SUMMARY:
The aim of the present study is to reveal hidden relationships between bone density, biogenic elements content, and the level of oxidative stress of female patients with osteoporosis and osteopenia. Additionally, specific links between the patients are sought in order to interpret different similarity patterns of objects (patients) helping to a better understanding of the significance of the clinical variables for each identified similarity pattern.

Material and Methods: The input dataset consisting of 59 objects (patients) and 11 experimentally determined variables (clinical parameters) was subject to intelligent data analysis including cluster analysis (hierarchical and non-hierarchical mode) and factor analysis.

Results and Discussion: In the hierarchical dendrogram for clustering of 11 variables are formed 3 major clusters. We could assume that three factors (impact) are linked to the structure of the data set: descriptors responsible for osteoporosis diagnosis; descriptors (essential elements) related to osteoporosis status; descriptors related to the "overall health status impact". The patients are clustered in 3 clusters corresponding to 3 different levels of health status (improving, worsening and intermediate) by K-means clustering. The specific descriptors are defined for each identified cluster.

Factor analysis shows that 3 latent factors explain nearly 70 % of the total variance of the system × each of them with respective clinical meaning. A relationship is proven between T-score, diagnosis, and antioxidant activity by a 3D plot of factor loadings.

Conclusion: The multivariate statistical data interpretation for patients with osteoporosis problems reveals hidden relationships between specific similarity clusters among all patients or between the clinical parameters experimentally measured. It helps to better distinguish the variations between the specific groups and to determine the indicators for the variability. All this helps for more individual approach to medical treatment.

Keywords: osteoporosis, calcium, oxidative stress, multivariate statistics,
ric height, weight and body mass index (BMI). Increased weight (BMI > 27) in post-menopausal women is considered a protective factor against osteoporosis [4].

A comprehensive assessment of the impact of factors affecting bone health is crucial for the reliable prevention and treatment of osteoporosis. Early menopause and heredity are major risk factors for primary osteoporosis. Coffee, alcohol, smoking, and physical activity are factors with a proven role in bone health with an unclear mechanism [4-6].

When assessing bone health status, in addition to the established clinical indicators and lifestyle factors, the presence of hitherto unknown variables that contribute to the difficult control of the disease and its complications cannot be ruled out.

Copper, zinc, and iron are trace elements that play an important role in the human body, and some research teams are focusing their research on establishing their contribution to the development of osteoporosis [7-13].

There are a limited number of studies in the literature on the role of oxidative stress in the pathogenesis of osteoporosis. Their results are contradictory. According to some authors, bone marrow stem cells in women with osteoporosis have a higher capacity to fight free radicals than those in the control group of women without osteoporosis [14]. Other researchers report that BMD increases significantly with increasing plasma antioxidant capacity [15, 16].

More clinical studies are needed to establish the cellular and molecular mechanisms between oxidative stress, plasma antioxidant levels, and bone metabolism. Evidence of such a link is the significant reduction in plasma antioxidants with age. Oxidative stress alters the process of bone remodeling, causing an imbalance between osteoclasts and osteoblasts, and leading to bone resorption and pathogenesis of the skeletal system characterized by low bone mass [17]. We consider the Total Antioxidant Activity (AOA) indicator to be an appropriate tool for bone protection and a useful strategy against osteoporosis.

The condition of patients with osteoporosis depends on many factors simultaneously. The generally accepted medical approach for the use of established clinical indicators is to compare the value of each individual indicator in an attempt to decide on treatment. In such diseases, it is essential for this one-dimensional medical approach to change. All indicators and their ratios are taken together should be considered to determine health status, and their simultaneous interpretation should be used to make diagnostic decisions. This is possible through the application of methods for the classification and interpretation of multivariate statistics such as chemometric methods of analysis [18]. Chemometrics is underrepresented in medical research, despite there being examples of its specific role as a tool in medical statistics. It allows for specific relationships between the measured clinical parameters to be determined, and for different patterns of similarity between patients in the observation groups to be discovered. The information obtained contributes to a more differentiated approach in the treatment of the identified different groups of patients [19, 20].

The epidemiology and pathogenesis of bone mass reduction have been monitored by cluster analysis primarily to examine the relationships between fractures or the risk of fractures and bone density [21, 22]. The analysis has also been used successfully to create various models of pain progression in patients, which facilitates the determination of their condition [23].

We did not find data in the world practice for the application of multivariate statistical analysis to determine the relationship between bone density, nutrient content, and the level of oxidative stress. This motivates us to conduct an experimental-statistical study with menopausal and post-menopausal women in search of new factors, clarifying specific relationships between clinical indicators for the detection of key variables that increase the risk of developing osteoporosis.

**MATERIALS AND METHODS**

**Patients**

The experimental-statistical study included 59 menopausal and post-menopausal women. A retrospective analysis of the medical records of the diagnosed patients was performed. The women studied are of the Caucasian race and have no blood relations with each other. They had not received for osteoporosis.

Excluding factors were a parathyroid disease, the intake of estrogen and essential elements supplements.

**Clinical analyses**

The total levels of calcium, magnesium, copper, zinc, and iron in blood serum in all patients were measured by atomic absorption analysis with a spectrophotometer Perkin-Elmer AAnalyst 300. The quality of the results was guaranteed by applying internal quality control schemes (ICQ), as well as by the laboratory’s participation in EQAS external quality assessment programs.

The antioxidant status of blood serum samples was determined by the ABTS test [24]. The Shimadzu UV-1601 spectrophotometer recorded the change in the absorption of the chromophore introduced into the sample as a result of the interactions between free radicals and the antioxidants from the blood serum. The stable cation-radical of 2, 2’-azinobis (3-ethylbenzothiazoline-6-sulfonic acid) (ABTS + •) was used in the analysis.

**Clinical and laboratory parameters**

Clinical and laboratory parameters used for multivariate statistical analysis:

1. BMI, body mass index, kg/m²;
2. Age, years;
3. BMD, bone mineral density, g/cm². It was determined by the Dual X-ray Absorption (DEX A) Method.

4. T-score, the number of standard deviations from the norm for a healthy person at the age of 30 of the same sex. The T-score measurement is the ratio of the measured bone density compared to standard bone density, determined by measuring a large group of healthy 30-year-olds. For T-score > -1, bone density is considered to be stand-
ard. More negative values mean that the bones have a lower density than the standard. Osteopenia is diagnosed with T-score $-1 \div -2.5$, and T-score $<-2.5$ means osteoporosis.

3. Pen_Por, diagnosis of osteopenia or osteoporosis. Based on the result (T-Score), patients are divided into three groups: with osteoporosis (below -2.5 SD), osteopenia (between -1.0 and -2.5 SD); Normal density control group (above -1.0 SD)

6. Ca, mmol/L
7. Mg, mmol/L;
8. Cu, µmol/L;
9. Zn, µmol/L;
10. Fe, µmol/L;
11. AOA, total antioxidant activity, %

The multivariate statistical methods used for data mining and interpretations

Hierarchical cluster analysis (HCA)

Cluster analysis is a well-known multivariate statistical approach whose goal is to find patterns of similarity (clusters) in the dataset. These patterns are sought among the objects of the study (patients in the present study) as well as between the variables (clinical parameters). There are several obligatory steps in performing cluster analysis: data normalization (also known as z-transform, which turns the raw data into dimensionless units in order to avoid the dimensionality effect on the clustering process); selection of similarity measure (usually Euclidean distance as such a measure), which determine the spatial distance between the objects in a multidimensional space; choice of a linkage methods (very often Ward’s method) allowing to connect the close objects into clusters and to separate the distant objects from one another). The graphical output of the analysis is a tree-like plot called a dendrogram. All these steps are characteristic of the hierarchical mode of cluster analysis known as a typical unsupervised pattern recognition method (clustering without preliminary conditions). Finally, the significance of the clusters formed is determined (using the test of Sneath) [25].

K-means supervised clustering (K-means) - nonhierarchical clustering

K-means clustering is a typical supervised pattern recognition method, i.e., clustering according to a priori selection of the number of clusters which should contain all objects. The hypothesis for the preselected number of clusters is based on expert opinion or the need for confirmation of certain suggestion. Often K-means clustering is used to confirm or reject results from hierarchical cluster analysis.

Factor analysis

Factor analysis is also well documented chemometric approach leading to an explanation of the dataset structure and revealing latent (hidden) factors as new variable space directions able to replace the original variables. That is why the method is also known as the variable reduction or projection method [26]. It is based on the decomposition of the original data matrix into two new matrices known as factor loadings matrix (factor loading are related to the relative statistical significance of each old variable in the space of the new latent factors) and factor scores matrix (representing the new coordinates of the objects in the new space of the latent factors). These two matrices help in the interpretation of the similarity between the objects or variables in the reduced variable space. The number of the selected new latent factors depends mostly on the percentage of total variance explained by each latent factor. Usually, an explanation of a total 70% variance is accepted as normal.

All calculations are performed by the software package STATISTICA 8.0.

The input data set consist of 59 patients (objects) characterized by 11 clinical parameters (variables) or matrix with dimensions [59x11].

The goals of the multivariate statistical data interpretation are as follows:

- To detect patterns of similarity between the objects and between the variables;
- To search for specific parameters (descriptors) responsible for the formation of the patterns of objects similarity;
- To reveal hidden factors responsible for the data set structure.

RESULTS AND DISCUSSION

Hierarchical cluster analysis (HCA)

In Fig. 1 the hierarchical dendrogram for the clustering of 11 variables is presented.

Fig. 1. Hierarchical dendrogram for 11 variables

Three major clusters are formed:

- C1 (BMD, BMI, AOA, Age, Fe, Zn)
- C2 (T score, Pen_Por)
- C3 (Ca, Mg, Cu)

It could be assumed that C1 corresponds to the anthropometric impacts like BMI, BMD, Age, AOA) and to the levels of two essential blood components (Fe, Zn) – “overall health status impact”; C2 is obviously related to the parameters responsible for osteoporosis diagnosis and C3 – to the levels of essential elements related to osteoporosis status.
These results are confirmed by the application of K-means clustering which follows the hypothesis for the a priori existence of 3 clusters of variables.

The only difference to HCA is the numbering of the clusters. The membership to the cluster, however, is one and the same. Therefore, we could assume that three factors (impact) are linked to the structure of the data set:

- descriptors responsible for osteoporosis diagnosis;
- descriptors (essential elements) related to osteoporosis status;
- descriptors related to the “overall health status impact”.

When 59 patients (objects) are clustered with an a priori hypothesis of the formation of 3 clusters corresponding to 3 different levels of health status (improving, worsening and intermediate) by K-means clustering, the following patterns are formed:

Table 1. Members of clusters of objects (patients)

<table>
<thead>
<tr>
<th>Cluster 1 (12 members)</th>
<th>Cluster 2 (29 members)</th>
<th>Cluster 3 (18 members)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C_1</td>
<td>C_4</td>
<td>C_2</td>
</tr>
<tr>
<td>C_10</td>
<td>C_5</td>
<td>C_3</td>
</tr>
<tr>
<td>C_11</td>
<td>C_7</td>
<td>C_6</td>
</tr>
<tr>
<td>C_12</td>
<td>C_8</td>
<td>C_18</td>
</tr>
<tr>
<td>C_13</td>
<td>C_9</td>
<td>C_21</td>
</tr>
<tr>
<td>C_15</td>
<td>C_14</td>
<td>C_23</td>
</tr>
<tr>
<td>C_16</td>
<td>C_20</td>
<td>C_24</td>
</tr>
<tr>
<td>C_17</td>
<td>C_22</td>
<td>C_25</td>
</tr>
<tr>
<td>C_19</td>
<td>C_27</td>
<td>C_26</td>
</tr>
<tr>
<td>C_30</td>
<td>C_28</td>
<td>C_29</td>
</tr>
<tr>
<td>C_45</td>
<td>C_31</td>
<td>C_33</td>
</tr>
<tr>
<td>C_47</td>
<td>C_32</td>
<td>C_35</td>
</tr>
<tr>
<td>C_34</td>
<td>C_38</td>
<td></td>
</tr>
<tr>
<td>C_36</td>
<td>C_41</td>
<td></td>
</tr>
<tr>
<td>C_37</td>
<td>C_42</td>
<td></td>
</tr>
<tr>
<td>C_39</td>
<td>C_43</td>
<td></td>
</tr>
<tr>
<td>C_40</td>
<td>C_44</td>
<td></td>
</tr>
<tr>
<td>C_48</td>
<td>C_46</td>
<td></td>
</tr>
<tr>
<td>C_49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C_50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C_51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C_52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C_53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C_54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C_55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C_56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C_57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C_58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C_59</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It is very important to define the specific descriptors for each identified cluster and, thus, to understand the reason for the formation of the different patterns of patients.

In fig. 2 the plot of mean values of each variable for each identified cluster of patients.

Fig. 2. Plot of means of variables for each cluster (normalized values)

![Plot of Means for Each Cluster](image)

The number of each cluster on the plot corresponds to the number given in the tables with objects membership. This, cluster 1 has 12 members, cluster 2 – 29 members and cluster 3 – 18 members. The sequence of variables is: Ca, Mg, Cu, Zn, Fe, AOA, Age, T-score, Pen_Por, BMI, BMD.

Cluster 1 is characterized by the highest levels of Ca, Mg, Cu, lowest levels of Zn, Fe, AOA, BMI, BMD, and parameters Age, T-score and Pen-Por are close to those of members of cluster 2. It could be assumed that this pattern (phenotype) is of endangered osteoporosis patients with acceptable anthropometric indicators and well supported levels of essential osteoporosis elements.

Cluster 2 is the pattern with the most slightly negative osteoporosis impact – lowest levels of Ca, Mg, Cu, relatively high levels of T-score and Pen Por, high level of BMI and BMD. This group of patients requires higher attention and medical interventions (this is the biggest pattern of patients).

Cluster 3 patients are in an intermediate position with respect to Ca, Mg, Cu levels with very low T-score and Pen-Por indicators and still high BMI and BMD levels. This is probably the pattern not yet diagnosed with serious osteoporosis damages but requiring alert.

Factor analysis

In the table below the factor loadings of the variables for 3 latent factors are shown.
Three latent factors explain nearly 70% of the total variance of the system. Factor 1 (over 28% explanation of the total variance) is related to the levels of the essential elements responsible of the osteoporosis status (like cluster 3 in HCA or cluster 2 in K-means clustering). Factor 2 (over 20% of the total variance) could be conditionally named “osteoporosis diagnostic factor” (complete resembles results in both clustering procedures).

Factor 3 is the conditional “overall health status” (including anthropometric indicators) and resembles the clusters in HCA and K-means clustering.

It could be concluded that the data structure of the set is determined by three hidden factors each of them with respective clinical meaning.

In Figs. 3 and 4 these results of the factor analysis are completely confirmed. The grouping of the variables into three factors is well indicated both in 2D and 3D plots.

CONCLUSIONS
The major contributions of the present study could be summarized as follows:

- **Clarification of the data set structure** by identification of three latent factors, which determine over 70% of the total variance of the system; these variables reduction by principal components analysis makes it possible to define the most important factors affecting the system and to determine them conditionally as osteoporosis status factor (high factor loadings for mineral components like Ca and Mg), osteoporosis diagnosis factor (high factor loadings for osteoporosis diagnostic indicators like T_score, Pen-Por) and general health status factor (high factor loadings for anthropometric indicators like BMI, BMD, Age);

- **Identification of three patterns** of similarity (clusters) between the patients investigated; each one of
the clusters includes a certain number of patients, which supposes variations of the health status; this partitioning allows a more specialized medical approach to each specific group (cluster) depending on the clinical parameters characterizing each cluster of patients.

- Determination of specific indicators for each identified pattern of patients in order to reveal important markers for the different partitioning groups: cluster 1 has 12 members, cluster 2 – 29 members and cluster 3 – 18 members. Cluster 1 is characterized by the highest levels of Ca, Mg, Cu, lowest levels of Zn, Fe, AOA, BMI, BMD, and parameters Age, T-score and Pen-Por are close to those of members of cluster 2. It could be assumed that this pattern (phenotype) is of endangered osteoporosis patients with acceptable anthropometric indicators and well supported levels of essential osteoporosis elements. Cluster 2 is the pattern with the most negative osteoporosis impact – lowest levels of Ca, Mg, Cu, relatively high levels of T-score and Pen_Por, high level of BMI and BMD. This group of patients requires higher attention and medical interventions (this is the biggest pattern of patients) Cluster 3 patients are in an intermediate position with respect to Ca, Mg, Cu levels with very low T-score and Pen_Por indicators and still high BMI and BMD levels. This is probably the pattern not yet diagnosed with serious osteoporosis damages but requiring additional attention.

In general, the multivariate statistical data interpretation for patients with the potential danger of osteoporosis (ostepenia) problems reveals hidden relationships between specific similarity clusters among all patients or between the clinical parameters experimentally measured. It helps to better distinguish the variations between the specific groups and, additionally, to determine the indicators for the variability. All this helps for a more engaged and individual approach to medical treatment.

The trials to reach modeling of the relationship between, for instance, AOA and some other parameters like BMI, BMD, Ca, Mg, T_score, do not lead to reliable regression models and it hinders the option for prognostic conclusions.

REFERENCES:

7. Pepa GD, Brandi ML. Micronutrients for bone boost: The last but not the least. Clin Cases Miner Bone Metab. 2016 Sep-Dec;13(3): 181-185. [Crossref]
11. Okyay E, Erturgul C, Acar B, Sisman AR, Onural B, OzaksoyD. Comparative evaluation of serum levels of main minerals and post-menopausal osteoporosis. Maturitas. 2013 Dec;76(4):320-5. [Crossref]
mentation of Ascorbic Acid and Bone Mineral Density. Nutrients. 2021Mar 21;13(3):1012. [Crossref]


DOI: https://doi.org/10.5272/jimab.2022284.4742

Received: 10/10/2022; Published online: 19/12/2022

Address for correspondence:
Radka Tomova
Faculty of Pharmacy, Medical University of Pleven,
Address:1, Kliment Ohridski Str., Pleven 5800, Bulgaria.
E-mail: rtomova@mail.bg,