The Science of the Total Environment, 80 (1989) 167–174
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

TRACE ELEMENT ANALYSIS IN ENVIRONMENTAL AND OCCUPATIONAL HEALTH: BOX PLOT REPRESENTATION OF ELEMENTAL COMPOSITION RESULTS

G. NICOLAOU, R. PIETRA and E. SABBIONI

Commission of the European Communities, Joint Research Centre – Ispra Establishment Radiochemistry Division, 21020 Ispra (Va) (Italy)

R.M. PARR

IAEA, Medical Applications Section, Department of Research and Isotopes, P.O. Box 100, A-1400 Vienna (Austria)

(Received October 3rd, 1988; accepted November 26th, 1988)

ABSTRACT

Box plots are used in the visual representation of large data sets and in exploratory data analysis. They display batches of data with five values being used to describe the data set. These are the median, the upper and lower extremes of the range of values and the 75 and 25 percentiles. A notch about the median, e.g. at the 95 percent level of significance, can be incorporated in the display allowing the difference between the medians of different sets to be established. The method, although not recently established, has so far found little application in the analytical field. Hence, in an effort to strengthen its applicability, the features and capabilities of box plots, in terms of data reporting and insight into the data set, are here described through elemental composition studies in relation to environmental and occupational health.

INTRODUCTION

Many sensitive analytical techniques, e.g. neutron activation, atomic absorption spectrometry, inductively-coupled plasma mass spectrometry, have become available allowing the research and routine measurement of trace elements in fields such as biology, environmental science and industry. In particular, the interest in the application of these techniques in trace element analysis of biological samples, e.g. human tissues and fluids, has been stimulated by the discovery of additional essential elements [1], as well as by the recognition of environmental toxicity and occupational hazards of some metals and their compounds [2]. The definition of trace element concentrations in both normal and diseased human tissues and body fluids is the fundamental parameter to assess the biological effects of trace metals in exposed humans [3]. These studies are ideally performed on a large population so that statistically accurate description of the situation is achieved. This results in the collection of a large amount of data and, it is in this concept, that the issue of data reporting becomes extremely important, so that maximum information of the data set, on a reasonable scientific basis, is revealed.

0048-9697/89/$03.50

This paper focuses on a method of data reporting based upon the graphical representation of the results in the form of box plots [4, 5]. The method, although not recently established, has so far found little application in the analytical field. Thus, it is the purpose of this work to demonstrate the features and capabilities of the method, in terms of data reporting and insight into the data set, in an attempt to strengthen its validity and applicability. This is achieved through elemental composition studies of human samples in relation to environmental and occupational health research. Results from the following two studies, carried out by neutron activation analysis [6], have been used to demonstrate box plots:

(a) elemental analysis of normal human lung tissue, as a basis to establish reference data on its composition [7] and hence assess the health impact of toxic elements and compounds. The box plot representation is demonstrated for the case of cobalt concentration in the samples analysed;

(b) determination of cobalt concentration in the urine of subjects, following their occupational exposure in the hard metal industry [8].

MATERIALS AND METHODS

Sample collection

The following procedures, corresponding to the two investigations considered, were utilised for the sample collection:

(i) lung samples were collected from autopsy under a strict protocol in order to avoid discrepancies in the results arising from analytical inconsistencies, personal habits, physiological factors, drug consumption and environmental and occupational conditions [7]. The existing data in the literature on trace metals in human lung [9, 10], are subjected to discrepancies, which are probably the result of analytical errors and biological variations, thus rendering them unreliable in the evaluation of the human health impact. With these points under consideration, the lungs were removed during autopsy, at the Anatomia Patologica of the University of Sacred Heart, Rome, using TiN-coated surgical instruments to avoid contamination during collection. Nine non-smoking subjects, deceased from non-pathological causes, who during their life-time had not been occupationally exposed to contaminating elements have been analysed. All the subjects were inhabitants of the same urban area of Rome, in order to minimise inconsistencies arising from differences in diet and ambient air composition. Samples, of ~ 2 g in weight, were collected from the same six sampling points from each pair of lungs [11].

(ii) a population group, consisting of subjects working in the hard metal industry in North Italy, was investigated. Twenty-seven subjects, from two different occupational environments, here named A and B, with ages between 15 and 50 years and working period from 2 months to 10 years were screened. Urine was collected from each subject before-shift, and following at least two weeks at work, so that the effect of any period away from work on urine composition is minimised [12]. Urine was also collected from control subjects not working in this type of environment; these were three subjects from the medical personnel at the Hospital of Bergamo (Italy), responsible for the collection of the samples.

Lung tissues or urine samples upon collection were immediately transferred to polyethylene vials, previously washed with pure HNO_3 and Millipore water. The vials were sealed thermally and stored at −20°C before the analysis.

Neutron activation analysis

Prior to irradiation, the samples, in the vials used in the sampling procedure, were frozen to −80°C for 24 h and then lyophilised for 48 h. Both lung tissue and urine samples were irradiated for 8 h in a thermal neutron flux of 10^{13} neutrons $cm^{-2} sec^{-1}$ at the University of Pavia TRIGA MARK II reactor (Italy). The irradiated samples were then submitted to radiochemical separations involving mineralisation by the teflon bomb procedure and absorption on a set of chromatographic columns, filled with tin dioxide (TDO), CHELEX 100 or DOWEX 1–X8 resin, in order to remove strong interfering radionuclides, such as ^{24}Na, and hence improve the sensitivity for the detection of the other elements present [6]. Each eluted fraction and/or the columns were then counted by high resolution computerised gamma-ray spectroscopy using a Ge(Li) detector.

Box plot representation

The basic features of the configuration are shown in Fig. 1. To describe the data sets from the studies performed, box plots have been used according to McGill et al. (1978). Box plots display batches of data with five values used to

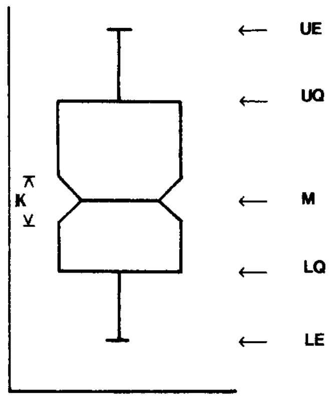

Fig. 1. Configuration of a box plot (M = median; UQ, LQ = 75th and 25th percentiles; UE, LE = upper and lower extremes of the range of values; K = notch around the median).

describe the data set. These are the median (M), the upper and lower extremes of the range of values (UE, LE) and the 75 and 25 percentiles (UQ, LQ). The median is used instead of the mean, as it has the advantage of being a more robust variable, less affected by any extreme values present in the displayed data set. The range between the 75 and 25 percentiles, termed interquartile range, indicates the midspread of the data set containing 50 percent of the values of the set.

This form of data reporting allows the comparison of different data sets. In this case, the notch (K) around the median is of importance [4, 13]. The notch may be calculated as M $\pm$ Cs, where s is the standard deviation of the median (M) and is given by the equation:

$$s = 1.25 \cdot R/1.35 \cdot N^{1/2}$$

and R is the interquartile range of the data, N is the number of observations for each group of data and C is a constant. The notch can be used either to test whether or not the medians for two different data sets are significantly different or to represent a confidence interval about the median, with C being 1.7 and 1.96 respectively at the 95 percent level of significance [4, 13].

RESULTS AND DISCUSSION

Box plot representation in the study of elemental composition of 'normal' human lung

The objective of this study is to define reference values, concerning the range of metal concentrations in unexposed ('normal') human lung, to be subsequently used in toxicological assessment studies.

The box plot representation is shown for the case of cobalt in 'normal' human lung tissue. The results from the neutron activation analysis of cobalt in lung tissue and the corresponding data available in the literature [9, 10] with means ($\pm$ standard deviation) of 23.2 $\pm$ 1.8 and 62.3 $\pm$ 13.2 respectively, are displayed in the form of box plots in Fig. 2. A comparison of Figs 2a and 2b reveals a much wider range of literature values, more precisely three orders of magnitude, as opposed to one order of magnitude for the values obtained in this study. Furthermore, the interquartile range of the literature values is twice that for the results of the present study. The wide spread of the literature values may be attributed to analytical inconsistencies, differences in age, smoking habits and diet, biological variations between individuals and variation in their urban and occupational ambient environment. All these factors were probably not considered in the collection and analysis of the samples that generated the literature values. The strict protocol followed in the sample collection and analysis in the present work is probably responsible for the minimisation of the effect of the above factors on the elemental composition data, clearly demonstrated by the narrow range of the analytical values obtained.

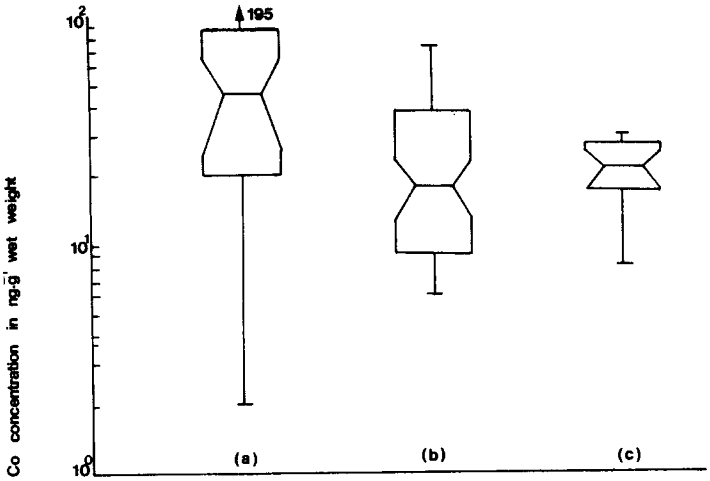

Fig. 2. Presentation, in the form of box plots, of cobalt concentrations in human lung tissue: (a) literature values; (b) and (c) values obtained from the analysis of nine pairs of lung, the samples taken from different sampling points and the same point respectively.

Neutron activation analysis showed a difference in cobalt concentration between the different segments of the lung [7]. The effect of this inhomogeneity on the data set obtained was then investigated by considering samples originating from the same sampling point, as in the case of the upper lobe of the left lung. The cobalt concentrations in these samples are displayed, in the form of box plots, in Fig. 2c.

The following two observations may be made. Firstly, the medians of the two data sets displayed in Figs 2b and 2c do not differ at the 95 percent level of significance. This is consistent with the fact that the concentrations from the same sampling point in the lungs form a sub-set of the population comprised of the concentrations from all the different sampling points analysed. Secondly, the range of concentrations, Fig. 2c, is now further decreased, with respect to the concentrations displayed in Fig. 2b, emphasizing the necessity for consistency in the choice of the sampling point considered in such a study.

Box plot representation in the study of the urine of hard metal workers

The analysis of the urine of hard metal workers showed elevated cobalt concentrations in comparison to controls [8], as shown in Table 1. The concentrations determined are presented, in the form of box plots, for the two occupational environments A and B, in Fig. 3. A large difference is observed between the mean and median values of urine cobalt, with the former being higher (Table 1). This is due to the presence of some extreme high values in the data sets, which influences the evaluation of the mean. This demonstrates the

TABLE 1

Mean and median of the data sets, on urine cobalt content of hard metal workers, displayed by box plots in Fig. 3.

Origin of samples	Elemental concentration ($ng\,g^{-1}$ wet weight)	
	Mean (± SD)	Median
Factory A	210 (± 55)	120
Factory B	1010 (± 530)	300
Controls (n = 3)	1.20 (± 0.25)	–

advantage of using the median, instead of the mean, as it is a more robust variable less affected by the presence of these extreme values.

Although there is an overlap between the two sets, a difference is observed between the corresponding medians, at the 95 percent level of significance. This indicates a possible difference in cobalt content of the ambient environment in the factories. Subsequent analysis of the cobalt content in the dust of the ambient environment of the two factories showed a higher content, by a factor of two, in factory B. This observation is consistent with the higher concentrations encountered in the workers of this factory, Fig. 3.

In seeking an explanation for the wide range of cobalt concentrations measured in the urine of the subjects, other information regarding the subjects is considered. It becomes evident that it is not only the difference in ambient

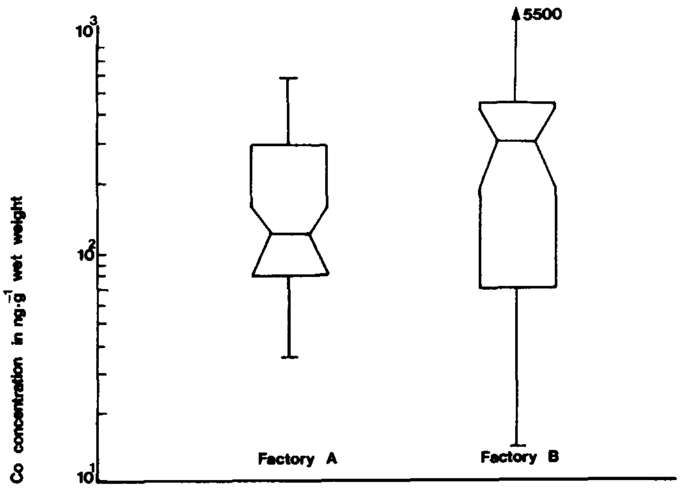

Fig. 3. Presentation, in the form of box plots, of cobalt concentrations in the urine of hard metal workers from two different occupational environments A and B.

air of the two factories, but also other factors, such as age and working period, which influence the results. In this context, higher cobalt concentrations occur for the younger workers with the smaller lifespan in the hard metal industry [14].

CONCLUSIONS

This work has focused on a data reporting method based upon the graphical representation of the results in the form of box plots. However, as box plots have not been extensively applied in analytical studies, an attempt was made to demonstrate their representation and hence strengthen their applicability in trace element analysis for data reporting and insight into the data sets. Results from two elemental composition studies of human samples, in relation to environmental and occupational health, were considered.

Box plots proved a most valuable tool in data reporting. They allowed the graphical representation of the data sets indicating important features such as the median, extreme values and spread of values about the median in the form of quartiles. Furthermore, a comparison of the medians of the data sets was achieved with the inclusion of a notch, e.g. at the 95 percent level of significance, around the median.

The features which characterize box plots also provided additional insight into the data allowing interpretation as has been demonstrated by the two studies carried out. Clearly, this aspect of box plots may be employed as a guideline for further data analysis and experimental investigations.

REFERENCES

1 A. Berlin and E. Di Ferrante, La recherche à l'appui de la réglementation sur les métaux lourds dans la Communauté Européenne in: The Scientific Bases for Environmental Regulatory Action 'Health Environment', E. Di Ferrante (Ed.), Report EUR 7952, Commission of the European Communities, Luxembourg, 1982.
2 Handbook on Toxicity of Inorganic Compounds, H.G. Seiler, H. Sigel and A. Sigel (Eds.), Marcel Dekker Inc., New York, 1986.
3 G. Nordberg, J. Parizek and M. Piscator, Factors influencing effects and dose-response relationships of metals, in: Handbook on the Toxicology of Metals, L. Friberg, F. Nordberg and V.B. Vouk (Eds.), Elsevier/North Holland Biomedical Press, Amsterdam, 1980.
4 R. McGill, J.N. Tukey and W.A. Larson, The American Statistician, 32 (1978) 12.
5 R.M. Parr, On the need of improved quality assurance in biomedical neutron activation analysis as revealed by the results of some recent IAEA intercomparisons, IAEA/TECDOC/323, Vienna (1984).
6 R. Pietra, E. Sabbioni, M. Gallorini and E. Orvini, J. Radioanal. Nucl. Chem., 102 (1986) 69.
7 G. Nicolaou, R. Pietra, E. Sabbioni, A. Alimonti, S. Caroli, E. Coni, W. Altaf and N.M. Spyrou, Symp. Environmental Aspects of Trace Elements, Paris, 1–4 Dec., 1987.
8 G. Nicolaou, R. Pietra, E. Sabbioni, G. Mosconi, G. Cassina and P. Seghizzi, J. Trace Elem. Electrolytes Health Dis., 1 (1987) 73.
9 G.V. Iyengar, M.E. Kollmer and H.I.M. Bowen, The elemental composition of human tissues and body fluids, Verlag Chemie, New York, 1978.
10 C. Vanoeteren, R. Cornelis and E. Sabbioni, Critical evaluation of normal levels of major and trace elements in human lung tissue, Report EUR 40440 EN, Commisson of the European Communities, Luxembourg, 1986.

11 S. Caroli, E. Coni, A. Alimonti, E. Beccaloni, E. Sabbioni and R. Pietra, Analusis, 16 (1988) 75.
12 G. Scansetti, S. Lamon, S. Talarico, G.C. Botta, P. Spinelli, F. Sulotto and F. Fantoni, Int. Arch. Occup. Environ. Health, 57 (1985) 19.
13 M.G. Kendall and A. Stuart, The advanced theory of statistics, Vol. 1, Hafner Publishing Co., New York, 1967.
14 G. Nicolaou, R. Pietra and E. Sabbioni, Symp. Trace Elements in Human Health and Disease, Odense, Denmark, 17–21 August, 1987.