

# Longitudinal study on potential neurotoxic effects of aluminium: I. Assessment of exposure and neurobehavioural performance of Al welders in the train and truck construction industry over 4 years

E. Kiesswetter · M. Schäper · M. Buchta ·  
K. H. Schaller · B. Rossbach · H. Scherhag ·  
W. Zschiesche · S. Letzel

Received: 10 August 2006 / Accepted: 1 March 2007 / Published online: 24 May 2007  
© Springer-Verlag 2007

## Abstract

**Objectives** Previous cross sectional studies on potential neurotoxic effects of long-term aluminium exposures by aluminium welders lack clear interpretable results for methodological reasons. The present longitudinal study examined on the one hand the reliability and representativity of Al-biomonitoring as indicator of individual long-term exposure and on the other hand the long-term changes of neurobehavioural performance in Al welders in relation to Al exposure and neurobehavioural performance changes of a non-exposed control group.

**Methods** The longitudinal study compared repeatedly measured exposure data and neurobehavioural data of 20, initially 44, male Al welders in the train and truck construction industry with data of a control group of similar age on the basis of three investigations over a period of 4 years. The repeated measurements of exposure included total dust in air as well as Al in pre- and post-shift plasma and urine samples. Neurobehavioural methods comprised

symptoms, verbal intelligence, logic thinking, psychomotor behaviour, memory, and attention. Computer-aided tests from the Motor Performance Series (MLS) and the European Neurobehavioural Evaluation System (EURO-NES) were used. The characteristics of the biomonitoring data and the relationship to neurobehavioural data were examined with methods of correlation and regression analysis. The courses of neurobehavioural changes were analysed with multivariate covariance-analytical methods (MANCOVA) considering the covariates age, indicators of 'a priori' intelligence differences (education or 'premorbid' intelligence), and alcohol consumption (carbohydrate-deficient transferrin in plasma, CDT).

**Results** The mean total dust load during welding, near to the routinely worn ventilated helmets, was in the range of 5–8 mg/m<sup>3</sup>. The biomonitoring data of the welders (pre-shift: 88–140 µg Al/g creatinine in urine; 13–16 µg Al/l plasma) showed a high long-term stability but also sensitivity to acute shift dependent exposure changes. The Al welders who had been working in this profession at an average of 15 years showed no significantly increased symptom levels compared with the control group. Explorative regression and covariance analyses revealed neither a correlation between biomonitoring and performance variables nor a significant difference between Al-exposed and controls in the performance courses during the 4 years period. Explorative modelling indicated that the structure of neurobehavioural outcomes could be determined by possible indicators of intellectual 'a priori' (premorbid) differences between subjects but not by their exposure information.

**Conclusions** Compared to studies in the literature this study is characterized by relatively high and non-confounded Al exposure of the welders, a repeated-measurement design, and multivariate analyses. However,

---

E. Kiesswetter (✉) · M. Schäper  
Institute for Occupational Physiology,  
University of Dortmund, Ardeystr. 67,  
44139 Dortmund, Germany  
e-mail: kiesswetter@ifado.de

M. Buchta · B. Rossbach · H. Scherhag · S. Letzel  
Institut für Arbeits- Sozial- und Umweltmedizin,  
Universität Mainz, Mainz, Germany

W. Zschiesche  
Berufsgenossenschaft der Feinmechanik und Elektrotechnik,  
Köln, Germany

K. H. Schaller  
Institut und Poliklinik für Arbeits- Sozial- und Umweltmedizin,  
Universität Erlangen-Nürnberg, Erlangen, Germany

the long-term stable interindividual differences of internal Al exposure were not related to interindividual differences in neurobehavioural performances. Additionally, the lack of processual changes of neurobehavioural performances during the observation phase and the insignificant group differences do not make it very probable that degenerative processes caused by Al had happened before study onset or stopped just at this time point.

**Keywords** Aluminium welders · Biomonitoring · Neurobehavioural performance · Longitudinal study

## Introduction

### Background of the study

The aim of this study is to examine whether neurotoxic effects of long-term aluminium (Al) loads can be shown in Al welders. The special features of this study are its longitudinal design and the relatively high Al load. Both exposure and neurobehavioural health of welders in train and truck construction industry were repeatedly recorded over 4 years. Intermediate results of this longitudinal study after 2 years (Buchta et al. 2005) showed that the welders of this industrial sector were considerably higher (90–150 µg Al/g creatinine in urine) exposed than welders of a parallel study in the automobile industry (30–40 µg Al/g creatinine in urine; Buchta et al. 2003).

In the modern vehicle engineering, technically high-quality products are manufactured of Al mainly to minimize vehicle weight. During welding activities the route of Al uptake is essentially inhalation and resorption through the lungs (Letzel et al. 1999). Although Al-welding activities have been carried out in general for more than a decade with ventilation systems and breathing protection (e.g. excess-pressure helmets, Hänninen et al. 1994), the working conditions led to significant internal Al loads documented in several neurobehavioural studies (Table 1).

Inhalation loads arise from Al containing dust and welding fumes containing particles and Al oxide (BGI 616 2005; Antonini et al. 2006a, b). Total dust concentration, Al portion of dust particles, particle size, respirability, and resorption are some of the relevant parameters determining Al uptake.

Animal studies, modelling mainly oral or intraperitoneal Al uptake, show that small portions of Al reach the brain even in the case of intact blood brain barrier (e.g. Colomina et al. 1999), a necessary but not sufficient prerequisite for central neurobehavioural effects. A relatively direct admission of Al from the air to the brain is also possible over olfactory paths, whereas the significance of the uptake is uncertain.

The current German occupational threshold limit value for aluminium in urine is 200 µg Al/l urine (DFG 2006; Biological Tolerance value, BAT). The maximum allowable concentration (MAK) of airborne aluminium is founded on dust measurements, as sufficiently reliable methods to determine the aluminium portion are lacking. The limits are 1.5 mg/m<sup>3</sup> dust in the alveolar dust fraction and 4 mg/m<sup>3</sup> dust in the respirable fraction. Deduced from the exposure data base (MEGA) of the BIA (German institutions for statutory accident insurance and prevention) the aluminium content of dust in workplaces with aluminium exposure, inclusive welding, seems to be in worst cases maximally 30%.

The Al levels in the urine of Al welders, observed in previous neurobehavioural studies, reached up to 270 µg Al/l urine (Table 1). This is roughly 18 times higher than reference values in the general population (<15 µg Al/l urine, Human Biomonitoring 1998). The natural ubiquitous occurrence of Al in the environment, the oral Al uptake by water, food, medications containing Al, antiperspirants, and cosmetics leads to a more or less (un-)avoidable basic Al burden in the general population.

### Models of Al neurotoxicity

Generally, studies on the questions whether and to what degree Al is neurotoxic show a non-uniform pattern. As the few neurobehavioural field study results on welders represent a limited perspective and unclear results (Table 1 and next chapter), other models of Al neurotoxicity might be relevant to deduce and narrow down the critical effect levels of Al exposure.

Mainly three models of Al neurotoxicity are described in the literature: (1) medical models of dialysis encephalopathy, (2) epidemiologic models of possible Al-related illnesses, and (3) experimental animal models with mainly oral and intestinal routes of Al uptake.

1. The therapeutically caused dialysis encephalopathy with potential renal damage of Al elimination is associated with very high plasma or serum levels considerably above occupationally caused levels. Lowest observable effect levels (LOELs) are suspected between 50 µg Al/l and 100 µg Al/l serum (Altmann et al. 1989; Flaten et al. 1996). This model of Al toxicity describes a very specific encephalopathy with speech disorders and seizures which were never observed in examinations of welders. The characteristics are neither comparable to Al-induced responses in animal studies nor to features of solvent induced encephalopathy. It is questionable whether the extremely high Al-load or a combination of factors connected with dialysis is responsible for these specific characteristics.

**Table 1** Neurobehavioural studies of Al welders

Study	<i>n</i> exposed (controls)	Age exp. (controls)	Years Al- welding (cont.)	Co-pre- expo.	Work exposed (controls)	Respiratory protect.	Al $\mu$ g/l urine	Al $\mu$ g/l blood	Added variable	Model premo. intellig	Multivar $\alpha$ correct.	Group- difference <i>n</i> sign.tests	Tests	Exposure- relation <i>n</i> sign.	Tests
Sjögren et al. (1990)	65 (217)	20–69 1–39	1–39	Mn Pb	Al-welding (railroad track weld.)		250		age			2 of 16 odds ratio for 13–38 years	Q16 concentrating <sup>a</sup> , depressed <sup>a</sup>		
Hämminen et al. (1994)	17 (no)	37	last 4(11)	pre weld. others metals	Al, shipyard (no)	yes	76	5.7	no	no	no	0 of 12 (norms)	simple reaction motor tests block design verbal ability	4 of 12 Al-urine (1 inverse <sup>a</sup> )	memory f design <sup>a</sup> digit span <sup>a</sup> symbol learning <sup>a</sup> associative learn <sup>a</sup> mood state motor <sup>2</sup> (Luria <sup>4</sup> , tapping)
Sjögren et al. (1996)	38 12 (39)	39 40 (40)	>5/17 19 (14)	(Pb) Mn (Mn)	Al-welding Mn alloy w. (railroad track welders)	?	§ 22 § 2 § 3 § estimates, upto 165 ds after last shift	§ 3 § <1 § 1	no	no	no	4 of 19	motor tests, <sup>4</sup> SDS simple reaction memory(DS) synonyms CNS <sup>a</sup> , Q16	2 of 15	
Akila et al. (1999)	27 24 (28)	37 42 (38)	? ? (38)	? (Al)	Al, shipyard Al, shipyard (steel welders)	?	61 270 (12)	? ? ?	age edu	no	no	4 of 96	DSS <sup>a</sup> synonyms <sup>a</sup> dual t. count, <sup>2</sup> block design memory f design simple reaction motor tests	5 of 212	DSS <sup>a</sup> syno. (t-limit) <sup>a</sup> dual t. counting 3 <sup>a</sup>
Riihimäki et al. (2000)	29 30 (25)	36 42 (38)	5 12 1	? (Al)	Al, shipyard Al, shipyard (steel welders)	?	49 192 (11)	4 12 2	age	no	no	4 of 24	stroop DSS synonyms <sup>a</sup> counting <sup>a</sup> block design memory f design <sup>a</sup> cancellation <sup>a</sup> memory(DS)	4 of 96 Al-urine (1 inverse <sup>a</sup> )	speed (cancell. <sup>a</sup> , counting <sup>a</sup> ) DSS <sup>a</sup> syno. (t-limit) <sup>a</sup> digit span
Bast-Pettersen et al. (2000)	20 (20)	33 (34)	8	? (34)	Al, rail wag. (construction)	4 years	50	Al-air 1.2 mg inside helmet	age alcohol edu	edu	no	0 of 6 (4 inverse) 1 of 1	steadiness continuous perform. Q16 <sup>a</sup>	4 of 18 1 duration 1 Al-air 0 of 1	steadiness 3 <sup>a</sup> simple reaction <sup>a</sup> Q16

Table 1 continued

Study	<i>n</i> exposed (controls)	Age exp. (cont.)	Years Al- welding	Co-pre- expo.	Work exposed (controls)	Respiratory protect.	Al $\mu\text{g/l}$ urine	Al $\mu\text{g/l}$ blood	Added variable	Model premo. intellig	Multivar $\alpha$ correct.	Group- difference <i>n</i> sign.tests	Tests	Exposure- relation <i>n</i> sign.
Iregren et al. (2001)	38 119	38 46	>5/15 >5/15	Pb? Mn?	Al-welding smelters		§ 22 § 4	§ 3 § 1	age	no	yes Bonferr	0 of 19	s. Sjögren (1996) motor tests n.s.	0 of 57
	16 (39)	(36) (39)	8 12	w.spirit	flake powder (mild steel welders)		§ 83 § 3 § estimates	§ 9 § 1						
Giorgianni et al. (2003)	50 (?)	38 (?)	15	?	shipyard (?)	?	?		no	no	no	3 of 3	memory <sup>a</sup> color word <sup>a</sup> matrices attent <sup>a</sup>	

<sup>a</sup> Sign

- Epidemiological studies which examine the prevalence of a toxic encephalopathy as well as Alzheimer and Parkinson diseases in relation to lifetime environmental and occupational Al exposures hardly give references to significant outcomes (e.g. Antonini 2003; Fored et al. 2006; Fryzek et al. 2005; Graves et al. 1998; Salib and Hillier 1996; Savory et al. 1996).
- Models of Al neurotoxicity from experimental animal studies are limited to the routes of oral, intravenous or intraperitoneal Al uptake under application of Al salts. A review of the more than 50 studies in this research field cannot lead to meaningful conclusions referred to the potential Al neurotoxicity of welding primarily for four reasons:  
Firstly, the absorption and bioavailability of Al depend strongly on both the route of Al uptake and the type of Al compound (Berthon 2002). Animal models studying neurobehavioural effects of inhaled Al in analogy to the exposure of welders (Al dust, Al oxide) are lacking. However, there are new animal-exposure models to simulate specifics of welding fumes (Antonini et al. 2006a). Secondly, food aversion and weight retardation ('systemic effect') were checked neither experimentally nor statistically in the majority of the neurobehavioural studies so that it cannot be ruled out that potential neurobehavioural differences described are of secondary nature. Thirdly, in neurobehavioural animal studies details on internal Al loads are missing which would allow analogies between oral uptake models (of various Al salts) and pulmonary uptake models (resorption of Al dust, Al oxide). Fourthly, the neurobehavioural studies model mainly acute high (e.g. Colomina et al. 1999; Yen-Koo 1992) but not long-term and moderate Al exposure representing a model of occupational life time loads.

After closer examination of the often-cited models it seems that none of the three models is suited to deduce potential neurotoxic effects with reference to exposure levels and temporal conditions of Al welding. However, it cannot be ruled out that lack of significant outcomes could be due to exposure imprecision (e.g. food Al-uptake in animal studies), insensitive outcome variables, and lack of statistical power.

#### Neurobehavioural studies on Al welders

The following short review is confined to neurobehavioural studies on Al welders because of the specificity of Al exposure. Furthermore, it is reduced to a few general aspects concerning content and methods without going into detail. Of primary interest are exposure assessments, pos-

sible affected cognitive domains, study designs, and statistic analyses. Some results and criteria of the review are summarized in Table 1.

#### *Exposure monitoring, internal load*

Table 1 shows that mainly Al welders of shipyards were examined in these neurobehavioural studies. This is of interest because it is difficult to find in this and in similar industrial sectors ‘pure Al welders’ with no former or current welding with other metals (Hänninen et al. 1994; Sjögren et al. 1990; Sjögren et al. 1996; Iregren et al. 2001). The welding of other metals requires as additive predominantly manganese (Mn), which presents a potential neurotoxic co-exposure (Antonini et al. 2006b; Erikson et al. 2005). Some studies controlled the different types of welding by category formation (Sjögren et al. 1996; Iregren et al. 2001).

Only one study informs whether protection equipments (Hänninen et al. 1994) were used and only one informs on external (dust) exposure (Bast-Pettersen et al. 2000). The specimen sampling for biomonitoring differs considerably between studies. Measures represent either single actual samples with no information about time of day or shift, means of several weekly exposure samples, or samples with varying distances to the last exposure up to several months.

The reported group means of the internal Al loads of welders varied in the different neurobehavioural studies between 22 and 270 µg Al/l urine (Table 1). However, the studies of Akila et al. (1999) and Riihimäki et al. (2000) seem to refer to the same sample of workers. Also the studies of Sjögren et al. (1996) and Iregren et al. (2001) used the same group of welders for different comparisons. Thus there is only one examination referring to ‘high’ exposure  $\geq 192$  µg Al/l urine (Akila et al. 1999; Riihimäki et al. 2000), three studies refer to ‘low’ exposure between 22 and 76 µg Al/l urine, and no study represents the wide gap between 76 and 192 µg Al/l urine. The study of Giorgianni et al. (2003) is a short communication without relevant exposure, demographic, and statistical data, and therefore could not be included for further evaluation.

A generalizing comparison of exposure levels between welder studies is risky as the levels depend on sampling. Additionally, information about the representativity, validity, and long-term stability of the used biomonitoring data is lacking in these studies. But such information is necessary if it is assumed, that urinary Al and the variation between pre- and post-shift measures are mainly related to current exposure, whereas serum Al may be related more to prolonged exposure.

Nevertheless, in some studies single significant relationships between exposure data and neurobehavioural

performance parameters were reported (Table 1). The findings are results of multiple testing of different exposure parameters like Al-urine, Al-plasma, and Al-exposure duration. Systematic patterns of most sensitive exposure measures with strongest relation to neurobehavioural performances were not identified.

#### *Potential effects in different cognitive domains*

The neurobehavioural studies of Al welders are exclusively cross-sectional, comparing the performances of Al welders with performances of non-Al welders, unexposed control groups, or norm values (Table 1).

In two studies Al welders showed generally normal performance or performed better than the referents (Hänninen et al. 1994; Bast-Pettersen et al. 2000). Nevertheless, both studies (Table 1) found a few correlations between momentary Al concentrations in urine and psychological performance parameters, but in different domains. According to the authors, their study suggests an inverse relationship of urine Al concentrations to memory and learning (Hänninen et al. 1994) or to motor steadiness (subclinical tremor) (Bast-Pettersen et al. 2000). In one study potential age influences were not included in the exposure effect regression analyses (Hänninen et al. 1994).

Sjögren et al. (1996) stratified welders with different mixed exposures. The interval between last Al-exposure and urine-sampling varied between 20 h and 166 days. The welders mainly exposed to aluminium showed levels of 36 µg lead/l blood and 8 µg manganese/l blood in the mean. The study of Sjögren et al. (1996) suggests motor impairment of Al welders compared with steel welders. However, reanalysis of the data together with two other Al-exposed groups (Iregren et al. 2001) using an approach controlling for age and multiple comparisons (Bonferroni) did not confirm this difference (Table 1).

The high exposure studies of Akila et al. (1999) and Riihimäki et al. (2000) seem to be based on the same samples with little differences in grouping characteristics (Table 1). Both published designs compare steel welders with selected Al welders of low and high exposure, classified based on urinary Al concentrations. In the publication of Akila et al. (1999) more than 200 significance tests have been performed. Only in two group comparisons main effects of the exposure factor are significant ( $P \leq 0.05$ ; memory for design, backward counting) and three are borderline significant ( $P > 0.05$  and  $< 0.10$ ; digit symbol, block design, forced choice synonyms). Ex-post-analyses of single group differences with adjustment for significant covariates (e.g. age or education) show two times significant differences only between referents and low exposure group (memory for design, block design) and three times

only between referents and high exposure group (digit symbol, backward counting, and synonyms task). Mainly speed parameters of tasks (synonyms task and counting) seemed to be correlated with exposure measures. Accuracy of task performances did not show any relationships. Neither Akila et al. (1999) nor Riihimäki et al. (2000) provide information whether verbal differences (synonyms task) were modelled as confounders, e.g. as indicator of pre-morbid intelligence, in statistical analyses (Table 1). According to the authors the neuropsychological testing revealed a “circumscribed effect of aluminium” (p. 180, Riihimäki et al. 2000) concerning complex attention, working memory processing, and visuospatial information. However, digit span showed a paradox positive correlation with urine Al-concentration. Furthermore, there is a paradox shift in affected parameters, e.g. either accuracy is significant (cancellation single task) or speed is significant (cancellation dual task). Riihimäki et al. (2000) suppose a threshold for the adverse effects on the function of the central nervous system in the concentration range of 108–162 µg Al/l urine.

### Conclusions and methodological problems

*Diversity of domains, no observable effect level* The studies of Al welders reflect important steps to isolate psychic domains that might be affected by Al exposure as a consequence of welding. However, the pattern of potential effects is diverse, shifting between verbal, visuospatial, memory, and motor aspects (Table 1). Additionally, there is no trend deducible, that the number or strength of changes increases with exposure. Therefore, across studies neither a sensitive domain nor a ‘lowest observable adverse effect level’ (LOAEL) or ‘no observable adverse effect level’ (NOAEL) is deducible.

*Multiple testing and significance* The group differences, described as significant in neurobehavioural studies of Al welders, refer to a small proportion of multiple tests, are sometimes of borderline significance, sometimes paradox, inconsistent across studies, and sometimes not matching dose–response relationships. Furthermore, parts of complex or correlated tasks were not analysed and evaluated simultaneously ignoring, for example, trade-off effects between speed and accuracy, related effects in dual tasks performances, and performance relations between dominant and non-dominant hand.

Corresponding to these interpretation difficulties, without exception, the studies of welders used univariate statistics, studying each single dependent variable separately without adjusting the significance of thresholds for multiple testing. This overestimates the significance of

effects compared to insignificant effects. Table 1 shows the proportion of significant tests per number of minimally performed tests (lowest limit). If the test-wise error rate ‘alpha’ for a wrong rejection of hypothesis  $H_0$  is fixed to 5%, for example, with 10 univariate tests the examination-wise error rate inflates to 40%. In this case, the adjustment of ‘alpha’ to 0.005 would lead to statistically correct decisions. With adjustments of e.g. Bonferroni type—less conservative corrections of  $P$ -values for multiple comparisons are from Holm (1979) and Hommel (1988)—most tests in the welder studies would not even reach borderline significance. On the other hand, adjustments which are too conservative (unnecessarily small  $\alpha$ ) can lead to the failure not to reject the null hypothesis when the alternative hypothesis is true (Type II error). In this conflict ‘familywise’ error rates, referring to tests with similar hypotheses, or stepwise statistical test procedures are seen as solution.

However, correction methods have generally a basic disadvantage of not providing statistical information on the underlying data structure. Multivariate statistics are the only adequate techniques for analysing complicated data with multiple dependent variables. It allows analysing simultaneously different parameters of a single task, or correlated tasks of a single domain, together with multiple independent factors, including possible confounders, and provides relevant statistics for interpretation.

Across the available Al welder studies (Table 1), firstly the background of univariate statistics without significance adjustment or ‘familywise hypothesis testing’, and secondly the diversity, instability, and weakness of deduced effects suggest accidental results leading to a high risk of causal misinterpretation (Type I error).

*Matching, confounder, premorbid intelligence* Examinations in the industry are subject to restrictions in the pre-planning. Already the non-random worker selection represents a problem which limits the generalisation of results.

With small groups it is impossible to match exposed and controls individually for combinations of factors like age, intelligence, and alcohol consumption. Therefore, it cannot be excluded that the performance differences found occasionally represent ‘a priori’ differences in the intellectual abilities which do not depend on exposure influences but on the interaction with confounders.

However, no Al welder study models explicitly possible ‘a priori’ differences between groups concerning intellectual abilities (e.g. concept of premorbid intelligence) or integrates confounders, like age or alcohol consumption, in a systematic way including the presentation of adequate statistics,  $P$ -values, and explained variance (Table 1).

Adequate covariance models may increase the statistical sensitivity of the main effect test and, in addition, serve to affirm the sensitivity of neurobehavioural measures, e.g. for age, if exposure effects are non-significant.

**Conclusion** A conclusive comparison of neurobehavioural Al welder studies (Table 1) and aggregation of results is not possible as ‘a priori’ differences between groups are not modelled, the quality of confounder control is not recognisable, and the validity and reliability of biomonitoring is not shown. The broadly based search for significant effects contrasts with the focus on singular apparently significant results, ignoring the frequency of numerous insignificant and controversial results. It can be assumed that the diversity of results reflect either accidental differences, independent of Al load, or shifting effects depending on the insufficient control of confounders and the instability of biomarkers.

#### Modelling and objectives of the present study

In contrast to previous studies of welders, the present study will investigate potential neurobehavioural effects of Al using a longitudinal design. Welders of the train and truck construction industry were repeatedly examined with neurobehavioural methods within a period of 4 years. The working conditions were characterized by long-term ‘pure’ Al welding and higher Al exposure than in most of the previously cited welder studies.

In a first step it was intended to test the stability, reliability, and representativity of different exposure measures. These features are preconditions for a meaningful use of exposure assessments in relation to neurobehavioural data. The aims were to test whether there is a positive association between internal and external exposure measures, whether biomonitoring measures are sensitive to acute exposures (pre-, post-shift differences), and whether biomonitoring measures are stable indicators of individual exposure in the long term.

In a second step there will be checked whether Al welders and non-exposed referents show different trends in neurobehavioural performances over the examination period of 4 years. The neurobehavioural methods will include all intellectual domains suspected to be affected by Al in previous studies (Table 1). Adequate statistical methods are the analysis of repeated measurements, simultaneous analyses of multiple dependent variables, and systematic control of relevant covariates known to affect neurobehavioural performance.

The last issue will deal with explorative models explaining ‘a priori’ differences between groups and potential influences of so-called ‘premorbid’ intelligence.

## Methods

### Longitudinal study design

In the longitudinal study design Al welders of the train body and truck trailer construction industry and production workers of a control group were examined three times in a period of 4 years with neurobehavioural methods covering neuropsychologic tests and a symptom survey. The intervals between the examinations amounted to 2 years (Table 2). Corresponding to neurobehavioural measures the external and internal Al loads of welders were recorded. The background Al load of the control group was investigated for comparison purposes.

### Cohorts

The starting sample at the first examination comprised 81 male workers, 44 Al welders and 37 assembly workers as controls, from the same enterprises. Exposed and controls were employees of five companies of the train body and truck trailer construction industry. The enterprises are examples of the Al-processing industry from different regions of Germany belonging to the category of small or medium sized firms, contrasting to the large company of our parallel study in the automobile industry (in preparation). The groups of Al welders and assembly workers were comparable with regard to gender (male), age, education, physical work environment and social environment.

Both groups worked in a weekly rotated three-shift system. This shift system (morning, afternoon, night shift; 8 h) requires a weekly new adjustment of the circadian system of the shiftworkers with special adaptation difficulties during and after the night shift week (Kiesswetter et al. 2000). To avoid circadian and adjustment effects, the examinations were carried out standardized in the second day shift week between 8:00 and 13:00 h. All participants were repeatedly medically examined. Subjects with neurological diseases related to injuries or relevant metabolic illnesses were excluded as well as subjects with insufficient knowledge of German language. Only ‘pure’ Al welders with at least 2 years Al exposure and no current or former exposure to other potential neurotoxic substances (solvents, metals) were included. The control group comprised workers with no known neurotoxic exposures at work.

The study was reviewed by the appropriate ethics committee and was performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki. All study participants gave their informed consent prior to their inclusion in the study.

The present study showed in contrast to the parallel study in the automobile industry (in preparation) a steady diminution of the sample (Table 2) caused by economic

problems, close down of one firm, and lay-offs in the others. As the neurobehavioural analyses aimed on the repeated measurement with the same subjects the study sample reduced markedly. While the repeated-measurement analyses of the first two examinations could be based on 75% of subjects in the exposed sample and 70% in the control sample (Buchta et al. 2005), the present final analyses were based on 45% of the welder sample and 32% of the control sample, resulting in a total sample of 32 workers who had taken part in all three examinations (Table 3).

The Al welders of this final analysis had worked on average 15 years in their job at the time of the last examination (Table 3). The average age of both Al welders and assembly workers was about 43 years. The average values of the alcohol marker carbohydrate deficient-transferrin (CDT) were below 5 U/l in both groups and indicated moderate alcohol consumption. The education index indicated elementary school education in both groups. The statistical examination did not show any significant differences between both groups concerning age (ANOVA,  $P = 0.88$ ), CDT (ANOVA,  $P = 0.41$ ), and education (Man-Whitney,  $P = 0.27$ ).

#### Exposure assessment

For each of the three examinations internal and external exposures of welders were monitored. Total dust in air (see [Background of the study](#)) served as indicator of external exposure and Al in urine and plasma as indicators of internal exposure.

The welders routinely wore forced-air-ventilated helmets during welding. A membrane filter attached in the 'breathing zone' outside the helmets served as the collection of dust. The filter samples were analysed regarding the quantity of total dust but without determining the proportion of Al.

The biomonitoring of welders was carried out pre- and post-shift. Reference values of the Al load in the control

**Table 2** Longitudinal study design with three examinations (E1, E2, E3) within 5 years and corresponding sample sizes

	E1 Start	E2 After 2 years	E3 After 4 years
Exposed ( <i>n</i> )	44	33	20
Controls ( <i>n</i> )	37	26	12
Total ( <i>n</i> )	81	59	32
Dust	x	x	x
Al-urine	x	x	x
Al-plasma	x	x	x
Neurob.-tests	x	x	x
Symptoms	x	x	x

**Table 3** Background characteristics of subjects with repeated neurobehavioural measurements ( $n = 32$ )

	Exposed $n = 20$		Controls $n = 12$	
	Mean	SD	Mean	SD
Age (years)	43.3	7.4	42.9	5.7
Education (index)	1.4	0.4	1.2	0.4
CDT (U/l)	4.3	4.2	2.9	5.5
Al-welding (y)	14.8	4.1	–	–

The data of age and Al welding refer to the time of the third examination

Education index: 0 = no elementary school, 1 = primary school, 2 = secondary school, 3 = high school graduation

CDT Carbohydrate-deficient transferrin used as a biomarker for alcohol consumption

group were determined exclusively pre-shift. The Al urine values were referred to the creatinine concentrations in urine intending to consider the dilution depending on the quantity of fluid intake.

Quantitative determinations of Al in plasma and urine samples were carried out by graphite furnace—atomic absorption spectrometry with specific calibration technique according to Angerer and Schaller (1982), Fleischer and Schaller (1999), and internal and external quality assessment (Schaller et al. 2002).

#### Medical examination

Before the first examination, a standardized medical interview including occupational history was carried out. Additionally, a physical examination, including neurological status and pulmonary function tests, was performed. Relevant parts of the interview and examination were repeated in the following two examinations.

#### Neurobehavioural methods

##### Schedule

The neurobehavioural methods contained a questionnaire for the recording of neurotoxic symptoms and a number of psychological tests assigned to different functional domains that included approaches to the modelling of pre-morbid intelligence.

The schedule of neurobehavioural methods (Table 4) shows the application scheme of questionnaire and tests across the three examinations.

The Standard Progressive Matrices Test, a test of logic thinking, was carried out only in the first examination, and a test of verbal intelligence (WST) only in the second examination.



The computer-aided European Neurobehavioural evaluation system (EURO-NES) was used in the second and third examinations. The new tests were introduced by a new cooperation partner who joined the research at the beginning of the second examination. All other methods of testing and the symptom questionnaire Q16 were employed in all three examinations.

#### Symptom questionnaire

For symptom monitoring a German version of the well-known Swedish questionnaire Q16 was applied (Hogstedt et al. 1984; Ihrig et al. 2001). The questionnaire focuses on psychological and neuropsychological symptoms related to neurotoxic exposures.

#### Neuropsychological tests

The tests can be associated with four neuropsychological domains.

**Cognitive abilities** Four tests represent intelligence tests and measure different forms of cognition. Typically they are related to unique parts of cognition and to a general factor of intelligence. The four tests are the German multiple choice vocabulary test (Wortschatztest, WST; Schmidt and Metzler 1992), the German version of the Standard Progressive Matrices Test (SPM, Raven; Heller et al. 1988), the block design test (HAWIE, Wechsler 1991), and the Trail Making Test (TMT, German: Zahlenverbindungstest; ZVT, Oswald and Roth 1997).

The multiple choice vocabulary test WST is a measure of verbal intelligence. In vocabulary tests memory contents are checked which remain available due to their early acquisition and their frequent activation for a long time. The vocabulary test (WST) contains norms to transform the test scores in IQ estimates. However, this test can only be carried out in workers whose mother tongue is German.

The German version of the Standard Progressive Matrices Test (SPM) is a verbal-free test of deductive and inductive thinking and general intelligence.

The block design test (BDT) is a verbal-free test of visuospatial thinking and reasoning.

The trail making test (TMT) comprises intelligence and speed aspects. The tests measures the average time it takes to connect sequentially a series of numbers, randomly distributed on a sheet of paper. According to the authors the outcomes are positively related to IQ-measures.

**Psychomotor performance** Four tests cover the domain of psychomotor performance ('steadiness', 'line tracing', 'aiming', 'tapping') and are parts of a computerized test

**Table 4** Schedule of neurobehavioural methods during the three examinations of the longitudinal AI study

	Examination		
	E1	E2	E3
<i>Questionnaire</i>			
Symptoms Q16	x	x	x
<i>Conventional tests</i>			
Stand. progressive matrices	x		
Verbal intelligence (WST)		x	
HAWIE			
<i>1. Recall of digits</i>			
<i>a. forward</i>	x	x	x
<i>b. backward</i>	x	x	x
<i>2. Block design</i>	x	x	x
<i>Computer tests</i>			
Psychomotor performance			
<i>1. Steadiness</i>	x	x	x
<i>2. Line tracing</i>	x	x	x
<i>3. Aiming</i>	x	x	x
<i>4. Tapping</i>	x	x	x
Simple reaction time	x	x	x
EURO-NES			
<i>1. Symbol-digit substitution</i>		x	x
<i>2. Digit span</i>		x	x
<i>3. Switching attention</i>			
<i>a. Block</i>		x	x
<i>b. Arrow</i>		x	x
<i>c. Mixed</i>		x	x

battery (Motorische Leistungsserie, MLS, 1994). The tests were performed with each hand separately, the dominant and non-dominant one. The different computer-measures include accuracy and speed parameter. The steadiness-test, and to a lower degree the line tracing tests, are indirect measures of tremor, a variable that seemed to be related to AL exposure according to the study of Bast-Pettersen et al. (2000).

**Short-term memory, working memory** Short-term memory or working memory were measured with the well-known tests 'recall of digits' forward and backward of the Wechsler Adult Intelligence Scale (WAIS, German: Hamburg-Wechsler-Intelligenztest für Erwachsene, HAWIE, 1991) but also with the computer test 'digit span' (DS) of the European Neurobehavioural Evaluation System (EURO-NES, see Gilioli 1993). The latter is an intensive adaptive test of short-term memory with steps of increasing or decreasing difficulty approximating the individual reproduction ability. Additionally the symbol-digit substitution test (SDS) of the EURO-NES was subsumed in this

category as working memory plays an important role in the computerized test version.

**Attention** Attentional processes are partially involved in simple reaction time (SRT) tasks but are of dominant importance in the ‘switching attention’ task of the EURO-NES. The simple reaction time test (SRT, Wiener Reaktionstest, 1997) includes in this form not only the reaction time parameter but also a movement time of the index finger between resting and target key. The ‘switching attention’ task was introduced in the EURO-NES by Letz and is comparable with the test of the same name in the Neurobehavioural Evaluation System (NES, Letz et al. 1996). The reaction time and error measures correspond to the increasing difficulty of the three subtests. In the simple forms attention has to be directed to a few constant features (side, arrow) of the stimuli. In the complex form, switching between stimulus response rules is needed (mixed). Tests like ‘switching attention’ seem suitable to investigate higher level processes, so-called ‘executive’ functions.

#### *Models of ‘crystalline’, ‘fluid’, and ‘premorbid’ intelligence*

Generally, in neurobehavioural studies there is a problem to model and interpret performance differences between exposed and non-exposed workers. Group differences might reflect both exposure effects and a priori existing intellectual differences.

General opinion is that verbal intelligence is hardly affected by exposure influences (‘crystalline’ intelligence). Therefore, present performance levels in vocabulary tests are considered as a mean for assessing former intellectual levels, which, for example, existed before a possible neurotoxic damage (‘premorbid’ intelligence).

Tests which primarily measure attention, perception speed, short-term memory, working memory are usually categorized as measures of ‘fluid’ intelligence. Together with psychomotor tests they are seen as sensitive methods for the recording of neurofunctional impairments. They were used in neurobehavioural examinations of solvent effects (Bolla 1991), in examinations of metabolism diseases, brain lesions, and dementia (Lezak 1995), and as an indicator of age-related changes (Letz et al. 2003; Kiesswetter et al. 2000).

It was planned to test exploratively whether so-called non-verbal or culture-fair tests, like the block design test, might be an appropriate mean to control for intellectual differences between groups. It was intended to test its dependency from exposure and in the case of no exposure dependency to use it in alternative covariance models instead of education or verbal intelligence for the control of a priori differences.

## Statistical methods

### *Exposure data*

The repeated measurements of external and internal AI exposures allowed the investigation of several characteristics of exposure measures. Correlation and regression methods were used to investigate the interdependency, reliability, and stability (2-, 4-year intervals) of external and internal exposure measures. Additionally, with ANOVA for repeated measurements it was tested, whether AI biomonitoring data are also sensitive to acute exposure changes (pre-, post-shift differences).

### *Neurobehavioural data*

The neurobehavioural data of the participants who had taken part in all examinations were analysed with multivariate models of covariance (MANCOVA) for repeated measurements (Tables 9, 10, 11, 12). Depending on theoretical considerations, different components of a test (e.g. speed and errors), tests within one domain, as well as motor performance of both hands, were analysed simultaneously in multivariate analysis models. Psychomotor data of the right and left hand were reorganized to data of dominant and non-dominant hand.

The statistical MANCOVA models which were used routinely included a grouping factor (welders vs. control), a repetition factor (examination, three levels), an interaction factor ‘examination  $\times$  exposure’ and as covariates age, education index, and CDT (Carbohydrate Deficient Transferrin, alcohol marker). The interaction term is of highest importance, because it is used to test the question whether neurobehavioural performances of exposed and controls changed differently across the examinations. The variances explained by the respective model components ( $\eta^2$ ) are given in the Tables 10 and 12 in addition to the error probability  $P$ .

In a further statistical approach, the association between exposure and neurobehavioural measures was tested by stepwise regression.

Finally, MANCOVA models were used to examine moderating effects of different potential indicators of ‘premorbid’ intelligence.

## Results

### Ambient and biological monitoring

The dust and biomonitoring data of the three examinations are represented in tabular form using the maximum samples at a certain time (pre-/post-shift) and examination

**Table 5** Exposure data of the group of Al welders, total dust and pre-/post-shift biomonitoring (Al in urine, in urine per creatinine, Al in plasma) for examination 1–3

		Median	Min	Max	Mean	SD	SE	<i>n</i>
Total dust—air (mg/m <sup>3</sup> )	E1	5.7	0.0	31.5	7.3	6.8	1.1	36
	E2	4.5	1.3	15.6	5.8	4.1	0.9	21
	E3	6.8	1.9	29.7	8.6	7.7	1.8	19
Al-U-pre-shift (µg/l)	E1	136.6	24.8	540.4	185.9	144.2	25.1	33
	E2	163.9	2.9	656.3	184.5	146.7	25.9	32
	E3	97.7	9.9	801.4	158.9	174.2	31.3	31
Al-U-post-shift (µg/l)	E1	130.0	22.8	810.0	210.0	217.3	39.0	31
	E2	145.5	5.0	656.3	191.5	154.6	30.9	25
	E3	93.7	26.8	568.6	155.7	147.4	31.4	22
Al-U-pre-shift (µg/g creat.)	E1	92.1	17.9	292.2	110.7	75.9	13.2	33
	E2	90.1	7.6	420.6	120.0	97.0	17.1	32
	E3	58.8	10.7	276.4	81.5	74.4	13.4	31
Al-U-post-shift (µg/g creat.)	E1	97.0	17.9	399.0	135.5	103.8	18.6	31
	E2	143.9	8.9	431.8	153.0	124.2	24.8	25
	E3	64.5	23.9	560.0	113.5	126.1	26.9	22
Al-P-pre-shift (µg/l)	E1	9.6	4.1	31.0	12.5	7.5	1.3	32
	E2	11.1	3.9	40.3	13.8	9.3	1.6	32
	E3	10.8	4.0	39.3	13.9	8.8	1.7	28
Al-P-post-shift (µg/l)	E1	11.6	5.0	39.6	14.8	9.7	1.7	31
	E2	14.3	3.8	51.0	18.6	13.3	2.8	22
	E3	13.2	6.6	44.3	17.8	11.6	2.6	20

Maximum samples, including subjects without complete sets of repeated measurements

**Table 6** Pre-shift biomonitoring of the control group (Al in urine, in urine per creatinine, and Al in plasma) for examination 1–3

		Median	Min	Max	Mean	SD	SE	<i>n</i>
Al-U-pre-shift (µg/l)	E1	5.8	1.9	148.3	14.5	29.7	5.7	27
	E2	6.0	1.6	88.8	11.5	18.0	3.7	24
	E3	8.3	4.4	41.2	11.4	9.3	2.3	17
Al-U-pre-shift (µg/g creat.)	E1	4.0	1.6	78.9	9.0	15.6	3.0	27
	E2	4.5	1.6	86.2	9.6	17.0	3.5	24
	E3	8.5	1.8	37.5	10.0	8.4	2.0	17
Al-P-pre-shift (µg/l)	E1	3.5	1.0	8.2	4.1	2.3	0.4	27
	E2	2.8	1.3	5.9	2.8	1.1	0.2	23
	E3	4.5	3.3	5.9	4.5	0.8	0.2	17

Maximum samples, including subjects without complete sets of repeated measurements

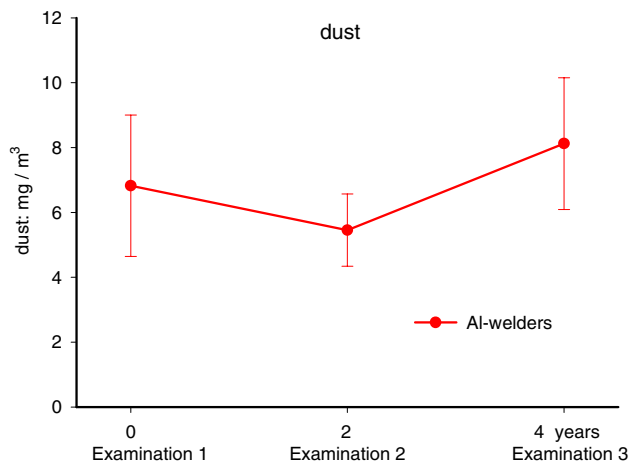
(Table 5). These results reflect the Al loads in this industrial sector and are shown for comparison with other studies.

Additionally, the exposure values restricted to the sample with a complete set of repeated neurobehavioural and biomonitoring measurements will be represented graphically and analysed statistically. The latter approach permits better examination of regular changes (stability, sensitivity) in the outlined data and shows moreover that the biomonitoring values of the neuropsychologically

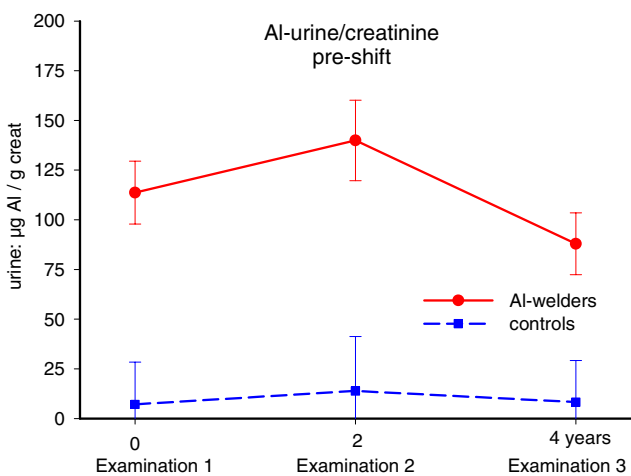
examined Al welders are higher on average than the exposure values of the maximum samples biomonitoring per examination.

#### *Exposure data of the maximum samples biomonitoring per examination and time*

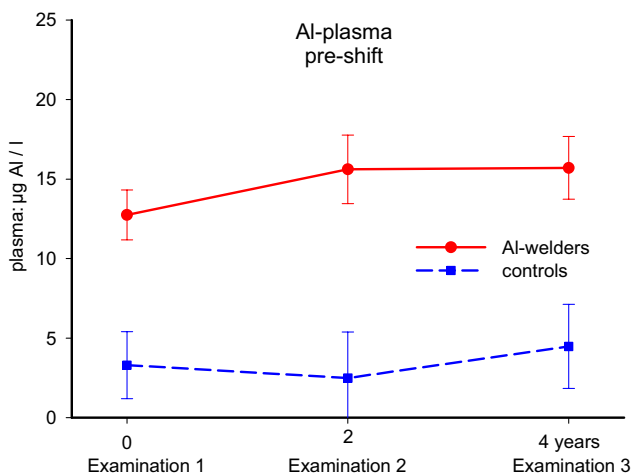
The medians of total dust concentration measured during the shift near the breathing zone of welders showed modest changes between 4.5 and 6.8 mg/m<sup>3</sup> in the course of the



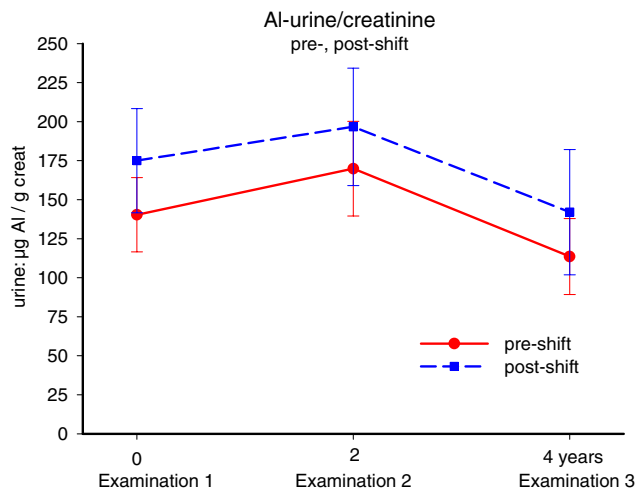
**Fig. 1** Total dust in the air during Al welding (Mean, SE)



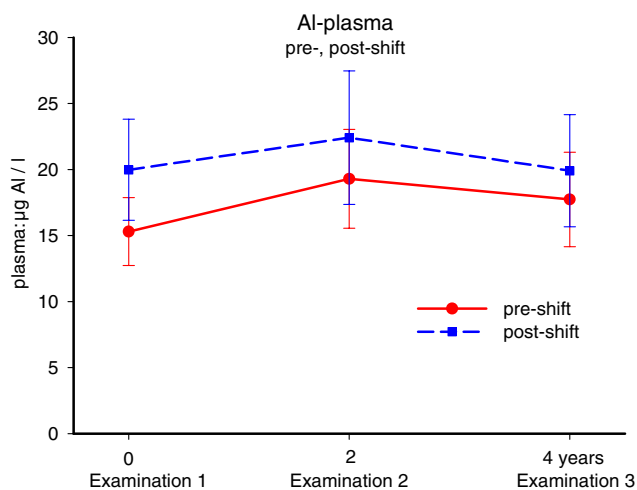
**Fig. 2** Biomonitoring, Al welders and controls: pre-shift aluminium concentrations in urine referred to creatinine concentrations (Mean, SE)



**Fig. 3** Biomonitoring, Al welders and controls: Pre-shift aluminium concentrations in plasma (Mean, SE)



**Fig. 4** Biomonitoring, Al welders: pre-shift and post-shift aluminium concentrations in urine referred to creatinine concentrations (Mean, SE)



**Fig. 5** Biomonitoring, Al welders: pre-shift and post-shift aluminium concentrations in plasma (Mean, SE)

longitudinal study (Table 5). The medians of Al in urine vary between 98 and 164  $\mu\text{g Al/l}$  urine, pre-shift, and between 94 and 145  $\mu\text{g Al/l}$  urine, post-shift. It seems paradoxical that the pre-shift concentrations of this parameter are higher than the post-shift, but this is due to different numbers of pre- and post-shift specimens and contrasts to the repeated measurement results based on the identical workers at each sampling (next chapter).

The medians of Al concentrations in urine related to creatinine vary across the study between 59 and 92  $\mu\text{g}$ , pre-shift, and between 64 and 144  $\mu\text{g Al/g creatinine}$ , post-shift (Table 5).

The numeric values of the unadjusted Al-urine concentrations are by about 1.5 times higher than the creatinine-related concentrations.

The medians of Al concentrations in the plasma varied pre-shift between 9.6 and 11.1 µg Al/l and post-shift between 11.6 and 14.3 µg Al/l, showing higher post-shift values comparable to creatinine-related Al concentrations in urine.

Descriptive biomonitoring data of the control group exist exclusively for pre-shift (Table 6). The medians of Al concentrations in urine with and without correction by creatinine concentrations are below 9 µg Al. The Al concentrations in the plasma are below 5 µg of Al/l. While the urine values of welders and controls differed approximately around the factor 20 the plasma values differed only by the factor 3.

*Exposure data of the subsample analysed neurobehaviourally: course and statistical analysis*

Figure 1 shows the course of total dust load across examinations for the group of Al welders with neuropsychological measures.

The u-shaped course has a minimum of 5.5 mg/m<sup>3</sup> total dust at the second examination and a maximum of 8.1 mg/m<sup>3</sup> at the last examination. The differences are statistically not significant ( $P = 0.44$ ).

In contrast to the dust trend the biomonitoring data of welders show primarily an inverse u-shaped trend (Figs. 2, 3).

The creatinine-related Al concentrations reveal a maximum (140 µg Al/g creatinine) at the second examination and a strong decrease of 52 µg Al/g creatinine from the second to the third examination. The Al-plasma values rise from the first to the second examination and remain on a level (15.7 µg Al/l) that is by 2 µg above the average values of the maximum sample (13.8, 13.9 µg Al/l, Table 5). While the creatinine-related Al concentrations reveal a significant difference over time ( $P < 0.001$ ) the changes of the Al-plasma concentrations do not reach level of significance ( $P = 0.22$ ).

The comparison between pre-shift and post-shift refer to neurobehaviourally examined welders with data at all examinations. Figures 4 and 5 show systematic relationships across examinations. The post-shift Al concentrations are higher than pre-shift concentrations by 30 µg/g creatinine in urine, and by 3.5 µg/l in plasma.

An ANOVA model for repeated measurements was used with the factors, examination (three levels) and shift time (two levels) to test the sensitivity of the biomonitoring to acute shift Al exposure. The creatinine-related Al concentrations in urine showed significant differences for examination ( $P = 0.003$ ) and shift time ( $P = 0.038$ ) but no significant interaction examination × shift ( $P = 0.88$ ). For Al-plasma data only a borderline significance for shift time was found ( $P = 0.062$ ).

The pre-shift biomonitoring data of the welder and control samples used in the neurobehavioural analyses

**Table 7** Stability, retest reliability of exposure measures

Exposure variable	Sampling	Pearson correlation $r_{tt}$		
		E1–E2 2 years	E2–E3 2 years	E1–E3 4 years
Al-urine	Pre-shift	0.85**	0.62**	0.67*
	Post-shift	0.71**	0.74**	0.66*
Al-urine/crea	Pre-shift	0.94**	0.94**	0.93**
	Post-shift	0.95**	0.73**	0.76**
Al-plasma	Pre-shift	0.84**	0.66**	0.59*
	Post-shift	0.88**	0.80**	0.71**
Dust	Shift	0.70**	0.44	-0.02

Pearson correlation between repeated measurements ( $n = 20$  Al welders)

\*  $P \leq 0.05$

\*\*  $P \leq 0.01$

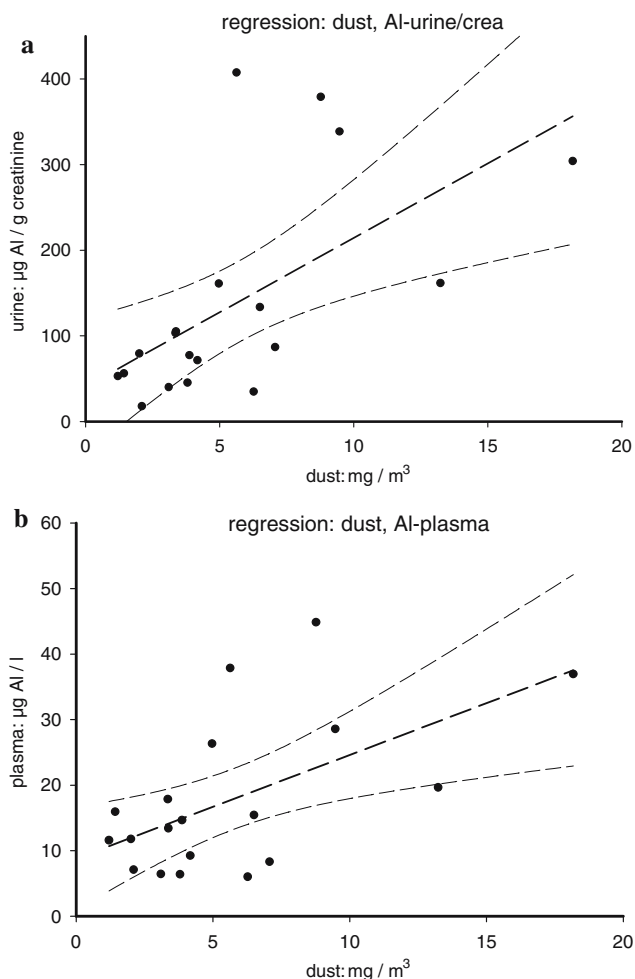
**Table 8** Intercorrelations between measures of average exposures (means of examination 1, 2, 3)

Exposure variable	Sampling	Pearson correlation, $r$					
		Al-urine pre-shift	Al-urine post-shift	Al-u/cre pre-shift	Al-u/cre post-shift	Al-plasma pre-shift	Al-plasma post-shift
Al-urine	Pre-shift						
	Post-shift		0.95**				
Al-urine/crea	Pre-shift	0.96**	0.96**				
	Post-shift	0.88**	0.92**	0.96**			
Al-plasma	Pre-shift	0.83**	0.88**	0.89**	0.94**		
	Post-shift	0.81**	0.90**	0.88**	0.95**	0.95**	
Dust	Shift	0.52*	0.58**	0.57*	0.60**	0.49*	0.58*

$n = 20$  Al welders

\*  $P \leq 0.05$

\*\*  $P \leq 0.01$



**Fig. 6** **a** Linear regression (95% CI): total dust, aluminium in urine per creatinine. **b** Linear regression (95% CI): total dust, aluminium in plasma

differed significantly concerning creatinine-adjusted Al-urine concentrations ( $P \leq 0.001$ ) as well as to Al-plasma concentrations ( $P \leq 0.001$ ).

#### Intercorrelation and temporal stability of biomonitoring data

Table 7 shows the temporal stability of the biomonitoring data for the cohort of neurobehaviourally investigated welders. The correlations of the individual measurements are shown for the two 2-year intervals (E1–E2, E2–E3) and for the 4-year intervals (E1–E3). The biomonitoring data reveal a significant temporal stability of individual data. The dust data show a moderate stability for 2-year intervals but no stability over 4 years.

The biomonitoring parameters of welders averaged across the three examinations reveal a high intercorrelation (Table 8) as well as a significant moderate correlation to the dust measures.

The linear regressions of the averaged dust values on the creatinine related post-shift Al concentrations and on the post-shift Al plasma values are shown in Fig. 6a, b. Generally, the linear regression models with post-shift data explain more variance than models with pre-shift data. This is valid both for the creatinine-related post-shift Al concentrations ( $R^2$ : 0.37 vs. 0.32) and for the Al plasma values ( $R^2$ : 0.33 vs. 0.24). From the regression analyses it can be deduced that the relation between dust- and creatinine-related Al concentrations in urine is a little more systematic than between dust and Al plasma values.

#### Neurobehavioural results

Descriptive data for the three neurobehavioural examinations and related MANCOVA analyses are shown in Tables 9, 10, 11, and 12. Most important term is the interaction between examination and exposure (exam  $\times$  expo), indicating whether exposed and non-exposed show diverging changes of the neurobehavioural parameters in the course of 4 years.

Additionally, in explorative models it was examined to what degree the general structure of neurobehavioural data is determined by exposure or (and) individual cognitive abilities.

#### Questionnaire Q16

Only few symptoms were indicated by welders and control persons in the symptom questionnaire Q16 (Table 9). In the examination course a parallel reduction of the average symptom score was seen: from less than 3 to less than 2 in welders and from less than 2 to less than 1 in the control group (Fig. 7). Neither the group difference nor the interaction term (exam  $\times$  expo) was significant under consideration of covariates (Table 10:  $P = 0.08$ ,  $P = 0.45$ , respectively).

#### Cognitive abilities

**Verbal IQ test (WST)** The vocabulary test (WST) was carried out at the time of the second examination and served for assessing premorbid intelligence in workers with German mother tongue. Related to the cohort with repeated neurobehavioural measurements 75% ( $n = 15$ ) of the exposed and 58% ( $n = 7$ ) of the controls could be examined (Table 11). The IQ values of both groups, estimated from the WST, indicate average intelligence. The mean score of welders (95.3) is lower than the score of controls (98.7). However, considering a potential influence of age no significant difference could be stated (Table 12,  $P = 0.263$ ). CDT had a significant influence in this model ( $P = 0.027$ ).

**Table 9** Means and standard deviations of motor performance (dominant and non-dominant hand), simple reaction, and symptoms (Q16) for AI exposed and non-exposed workers, examination 1–3 (repeated measurements)

n		Exposed E1		Exposed E2		Exposed E3		Non-exposed E1		Non-exposed E2		Non-exposed E3		
		20		20		20		12		12		12		
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
Motor performance														
Steadiness														
	Errors (n)	dom.hand	9.9	21.8	2.2	3.4	5.6	9.3	6.8	6.8	1.4	1.7	3.8	3.7
	Error time (ms)	dom hand	0.6	1.4	0.1	0.2	0.3	0.6	1.5	4.1	0.1	0.3	0.2	0.3
Line tracing														
	Errors (n)	dom. hand	22.8	7.5	22.4	7.7	22.3	6.3	21.9	7.5	19.1	6.7	20.8	6.7
	Error time (ms)	dom hand	2.3	1.1	2.2	0.9	2.1	0.8	1.9	0.9	1.6	0.7	1.9	0.6
Aiming														
	Hits (n)	dom. hand	19.9	0.7	19.7	0.6	19.9	0.4	19.8	0.4	19.9	0.3	19.9	0.5
	Total time (ms)	dom hand	6.6	1.0	6.9	1.0	6.3	1.6	6.8	0.6	7.0	0.7	6.9	0.8
Tapping														
	Hits (n)	dom. hand	201.3	22.9	195.3	25.1	194.5	24.3	207.6	15.4	201.2	12.4	199.0	17.9
Steadiness														
	Errors (n)	nd. hand	12.2	19.1	2.7	2.9	8.2	8.5	9.8	11.1	2.6	3.0	5.9	5.5
	Error time (ms)	nd. hand	0.8	1.3	0.2	0.2	0.6	0.7	0.9	1.6	0.1	0.2	0.5	0.7
Line tracing														
	Errors (n)	nd. hand	28.3	6.8	29.4	9.0	26.2	7.3	24.9	9.5	23.7	6.2	21.8	7.2
	Error time (ms)	nd. hand	2.8	1.0	2.9	0.8	3.0	1.0	2.6	1.0	2.6	1.0	2.6	1.0
Aiming														
	Hits (n)	nd. hand	19.6	1.5	19.2	4.7	19.2	4.7	20.1	0.8	20.0	0.0	18.5	5.9
	Total time (ms)	nd. hand	7.6	1.4	8.0	1.5	7.7	1.5	7.5	0.9	7.6	0.8	7.4	0.9
Tapping														
	Hits (n)	nd. hand	182.8	21.0	176.6	19.2	178.9	21.1	183.3	23.3	186.9	18.1	185.3	19.2
Simple reaction														
	Reaction t (ms)		270.0	39.2	267.6	40.7	283.0	58.9	263.8	38.4	279.0	43.7	281.6	44.6
	Movement t (ms)		171.2	33.4	161.3	35.6	156.1	34.4	158.9	36.6	156.4	38.5	168.1	44.3
Q16 sum (n)														
			2.7	2.2	2.0	1.8	1.8	1.9	1.9	1.8	1.6	1.6	0.9	0.8

**Standard progressive matrices test (SPM)** The standard progressive matrices test (SPM), a verbal-free test of logic thinking, was carried out only at examination 1. The welders showed insignificantly weaker performances than controls (Table 11). No significant group difference could be established in the covariance analysis (Table 12,  $P = 0.299$ ). The covariate age showed a significant influence ( $P < 0.001$ ).

**Block design** In the block design test (HAWIE), a measure of visuospatial thinking and general intelligence, the group of the welders revealed significantly lower scores than controls (Table 12,  $P = 0.033$ ). The covariate age was significantly related to the outcome variable ( $P = 0.026$ ). However, the time courses of group performances across examinations showed no critical diverging trends but a

tendency to convergence (Fig. 8). The interaction between examination and group factor was insignificant (Table 12,  $P = 0.51$ ).

**Trail making test** The trail making performance is also used to estimate general intelligence. However, it is not comparable to the aforementioned methods, because the speed factor is of high importance. Therefore, processing and faults were analysed together multivariately. There is an insignificant tendency to poorer performance in welders (Table 12,  $P = 0.09$ ) as well as a borderline influence of age ( $P = 0.075$ ). The interaction factor ‘examination  $\times$  group’ is highly significant ( $P \leq 0.001$ ). However, in contrast to diverging trends, possibly determined by exposure, the trends are converging (Fig. 9).

**Table 10** Results of multivariate and univariate analyses of covariance for repeated measures of motor performance, simple reaction and symptoms (Q16)

	Exposure			Age			Education			CDT			Examination			Examination × exposure			
	F	P	$\eta^2$	F	P	$\eta^2$	F	P	$\eta^2$	F	P	$\eta^2$	F	P	$\eta^2$	F	P	$\eta^2$	
<b>Steadiness</b>																			
Errors (n)	dom. hand	0.789	0.544	0.116	0.573	0.685	0.087	0.275	0.891	0.044	0.712	0.592	0.106	1.968	0.105	0.441	0.627	0.746	0.201
	nd. hand																		
Error time (ms)	dom hand																		
	nd. hand																		
<b>Line tracing</b>																			
Errors (n)	dom. hand	2.563	0.066	0.308	5.938	0.002	0.508	1.270	0.310	0.181	0.309	0.869	0.051	0.921	0.521	0.279	0.593	0.772	0.200
	nd. hand																		
Error time (ms)	dom hand																		
	nd. hand																		
<b>Aiming</b>																			
Hits (n)	dom. hand	1.187	0.321	0.084	1.919	0.167	0.129	3.616	0.041	0.218	1.010	0.378	0.072	2.624	0.060	0.304	0.594	0.671	0.090
	nd. hand																		
Total time (ms)	dom hand																		
	nd. hand																		
<b>Tapping</b>																			
Hits (n)	dom. hand	0.171	0.843	0.013	1.807	0.184	0.122	1.348	0.277	0.094	0.092	0.912	0.007	1.985	0.129	0.249	0.952	0.452	0.137
	nd. hand																		
<b>Simple reaction</b>																			
Reaction t (ms)	dom. hand	0.112	0.894	0.009	1.710	0.201	0.116	1.757	0.193	0.119	0.625	0.543	0.046	0.797	0.539	0.117	2.142	0.107	0.263
	dom. hand																		
Movement t (ms)	dom. hand																		
	dom. hand																		
Q16		3.288	0.080	0.099	3.118	0.088	0.094	2.616	0.116	0.080	0.000	0.995	0.000	0.178	0.786	0.006	0.746	0.450	0.024

Model: exposure (grouping factor), covariates [age, education, and carbohydrate-deficient transferrin (CDT)], repetition factor (examination E1–3)

F F-value, P P-value,  $\eta^2$  portion of explained variance



**Table 11** Means and standard deviations of cognitive performance tests for exposed and non-exposed workers, examination 1–3 (repeated measurements)

	Exposed E1		Exposed E2		Exposed E3		Non-exposed E1		Non-exposed E2		Non-exposed E3	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Cognitive performance												
HAWIE												
<i>n</i>	20		20		20		12		12		12	
Digit span												
Forward (score)	5.8	2.0	8.7	2.5	9.7	2.2	6.6	2.0	9.8	2.5	9.8	2.2
Backward (score)	5.3	1.7	8.1	1.8	8.8	2.1	6.0	1.5	8.0	2.0	9.1	2.7
Block design (score)	24.9	5.7	31.0	3.6	33.8	4.0	29.1	5.4	34.6	4.8	35.5	4.4
<i>n</i>	11		11		11		9		9		9	
Trail making												
Errors ( <i>n</i> )	2.8	3.5	2.2	7.2	0.0	0.0	2.1	1.6	0.0	0.0	0.0	0.0
Time (s)	86.7	15.7	85.3	14.9	83.3	17.0	81.7	17.5	81.9	15.7	79.8	18.0
<i>n</i>			15						7			
Verbal IQ			95.3	6.0					98.7	7.6		
<i>n</i>	20						12					
S. P. Matrices	44.2	7.9					47.4	7.5				
EURONES												
<i>n</i>			23		23				13		13	
SDS (ms)			3209.2	648.8	2914.4	557.4			2708.3	420.3	2911.7	719.5
Digit span												
Forward ( <i>n</i> )			5.3	0.8	5.4	0.9			6.0	1.2	6.0	1.2
Backward ( <i>n</i> )			4.1	1.8	3.6	2.0			5.0	2.3	4.5	3.1
Switching attention												
Reaction t (ms)	Side	411.7		73.7	400.3	79.6			374.2	46.6	371.5	64.6
	Arrow	805.5		266.3	765.7	242.7			627.4	103.0	613.3	150.3
	Mixed	766.3		229.8	723.8	244.0			653.2	247.9	604.2	210.6
Errors (n)	Side	0.5		0.7	0.4	0.7			0.3	0.9	0.3	0.5
	Arrow	2.5		4.1	1.5	2.9			0.5	0.8	0.3	0.5
	Mixed	6.3		7.8	4.9	7.2			1.7	1.6	2.4	2.0

### Psychomotor performance

The descriptive and statistical data for the four psychomotor performance tests are shown in Tables 9 and 10, respectively. Time for processing and faults of the dominant and non-dominant hand were analysed simultaneously in multivariate statistical approaches.

**Steadiness, line tracing, aiming, tapping** In the four subtests of the motor performance series no significant group differences were found (Table 10). Merely in the 'line tracing' test welders revealed a borderline tendency to higher error numbers and longer error times (multivariate,  $P = 0.066$ ). However, the age factor showed a strong influence on 'line tracing' performance ( $P = 0.002$ ). None of the psychomotor performance tests did show a significant interaction term 'examination  $\times$  exposure' (Table 10;

steadiness  $P = 0.746$ , line tracing  $P = 0.772$ , aiming  $P = 0.671$ , tapping  $P = 0.452$ ) indicating diverging courses of exposed and controls. Figure 10 shows the mostly parallel change in the number of 'line tracing' errors of both groups across the examinations.

### Digit span, working memory

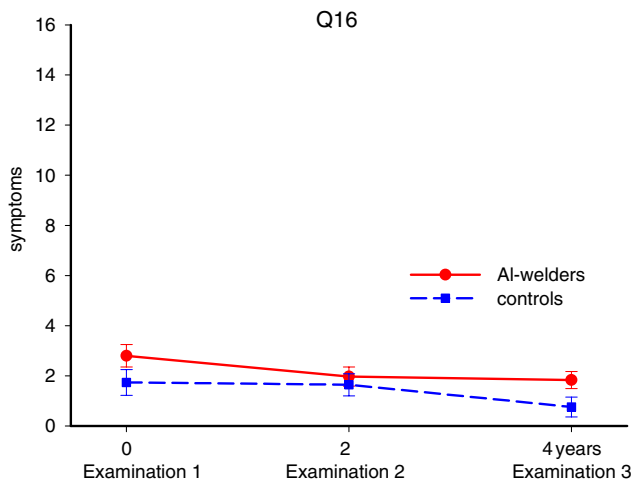
**Recall of digits (HAWIE)** With regard to the short-term memory performance in forward and backward reproduction of short number sequences, no significant difference between welders and controls could be stated (Tables 11, 12). Figure 11 illustrates the convergence of backward recall performance of both groups over the 4-year period. The interaction of forward and backward recall was analysed multivariately revealing no significance ( $P = 0.661$ ).

**Table 12** Results of multivariate and univariate analyses of covariance for repeated measurements of motor performance, simple reaction, and symptoms (Q16)

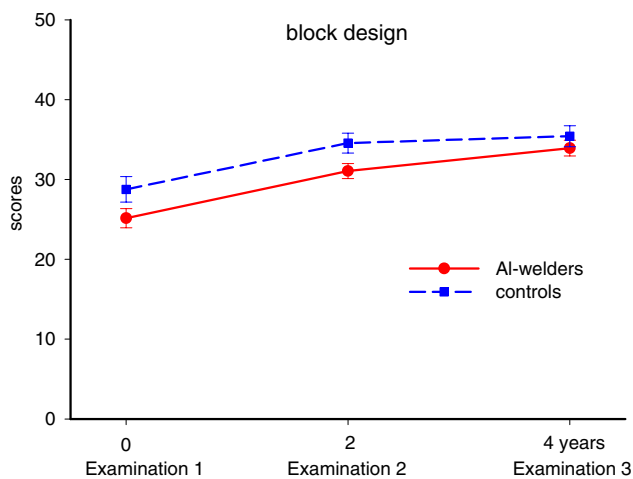
	Exposure			Age			Education			CDT			Examination			Examination*exposure			
	F	P	$\eta^2$	F	P	$\eta^2$	F	P	$\eta^2$	F	P	$\eta^2$	F	P	$\eta^2$	F	P	$\eta^2$	
HAWIE																			
Digit span																			
Forward (score)	0.790	0.465	0.057	0.985	0.387	0.070	0.314	0.733	0.024	1.484	0.245	0.102	0.736	0.577	0.109	0.607	0.661	0.092	
Backward (score)	5.065	0.033	0.168	5.619	0.026	0.184	0.051	0.822	0.002	0.391	0.538	0.015	0.200	0.789	0.008	0.636	0.513	0.025	
Block design (score)	2.827	0.093	0.288	3.144	0.075	0.310	1.605	0.236	0.186	0.120	0.888	0.017	16.770	0.000	0.848	9.520	0.001	0.760	
Trail making																			
Errors (n)	1.334	0.263	0.069	0.686	0.419	0.037	–	–	–	5.829	0.027	0.245	–	–	–	–	–	–	
Time (ms)	1.093	0.299	0.014	25.094	0.000	0.248	2.131	0.148	0.027	0.343	0.560	0.004	–	–	–	–	–	–	
Verbal IQ																			
(E2)	1.068	0.380	0.114	3.164	0.042	0.275	1.641	0.205	0.165	0.942	0.435	0.102	0.122	0.946	0.014	0.981	0.417	0.105	
S. P. Matrices																			
(E1)	0.694	0.657	0.138	0.688	0.661	0.137	0.534	0.777	0.110	0.183	0.979	0.041	0.680	0.667	0.136	1.004	0.444	0.188	
EURONES																			
(E2, E3)																			
SDS (ms)																			
Digit span																			
Forward (n)																			
Backward (n)																			
Switching attention																			
Reaction t(ms)																			
Side																			
Arrows																			
Mixed																			
Side																			
Arrows																			
Mixed																			

Model: exposure (grouping factor), covariates [age, education, and carbohydrate-deficient transferrin (CDT)], repetition factor (examination, E)

F F-value, P = P-value,  $\eta^2$  portion of explained variance



**Fig. 7** Number of symptoms (Q16) of AI welder and controls adjusted for education, age, and alcohol consume (CDT) (Mean, SE)



**Fig. 8** Block design scores (HAWIE) of AI welder and controls adjusted for education, age, and alcohol consume (CDT) (Mean, SE)

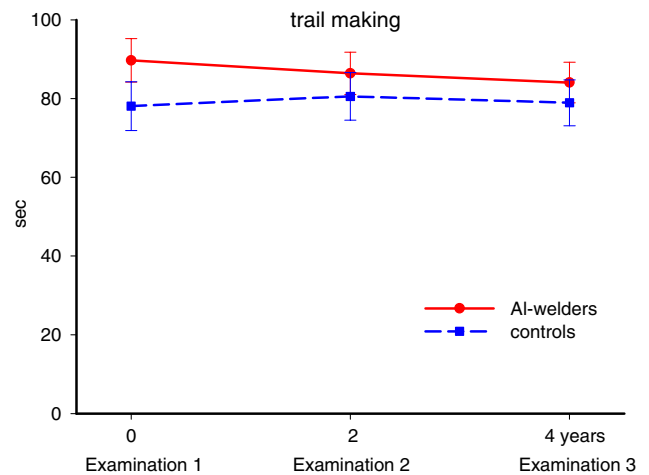
*Digit span, symbol-digit substitution (EURO-NES)* The joint multivariate covariance analysis of the EURO-NES computer tests ‘digit span’ and ‘symbol-digit substitution’ did neither show a significant difference between welders and controls (Table 11; 12,  $P = 0.38$ ) nor a significant interaction ( $P = 0.417$ ) regarding the temporal trends.

The performances in digit span forward showed a parallel change in both groups (Fig. 12).

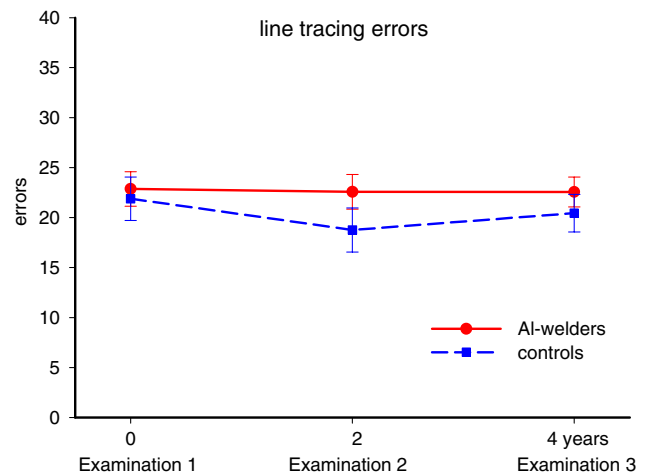
The performances in symbol-digit substitution (Fig. 13) show converging trends.

#### Attention

*Simple reaction time* The multivariate analysis of covariance in the simple reaction time task included reaction time (decision time) and motor movement time in one model (Figs. 14, 15). Neither a significant group difference



**Fig. 9** Trail making performance of AI welder and controls adjusted for education, age, and alcohol consume (CDT) (Mean, SE)



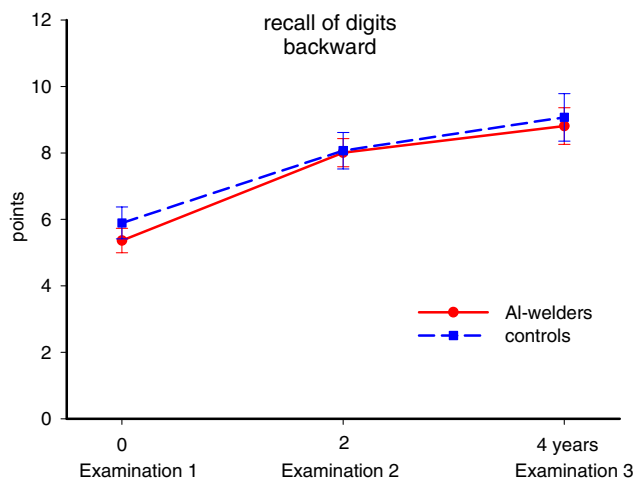
**Fig. 10** Line tracing errors (dominant hand) of AI welder and controls adjusted for education, age, and alcohol consume (CDT) (Mean, SE)

(Table 10,  $P = 0.894$ ) nor a significant interaction ‘examination  $\times$  exposure’ was found (Table 10,  $P = 0.107$ ).

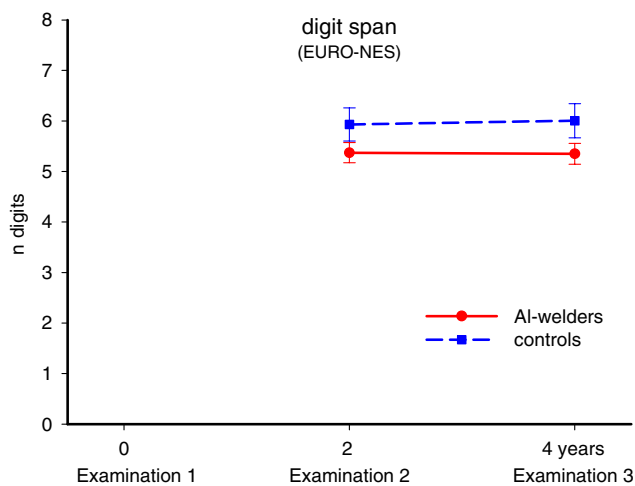
*Switching attention* The common, multivariate analyses of reaction times and errors in the three subtests of the ‘switching attention’ test (side, arrow, mixed) did not indicate group differences or interaction effects (Tables 11, 12). The Figures of the simple switching attention task ‘side’ (Fig. 16) and the difficult switching attention task ‘mixed’ (Fig. 17) show parallel performance courses for both groups.

Explorative modelling: a priori differences and premorbid differences

The results of the study show only one significant difference, the group difference in the block design test

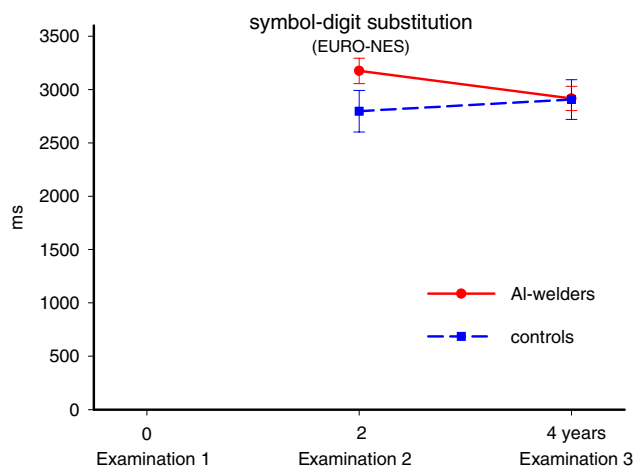


**Fig. 11** Recall of digits backward (HAWIE) of AI welder and controls adjusted for education, age, and alcohol consume (CDT) (Mean, SE)

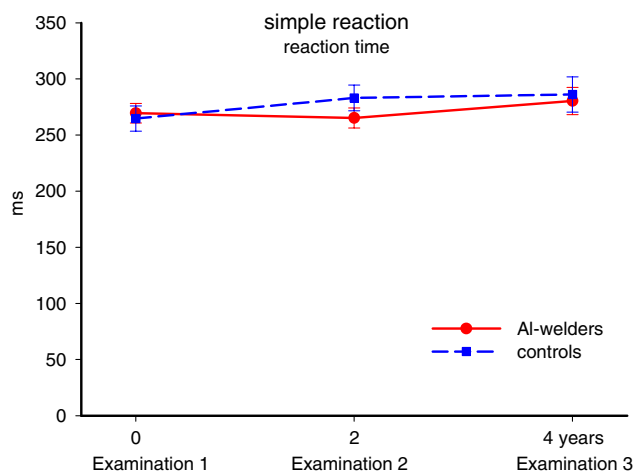


**Fig. 12** Digit span performance (EURO-NES) of AI welder and controls adjusted for education, age, and alcohol consume (CDT) (Mean, SE)

( $P = 0.026$ ). Trail making ( $P = 0.09$ ), and line tracing ( $P = 0.066$ ) only reveal tendencies to differences. Taking into account a necessary correction for multiple testing, these test results would be far from significance (see [Introduction](#)). The following analyses have the purpose to investigate the structural relationships between performance and exposure. Therefore, in different explorative models intelligence performance was examined either as dependent variable in relation to exposure measures, or as independent variable and possible moderator in relation to neurobehavioural performances.



**Fig. 13** Symbol-digit substitution performance (EURO-NES) of AI welder and controls adjusted for education, age, and alcohol consume (CDT) (Mean, SE)

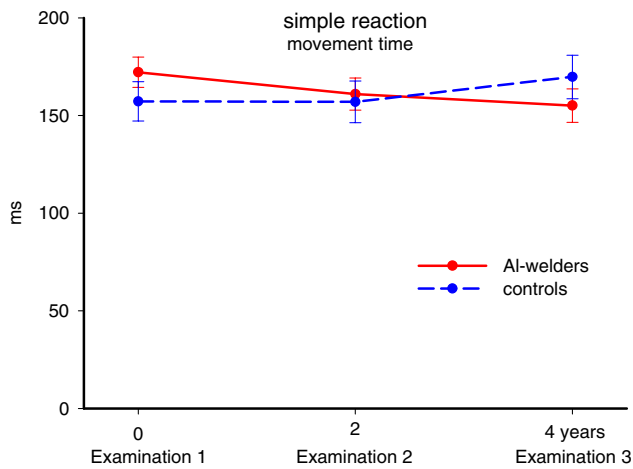


**Fig. 14** Simple reaction 'reaction time' of AI welder and controls adjusted for education, age, and alcohol consume (CDT) (Mean, SE)

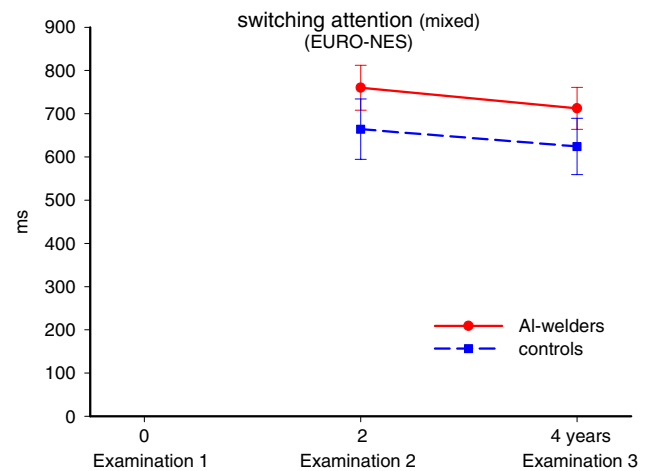
#### *Stepwise regression: exposure and general intelligence as main predictors of neurobehavioural performances*

In a first explorative model it was examined, whether exposure parameters (diverse biomonitoring measures, dust, and exposure duration) explain variance in 'block design' performance. However, stepwise regression did not show that 'block design' performance was dependent on exposure parameters ( $P > 0.05$ ).

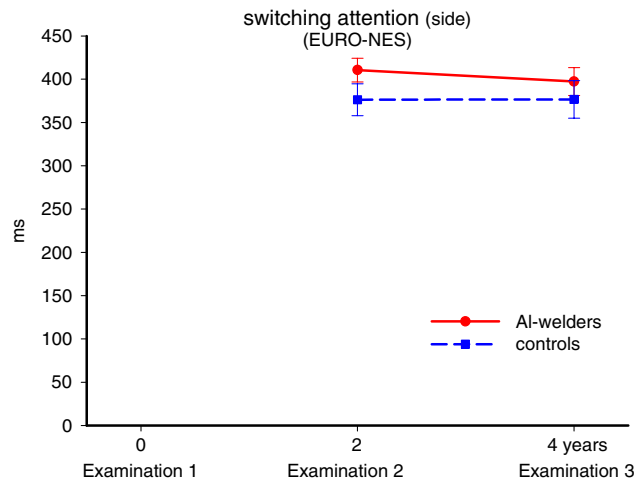
In a second step it was tested which variables show power to predict performances in the several other neurobehavioural tests. Block design, assumed as representing a general factor of intelligence, was integrated in these explorative models as an independent variable predicting performance. Besides block design, several exposure parameters and exposure duration, education, age, and



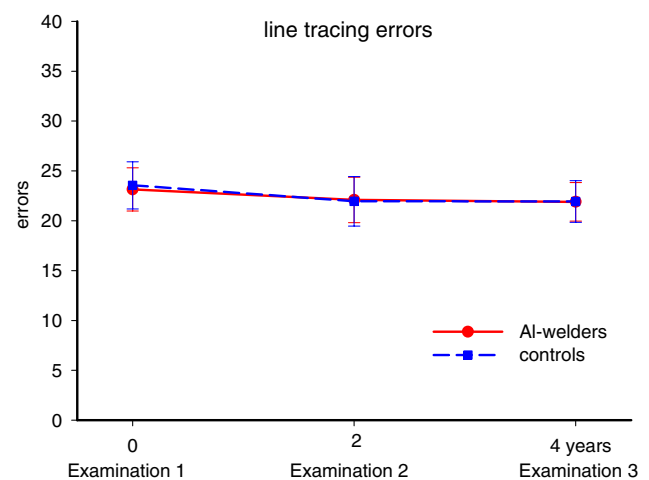
**Fig. 15** Simple reaction ‘movement time’ of AI welder and controls adjusted for education, age, and alcohol consume (CDT) (Mean, SE)



**Fig. 17** Switching attention performance ‘mixed’ (EURO-NES) of AI welder and controls adjusted for education, age, and alcohol consume (CDT) (Mean, SE)



**Fig. 16** Switching attention performance ‘side’ (EURO-NES) of AI welder and controls adjusted for education, age, and alcohol consume (CDT) (Mean, SE)



**Fig. 18** Line tracing errors (dominant hand) of AI welder and controls adjusted for verbal intelligence, age, and alcohol consume (CDT) (Mean, SE)

CDT were used as potential predictors. For each of the 32 neurobehavioural test parameters, whose number arose from ‘examination  $\times$  test  $\times$  parameter’, the significant predictors of neurobehavioural performance were determined by stepwise regression. The models were restricted to exposed workers.

The regressions revealed that education and exposure parameters were without any meaning in all 32 models. These results correspond to the results of multivariate analyses of covariance. An important variable, significantly associated with neurobehavioural test performance, was ‘block design’, which explained in 11 different regressions 13–48% of the variance of the respective dependent neurobehavioural parameters.

#### *Explorative models analysing concepts of pre-morbid intelligence: education versus block design versus verbal IQ*

Education, being looked at as a surrogate of ‘pre-morbid intelligence’, was an ineffective covariate both in the regression analyses and multivariate analyses of covariance. According to the results of this study, ‘block design’ is a stable variable with trait character in the examination course and is not associated to exposure. Block design therefore was used as covariate instead of school education in the two covariance analyses which had shown borderline differences between groups. For comparison purposes, additionally covariance models with ‘verbal IQ’ were checked using the reduced German-speaking samples.

The MANCOVA with age, CDT, and ‘block design’ as covariate, reduced the error probability for group differences in the ‘trail making’ test from  $P = 0.09$  to a new adjusted probability of 0.42. The same approach with ‘verbal IQ’ instead of block design led to an error probability of 0.21.

An analogous MANCOVA of ‘line tracing’, with ‘block design’ as covariate instead of ‘education’, reduced the error probability from  $P = 0.066$  to the new adjusted error probability of 0.20. The same approach with ‘verbal IQ’ instead of block design led to a similar increase of error probability to  $P = 0.22$ . Figure 18 shows the reduced group differences in ‘line tracing’ after adjustment with verbal IQ. The change in difference magnitude is remarkable compared to Fig. 10 showing ineffective intelligence adjustment by education.

## Discussion

The present study is characterized by comparatively pure Al exposure conditions, relatively high Al exposures, a long Al-exposure period of 15 years, repeated-measurement designs, and hypothesis conducted multivariate analyses examining simultaneously interrelated neuropsychological tests or test parameters. Additionally, potential essential covariates like education, intellectual ability level, age, and objective indicators of alcohol consumption (CDT) were taken into account in statistical and interpretative judgements.

Although the biomonitoring data were reliable and there was a high Al load compared with previous studies, no association between exposure and neurobehavioural performances could be affirmed. Fitting to the insignificant neurobehavioural performance results, the number of reported symptoms in the present study is small and clinically inconspicuous.

### Al-biomonitoring features

The nature of variation of Al-biomonitoring scores and the factors which may cause such variation determine the usability of Al-biomonitoring measures as an indicator of individual exposure in neurobehavioural studies. Relevant features which were tested in this context were the external validity of Al-biomonitoring parameters, the acute exposure dependent variability, and the long-term reliability and stability.

Individual Al-biomonitoring measures showed a moderate external validity because the Al-urine and Al-plasma scores correlated only modestly with dust exposures. Additionally, within the observation period of 4 years the trend of biomonitoring group means deviated from the

trend of dust means. Obviously, biomonitoring means revealed a stronger variation between examinations than dust means. Mainly four factors may be responsible for the discrepancies between external and internal exposure measures. Firstly, dust measures as external criterion were less reliable than biomonitoring measures. This was shown in the present study by comparative retest reliability statistics. Secondly, dust measures are of limited validity as measure of Al exposure, if there is a relevant background variation of Al proportions in the dust (see [Background of the study](#)). A third factor should not be disregarded. A general problem of laboratory data of longitudinal studies is that they are sensitive to minor changes in the analytical methods (Lehnert et al. 1999; Schaller et al. 2002). This factor might contribute to level differences of measured concentrations between examinations. A fourth factor that might weaken the association between measures of internal and external exposure is related to individual differences between workers, comprising behavioural aspects during exposure as well as biological aspects.

Studies of Al biomonitoring in the literature suggest that Al-urine and Al-plasma data are not suitable as markers of acute shift-dependent variations of Al exposure (Gitelman 1995, Gitelman et al. 1995, Rossbach et al. 2006). However, these studies found consistently, although statistically insignificant, that the biomonitoring scores increased depending on shift exposure. By contrast, the present study indicated significant pre-, post-shift differences despite a comparatively small sample. The significant results of the present study may be due to the following preconditions: the restriction to persons with repeated measurements, the use of statistical repeated measurements approaches (reduction of error variance), and the high exposures leading to relatively strong changes from pre-shift to post-shift.

In the literature there are only few approaches examining the long-term stability of Al biomonitoring data in welders. A study of Letzel et al. (1999) suggested long half-lives of aluminium in Al welders with a remarkable interindividual variation of 13–215 days, estimated on the base of Al concentrations in urine adjusted for creatinine. In the present study the biomonitoring data are characterized by high retest reliability and long-term stability across 4 years. The finding documents that biomonitoring measures are suitable to represent interindividual exposure differences in the long term. The relative contribution of biological and behavioural factors to the long-term stability is unclear. Slow biological processes of Al uptake and elimination are certainly involved (Letzel et al. 1999; Röllin et al. 2001; Sjögren et al. 1988). However, to what degree biological features of the workers (accumulation, elimination), individual behaviour when welding (breath protection), or job differences contribute to the stable interindividual differences cannot be answered.

Summing up, the Al-biomonitoring variables are reliable and stable measures showing high intercorrelations. In detail, the best distinctions between individuals, groups, or exposure conditions were found with Al-urine values related to creatinine. This corresponds to insights into other studies (e.g. Gitelman et al. 1995; Letzel et al. 1999; Rossbach et al. 2006). The results generally show that the Al-biomonitoring measures are indicators of a stable and continuing internal exposure characteristic.

#### Neurobehavioural results

The check of the biomonitoring data provided evidence for both significant internal Al load of the investigated welders and stable exposure differences between welders in the long term. On the one hand, this was the necessary precondition for a meaningful check of neurobehavioural group differences between welders and controls concerning the performance development during the observation period of 4 years. On the other hand, this was the necessary precondition to investigate neurobehavioural performance differences within welders in relation to internal exposure measures. Both models have to be discussed against the background of possible ‘a priori’ performance differences existing between groups or individuals independent of exposure.

#### Trends across 4 years, interaction between Al-exposure and time, different domains

Repeated measurement designs are predestined to test data for trend differences between groups across the examination period. Furthermore, it can be assumed, that the employed examination methods are sensitive to neuropsychological changes as they are appropriate means to portray ageing processes (Kiesswetter et al. 2000; Letz et al. 2003).

#### Symptom questionnaire

The symptom questionnaire Q16 revealed a comparable small number of symptoms in exposed and non-exposed subjects. No specificity of the symptoms in the group of the welders could be detected indicating exposure effects in so-called cognitive domains. This is in accordance with the study results of Bast-Pettersen et al. (2000) and Sjögren et al. (1996) who found low sum scores (<3) and no indication of obvious clinical signs. Only Sjögren et al. (1990) found that Al welders reported more often ‘concentration problems’ and ‘depression feelings’. However, these study results might be caused by high Al exposure in the previous decade (estimated as 250 µg Al/l urine) and potential neurotoxic co-exposures (Mn, Pb).

#### Domains

*Cognitive performance* The block design test and the trail making test did not show any trends towards performance diminution in the group of welders. On the contrary, and typical for other performance domains, exposed and controls showed parallel changes or even convergence of neurobehavioural performances. However, the block design test revealed significant level differences between exposed and controls not comparable to results of other studies (Hänninen et al. 1994; Akila et al. 1999; Riihimäki et al. 2000; see below: discussion of a priori differences).

*Psychomotor performance* Sjögren et al. (1996) and Iregren et al. (2001) reported discrepant results concerning motor performance in the same sample of welders depending on statistical design, control of factor age, and correction of multiple comparisons (see Introduction). The study using the controlled model produced no exposure effect. Bast-Pettersen et al. (2000) reported the paradoxical result that welders with significantly ‘better’ motor performance compared to controls revealed associations between ‘steadiness’ performance and internal exposure. According to the evaluation of the authors, a random finding could not be excluded and, furthermore, no tremor or clinical impairment of motor performance could be observed. Therefore, the three studies provide no convincing evidence of motor effects of Al exposure. In the present study, fitting to these negative results, no significantly diverging performance trends across the examinations for Al-exposed and non-exposed workers were observed. This is also true for this test of the motor test battery, ‘line tracing’, where welders showed a slight tendency to generally worse performance (see discussion of a priori differences).

*Memory: digit span, working memory* The model of Hänninen et al. (1994) revealed an unfavourable relationship of urine Al concentrations to memory and learning parameters, however, at a ‘normal’ (average) test performance and without statistical control of age. The model of Akila et al. (1999), adjusted for age but not for premorbid intelligence, did show significant differences only between referents and the low Al-exposure group in memory for design, not conform to dose–response relationships, and between referents and the high Al-exposure group in the digit symbol test. Paradoxically, digit span showed a non-hypothesized positive correlation with Al in urine (Riihimäki et al. 2000). Insignificant group differences for digit span and symbol digit coding were found in the study of Iregren et al. (2001).

The present multivariate analyses of covariance of forward and backward reproduction (HAWIE), as well as the

joint multivariate analysis of covariance of the computer tests ‘digit span forward and backward’ and ‘symbol-digit substitution’ (EURO-NES), did not show a significant group difference regarding the temporal trends across examinations. Furthermore, this model did not show level differences in memory performance between groups.

**Attention** Akila et al. (1999) and Riihimäki et al. (2000) assumed that AI exposure might also affect complex attention and visuospatial information, additional to working memory processing.

However, neither Akila et al. (1999) and Riihimäki et al. (2000) nor Bast-Pettersen et al. (2000) and Iregren et al. (2001) used explicitly tests of complex attention processes. The simple reaction test (SRT), employed in these studies involves only simple attention processes, and did not result in any significant exposure effects.

Contrary to the introduced studies the present longitudinal study analysed complex attention performance using the ‘switching attention test’. The group comparisons revealed no significant deficits in the attention control functions of welders. Additionally, trend analyses of the simple and the complex form of attention tests of our study did not yield any indication that the attention performance of AI welders and controls developed in different directions over the examination period of 4 years.

#### Correlation between biomarkers of exposure and neurobehavioural performance

Although the biomonitoring variables showed stable interindividual differences, no significant correlations were detected between measures of internal AI exposure and the diverse neurobehavioural variables. This was true for the analyses of each of the three examinations separately and also for the multiple regression analyses of averaged parameters. The regression analyses confirmed the results of covariance analyses (MANCOVA) revealing no effects of exposure. These results do not fit to isolated results in literature which were found with incomplete multivariate modelling of corresponding dependent parameters and independent confounders (see [Introduction](#), Table 1). However, in the case of block design the present findings are in agreement with other welder studies. Block design was not related to AI exposure measures, neither in the present study nor in the studies of Hänninen et al. (1994), Akila et al. (1999), and Riihimäki et al. (2000).

#### ‘A priori’ and ‘premorbid’ differences

In the above-reviewed welder studies (Table 1) primarily the variable age is used to match groups of AI-exposed and

non-AI-exposed workers or to adjust performance data. In these studies the samples are generally seen as homogeneous concerning the intellectual level. Only few studies (Table 1) used an education index to control for intellectual imbalance between groups. However, it remains obscure whether the used surrogate measures of premorbid intelligence, like education indices, were able to clear up variance. Statistical references, showing the quality of covariates (premorbid intelligence, age, alcohol consumption) in specific covariance models, were not or incompletely published. In one welder study (Akila et al. 1999; Riihimäki et al. 2000) a verbal test, synonyms test, indicated significant group differences but this asymmetry was not used for structural analyses of related intellectual abilities or the adjustment of other performance results. Generally, in all welder studies the possibility that the non-random samples differed in their general intellectual ability or specific cognitive abilities ‘a priori’ and ‘premorbid’, not affected by exposure, was not sufficiently taken into consideration, neither in the analyses nor in the interpretation models.

If so-called ‘crystalline’ or ‘premorbid’ intelligence is controlled in neurobehavioural toxicological studies, it is done conventionally with tests of verbal intelligence. However, the conventional concept of ‘premorbid’ intelligence has the great practical weakness that verbal tests cannot be meaningfully used for comparisons of workers with different native language. Therefore, in the present study ‘reasoning tests’ were examined as possible indicators of a priori differences between groups in competition with verbal tests. Reasoning tests with visuospatial elements like ‘progressive matrices’ and ‘block design’ on the one hand fall into the category of ‘non-verbal’ and ‘culture-fair’ tests but on the other hand they are often judged as tests of ‘fluid’ intelligence implicating sensitivity to exposure. However, in contrast to the latter appraisal, it is numerously documented that reasoning performances are related to verbal performances, reflecting—like verbal performances—an underlying factor of ‘general’ intelligence (Vernon 1949; Guilford 1972) with temporal stability and fundamental hereditary influences (Burt 1958). The adequacy of a categorical model of ‘crystalline’ or ‘fluid’ intelligence instead of a model with sensitivity continuum between poles is not well studied in neurobehavioural toxicological research.

In the present study the use of the education index as covariate did not lead to any important variance explanation. In contrast, the explorative and comparative use of a ‘verbal-free’ reasoning test and a ‘verbal’ test as covariates led to a parallel and complete diminution of ‘borderline’ group difference indicated in previous unadjusted data analyses.



Since the only significant performance difference (block design; univariate analyses without correction for multiple testing) between groups in the present study existed already in the first examination, its meaning and cause cannot be clarified clearly. However, the critical consideration of the different results of the present study does not support the hypothesis of an AI-exposure effect.

Firstly, the group differences in block design scores indicate similar asymmetry and adjustment as the verbal test scores. Secondly, block design and other test performances showed neither a significant association with external nor with internal AI loads. However, it cannot be excluded that exposure indicators do not reflect the biologically effective dose at the target organ.

Thirdly, and this is the most meaningful argument, for at least a period of roughly 4 years of high AI exposure no evidence was found that the performance courses of exposed and non-exposed differed significantly. Fourthly, the latter findings allow a cautious conclusion according to the processes before the start of the longitudinal study. If a neurodegenerative process existed in the pre-phase leading to growing differences between groups, it is improbable that this was interrupted by chance in the 4-year observation phase of the longitudinal examination. Additionally, with no exception, all other performance variables show comparable results.

#### Conclusion, limits, and perspectives

Generally, in the neurotoxicological literature comparable longitudinal examinations and explorative analyses are missing, controlling for potential influence of intellectual performance prerequisites, which would permit to assess the stability and sensitivity of cognitive functions under simultaneous consideration of the age factor (Kiesswetter et al. 2000).

Unlike the conditions of cross sectional studies, longitudinal neurobehavioural studies and specific exposure conditions (pure, stable, representative) permit to interpret data structures and 'causal' relations to a certain degree.

The explorative modelling carried out in the present study, pointed to a slight difference in the basic intellectual efficiency of AI exposed and non-exposed workers, which affected hardly noticeably the performance in some cognitive parameters.

The trend analyses, under consideration of different covariance models, did not yield any indication that the neurobehavioural performances of AI welders and controls developed over the examination period of 4 years in different directions. These non-effect structures were confirmed in the domains of psychomotor behaviour, short-term and working memory, and attention by different measures in each case.

On average, for every welder the time span from the 10th to the 15th year of his working life was examined in the present study. This short period allows only cautious backward extrapolation and long-term forecasting. According to models of time series analysis, only trends near to the observation phase can be predicted, backward as well as forward. Therefore, although the neurobehavioural parameters did not correlate with exposure during the study period, neither early AI-exposure effects before this prediction phase, e.g. during the first years of exposure, nor late exposure effects, e.g. after 20 years of exposure (model of 'delayed exposure effects', Kiesswetter et al. 2000), can be fully excluded.

Of course, the results of the examination are subject to a certain fortuity. The small sample size, reducing the statistical power, and possible selection tendencies are limiting factors.

However, with regard to the statistical power the longitudinal approach can more than compensate the handicap of small sample size by using repeated measurement analyses. The advantages of repeated measurement models are controlling the variability due to differences between workers and separating it from exposure effects and error variance. Additionally, measurements on covariates were introduced to increase the precision of the study.

With regard to possible selection effects, there exists the possibility of a comparison with intermediate results published after the second examination based on the hardly decimated starting sample (Buchta et al. 2005). These intermediate analyses were less detailed than the present final study, but did not show divergent results.

Albeit the drawback of sample size, it was considered as important, to show the present study separated from the parallel study in the automobile industry, because of differences concerning duration and level of exposure. The present study examined the neurobehavioural health of high and long-term exposed workers. The parallel study in the automobile industry complements the present study, as the neurobehavioural examinations were carried out in the early phase of working life of AI-welders with moderate AI exposure. Whether the existing results and conclusions will hold, the comparison with the study in the automobile industry will show. In this comparative study almost 150 workers could be followed up across the same time period with the same methods and a complete set of repeated measurements.

**Acknowledgments** We thank Hauptverband der gewerblichen Berufsgenossenschaften (HVBG), Vereinigung der Metallberufsgenossenschaften (VMBG) and Berufsgenossenschaft für Feinmechanik and Elektrotechnik for the financial support of the study. Furthermore we thank Mr. Thomas Rettig for his help with the personal air sampling.

## References

- Akila R, Stollery BT, Riihimäki V (1999) Decrements in cognitive performance in metal inert gas welders exposed to aluminium. *Occup Environ Med* 56:632–639
- Altmann P, Dhanesha U, Hamon C, Cunningham J, Blair J, Marsh F (1989) Disturbance of cerebral function by aluminium in haemodialysis patients without overt aluminium toxicity. *Lancet* 2:7–12
- Angerer J, Schaller KH (1982) Analytische Methoden zur Prüfung gesundheitsschädlicher Arbeitsstoffe, Bd. 2 Analysen in Biologischem Material, Lieferung (in German). Wiley-VCH, Weinheim, vol 6
- Antonini JM (2003) Health effects of welding. *Crit Rev Toxicol* 33:61–103
- Antonini JM, O'callaghan JP, Miller DB (2006a) Development of an animal model to study the potential neurotoxic effects associated with welding fume inhalation. *Neurotoxicology* (in press)
- Antonini JM, Santamaria AB, Jenkins NT, Albini E, Lucchini R (2006b) Fate of manganese associated with the inhalation of welding fumes: potential neurological effects. *Neurotoxicology* 27:304–310
- Bast-Pettersen R, Skaug V, Ellingsen D, Thomassen Y (2000) Neurobehavioral performance in aluminium welders. *Am J Ind Med* 37:184–192
- Berthon G (2002) Aluminium speciation in relation to aluminium bioavailability, metabolism and toxicity. *Coord Chem Rev* 228:319–341
- BGI 616 (2005) BG-Information 616. Beurteilung der Gefährdung durch Schweißrauche. Vereinigung der Metall-Berufsgenossenschaften (ed). Carl Heymanns Verlag KG, Köln
- Bolla KI (1991) Neuropsychological assessment for detecting adverse effects of volatile organic compounds on the central nervous system. *Environ Health Perspect* 95:93–98
- Buchta M, Kiesswetter E, Otto A, Schaller KH, Seeber A, Hilla W, Windorfer K, Stork J, Kuhlmann A, Gefeller O, Letzel S (2003) Longitudinal study examining the neurotoxicity of occupational exposure to aluminium-containing welding fumes. *Int Arch Occup Environ Health* 76:539–548
- Buchta M, Kiesswetter E, Schäper M, Zschiesche W, Schaller KH, Kuhlmann A, Letzel S (2005) Neurotoxicity of exposures to aluminium welding fumes in the truck trailer construction industry. *Environmental Toxicology and Pharmacology* 19:677–685
- Burt C (1958) The inheritance of mental ability. *Am Psychol* 13:1–15
- Colomina MT, Sanchez DJ, Sanchez-Turet M, Domingo JL (1999) Behavioral effects of aluminum in mice: influence of restraint stress. *Neuropsychobiology* 40:142–149
- DFG (2006) Deutsche Forschungsgemeinschaft. MAK- und BAT-Wert-Liste 2003 (List of MAK and BAT values 2006). In: Maximum allowable concentrations and biological tolerance values at the workplace. Wiley-VCH, Weinheim
- Erikson KM, Syversen T, Aschner JL, Aschner M (2005) Interactions between excessive manganese exposures and dietary iron-deficiency in neurodegeneration. *Environ Toxicol Pharmacol* 19:415–421
- Flaten TP, Alfrey AC, Birchall JD, Savory J, Yokel RA (1996) Status and future concerns of clinical and environmental aluminum toxicology. *J Toxicol Environ Health* 48:527–541
- Fleischer M, Schaller KH (1999) Aluminium. In: Angerer J, Schaller KH (eds) Analysis of hazardous substances in biological materials, Deutsche Forschungsgemeinschaft. Wiley-VCH, Weinheim, vol 6
- Fored CM, Fryzek JP, Brandt L, Nise G, Sjögren B, McLaughlin JK, Blot WJ, Ekblom A (2006) Parkinson's disease and other basal ganglia or movement disorders in a large nationwide cohort of Swedish welders. *Occup Environ Med* 63:135–40
- Fryzek JP, Hansen J, Cohen S, Bonde JP, Llabias MT, Kolstad HA, Skytthe A, Lipworth L, Blot WJ, Olsen JH (2005) A cohort study of Parkinson's disease and other neurodegenerative disorders in Danish welders. *J Occup Environ Med* 47:466–72
- Gilioli R (1993) EURONEST (1993) a concerted action of the European community for the study of organic solvents neurotoxicity. *Environ Res* 62:89–98
- Giorgianni C, Faranda M, Brecciaroli R, Beninato G, Saffiotti G, Muraca G, Congia P, Catanoso R, Agostani G, Abbate C (2003) Cognitive disorders among welders exposed to aluminium. *G Ital Med Lav Ergon* 25(Suppl):102–103
- Gitelman HJ (1995) Aluminum exposure and excretion. *Sci Total Environ* 163:129–135
- Gitelman HJ, Alderman FR, Kurs-Lasky M, Rockette HE (1995) Serum and urinary aluminium levels of workers in the aluminium industry. *Ann Occup Hyg* 39:181–191
- Graves AB, Rosner D, Echeverria D, Mortimer JA, Larson EB (1998) Occupational exposures to solvents and aluminium and estimated risk of Alzheimer's disease. *Occup Environ Med* 55:627–633
- Guilford JP (1972) Thurstone's primary mental abilities and structure-of-intelligence abilities. *Psychol Bull* 77:129–143
- Hänninen H, Matikainen E, Kovala T, Valkonen S, Riihimäki V (1994) Internal load of aluminium and the central nervous system function of aluminium welders. *Scand J Work Environ Health* 20:279–285
- Heller KA, Kratzmeier H, Lengfelder A (1988) Matritzen-Test-Manual zu den Standard Progressive Matrices von J.C. Raven. Beltz Test, Göttingen
- Hogstedt C, Andersson K, Hane M (1984) A questionnaire approach to the monitoring of early disturbances in central nervous functions. In: Aitio A, Riihimäki V, Vaninio H (eds) Biological monitoring and surveillance of workers exposed to chemicals. Hemisphere Publ Corp, Washington, pp 275–287
- Holm S (1979) A simple sequentially rejective multiple test procedure. *Scand J Stat* 6:65–70
- Hommel G (1988) A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika* 75:383–386
- Human Biomonitoring des Umweltbundesamtes (1998) Bundesgesundheitsblatt 41:271
- Ihrig A, Triebig G, Dietz MC (2001) Evaluation of a modified German version of the Q16 questionnaire for neurotoxic symptoms in workers exposed to solvents. *Occup Environ Med* 58:19–23
- Iregren A, Sjögren B, Gustafsson K, Hagman M, Nysten L, Frech W, Andersson M, Ljunggren KG, Wennberg A (2001) Effects on the nervous system in different groups of workers exposed to aluminium. *Occup Environ Med* 58:453–460
- Kiesswetter E, Blaszkewicz M, van Thriel C, Vangala R, Seeber A (2000) Shiftwork and neurotoxic occupational exposure: combined effects on tiredness and sleep and implications for working time arrangements. In: Marek T, Oginska H, Pokorski J, Costa G, Folkard S (eds) Shiftwork 2000. Implication for sciences, practice and business. Jagiellonian University, Krakow, pp 151–172
- Kiesswetter E, Sietmann B, Zupanec M, Seeber A (2000) Neurobehavioral study on the interactive effects of age and solvent exposure. *Neurotoxicology* 21:685–695
- Lehnert G, Schaller KH, Angerer J (1999) Report on the status of the external quality-control programs for occupational—medical and environmental—medical toxicological analyses in biological materials in Germany. *Int Arch Occup Environ Health* 72:60–64

- Letz R, Pieper WA, Morris RD (1996) NES test performance in a large US army veteran sample: relationship with both demographic factors and traditional neuropsychological measures. *Neurotoxicol Teratol* 18:381–390
- Letz R, DiIorio CK, Shafer PO, Yeager KA, Schomer DL, Henry TR (2003) Further standardization of some NES3 tests. *Neurotoxicology* 24:491–501
- Letzel S, Schaller KH, Hilla W, Windorfer K, Kraus T (1999) Untersuchung zur biologischen Halbwertszeit der renalen Aluminausscheidung bei Aluminiumschweißern: *Arbeitsmed Sozialmed Umweltmed* 11:456–460
- Lezak MD (1995) *Neuropsychological assessment*. Oxford University Press, New York
- Motorische Leistungsserie (MLS) V.3 (1994) Dr. G. Schuhfried, Mödling
- Oswald WD, Roth E (1997) *Der Zahlenverbindungstest, ZVT (trailmaking)*. Hogrefe, Göttingen
- Riihimäki V, Hänninen H, Akila R, Kovala T, Kuosma E, Paakkulainen H, Valkonen S, Engström B (2000) Body burden of aluminium in relation to central nervous system function among metal inert-gas welders. *Scand J Work Environ Health* 26:118–130
- Röllin HB, Theodorou P, Nogueir CM, Levin J (2001) Aluminium uptake and excretion in potroom workers of a new primary aluminium smelter during the construction stage. *J Environ Monit* 3:560–564
- Rosbach B, Buchta M, Csanady GA, Filser JG, Hilla W, Windorfer K, Stork J, Zschiesche W, Gefeller O, Pfahlberg A, Schaller KH, Egerer E, Pinzon LC, Letzel S (2006) Biological monitoring of welders exposed to aluminium. *Toxicol Lett* 162:239–245
- Salib E, Hillier V (1996) A case-control study of Alzheimer's disease and aluminium occupation. *Br J Psychiatry* 168:244–249
- Savory J, Exley C, Forbes WF, Huang Y, Joshi JG, Kruck T, McLachlan DRC, Wakayama I (1996) Can the controversy of the role of aluminium in Alzheimer's disease be resolved? What are the suggested approaches to this controversy and methodological issues to be considered? *J Toxicol Environ Health* 48:615–635
- Schaller KH, Angerer J, Drexler H (2002) Quality assurance of biological monitoring in occupational and environmental medicine. *J Chromatogr B Analyt Technol Biomed Life Sci* 778:403–417
- Schmidt KH, Metzler P (1992) *Wortschatztest (WST)*. Beltz, Weinheim
- Sjögren B, Elinder CG, Lidums V, Chang G (1988) Uptake and urinary excretion of aluminium among welders. *Int Arch Occup Environ Health* 60:77–79
- Sjögren B, Gustavsson P, Hogstedt C (1990) Neuropsychiatric symptoms among welders exposed to neurotoxic metals. *Br J Ind Med* 47:704–707
- Sjögren B, Iregren A, Frech W, Hagman M, Johansson L, Tesarz M, Wennberg A (1996) Effects on the nervous system among welders exposed to aluminium and manganese. *Occup Environ Med* 53:32–40
- Vernon PE (1949) The structure of practical abilities. *Occup Psychol* 23:81–96
- Wechsler D (1991) *Hamburg-Wechsler-Intelligenztest für Erwachsene (HAWIE)*. Hans Huber, Wien
- Wiener Reaktionstest V.3 (1997) Dr. G. Schuhfried, Mödling
- Yen-Koo HC (1992) The effect of aluminum on conditioned avoidance response (CAR) in mice. *Toxicol Ind Health* 8:1–7