Relating personality traits and mercury exposure in miners with machine learning methods

Bernard Ženko¹, Sašo Džeroski¹, Alfred B. Kobal², Darja Kobal Grum³, Niko Arnerič⁴, Joško Osredkar⁴, Milena Horvat¹

Abstract

We use machine learning/data mining methods to analyse scientific data in the area of environmental epidemiology, i.e., the study of the influence of environmental factors on human health. In particular, the aim of this study was an evaluation of the impact of long-term past occupational exposure to elemental mercury vapour (Hg⁰) on the mental health, i.e., personality traits of ex-mercury miners. Personality traits were defined by the Eysenck Personality Questionnaire (EPQ) and Emotional States Questionnaire (ESQ), which produced scores for traits such as depression and negative self-concept. Statistical analyses were performed to determine if there are significant differences between the values of the scores for ex-miners and controls. For the psychological traits for which significant differences were found between ex-miners and non-miners, we performed regression analysis. The target variables were the personality trait scores, while the independent/explanatory variables were the indices of previous occupational exposure to Hg⁰, medical history and lifestyle habits and some biological indices of actual non-occupational exposure. Regression/model trees were used to perform the analyses and revealed many interesting findings, e.g., that alcohol consumption and mercury exposure increase the depression score.

1. Introduction

The central nervous system is the critical organ for Hg⁰ vapour exposure [WHO 1976]. Post mortem studies [Kosta 1975, Byrne 1995, Falnoga 2000] have shown that the accumulation of mercury in the brain in ex-mercury miners was very high even several years after exposure. Long-term occupational exposure to Hg⁰ has also been found to be associated with symptoms of erethism, characterized by irritability, depression, introversion, apprehension, loss of self confidence and other non-specific symptoms [Smith 1970, Gerstner 1977, Piikivi 1989, Chapman 1990, Soleo 1990, Kobal 1975, Kobal 1991, WHO 1976, Echeverria 1995]. Only a few studies evaluated residual, mostly neurological, neurophysiological and neuropsychological effects associated

¹Jožef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia, email: bernard.zenko@ijs.si
²Idrija Mercury Mine, Idrija, Slovenia
³Department of Psychology, University of Ljubljana, Slovenia
⁴University Medical Centre, Ljubljana, Slovenia

13.8.2004, ZenkoB.tex

Within the research project “The impact of long-term past exposure to elemental mercury on antioxidative capacity and lipid peroxidation in mercury miners,” (financially supported by the Slovene Ministry of Education, Sciences and Sport and the Idrija Mercury Mine, Slovenia), data on miners employed in the Idrija mercury mine were collected, as well as data on a suitable control population. The data includes a number of indicators of exposure to mercury, as well as data on personal traits reported by ex-miners in the Eysenck Personality Questionnaire (EPQ) and Emotional States Questionnaire (ESQ) in the period after exposure.

An epidemiological study of the causes of death among miners in certain mercury mines [Boffetta 1998] also revealed an increased mortality rate due to suicides among the miners of the Idrija Mercury Mine (unpublished data). Even on the basis of long-term medical monitoring of workers exposed to Hg° [Hribernik 1950, Kobal 1975a, Kobal 1991], a direct connection between the increased suicidality of miners and poisoning with Hg° cannot be confirmed. The purpose of the present study is to evaluate the impact of long-term occupational exposure to Hg° on the personality traits reported by ex-miners in the Eysenck Personality Questionnaire and the Emotional States Questionnaire in the period after exposure.

In the rest of the paper we first describe the data set used (Section 2). Section 3 presents the methodology of the data analysis and Section 4 presents the results, which are then discussed in the final section (Section 5).

2. Data Set

The study population comprised 53 mercury miners and 53 workers in the control group. The miners were employed in the Idrija mercury mine from 7 to 30 years. The subjects in the control group were not occupationally exposed to Hg° or other heavy metals and organic solvents. They live in an environment without elevated concentrations of Hg°. Their eating and living habits are very similar to those of the ex-miners. Individuals with a medical history of neuropsychiatric disorders, alcoholism, use of analgetics and renal disease were excluded.

The study was conducted with the approval of the State Ethical Commission in accordance with the ethical standards laid down in the Declaration of Helsinki. All participants gave their informed consent before their inclusion in the study. The following types of data on selected subjects were collected:

- **Medical and psychological data.** The medical examination included a medical history, life style habits (smoking, alcohol consumption), an evaluation of non-specific symptoms of “micromercurialism” [WHO 1976], clinical neurological status and a selected psychological questionnaire (Slovene translation [Lojk 1981] of the Eysenck Personality Questionnaire – EPQ [Eysenck 1975] and Emotional States Questionnaire – ESQ [Lamovec 1989]).

- **Data on Hg exposure.** Environmental and biological data on the group of miners
studies were collected from 1959 onward from workload records, daily reports on Hg \textsuperscript{o} measurements in the workplace and personal medical records and biological monitoring data [Kobal 1997]. Since 1959, the miners have been biologically monitored by means of urine mercury (U-Hg) analyses. Biological past exposure indices were evaluated on the basis of 5452 U-Hg measurements in urine spot samples performed during past exposure of miners to Hg \textsuperscript{o}. This type of data was of course collected only on (ex-)miners, while the exposure of the control group was assessed to be zero.

- **Biological data.** In addition to basic routine biochemical and haematological parameters, which were determined by the usual clinical biochemical methods, the following parameters were also measured: B-THg and U-Hg (determined by cold vapour atomic absorption spectrophotometry [Horvat 1991]) and monomethylmercury in whole blood (B-MeHg) (determined by acid leaching/solvent extraction/aqueous phase ethylation/isothermal GC/CV AFS detection [Liang 1994]).

  All these data were described with the following features: groups (ex-miners - underground work; controls - work in the open), subgroups (active mines, retired miners), age, residence (municipality of Idrija - other location; town centre - hillside) dental amalgam score, cigarettes per day, alcohol consumption (ml/day), albumin in urine g/mol creatinine (potential marker of Hg exposure effect), years of work in mercury mine (years of exposure), work cycles of Hg exposure (number), the geometrical mean of cycles U-Hg level (g/L), the geometrical mean of cycles peak U-Hg level, cumulative U-Hg level (g/L), U-Hg level at the last exposure (g/L), time since last exposure in days (exposure free interval), and the selected personality traits score. Those miners with an exposure free interval below 12 months were excluded from the evaluations. To focus on miners with higher exposure we also ignored five miners with U-Hg value of the last exposure under 10 g/L.

  For a detailed description of the data, we refer the reader to Kobal Grum et al. [2004].

**2.1 Methodology**

Group differences of all observed parameters were evaluated by the application of an analysis of variance using one-way ANOVA software. The relationship between exposure and other variables was evaluated by means of Pearson’s correlation coefficient, which reflects the degree of linear relation between two sets of data. For all computations we used the SPSS for Windows. To find possible explanations of associations between the target variables (personality traits) and biological indicators of occupational Hg \textsuperscript{o} exposure in combination with covariables, we used machine learning methods such as regression and model trees [Breiman 1984, Quinlan 1992].

Regression trees are a representation for piece-wise constant or piece-wise linear functions. Like classical regression equations, these predict the value of a dependent variable (called class) from the values of a set of independent variables (called attributes). Data represented in the form of a table can be used to learn or automatically construct a regression tree. In the table, each row (example) has the form \((x_1, x_2, \ldots, x_N, y)\), where \(x\) are values of \(N\) attributes (e.g., subjects’ age, daily con-
sumption of alcohol, etc.) and $y$ is the value of the class (e.g., the EPQ psychoticism score). Unlike classical regression approaches, which find a single equation for a given set of data, regression trees partition the space of examples into axis-parallel rectangles and fit a model to each of these partitions. A regression tree has a test in each inner node which tests the value of a certain attribute, and in each leaf a model for predicting the class: the model can be a linear equation or merely a constant. Trees that can have linear equations in leaves are also called model trees.

Given a new example for which the value of the class should be predicted, the tree is interpreted from the root. In each inner node, the prescribed test is performed and according to its result, the corresponding left or right sub-tree is selected. When the selected node is a leaf, the value of the class for the new example is predicted according to the model in the leaf. Tree construction proceeds recursively, starting with the entire set of training examples (entire table). At each step, the most discriminating attribute is selected as the root of the (sub)tree and the current training set is split into subsets according to the values of the selected attribute. For discrete attributes, a branch of the tree is typically created for each possible value of the attribute. For continuous attributes, a threshold is selected and two branches are created based on that threshold.

Technically speaking, the most discriminating discrete attribute or continuous attribute test is the one that reduces to the greatest degree the variance of values of the class variable. For continuous attributes, the values of the attribute appearing in the training set are considered thresholds. For the subsets of training examples in each branch, the tree construction algorithm is called recursively. Tree construction stops when the variance of class values of all examples in a node is small enough (or if some other stopping criterion is satisfied). Such nodes are called leaves and are labeled with a model (constant or linear equation) for predicting class value. An important mechanism used to prevent trees from over-fitting data is tree pruning. Pruning can be employed during tree construction (pre-pruning) or after the tree has been constructed (post-pruning). Typically, a minimum number of examples in branches can be prescribed for pre-pruning and the confidence level in error estimates in leaves for post-pruning.

There are a number of systems for inducing regression trees from examples, such as CART [Breiman 1984] and M5 [Quinlan 1992]. M5 is one of the most well-known programs for regression tree induction. We used the M5’ system [Wang 1997], a re-implementation of M5 within the software package WEKA [Witten 1999]. The parameters of M5’ were set to their default values, except where described differently in this text.

3. Results

Data from the Eysenck Personality Questionnaire (EPQ) and the Emotional State Questionnaire (ESQ) are presented in Table 1. A comparison of the group of miners and the control group revealed a lower mean score of extraversion ($p = 0.017$) and lie ($p = 0.003$) in the group of miners, while the average scores for depression and negative self-concept were significantly higher ($p<0.01$) in miners than in controls.
Table 1: Average EPQ and ESQ scores of miners and controls.

<table>
<thead>
<tr>
<th>EPQ or ESQ score</th>
<th>Miners</th>
<th>Controls</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>EPQ – Psychoticism</td>
<td>3.80</td>
<td>2.37</td>
<td>3.74</td>
</tr>
<tr>
<td>EPQ – Extroversion</td>
<td>12.09</td>
<td>3.81</td>
<td>13.93</td>
</tr>
<tr>
<td>EPQ – Neuroticism</td>
<td>8.14</td>
<td>4.27</td>
<td>7.55</td>
</tr>
<tr>
<td>EPQ – Lie</td>
<td>12.45</td>
<td>4.22</td>
<td>15.05</td>
</tr>
<tr>
<td>ESQ – Depression</td>
<td>20.33</td>
<td>5.07</td>
<td>17.73</td>
</tr>
<tr>
<td>ESQ – Contentment</td>
<td>30.52</td>
<td>4.97</td>
<td>31.05</td>
</tr>
<tr>
<td>ESQ – Aggressions</td>
<td>17.17</td>
<td>4.20</td>
<td>15.95</td>
</tr>
<tr>
<td>ESQ – Indifference</td>
<td>9.74</td>
<td>2.73</td>
<td>8.51</td>
</tr>
<tr>
<td>ESQ – Pos. self-concept</td>
<td>15.52</td>
<td>3.16</td>
<td>16.51</td>
</tr>
<tr>
<td>ESQ – Neg. self-concept</td>
<td>8.98</td>
<td>2.51</td>
<td>7.68</td>
</tr>
</tbody>
</table>

Table 2: Linear regression model constructed by M5*, describing the extroversion score (correlation coefficient = 0.43).

```
extroversion_score =
   2.9701 * group = retired_miners,controls +
   -0.1642 * age_in_years +
   0.1891 * years_of_work_in_mercury_mine +
   -0.0002 * cumulative_uhg_level_micro_g_L +
   -0.0198 * uhg_at_last_exposure_micro_g_L +
   0.15 * alcohol_consumption_ml_per_day + 17.7282
```

The indifference average score also tended to be higher in miners (p=0.025) in comparison to the controls. No correlation between scores of EPQ and ESQ variables, or between indices of past exposure or alcohol consumption evaluated by means of Pearson’s correlation coefficient were detected, probably due to the non-linear relationship between these variables.

The M5 models built for extroversion, lie, depression, indifference and negative self-concept score are presented in Tables 2, 3, 4, 5, and 6, respectively. The model describing extroversion score presented in Table 2 consists of a single linear equation (a model tree with a single leaf node) comprising the following features: groups, age, years of work in the mercury mine, cumulative U-Hg level, U-Hg level at last exposure, and alcohol consumption. Evidently, the “extroversion score” is associated with the sub-groups of retired miners and control groups. Years of work in the mercury mine and alcohol consumption increased the extroversion score. Age, cumulative U-Hg level and U-Hg level at last exposure tended to decrease the extraversion score.

The model describing lie score presented in Table 3 consists of a single linear equation comprising age and place of residence. We can see that the lie score increases with age and decreases in miners living in the center of Idrija (55 % of miners), but it is not related to the occupational exposure indices.

The model tree predicting the depression score presented in Table 4 contains four leaves, of which three contain constant predictions and one linear model. It is evident
Table 3: Linear regression model constructed by M5', describing the lie score (correlation coefficient = 0.45).

\[
\text{lie\_score} = 0.1374 \times \text{age\_in\_years} + 2.6639 \times \text{residence\_in\_town\_centre = not} + 5.9845
\]

Table 4: Model tree (with 4 linear models) constructed by M5', describing the depression score (correlation coefficient = 0.54). The number of subjects in each leaf is given in parenthesis.

```
from the LM1 model, which was based on a larger number of subjects (39 controls and 9 miners), that low alcohol consumption (<26.6 ml/day) at a lower level of occupational exposure (mean cycle U-Hg <38.7 µg/L) did not increase the depression score. Models LM2 and LM3 relate to an increased depression score in 28 miners at a higher level of exposure (male cycles U-Hg >38.7 µg/L). A higher consumption of alcohol (per se) (>26.6 ml/day) tends to increase the depression score in 14 miners and 10 controls.

The indifference score model is presented in Table 5 and consists of a single linear equation comprising group, age, the mean of cycles and the mean of cycles peak U-Hg level, U-Hg level at last exposure, and cigarette consumption per day. We can see that the indifference score is typical for the ex-miners’ group. All mentioned variables increased the observed score, except the mean of cycles peak U-Hg levels, which does not seem to represent the miners’ integral occupational exposure level.

The model tree predicting the negative self-concept score presented in Table 6 contains two leaves with one linear model. Model LM1 represents the controls (N=53) with a relatively lower score. Age and alcohol consumption partly increased their negative self-concept score. LM2, which represents the ex-miners (N=47, only miners with “last exposure” U-Hg >10 µg/L), relates the negative self-concept score to mean cycles of U-Hg level (>32.5 µg/L) and U-Hg level at last exposure. In these cases as well, the mean cycles peak U-Hg level did not increase the observed score.
Table 5: Linear regression model constructed by M5', describing the indifference score (correlation coefficient = 0.42).

\[
\text{indifference\_score} = 1.8823 \times \text{group\_miners,\_retired\_miners} + 0.0538 \times \text{age\_in\_years} + 0.0331 \times \text{cyc\_uhg\_level\_geom\_mean\_mic\_g\_L} + -0.0469 \times \text{cyc\_peak\_uhg\_level\_geom\_mean\_mic\_g\_L} + 0.0201 \times \text{uhg\_at\_last\_exposure} + 0.0607 \times \text{cigarettes\_per\_day} + 5.7682
\]

Table 6: Model tree (with 2 linear models) constructed by M5', describing the negative self-concept score (correlation coefficient = 0.50). The number of subjects in each leaf is given in parenthesis.

\[
\begin{align*}
\text{cycles\_peak\_uhg\_level\_geom\_mean\_µg\_L} & \\
\leq & 32.5 & >32.5
\end{align*}
\]

LM1: neg\_self\_concept\_score = 0.0465 \times \text{age\_in\_years} + 0.0254 \times \text{alcohol\_consumption} + 5.3991

LM2: neg\_self\_concept\_score = -0.0455 \times \text{cycles\_peak\_uhg\_level\_geom\_mean\_micro\_g\_L} + 0.0248 \times \text{uhg\_at\_last\_exposure\_micro\_g\_L} + 12.3991

4. Discussion

In our study, the lower EPQ score of extraversion found in mercury miners suggests that miners are less outgoing, less sociable and more introverted than the control group. However, it is evident from the regression model that extraversion seems to be properly associated with retired miners and controls, decreasing with age, cumulative U-Hg level and U-Hg at last exposure. Years of work and alcohol consumption slightly increase extroversion. The latter is apparently connected with the acute, short-term effects of alcohol. The influence of age to extraversion in males had also been reported in other studies [Eysenck 1975, Lojk 1981], but less is known about late impacts of occupational exposure. The lower score on the lie scale (EPQ) of mercury miners is not associated to Hg\textsuperscript{0} exposure, but suggests that the answers obtained from miners could be considered more valid in comparison to the controls. This could also mean that sincerity is one of the most important personality characteristics enabling miners to preserve their collaborative and team working spirit.

The results obtained from the ESQ showed some significant differences between miners and the control group. Mercury miners tend to be more depressive, more rigid in expressing their emotions, and are likely to have a more negative self-concept than the members of the control group. From the regression tree we can see that permanent increased alcohol consumption per se (> 26 ml/day) increases depression in both miners and controls, which is also known from other studies [Schuckit 1986,
Leibenuft 1993]. Lower permanent alcohol consumption (< 26 ml/day) associated with long-term higher occupational Hg° exposure (cycles U-Hg level > 38 µg/L) seems to increase the miners’ depression score.

The relative indifference (emotional rigidity) expressed by the indifference score is a common characteristic of all miners. This is also slightly increased by the level of internal Hg° doses received during previous occupational exposure (geometrical mean U-Hg level, U-Hg of last exposure). The indifference established in miners in the period after exposure is in genuine contradiction to the known emotional lability that is typical for the state of increased absorption and chronic occupational intoxication with Hg°. From the regression tree we can see that the internal doses received during occupational exposure, expressed by the geometrical mean U-Hg level (> 32 mg/L) and U-Hg at last exposure, appear to be the factor most strongly associated with miners’ negative self-concept.

Our results suggest that emotional rigidity, depression, negative self-concept, and partly also introversion, which characterise the personalities of miners, may be associated with long-term substantial occupational exposure to Hg°. We presume that the mutual interaction of long-term increased exposure to Hg° and long-term “moderate” alcohol consumption has had a decisive influence on the development of depression in the miners observed. This depressive mood itself could, as a result, increase the risk of suicide among miners of the Idrija Mercury Mine. This study thus further supports efforts to reduce the occupational and non-occupational exposure levels of Hg°, in our case the U-Hg level, to the lowest observed adverse effect level [UNEP 2002, WHO 2003] capable of preventing the late effects of Hg° exposure on the potential development of depression, emotional rigidity and negative self-concept, which can significantly decrease resistance towards psychosocial stress factors.

Bibliography


13.8.2004, ZenkoB.tex