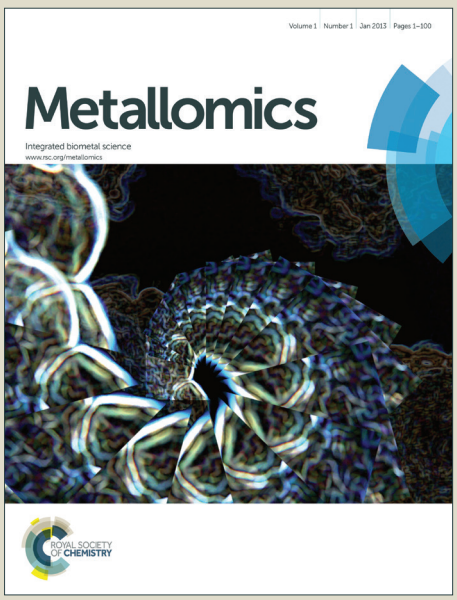
Metallomics

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/metallomics

Classification models based on level of metals in hair and nails of laryngeal cancer patients. Diagnosis support or rather speculation?

Magdalena Golasik,[a] Wojciech Jawień,[b] Agnieszka Przybyłowicz,[a] Witold Szyfter,[c,d] Małgorzata Herman,[a] Wojciech Golusiński,[c,e] Ewa Florek[f] and Wojciech Piekoszewski*[a]

[a] *Department of Analytical Chemistry, Faculty of Chemistry, Jagiellonian University in Kraków, Ingardena 3, 30-060 Kraków, Poland*
[b] *Department of Pharmacokinetics and Physical Pharmacy, Jagiellonian University School of Medicine, Medyczna 9, 30-688 Kraków, Poland*
[c] *Department of Otolaryngology and Laryngological Oncology, University of Medical Sciences, Przybyszewskiego 49, 60-355 Poznań, Poland*
[d] *Clinic of Phoniatrics and Audiology, University of Medical Sciences, Przybyszewskiego 49, 60-355 Poznań, Poland*
[e] *Greater Poland Cancer Center, Garbary 15, 61-866 Poznań, Poland*
[f] *Laboratory of Environmental Research, Department of Toxicology, University of Medical Sciences, Dojazd 30, 60-631 Poznań, Poland*

Corresponding author: Prof. Wojciech Piekoszewski
Department of Analytical Chemistry, Faculty of Chemistry, Jagiellonian University
R. Ingardena 3, 30-060 Kraków, Poland
Phone.: +(48 12) 663 56 00, Fax: +(48 12) 663 56 01
e-mail: wpiekosz@tlen.pl

Abstract

The relationship between the concentration of chemical elements and cancer is regarded as one of many factors taken into account in carcinogenesis. The aim of this study is to investigate some relations between cancer risk and element status as well as cancer risk and external factors, such as diet, smoking and drinking habits, in order to support diagnosis of cancer. The samples of hair and nails obtained from patients with larynx cancer and healthy subjects were analyzed. Essential elements (Ca, Cr, Mg, Zn, Cu, Mn, Fe), besides toxic metals (Cd, Co, Pb), were determined with an inductively coupled plasma atomic emission spectrometry (ICP-OES) and mass spectrometry (ICP-MS) techniques. The concentration of essential elements was from 1.5- (Zn) to 4.7-fold (Fe) higher in hair and from 2.4- to 3.3-fold higher in the nails of the control group compared to the patients, while the opposite trend was observed for the heavy metals. The differences between two groups in the level of metals (except for Zn) were statistically significant ($p<0.05$). The association of cancer with metals and other factors was evaluated with various statistical methods, among whom the best predictions were obtained with the logistic regression, artificial neural networks and canonical discriminant analysis. The classifiers constructed using the data from a survey of diet and lifestyle, and analysis of elements in hair and nails, can be useful tools for estimating cancer risk and early screening of the disease.

Key words: laryngeal cancer, metals, hair, nails, statistical models

Introduction

The latest cancer statistics reported that the overall cancer mortality has steadily declined over the past two decades. Nevertheless, in 2014, deaths from cancer are projected to reach about 585 720 in the United States and about 1 323 600 in the European Union.[1,2] Cancer is characterized by the uncontrolled growth of an abnormal cell and is caused by a number of factors. The process of carcinogenesis is quite complex and multi-directional, and therefore it is extremely difficult to find reliable markers to facilitate the identification of cancer at an early stage.[3] In the case of larynx cancer, the most important risk factor is associated with lifestyle and is definitely long-term exposure to tobacco smoke.[4] Conducted studies also confirm that alcohol drinking (especially strong alcohol) is a second major factor that has an important role in laryngeal carcinogenesis.[5] Additionally, tobacco and alcohol altogether exhibit the synergistic effect with regard to larynx cancer.[6] Another factor which is involved in the pathogenesis of cancer of the larynx is diet. Fresh fruit, vegetables, and legumes have a favourable effect on laryngeal cancer, and poor nutrition is associated with an increased cancer risk.[7,8] Moderate consumption of olive oil was reported to be associated with reduced risk as well.[9]

Lifestyle and dietary habits have undoubtedly impacted on the level of the elements in the body which are involved in the biochemical processes that occur in the human organism. The significance of trace elements in health and disease is undisputed because of their essential role in specific concentration ranges and a toxic role at relatively higher levels.[10] The importance of micronutrients and toxic elements in changes in the human body indicate that they may be important as inhibitors or promoters of the development of malignancies. The correlation between the levels of metals and the development of neoplastic processes has been widely studied for use in medical diagnosis. Even though the biochemical mechanism of these elements being the cause of cancer in the human body is not very clear at the present stage, more research has to be done to get a better understanding of the relationship, and further clinical investigations should be given to make the correlation more useful and reliable. The elevated levels of heavy metals may be associated with a number of physiological disorders in humans. It has been reported that cadmium is a mutagen in mammals, and that an elevated concentration of this element may result in prostate and lung cancers.[11,12] Messner et al. suggested that atherosclerosis might be initiated by cadmium.[13] In prostate cancer, a high concentration of iron and copper was reported.[14]

Classical biological materials are usually used in the diagnosis of various diseases. Blood and sometimes urine are widely used specimens for trace metals analyses. But the obtained information

does not reflect the actual concentration of elements in the whole body. This restriction is associated not only with the pharmacokinetics/toxicokinetics of metals, but also with the compensation abilities of the body. Thus, in order to obtain meaningful results, it is important to select a biological material which is free from changes caused by the body to maintain homeostasis.[15] Such opportunities are provided by alternative materials, including hair and nails. In contrast to blood and urine, analysis of hair and nails reflects the real concentration of elements in the whole body over several months. This retrospective information allows for the assessment of nutritional status, environmental and occupational exposure to metals, and body load.[16-19] Unfortunately, despite the many positive attributes of alternative materials, there are some difficulties associated primarily with the interpretation of the results. The biggest limitation is a properly defined range of reference values of concentration, especially in relation to hair. It is a consequence of the naturally occurring diversity of hair composition related to age, sex, ethnicity, latitude, natural conditions, the composition of the soil and water, characteristic for the region's diet and nutritional habits. Not without significance is also the difference in thickness, length or growth rate and hair color as well.[20] Despite these limitations, many attempts have been taken to use elemental analysis of hair in order to both assess the exposure to metals and for diagnosis of certain diseases, including cancer. Several studies have focused on the relationship between the level of trace metals in scalp hair and cancer in humans, but the results were inconsistent, thus prompting further investigations in this field.[21-23] Due to the difficulties connected with analysis of hair, in recent years simultaneous analysis of both hair and nails has become a very common laboratory practice.[17,24]

The aim of the study was the connection of the determination of selected metals: calcium (Ca), magnesium (Mg), zinc (Zn), copper (Cu), iron (Fe), manganese (Mn), cobalt (Co), chromium (Cr), cadmium (Cd), and lead (Pb), in two different materials: hair and nails, the diet and lifestyle survey, and the application of advanced statistical methods (logistic regression, canonical discriminant analysis, classification trees, artificial neural networks and support vector machines), in order to discriminate between healthy volunteers and patients diagnosed with larynx cancer, and then apply the results in the diagnosis of this disease.

Materials and methods

Study groups

The hair and nail samples for this study were collected from two groups: 68 patients with cancer of the larynx, and a group of 73 healthy volunteers. The cases considered in the present report were 63 men

and 5 women with confirmed laryngeal cancer. Thirty-eight people (71%) were located in the 40-59-year-old age group, while 30 (29%) exceeded 60 years. The patients were hospitalized in the Otolaryngology and Laryngological Oncology Clinic and Head and Neck Clinic and Oncological Laryngology of the Poznań University of Medical Sciences, Poland. The control group included 22 men and 51 women. In this group, 15 subjects (21%) were 20-39-years-old, 18 subjects were 40-59-years-old, and 40 (54%) were above 60 years old.

The protocol of the study was approved by the Bioethics Committee of the Poznań University of Medical Sciences, Poland. All subjects signed an informed consent form.

Questionnaire

All subjects completed a survey that assessed sociodemographic characteristics, lifestyle habits such as smoking and alcohol consumption, and intake of selected foods.

Chemicals

Reagents used in this study were of the highest purity available. Solutions were prepared with ultrapure, deionized water (18.2 MΩ·cm). Ultrapure, concentrated nitric acid (65%, Merck, Germany) was used for microwave-assisted digestion. Calibration solutions were prepared from multi-element ICP-MS Calibration Std 3 (Perkin Elmer, USA), and multi-element standard solution VI (Merck, Germany) served as a control solution. The daily test was performed with the use of Smart Tune Solution Std ELAN&DRC-e (Perkin Elmer, USA).

Sample collection and preparation

Seventy-three and sixty-eight samples of hair and nails were collected from healthy people and patients, respectively. The collection method and washing procedure were described in detail previously.[25] Digestion was performed in a microwave system (MARS 5X CEM, Matthews, USA). Seven mL of concentrated HNO_3 were added to about 0.1-0.2 g of the sample of hair or nails and left to pre-digest overnight at room temperature. After that the samples were subjected to a two-step digestion program. The digests were quantitatively transferred to volumetric flasks and diluted with deionized water. The digests were stored at 4°C until analysis.

Determination of metals

An ICP-MS instrument, Elan DRC-e (Perkin-Elmer, USA), was used for the determination of Cd, Co,

Cu, Fe, Mn, Pb, Zn, and the measurement of Ca and Mg was carried out by an Optical Emission Spectrometer (ICP-OES) Optima 2100DV, (Perkin-Elmer, USA). The certified reference material of human hair (GBW07601, GSH-1, China) and a lab-sample of nails prepared by grinding material collected from healthy volunteers were used for validation of the analytical method. Details about both methods and the validation parameters were described elsewhere.[25] Additionally, serum samples were also analyzed, but the results were not statistically significant and therefore excluded from statistical models.

Statistical methods

The statistical analysis was done with the two statistical software package. The Statistica 10 program (StatSoft, Poland) was used to perform the Mann-Whitney U test, and calculate Spearman correlation, and the rest of the calculations were performed with SAS® 9.4 software (SAS Institute Inc, Cary, NC).

A variety of statistical data mining procedures has been used, including logistic regression, canonical discriminant analysis, classification trees and artificial neural networks, in order to interpret the results from elemental analysis in a few different perspectives. Because there were numerous missing observations, two paths of analysis were chosen: with and without data imputation. Also, to evaluate and eliminate the possibility of the overfitting, the data pool has been randomly divided into two or three datasets. Two new datasets: train (80% of cases from original dataset) and test (20% of cases), were created when imputation was not applied. The train dataset was used to establish the model and estimate its parameters, and the test dataset was used to obtain an unbiased estimate of the misclassification rate resulting from application of that model. For the complete dataset (i.e. after imputation), the data was split into three instead of two parts: train (60%), validation (20%) and test (20%). The additional validation dataset was intended to be used with stepwise procedures. Stepwise procedures were applied whenever possible to pick important covariates. While generating candidate models, an inclusion criterion of $p<0.5$, and $p<0.05$ was assumed as a stay criterion. When a validation set was available, the misclassification rate was estimated with it for each candidate model, and the model with the lowest error count was chosen. In the absence of the validation set, the cross-validation procedure (leaving one out) was used. The complete (non-split) data were also fitted in another run to reveal additional factors which could be overlooked due to insufficient sample size. The factors potentially taken into account included binary (various ingredients of diet), ordinal (age group, cigarette smoking) and interval (concentration of elements in alternative materials) variables.

Canonical discriminant analysis. Information obtained from element analysis of hair and nails was also used to identify the association of the selected elements with cancer. The equation of the discriminant function was constructed using the stepwise forward method.

Logistic regression. A probability of larynx cancer in a pooled population of patients and controls was modeled using logistic regression. Stepwise regression was applied.

Logistic regression with interval data replaced by the corresponding principal components. The original set of interval factors (i.e. concentration of elements) was replaced by the set of principal components. This enables a relinquishing from correlations that existed in the original data, and creates an opportunity to reduce the number of necessary interval factors. These new variables, along with nominal and ordinal factors that were taken intact, formed input to the stepwise regression.

Classification trees. This technique yields results that have probably the most straightforward interpretation. Based on various statistical tests, minimal trees with satisfactory classification power are constructed.

Artificial Neural Networks (ANN). This method, which borrowed its concepts and methods from human nervous system physiology, proved to provide exceptionally flexible tools for approximation and classification.

There is, however, an increased risk of overfitting – the network may adapt itself to an actual data rather than to generally observed rules. It is also difficult to find a simple interpretation of the resulting network parameters and to indicate factors that most influence the result.

Support Vector Machines (SVM). This is yet another approach to machine learning. It shares is flexibility with ANN. It also potentially has the same deficiencies as ANN: lack of proper generalization because of the overfitting tendency, and inability to recognize important factors.

Results and discussion

Lifestyle and diet characteristics

The results of the surveys regarding smoking and drinking habits, and diet, are presented in Table 1 and 2, respectively.

The group of patients was characterized by a higher percentage of people heavily addicted to nicotine, who smoked over eleven cigarettes a day. Only 3% of the patients have never smoked, while in the control group the figure is 44%. Similar conclusions can be drawn from alcohol. In patients, significantly more people drank alcohol compared with healthy subjects. These data confirmed the fact that long-term exposure to tobacco smoke and alcohol abuse predisposes to the development of cancer of the larynx.[4,5] Simultaneously, exposure to them causes a multiplicative increase in risk.[26]

Acetaldehyde, which is produced from ethanol through bacterial and/or mucosal alcohol dehydrogenase and cytochrome P-4502E1, plays a significant role in developing cancer. Additionally, ethanol eases mucosal penetration by carcinogens contained in tobacco smoke.[27] Other factors that were considered in terms of the influence on the level of elements are diet and intake of supplements containing trace elements and vitamins. A defective diet may contribute to the development of cancer. Reiss *et al.*[28] suggested that increasing the amount of fruit and vegetables in the diet may reduce the risk of cancer. Comparing the consumption of meat in both groups, there are no significant differences in the amount of eaten poultry, beef and pork. It can therefore be assumed that the impact of micronutrients in this kind of food was similar in both groups. A similar situation existed with regard to whole grains and dairy products. In contrast, the only difference was in the amount of wheat germ and cornflakes. Definitely more people in the control group declared the presence of these products in their diet. Due to the fact that they are a rich source of magnesium, calcium, iron, zinc and copper, the concentration levels of these elements may be higher in a group of healthy patients, provided that they are consumed on a regular basis and for a long time[29]. On the other hand, Edefonti *et al.*[30], who analyzed the dietary patterns of 2452 head and neck cancer cases, and 5013 controls, concluded that a diet rich in cereals, animal products and monounsaturated, polyunsaturated and saturated fatty acids, and vitamin E, was positively associated with laryngeal cancer, and those rich in vegetables and fruits is negatively related to oral and pharyngeal cancer.

The elements concentration profile in hair and nails in the control and case groups

Nowadays, considering the enormous number of deaths from cancer, many epidemiologic studies have been undertaken to identify potential risk factors for this disease. Some attention has been paid to the association of cancer with trace elements.[31] The clinical application of metal analysis of hair has tried to investigate the association of some diseases/symptoms with trace bioelements (including essential and toxic) like autism,[32] other neurodegenerative and neuropsychiatric disorders,[33] and cardiovascular disease.[34] Many researchers have tried to evaluate the potential diagnostic values of metals to predict different types of cancer: lung,[35,36,37] breast,[38,39] prostate,[40] liver and stomach.[37]

In the present study, ten elements were chosen to be analysed in the hair and nails. The selection of essential metals (Ca, Cr, Cu, Fe, Mg, Mn and Zn) were based on the hypothesis that if they are involved in important metabolic functions, then both shortage and excess of trace elements leads to physiological disturbances. Heavy metals (Cd, Co, and Pb) are well-known carcinogenic agents whose role in the development of head and neck cancer has been extensively examined.[41] The median concentration of metals in the hair and nails of cancer patients and controls, along with the molar

concentration and interquartile range, are presented in Table 3. The mean levels are not shown because they are not representative of the centre of the groups (the distribution of the data is not symmetric). Statistical evaluation of the differences between the groups was based on the Mann-Whitney U test. The results of the serum analysis are not presented.

The biggest differences, when comparing the mean level of essential elements in the hair of the control group to those of the patients, were observed for Fe (4.7 times higher), Mg (4.6 times higher) and Ca (3.9 times higher), following by Mn (3.0 times higher), Cu (2.2 times higher) and Zn (1.5 times higher). Only Cr was present in 5.6-fold higher level in the patients than in the healthy subjects. In case of heavy metals, the mean concentration of Cd and Pb was 3.4- and 8.3-fold higher, respectively, in the patient group than in the controls, whereas the mean level of Co in the hair samples of the healthy volunteers was 7.0 times higher than that of the cancer patients. The comparison of the results to those reported by Cihan *et al.*[23] reveals some similarities. The level of Ca and Mg was higher in the controls than in the patients with breast cancer, whilst the concentration of Cd and Pb was lower. In the study by Qayyum and Shah[24] the healthy subjects had the increased concentration of Ca, Cr, Mg and Zn, whereas the patients with lung cancer – Pb, Cd, Co, Cu and Pb.

In case of the nails, the mean concentration of Ca, Mg, Cu, Fe and Mn was found to be 2.4-, 3.3-, 3.0-, 2.9-, and 2.6-fold higher, respectively, in the controls than in the patients, while the level of Cr and Zn was respectively 6.6- and 2.0-fold higher in the cancer group than in the control group. Two heavy metals were present in 6.4- (Cd) and 4.9-fold (Pb) higher concentrations in the nails of the patients than in the healthy subjects, and the mean level of Co was 3.0 times higher in the controls than in the cancer patients. Data on the metals content in cancer patients is scarce and not completely consistent with the results from present study. Qayyum and Shah[24] reported that the level of Ca, Cd, Co, Cr, Cu, Fe, Mn, Pb and Zn was increased in the cancer patients when compared to the control group, and only the concentration of Mg was decreased. Similar observations were made by Karimi *et al.*[42] – the concentration of Cu, Fe and Mn was higher in the cancer group, except for Zn.

The median concentration of all of the essential metals in the hair and nails (except for Cr in the hair and nails and Zn in the nails) was higher in the healthy donors in contrast to toxic metals (Cd, Pb), which were at a higher level in both biological materials in the cancer patients (with the exception of Co). Except for Zn, there was a statistically significant difference between the level of all metals both in the hair and nails of the cancer patients and controls ($p<0.05$). The basic interpretation of the results for the alternative materials provided a basis for the preliminary conclusion that it is possible to distinguish the studied populations with the use of elemental analysis.

The present study evidences that metabolism in cancer patients may be affected by toxic metals. Absorption and distribution of essential elements is regulated by a number of homeostatic processes, some of which are highly selective. Through the mechanisms of molecular and ionic mimicry heavy metals compete with essential elements for the site on the transporter proteins and enzymes.[43] For example, the divalent metal transporter 1 (DMT1), which is the major carrier of Fe, can also transport Cd, Co or Pb into the cells.[44] Lead enters the cell through the same plasma membrane ion channels as Ca and impairs its transport.[45] Interactions between toxic and essential metals are thought to be involved in the development of cancer. Carcinogenesis is the multistep process, which includes many mechanisms, such as DNA damage, dysregulations of cell signalling pathways, cellular proliferation and programmed cell death.[46] Metals, both essential and toxic, play different roles in these processes. Chromium(VI), in contrast to essential Cr(III), reacts with cellular reductants (e.g. glutathione), and forms Cr(V) and reactive oxygen species (ROS) which react with DNA.[47] In present study, the patients had the increased concentration of Cr in hair as well as Cd and Pb. Cadmium probably induces oxidative stress by replacing Cu and Fe in membrane proteins, what increases the concentration of unbound metal ions, capable of Fenton reactions, where ROS are generated.[48] Similar mechanism is proposed for Pb. This metal also inhibits the activity of δ-aminolevulinic acid dehydratase (ALAD), causing the increase in the level of δ-aminolevulinic acid (ALA), which promotes the generation of ROS.[49] Among studied heavy metals, only the level of Co was increased in the control group. Even though this element was proved to directly induce oxidative stress and react with DNA, it is classified as a Group 2B carcinogen by the International Agency for Research on Cancer (IARC) due to the lack of human data proving its cancerogenity.[50] In case of essential elements, they should be considered in the context of their deficiencies. Ions of Zn and Cu are a part of the enzymes forming the antioxidant barrier in the body, so a lack of these elements causes its weakening, and promotes oxidative stress.[51] Mahabir *et al.*[52] indicated that increased dietary intake of Cu and Zn decreased the risk of lung cancer, and Lappe *et al.*[53] suggested that high Ca intake may decrease the incidence of different cancers. Ions of this metal serve as messengers in many cellular processes (e.g. proliferation, apoptosis) that are disrupted by cancer.[54] Magnesium is a cofactor in enzymes involved in the replication and repair of DNA. There are many studies that show the level of Mg in diet is inversely correlated with the incidence of cancer.[55] Although Fe is essential for many cellular processes, as cell proliferation and DNA synthesis, it is believed that the excessive amount of metal might have a cancerogenic effect by the ability to produce ROS via Fenton reactions.[56,57]

Data mining

Apart from the fact that the analysis of elements in hair and nails gives an opportunity to discriminate both groups, the results can be used as variables, along with the information obtained from questionnaires, to construct statistical models. There are several reports which describe the employment of multivariate analysis of the concentration of trace elements for classification purposes.[34-40,58,59] However, all previous studies have taken into account only the level of elements in one biological material.

The results of all data mining procedures are compiled in Table 4. "None" in the column "Criterion of model choice" indicates that the model finally chosen by the stepwise procedure was used. The elements, which diversified the two groups the most, were Ca, Pb and Mg in hair and Mg, Zn, Cr and Ca in nails.

Logistic regression. The fitted equations of models are presented below:

Model no. 1: $\text{logit}(P) = 15.24 - 2.84\log(Ca_{hair}) + 2.87\log(Pb_{hair}) - 8.72diet_{cereal}$

Model no. 2: $\text{logit}(P) = -0.98 - 1.64\log(Ca_{hair}) + 1.91\log(Pb_{hair}) + 1.77\log(Pb_{serum})$

Model no. 3: $\text{logit}(P) = 9.60 - 2.27\log(Ca_{hair}) + 2.38\log(Pb_{hair}) + 1.31\log(Cr_{nails}) + 2.88cig_B + 5.17cig_C - 4.10diet_{cereal}$

Model no. 4: $\text{logit}(P) = 10.30 - 2.27\log(Ca_{hair}) + 2.39\log(Pb_{hair}) + 1.61\log(Cr_{nails}) - 4.40diet_{cereal}$

Model no. 5: $\text{logit}(P) = -2.53 - 2.38\log(Ca_{hair}) + 1.63\log(Pb_{hair}) + 1.51\log(Cr_{nails}) + 15.79age_B + 13.62age_C$

Model no. 6: $\text{logit}(P) = 8.68 - 1.81\log(Ca_{hair}) + 1.38\log(Pb_{hair}) + 1.12(Cr_{nails})$

The classification error of Models 1-4 for the test set ranged from 6.9% to 13.8%. Hernández-Caraballo *et al.*[58] achieved the sensitivity values ranging from 75% to 100% in the test set.

All models contained Ca and Pb in hair as important factors. Models with a greater data pool also included Cr in nails. A cereal diet appeared in three regression models.

The probability of cancer always decreased with an increase of Ca_{hair}. An increase of Pb_{hair} or Cr_{nails} caused a greater risk of cancer. Model 5 suggested that middle age (B) individuals are most likely to be ill, and Model 3 confirms the more cigarettes one smokes, the greater the risk of cancer is.

Decision trees. Classification trees (Models 7-9) are shown in Figures 1-3. Factor *diet_cereal* means that cereals were eaten at least once a week (path "A") or not (path "B"). Factor *cig_sum* is the number of cigarettes smoked per day (path "<B" – less than 20, path ">=B" – 20 or more).

The accuracy of Models 7 and 9 for test set was 79.3% and 83.9%, respectively. Chen *et al.*[60] successfully classified of normal and cardiovascular disease groups according to concentration of nine metal in blood/urine samples. Their classifiers performed with the accuracy of 97.6% for blood and 96.4% for urine.

Canonical discriminant analysis. The classification ability of the canonical variables is shown in Figures 4-7.

The classification error of the classifiers for the test set ranged from 6.7% to 17.1%. Most models contained Ca, Pb and Cu in hair and Zn or Cr in nails, although the simplest model only used Ca and Pb in hair. The only serum element that was included in one of the models was Pb.

Tan *et al.*[59] presented the classification power of linear discriminant analysis in the case of lung cancer. Both the FDA (Fisher's discriminate analysis) classifier and the ELDA (ensemble linear discriminant analysis) classifier achieved an accuracy of 92.9% for the test set.

Artificial neural networks. The perfect fit was reached for the train data, but the validation/test data revealed overfitting. The accuracy of Models 16 and 17 for the test set was 86.2% and 87.1%, respectively. Serum concentrations had to be excluded from the models without imputation (Models 17 and 18) due to numerous missing data. This way all subjects could be included in the study; otherwise the results proved to be extremely poor.

Artificial neural networks were successfully used by Hernández-Caraballo *et al.*[58] to distinguish both healthy and cancer groups on the basis of the concentration of Fe, Cu, Zn and Se in serum.

Support vector machines with linear kernel. This method performed similarly to ANN. The classification errors of Models 19 and 20 for the test set were 10.3% and 12.9%, respectively. Also, missing data had to be handled in the same manner as for ANN (serum concentrations excluded from Models 20 and 21).

Logistic regression of principal components. Principal components were built only for interval variables; i.e. concentration of elements. However, nominal and ordinal data were also taken into account in the regression. Among them, the cereal diet appeared to be the most important; it was present in all but one model (Model 28). Cigarette smoking was present in two models. Each model contained the first principal component (PC1). Other components that were used by at least one model were PC2, PC3, PC4 and PC8. No model used more than four components, and the simplest one used only PC1. The regression equations were not presented, because the concentrations of elements are hidden in principal components. A cereal diet, if present, always decreased the probability of cancer while smoking cigarettes worked in the opposite manner. The accuracy for the test set was 93.1% (Models 22 and 23) and 83.9% (Models 26 and 27).

Conclusions

The present study examined the usability of the statistical models constructed to identify the patients

with larynx cancer. We analysed the nails and hair samples, and apart from the concentration of 10 metals, we also considered factors connected with diet and lifestyle in statistical calculations. The results of our study showed that the distribution of nine from ten tested variables allowed the basic distinction between the two groups. The mean levels of all essential metals (Ca, Mg, Cu, Zn, Fe, Mn) are higher in the hair (except for Cr) and nails (except for Cr and Zn) of the normal donors when compared with the cancer patients, while toxic metals (Cr, Cd, Pb) showed the opposite pattern - except for Co. They were estimated to be higher in the hair and nails of the patients. Statistically significant differences in the level of all elements were found except for Zn.

Many statistical methods were evaluated as possible screening tools. Among the various data mining techniques used, logistic regression (Models 1 and 3-5), ANN (Models 16 and 17) and Model 13 of canonical discriminant analysis seemed to have the most balanced results between the train, validation (if present) and test datasets. The logistic regression proved to be most useful as it indicates significant factors and their role in the prediction of cancer probability. These factors were consistent with findings of other models. In most of them, Ca and Pb in hair (less often in nails) appeared. Calcium was negatively related to larynx cancer.

Our findings indicate that there is a relationship between cancer risk and elements, including toxic as well as essential metals. The obtained results suggest that using elemental analysis of hair and nails, information about lifestyle and diet, and appropriate statistic methods, it is possible to create a tool for estimating cancer risk and screening individuals. The constructed classifiers can be regarded as an useful addition to the common methods to diagnose laryngeal cancer.

Notes and references

1 R. Siegel, J. Ma, Z. Zou and A. Jemal, *CA Cancer J. Clin.*, 2014, **64**, 9-29.

2 M. Malvezzi, P. Bertuccio, F. Levi, C. La Vecchia and E. Negri, *Ann. Oncol.*, 2014, **25**, 1650-1656.

3 C. A. O'Brien, A. Kreso and J. E. Dick, *Semin. Radiat. Oncol*., 2009, **19**, 71-77.

4 E. P. Simard, L. A. Torre and A. Jemal, *Oral Oncol*., 2014, **50**, 387-403.

5 R. F. de Menezes, A. Bergmann and L. C. Thuler, *Asian Pac. J. Cancer Prev.*, 2013, **14**, 4965-4972.

6 E. S. Peters, M. D. McClean, C. J. Marsit, B. Luckett and K. T. Kelsey, *Cancer Epidemiol. Biomarkers Prev.*, 2006, **15**, 2196-2202.

7 W. Garavello, E. Lucenteforte, C. Bosetti, R. Talamini, F. Levi, A. Tavani, S. Franceschi, E. Negri and C. L. Vecchia, *Oral Oncol*., 2009, **45**, 85-89.

8 V. Wünsch Filho, *Sao Paulo Med. J.*, 2004, **122**, 188-194.

9 T. Psaltopoulou, R. I. Kosti, D. Haidopoulos, M. Dimopoulos and D. B. Panagiotakos, *Lipids Health Dis.*, 2011, **10**, 127.

10 O. A. H. Jones, D. A. Dias, D. L. Callahan, K. A. Kouremenos, D. J. Beale and U. Roessner, *Metallomics*., 2014, DOI: 10.1039/C4MT00123K.

11 G. Drasch, J. Schopfer and G. N. Schrauzer, *Biol. Trace Elem. Res*., 2005, **103**, 103–107.

12 R. Beveridge, J. Pintos, M.-É. Parent, J. Asselin and J. Siemiatycki, *Am. J. Ind. Med.*, 2010, **53**, 476-485.

13 B. Messner, M. Knoflach, A. Seubert, A. Ritsch, K. Pfaller, B. Henderson, Y. H. Shen, I. Zeller, J. Willeit, G. Laufer, G. Wick, S. Kiechl and D. Bernhard, *Arterioscler. Thromb. Vasc. Biol*., 2009, **29**, 1392-1398.

14 J. Guo, W. Deng, L. Zhang, C. Li, P. Wu and P. Ma, *Biol. Trace Elem. Res*., 2007, **116**, 257-271

15 H. Ka, *Eur. J. Clin. Invest*., 2011, **41**, 98-102.

16 M. Mikulewicz, K. Chojnacka, A. Zielińska and I. Michalak, *Environ. Toxicol. Pharmacol*., 2011, **32**, 10-16.

17 E. Oyoo-Okoth, W. Admiraal, O. Osano and M. H. S. Kraak, *Environ. Toxicol. Chem.*, 2012, **31**, 1461-1469.

18 J. P. Thyssen, A. Roeske-Nielsen and J. D. Johansen, *Contact Dermatitis.*, 2011, **65**, 125-137.

19 R. Mehra and M. Juneja, *J. Sci. Ind. Res*., 2005, **64**, 119-124.

20 K. Chojnacka, H. Górecka and H. Górecki, *Environ. Toxicol. Phar*., 2006, **22**, 52-57.

21 Q. Pasha, S. A. Malik, J. Iqbal, N. Shaheen and M. H. Shah, *Environ. Monit. Assess.*, 2008, **147**, 377-388.

22 Q. Pasha, S. A. Malik, N. Shaheen and M. H. Shah, *Clin. Chim. Acta*., 2010, **411**, 531-539.

23 Y. Benderli Cihan, S. Sözen and S. Oztürk Yıldırım, *Biol. Trace Elem. Res.*, 2011, **144**, 360-379.

24 M. Qayyum and M. Shah, *Biol. Trace Elem. Res.*, 2014, **158**, 305-322.

25 A. Przybyłowicz, P. Chęsy, M. Herman, A. Parczewski, S. Walas and W. Piekoszewski, *Cent. Eur. J. Chem*., 2012, **10**, 1590-1599.

26 K. D. Hunter, E. K. Parkinson and P. R. Harrison, *Nat. Rev. Cancer.*, 2005, **5**, 127-135

27 H. K. Seitz, F. Stickel and N. Homann, *Int. J. Cancer.*, 2004, **108**, 483-487.

28 R. Reiss, J. Johnston, K. Tucker, J. M. DeSesso and C. L. Keen, *Food Chem. Toxicol*., 2012, **50**, 4421-4427.

29 L. Stevenson, F. Phillips, K. O'Sullivan and J. Walton, *Int. J. Food Sci. Nutr.*, 2012, **63**, 1001-1013.

30 V. Edefonti, M. Hashibe, F. Ambrogi, M. Parpinel, F. Bravi, R. Talamini, F. Levi, G. Yu, H. Morgenstern, K. Kelsey, M. McClean, S. Schantz, Z. Zhang, S. Chuang, P. Boffetta, C. La Vecchia and A. Decarli, *Ann. Oncol*., 2012, **23**, 1869-1880.

31 S. N. Silvera and T. Rohan, *Cancer Causes Control.*, 2007, **18**, 7-27.

32 H. Yasuda and T. Tsutsui, *Int. J. Environ. Res. Public Health*., 2013, **10**, 6027–6043.

33 S. Pfaender and A. M. Grabrucker, *Metallomics*., 2014, **6**, 960-977.

34 C. Tan, H. Chen and C. Xia, *Biol. Trace Elem. Res.,* 2009, **129**, 9-19.

35 C. Tan, H. Chen and C. Xia, *J. Pharm. Biomed. Anal*., 2009, **49**, 746-752.

36 Y. Benderli Cihan and Oztürk Yıldırım, *Biol. Trace Elem. Res.,* 2011, **144**, 272-294.

37 Z. Zhang, H. Zhou, S. Liu and P. d. B. Harrington, *Chemom. Intell. Lab. Syst.*, 2006, **82**, 294-299.

38 Q. Pasha, S. A. Malik, N. Shaheen and M. H. Shah, *Biol. Trace Elem. Res.*, 2010, **134**, 160-173.

39 N. Gholizadeh, Z. Kabiri, O. Kakuee, M. Saleh-Kotahi, V. Changizi, V. Fathollahi, P. Oliaiy and R. Omranipour, *Biol. Trace Elem. Res.*, 2013, **153**, 105-110.

40 C. Tan and H. Chen, *Biol. Trace Elem. Res.,* 2011, **144**, 97-108.

41 R. Khlifi and A. Hamza-Chaffai, *Toxicol. Appl. Pharmacol*., 2010, **248**, 71–88.

42 G. Karimi, S. Shahar, N. Homayouni, R. Rajikan, N. F. A. Bakar and M. S. Othman, *Asian Pacific J. Cancer Prev.*, 2012, **13**, 4249–4253.

43 C. C. Bridges and R. K. Zalups, *Toxicol. Appl. Pharmacol.,* 2005, **204**, 274–308.

44 M. D. Garrick, S. T. Singleton, F. Vargas, H. C. Kuo, L. Zhao, M. Knöpfel, T. Davidson, M. Costa, P. Paradkar, J. A. Roth and L. M. Garrick, *Biol Res.*, 2006, **39**, 79-85.

45 E. J. Martinez-Finley, S. Chakraborty, S. J. B. Fretham and M. Aschner, *Metallomics*, 2012, **4**, 593–605.

46 A. Hartwig, *Free Radic. Biol. Med.*, 2013, **55**, 63–72.

47 M. Valko, C. J. Rhodes, J. Moncol, M. Izakovic and M. Mazur, *Chem. Biol. Interact.*, 2006, **160**, 1–40.

48 K. Jomova and M. Valko, *Toxicology*, 2011, **283**, 65–87.

49 M. H. Whittaker, G. Wang, X.-Q. Chen, M. Lipsky, D. Smith, R. Gwiazda and B. A. Fowler, Toxicol. Appl. Pharmacol., 2011, 254, 154–66.

50 D. Beyersmann and A. Hartwig, *Arch. Toxicol.*, 2008, **82**, 493–512.

51 C. G. Fraga, *Mol. Aspects Med*., 2005, **26**, 235–244.

52 S. Mahabir, M. R. Spitz, S. L. Barrera, S. H. Beaver, C. Etzel and M. R. Forman, *Int. J. Cancer*, 2007, **120**, 1108–1115.

53 J. M. Lappe, D.Travers-Gustafson, K. M. Davies, R. R. Recker and R. P. Heaney, *Am. J. Clin. Nutr*., 2007, **85**, 1586-1591.

54 G. R. Monteith, D. McAndrew, H. M. Faddy and S. J. Roberts-Thomson, *Nat. Rev. Cancer*. 2007, **7**, 519-530.

55 F. I. Wolf, J. A. M. Maier, A. Nasulewicz, C. Feillet-Coudray, M. Simonacci, A. Mazur and A. Cittadini, *Arch. Biochem. Biophys*., 2007, **458**, 24–32.

56 J. L. Heath, J. M. Weiss, C. P. Lavau, and D. S. Wechsler, *Nutrients*, 2013, **5**, 2836–2859.

57 S. Toyokuni, *Cancer Sci.*, 2009, **100**, 9–16.

58 E. A. Hernández-Caraballo, F. Rivas, A. G. Pérez and L. M. Marcó-Parra, *Anal. Chim. Acta.*, 2005, **533**, 161–8.

59 C. Tan, H. Chen and T. Wu, *Biol. Trace Elem. Res.*, 2011, **142**, 18-28.

60 H. Chen, C. Tan, Z. Lin, T. Wu and Y. Diao, *Comput. Biol. Med*., 2013, **43**, 865–869.

Table 1 Patterns of tobacco and alcohol use of cancer and control groups

		Controls (*n*=73) % (*n*)	**Cases (*n*=68)** % (*n*)
Smoking	Yes	25 (18)	62 (42)
	No	75 (55)	38 (26)
Number of cigarettes per day	0-5	57 (42)	4 (3)
	6-20	31 (23)	59 (40)
	>20	11 (8)	37 (25)
Drinking	Yes	42 (31)	65 (44)
	No	57 (42)	35 (24)
Number of drinks	1 drink per week	75 (55)	87 (59)
	2-5 drinks per week	9 (7)	12 (8)
	> 5 drinks per week	1 (1)	1 (1)
Type of alcohol	low-grade alcohol	86 (63)	56 (38)
	high-grade alcohol	14 (10)	44 (30)

Table 2 Dietary patterns of cancer and control groups

	Controls (*n*=73) % (*n*)		**Cases (*n*=68)** % (*n*)	
Eating at least once a week:	Yes	No	Yes	No
Poultry	90 (66)	10 (7)	91 (62)	9 (6)
Pork	67 (49)	33 (24)	88 (60)	12 (8)
Ham	90 (66)	10 (7)	87 (59)	13 (9)
Chicken liver	29 (21)	71 (52)	38 (26)	62 (42)
Pork or liver beef	22 (16)	78 (57)	34 (23)	66 (45)
Beef	36 (26)	64 (47)	46 (31)	54 (37)
Seafood	22 (16)	78 (57)	31 (21)	69 (47)
Corn flakes	44 (32)	56 (41)	12 (8)	88 (60)
Wheat sprouts	15 (11)	85 (62)	1 (1)	99 (67)
Buckwheat or oatmeal	51 (37)	49 (36)	44 (30)	56 (38)
Wholemeal bread	75 (55)	25 (18)	54 (37)	46 (31)
Eggs	92 (67)	8 (6)	88 (60)	12 (8)
White cheese	97 (71)	3 (2)	93 (63)	7 (5)
Spinach	27 (20)	73 (53)	28 (19)	72 (49)

Table 3 The concentration of metals (μg·g^{-1} dry weight tissue) in hair and nails of cancer patients (n=68) and healthy volunteers (n=73)

			Elements									
			Ca[a]	**Mg**[a]	**Cu**	**Fe**[a]	**Zn**[a]	**Mn**	**Co**	**Cr**	**Cd**	**Pb**
Hair	Control	Median	5.50	0.73	34.63	0.37	0.13	4.07	0.21	0.44	0.12	0.45
		Molar concentration[b]	137.23	30.42	0.55	6.63	1.99	0.07	0.004	0.008	0.001	0.002
		IQR	8.63	2.37	82.04	1.30	0.25	12.00	0.38	0.85	0.37	0.61
	Cancer	Median	0.76	0.10	19.37	0.09	0.15	1.29	0.03	1.40	0.48	1.94
		Molar concentration[b]	18.96	4.17	0.31	1.61	2.29	0.02	0.0005	0.03	0.004	0.009
		IQR	0.97[c]	0.19[c]	23.79[c]	0.17[c]	0.18	0.1.92[c]	0.05[c]	1.93[c]	1.51[c]	6.27[c]
Nails	Control	Median	2.27	0.69	15.78	0.22	0.11	3.09	0.06	0.64	0.06	0.24
		Molar concentration[b]	56.64	28.75	0.25	3.94	1.68	0.06	0.001	0.01	0.0005	0.001
		IQR	2.24	1.05	22.61	0.44	0.10	3.99	0.21	0.85	0.17	0.97
	Cancer	Median	1.32	0.24	9.89	0.08	0.12	1.15	0.02	2.85	0.28	0.91
		Molar concentration[b]	32.94	10.00	0.16	1.43	1.84	0.02	0.0003	0.05	0.002	0.004
		IQR	0.66[c]	0.02[c]	8.03[c]	0.08[c]	0.16	1.33[c]	0.06[c]	3.49[c]	1.08[c]	1.51[c]

[a][mg·g^{-1} dry weight tissue]; [b] [μmol·g^{-1} dry weight tissue]; [c] $p < 0.05$ with respect to control group; IQR - interquartile range

Figure captions

Fig. 1 Results for canonical discriminant analysis for Model 10

Fig. 2 Results for canonical discriminant analysis for Model 11

Fig. 3 Results for canonical discriminant analysis for Model 12

Fig. 4 Results for canonical discriminant analysis for Model 13

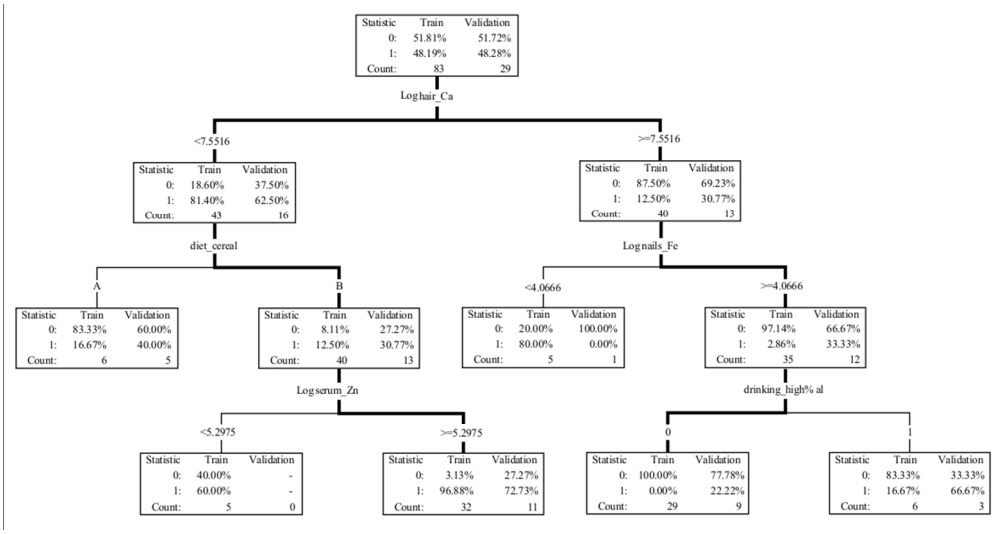
Fig. 5 Decision tree for Model 7

Fig. 6 Decision tree for Model 8

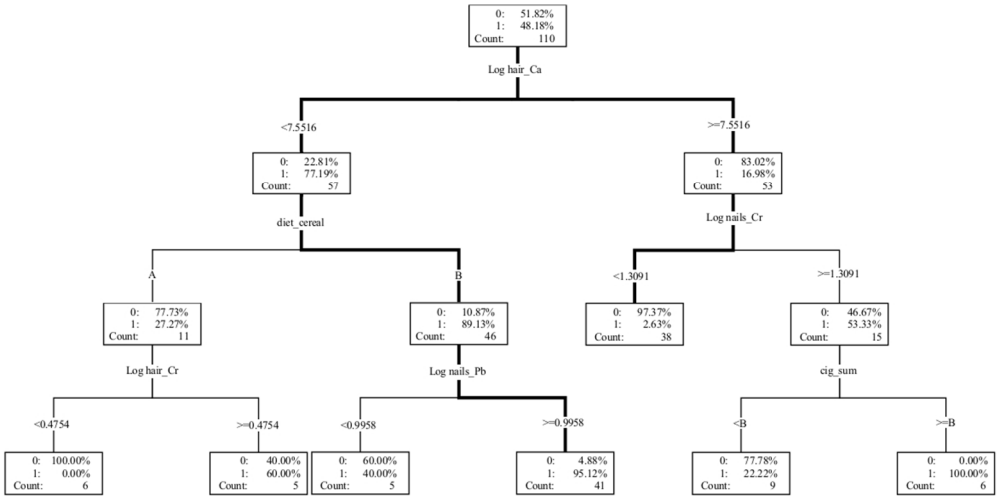
Fig. 7 Decision tree for Model 9

Table 4 Characterization of data mining models

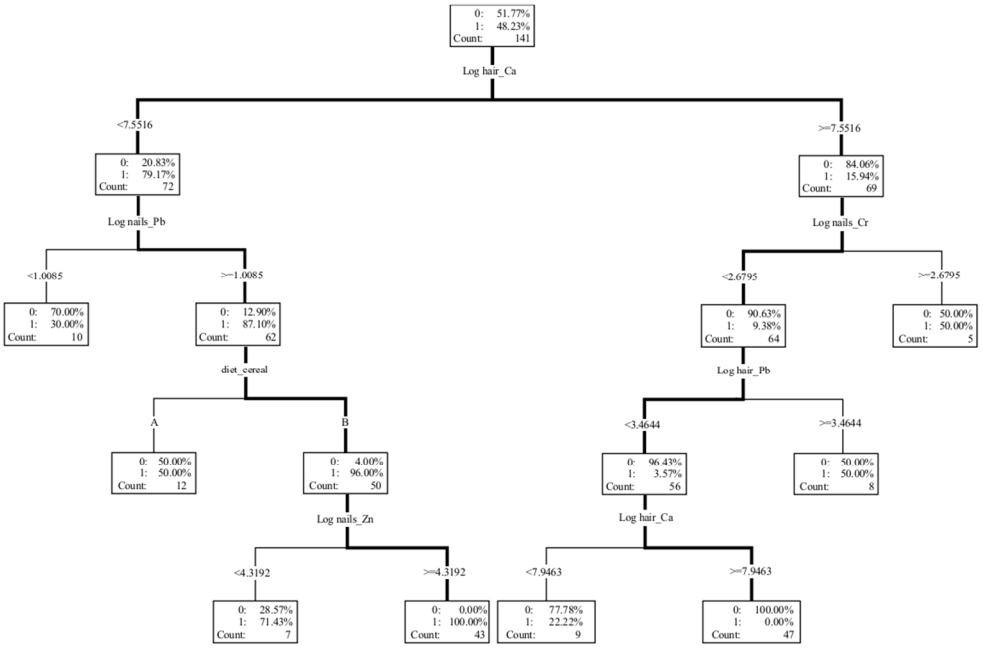
Method	Impu-tation	Model details			Misclassification [%] (dataset count)			
		No.	Criterion of model choice	Variables picked	Train	Valid.	Test	Cross. Valid.
Logistic regression	yes	1	Misclass. rate criterion	Ca and Pb in hair, cereal diet	4.8 (83)	17.2 (29)	13.8 (29)	-
		2	None	Ca and Pb in hair, Pb in serum	12.0 (83)	24.1 (29)	6.9 (29)	-
	no	3	Cross-validation	Ca, Pb in hair, Cr in nails, cereal diet, num. of cigar.	7.3 (110)	-	12.9 (31)	9.1
		4	None	Ca, Pb in hair, Cr in nails, cereal diet	9.1 (110)	-	12.9 (31)	-
		5	Cross-validation	Ca, Pb in hair, Cr in nails, age	8.5 (141)	-	-	10.6
		6	None	Ca, Pb in hair, Cr in nails	12.1 (141)	-	-	-
Decision trees	yes	7	Misclass. rate criterion	Ca in hair, Fe in nails, Zn in serum, cereal diet, alcoholic drinks	7.2 (83)	34.5 (29)	20.7 (29)	7.3
	no	8	Misclass. rate	Ca in hair, Pb and Cr in nails	12.1 (141)	-	-	14.3
		9	Misclass. rate	Ca and Cr in hair, Cr and Pb in nails, smoking, cereal diet	8.2 (110)	-	16.1 (31)	8.3
CDA	yes	10	Misclass. rate criterion	Ca and Pb in hair	9.6 (83)	20.5 (29)	6.7 (29)	9.6
		11	None	Ca, Pb, Cu, Mg in hair; Fe, Zn in nails, Pb in serum	4.7 (83)	34.5 (29)	6.9 (29)	6.0
		12	Cross-validation	Ca, Pb, Cu, Mg in hair; Fe, Zn in nails	4.7 (83)	34.5 (29)	17.1 (29)	5.9
	no	13	None	Ca, Pb, Mg, Cu in hair; Zn, Cr in nails	8.7 (92/110)	-	7.4 (27/31)	9.8
		14	Cross validation	Ca, Pb, Cu in hair, Cr in nails	8.9 (115)	-	-	9.6
		15	None	Ca, Pb, Cu in hair, Cr, Zn, Mg in nails	9.3 (115)	-	-	11.9
ANN	yes	16	Misclass. rate	-	0 (83)	13.8 (29)	13.8 (29)	-
	no	17	Misclass. rate	-	10 (110)	-	12.9 (31)	
		18	Misclass. rate	-	9.2 (141)	-	-	-
SVM with linear kernel	yes	19	Misclass. rate	-	0 (83)	24.1 (29)	10.3 (29)	-
	no	20	Misclass. rate	-	1.8 (110)	-	12.9 (31)	8.1
		21	Cross validate	-	2.8 (141)	-	-	10.6
PC with regression	yes	22	Misclass. rate	PC1,2,4; cereal diet	9.6 (83)	31.0 (29)	6.9 (29)	-
		23	None	PC1,4 cereal diet	7.2 (83)	37.9 (29)	6.9 (29)	-
	no	24	Cross valid.	PC1,2,3,4; cereal diet, number of cigar.	5.0 (141)	-	-	7.1
		26	Cross valid.	PC1,2,3,8; cereal. diet, number of cigar	3.6 (110)	-	16.1 (31)	9.2
		27	None	PC1,2,3,8; cereal diet	7.3 (110)		16.1 (31)	-
		28	Misclass. rate	PC1	25.5 (110)	12.9 (31)	-	-



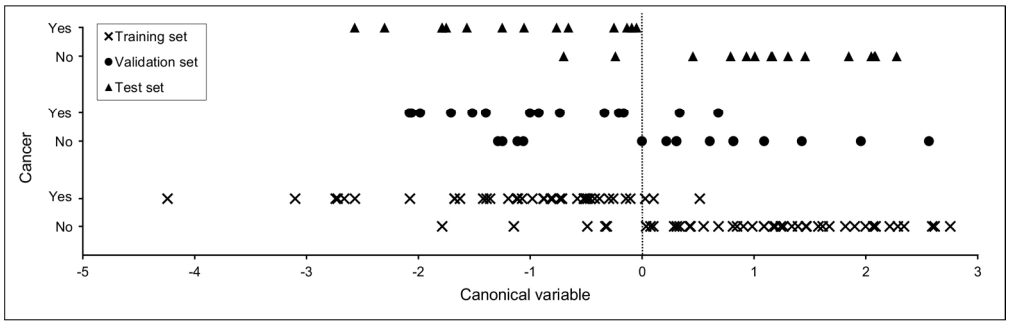
298x156mm (100 x 100 DPI)



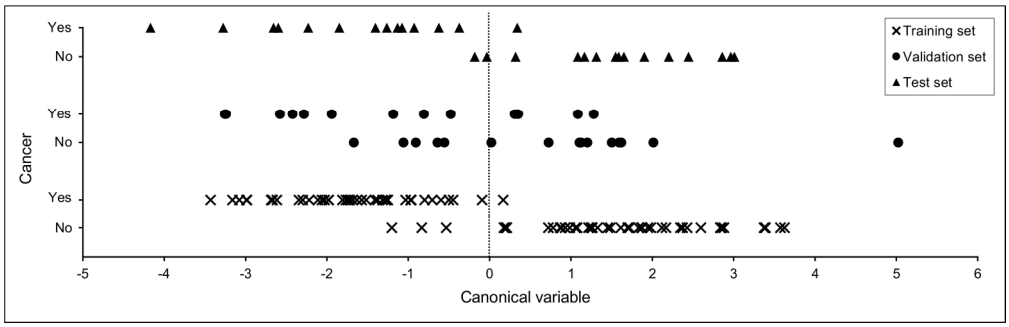
298x156mm (100 x 100 DPI)



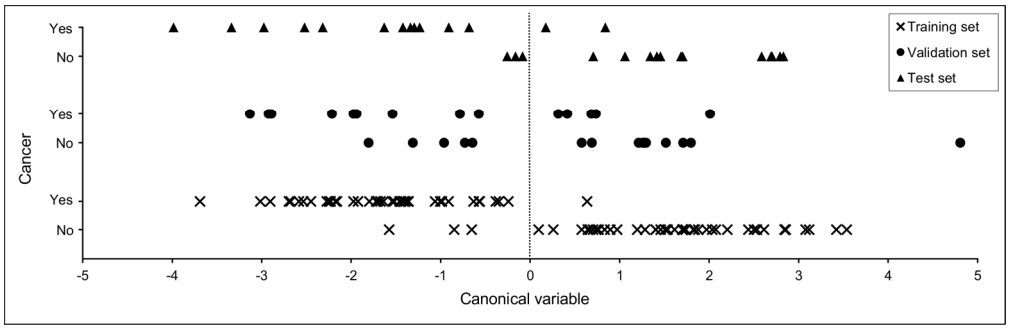
298x196mm (100 x 100 DPI)



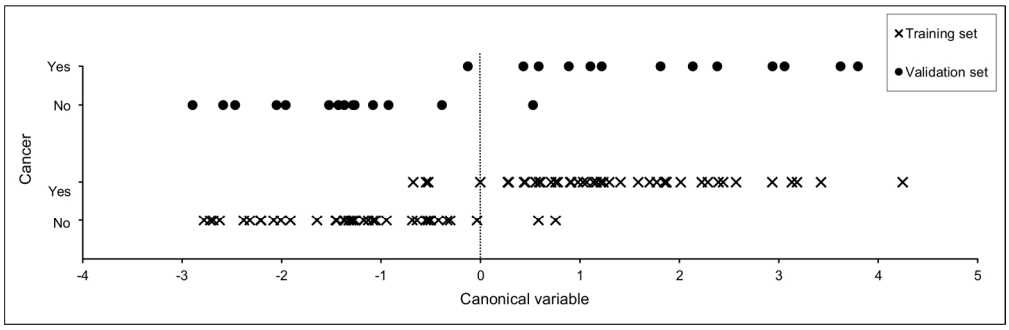
174x54mm (300 x 300 DPI)



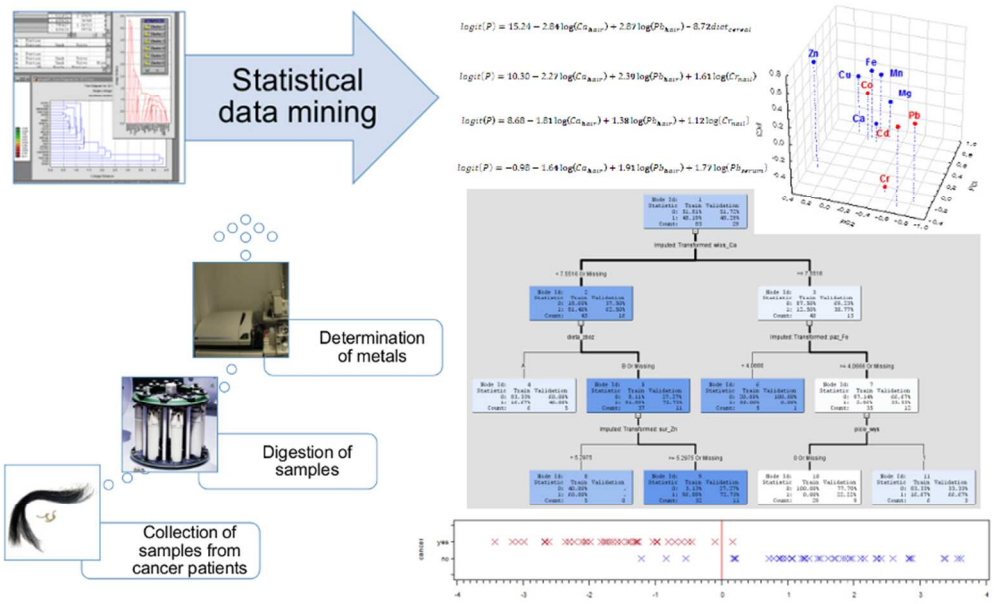
174x55mm (300 x 300 DPI)



174x55mm (300 x 300 DPI)



174x55mm (300 x 300 DPI)



Graphical abstract
248x152mm (96 x 96 DPI)