences in the means of the exposed participants might be related to exposure concentrations, the differences among the means of the reference participants question this explanation. Different age groups or intellectual capacities might explain different levels in both groups, but differences in the test material might also be of importance. Between 1955 and 2008 four different versions of the Wechsler Adult Intelligence Scale were published. As shown in Table 3 we hardly get any information which version was actually employed. Only where references to Wechsler were made, an unambiguous identification of the version was possible. The remaining uncertainties impede that differences