Non-Tuberculous Mycobacteria As Environmental Risk Factors For Human Infectious Diseases – Isolation Of Non-Tuberculous Mycobacteria From Patients In The Thessaly Region (Central Greece) And Their Correlation With Sarcoidosis Disease

E Dovriki, F Bardaka, I Gerogianni, E Petinaki, K I Gourgoulianis, A Papaioannou

Citation

DOI: 10.5580/IJMB.52861

Abstract
Mycobacteria are organisms that are of great interest in the field of medicine, particularly those causing tuberculosis and leprosy (e.g., mycobacterium tuberculosis). Conversely, the nontuberculous mycobacteria (NTM) that have been found to be widely dispersed in the environment are considered potentially pathogenic for humans and animals. Recently, NTM found in patients were correlated with sarcoidosis disease. The aim of the present study is to identify atypical mycobacteria, isolated from various samples of patients, and examine their possible correlation with sarcoidosis disease, for the period of 2010 to 2016. A total of 296 positive NTM patients (60% male) have been included in the study. Seventeen different NTM species were identified from the studied biological samples. The main species distribution among patients were as follows; M. fortuitum (n=107); M. gordonae (n=61); M. chelonae (n=44); M. intracellulare (n=27); M. avium (n=16) and M. abscessus (n=13). At the same period of time, a total of 124 sarcoidosis cases have been recorded from the same geographical area. The results of the present study showed that indeed there is a large dispersion of positive NTM patients in the environment. In addition, regarding the geographical distribution of positive NTM patients and patients with sarcoidosis disease in the same study area, there is evidence of a possible correlation between them.

1. INTRODUCTION
It is known worldwide that mycobacteria, which cause tuberculosis and leprosy (e.g. M. Tuberculosis), are not detected in the environment and are transmitted from human to human. Today, mycobacteria are of great interest to medicine industry because of human migration.

On the other hand, nontuberculous mycobacteria (NTM) that have been found to be widely dispersed in the environment are considered potentially pathogenic for humans and animals. Although NTM cause diseases in animals (e.g. M. avium - intracellulare causes disease in poultry and pigs), animal to human transmission is rare. Until today, more than 160 different species of NTM have been isolated and identified, while this number is increasing continuously due to improved determination and identification techniques of biological and environmental samples [Dantec et al. 2002; Epson and Einthrop 2012]. In the scientific literature, they are referred to as potential pathogens principally to immunosuppressed and immunocompetent patients (e.g. transplant recipients, AIDS and cancer patients), as well as to vulnerable groups such as children, chronically ill, and the elderly. Generally, populations at risk include individuals who suffer from lung disease and a weakened immune system. [Archibald and Jarvis 2011; Billinger et al. 2009; Epson and Einthrop 2012; Gerogianni et al. 2008; Gitti et al. 2011; Moore et al. 2010; Moorthy et al. 2012; Penn et al. 2011; Roux et al. 2009; Russell et al. 2014; Chou et al.]

DOI: 10.5580/IJMB.52861
Only few sporadic studies have been published, mainly showing that NTM cause lymphadenitis and infections to humans, mostly in the developed world, with increased frequency not geographically delimited. This increased incidence may be attributed to the interest obtained by the researchers, mainly due to AIDS patients with NTM infections. It is estimated that 25-50% of patients with AIDS have been infected by NTM [Dantec et al. 2002; Falkinham III 2009 and 2011; Gerogianni et al. 2008; Gill et al. 2011; Gitti et al. 2011; Moore et al. 2010; Griffith et al. 2007; Iseman et al. 2008; Jolobe 2011; Kaevska and Hruska 2010; Kankya et al. 2011; Lyu et al. 2011; Maekawa et al. 2011].

Another area of interest proved to be the study of possible correlation between the presence of NTM in patients and sarcoidosis disease, whereas recent research data suggest that there is a correlation between them [Ainslie and Benatar 1985; Olive and Kataria 1985; James 1986; James 1992; Oksanen 1994; Kon and Du Bois 1997; Sharma 1997; Am J Respir Crit Care Med. 1999; Dumouchel-Champagne et al. 2009; Keijzers et al. 2009; Costabel et al. 2010; Iannuzzi and Fontana 2011; Tavee and Culver 2011; Drent et al. 2012; Sobic-Saranovic et al. 2013; Blankstein et al. 2014; Cremers et al. 2014; Gupta et al. 2014; Mortaz et al. 2014].

Multivariate statistical methods are useful tools to examine possible correlations between different parameters and factors, as they give results when classical correlation analysis methods cannot. The use of these methods to investigate the association between NTM isolated in humans with other parameters has been considered crucial and necessary. Specifically, the use of Cluster Analysis and Factor Analysis to analyze large data sets in medicine research has proven valuable, giving results that the classical statistical analysis methods are not able to trace [Massart and Kaufman 1983; Vandergrinst et al. 1998; Poupard et al. 2002; Vogt and Nagel 1992; Nguyen et al. 2003; Ness et al. 2005; Papaioannou et al. 2010].

In the present study, multivariate statistical methods have been used to examine possible correlations among the parameters “NTM” isolated from human patients, “SEX” of positive NTM patients, “CLINIC” that came the positive NTM biological sample, “BIOSAMPLE” the kind of positive NTM biological sample and “RESIDENCE” of positive NTM patients”. Furthermore, for the first time in Greece, the present study investigates the possible correlation between the NTM isolated from patients and the cases of sarcoidosis disease, in the Thessaly region, Greece.

2. METHODS

2.1 General

Methods that were used for digestion and decontamination of clinical samples to recover from M. tuberculosis have also proved to be useful for the NTM. After sample preparation (liquidation, homogenization, decontamination and neutralization), staining and microscopy take place. Once decontaminated, samples may be grown on ordinary solid media for mycobacteria such as egg-based Lowenstein-Jensen (LJ), agar based medium (Middlebrook) or selective media (LJ Gruft; Difco Laboratories, Oxford, UK). Liquid media cultures include the automated Bactec MGIT 960 [Dantec et al. 2002; Gitti et al. 2011].


2.2 NTM cases

The study included patients admitted to the University Hospital of Larissa, from whom at least one biological sample was tested culture-positive for NTM. The study was approved by the Ethics Committee of the University of Thessaly. All patients were informed and provided their written consent.

NTM isolates were recovered from respiratory specimens, urine, sterile body fluids, and tissues. Multiple identical isolates related to one single individual, who was living at the same site and who was hospitalized during the same period, were counted as a one patient entry.

A total of 302 NTM species were isolated in the department of Microbiology, University Hospital of Larissa. All isolates included in the study were recovered from 296 patients who lived in the Thessaly region, central Greece, during a 7-year
period from January 2010 to December 2016. Subsequently, the medical records of these patients were reviewed, aiming to identify their locality of residence and relevant clinical characteristics. It was found that these individuals lived in 16 different locations in the region of Thessaly. The majority of them lived in the cities of Larissa, Trikala, Karditsa and Volos. The geographical distribution of these locations is presented in Figure 1.

**Figure 1**
Map of Thessaly region with the studied localities (municipalities) (Prefecture of Larissa: 1 = Agia; 2 = Elassona; 3 = Kileler; 4 = Larissa; 5 = Tempi; 6 = Tirnavos; 7 = Farsala. Prefecture of Karditsa: 8 = Karditsa; 9 = Palama; 10 = Sofades; 17 = Mouzaki. Prefecture of Trikala: 11 = Meteora; 12 = Pyli; 13 = Trikkaios; 14 = Farkadona. Prefecture of Magnesia: 15 = Almyros; 16 = Volos; and 18 = Riga Feraios).

### 2.3 Clinical specimens processing
Clinical specimens were stained by Ziehl-Nielsen and were grown as well. In more detail, a volume of 0.5 ml was inoculated into a Bactec MGIT 960 tube, while 0.25 ml were inoculated onto a commercially available Lowenstein-Jensen medium (BioMerieux, France). The Lowenstein-Jensen medium was incubated at 370°C and routinely examined for up to 6 and 8 weeks respectively.

All positive cultures were subsequently analyzed by the Genotype CM commercial kit (HainLife Science, Germany), a molecular genetic assay for identification of M. tuberculosis complex and 15 of the most common NTM species (Mycobacterium avium, Mycobacterium chelonae, Mycobacterium abscessus, Mycobacterium fortuitum, Mycobacterium gordonae, Mycobacterium intracellulare, Mycobacterium scrofulaceum, Mycobacterium interjectum, Mycobacterium kansasi, Mycobacterium malmoense, Mycobacterium peregrinum, Mycobacterium marinum / Mycobacterium ulcerans, Mycobacterium xenopi and MTBC). This essay is a recently developed commercial DNA-strip one about differentiating mycobacteria isolates from cultured media.

### 2.4 Sarcoidosis Cases
The study included 124 patients admitted to the University Hospital of Larissa and diagnosed with sarcoidosis disease for the same time period. The study was approved by the Ethics Committee of the University of Thessaly. All patients were informed and provided their written consent. Figure 1 presents the place of residence of positive NTM patients as well as the number of the sarcoidosis cases in Thessaly region for the period 2010 - 2016.

### 2.5 Statistical Analysis by Chemometric Methods
Two major classes of chemometric methods can be distinguished. The first comprises the methods concerned with the recognition of patterns (e.g. linear discriminant analysis). The second class of methods aims at the detection of patterns yet unnoticed in the field of medicine. A single pattern may be noticed, recognized as a highly abnormal one. Or a group of patterns belonging to a major category of patterns may be delimited, because the patterns of this group are in some sense distinct from all other patterns of the category – they form a “cluster”.

Cluster Analysis (CA) and Factor Analysis (FA) were used for multivariate statistical modeling of the input data. The calculation work was performed by the use of software package STATISTICA 10.0 for Windows.

#### Cluster Analysis (CA)
CA is a data reduction method that is used to classify entities with similar properties. The method divides a large number of objects into a smaller number of homogeneous groups on the basis of their correlation structure. The objective of CA is to identify the complex nature of multivariate relationships (by searching for natural groupings or types)
among the data under investigation, so as to foster further hypothesis development about the phenomena being studied. CA imposes a characteristic structure on the data analysis for exploratory purposes [Massart and Kaufman 1983; Vanderginste et al. 1998; Poupard et al. 2002; Vogt and Nagel 1992; Ness et al. 2005; Papaioannou et al. 2009].

B. Factor Analysis (FA)

FA is used to find associations between parameters, so that the number of measured parameters can be reduced. Known associations are then used to predict unmeasured parameters [Vanderginste et al. 1998; Poupard et al. 2002; Nguyen et al. 2003; Papaioannou et al. 2009].

FA is often used in data reduction to identify a small number of factors that explain most of the variance observed in a much larger number of manifest variables. Principal Component Analysis (PCA), as one subset of FA, extracts eigenvalues and eigenvectors (a list of loadings) from the covariance matrix of original variables to produce new orthogonal variables known as VariFactors (VFs), usually through Varimax rotation, which are linear combinations of the original variables. VFs provide information about the most meaningful parameters that describe a whole data set allowing data reduction with minimum loss of original information. Moreover, VFs include unobservable, hypothetical, latent variables.

3. RESULTS

The source of the NTM isolates were respiratory specimens from the hospital’s pulmonary clinic (53%), samples from the pathological clinic (30%), the rheumatology clinic (8%), and the gastrointestinal clinic (1.0%), while the remaining of the samples came from other clinics, presented in Figure 2.

Figure 2
Frequency of NTM positive patients per clinic

In the NTM isolates from the clinical specimens (Figure 3), the most frequently isolated organisms were M. fortuitum (n=107, 36%), M. gordonae (n=61, 21%), M. chelonae (n=44, 15%); M. intracellulare (n=27, 9%), M. avium (n=16, 5%) and M. abscessus (n=13, 4%). Additionally, M. tuberculosis was isolated in 1 patient. It is noteworthy that in 6 patients isolated 2 NTM species simultaneously.

Figure 3
Frequency of NTM species found in clinical specimens (1 = M. lentiflavum; 2= M. abscessus; 3 = M. avium; 4 = M. chelonae; 5 = M. fortuitum; 6 = M. gordonae; 7 = M. intracellulare; 8 = M. kansasii; 9 = M. arupense; 10 = M. conceptionense; 11 = M. peregrinum; 12 = M. smegmatis; 13 = M. chimaera; 14 = M. mucogenicum; 15 = M. yongonense; 16 = M. simiae; 17 = Mycobacterium spp; 18 = M. canariasense; 19 = M. avium & M. intracellulare; 20 = M. chelonae & M. gordonae; 21 = M. avium & M. chelonae; 22 = M. fortuitum & M. chelonae).

3.1 Box and Whisker plot analysis
When Box & Whisker plot analysis is contacted for all data set (Figure 4), it can be seen that the parameters “BIOSAMPLE”, “NTM” and “RESIDENCE” correlated each other.

**Figure 4**

Frequency of NTM species found in clinical specimens (1 = M. lentiflavum; 2 = M. abscessus; 3 = M. avium; 4 = M. cheloneae; 5 = M. fortuitum; 6 = M. gordonae; 7 = M. intracellularre; 8 = M. kansasii; 9 = M. arupense; 10 = M. conceptionense; 11 = M. peregrinum; 12 = M. smegmatis; 13 = M. chimaera; 14 = M. mucogencicum; 15 = M. yongonense; 16 = M. simiae; 17 = Mycobacterium spp; 18 = M. canariasense; 19 = M. avium & M. intracellularre; 20 = M. cheloneae & M. gordonae; 21 = M. avium & M. cheloneae; 22 = M. fortuitum & M. cheloneae).

**3.2 Cluster Analysis – Factor Analysis**

The studied parameters of the data set were:

- The “Sex” of the patient
- The “Biosample” the kind of biological sample from positive NTM patients
- The “Residence” of positive NTM patients
- The “NTM” that are the species of nontuberculous mycobacteria found in patients

**3.2.1 Implementation of CA and FA to all data set**

CA is conducted to all data set and in that case the dendrogram presented in Figure 5 is produced.

**Figure 5**

Hierarchical dendrogram for the five studied parameters of all positive NTM patients.

The produced dendrogram grouped the studied parameters into two main clusters, as follows:

- **1st Cluster** (two parameters included): “BIOSAMPLE”.
- **2nd Cluster** (three parameters included): “RESIDENCE”, “SEX” and “NTM”.

Consequently, CA results indicated that the parameters “NTM”, “SEX” and “RESIDENCE” correlated to each other.

Then FA is conducted to the whole data set and in Table 1(a, b) the loadings of the extracted new VFs of unrotated and varimax normalized factors are included. Finally, according to screeplot criterion, three eigenvalues are considered significant.
Table 1
Factor Loadings (Un = Unrotated and Var = Varimax normalized), Eigenvalues, and Total variance (%).

(a) When two (2) extracted Factors

<table>
<thead>
<tr>
<th>Parameter</th>
<th>V1 (Un)</th>
<th>V1 (Var)</th>
<th>V2 (Un)</th>
<th>V2 (Var)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEX</td>
<td>0.730629</td>
<td>0.512083</td>
<td>0.114005</td>
<td>0.472647</td>
</tr>
<tr>
<td>BIOSAMPLE</td>
<td>-0.134500</td>
<td>0.081920</td>
<td>-0.334582</td>
<td>0.112638</td>
</tr>
<tr>
<td>NTM</td>
<td>0.548855</td>
<td>0.907372</td>
<td>-0.304105</td>
<td>0.433437</td>
</tr>
<tr>
<td>RESIDENCE</td>
<td>0.527584</td>
<td>0.815768</td>
<td>-0.248338</td>
<td>0.090905</td>
</tr>
</tbody>
</table>

Eigenvalues: 1.362951
Total (Cumulative) variance (%): 31.61, 25.36 (76.97)

(b) When three (3) extracted Factors

<table>
<thead>
<tr>
<th>Parameter</th>
<th>V1 (Un)</th>
<th>V1 (Var)</th>
<th>V2 (Un)</th>
<th>V2 (Var)</th>
<th>V3 (Un)</th>
<th>V3 (Var)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEX</td>
<td>0.730629</td>
<td>0.512083</td>
<td>0.114005</td>
<td>0.472647</td>
<td>0.254300</td>
<td></td>
</tr>
<tr>
<td>BIOSAMPLE</td>
<td>-0.134500</td>
<td>0.081920</td>
<td>-0.334582</td>
<td>0.112638</td>
<td>-0.072800</td>
<td>0.059314</td>
</tr>
<tr>
<td>NTM</td>
<td>0.548855</td>
<td>0.907372</td>
<td>-0.304105</td>
<td>0.433437</td>
<td>0.500000</td>
<td>0.042567</td>
</tr>
<tr>
<td>RESIDENCE</td>
<td>0.527584</td>
<td>0.815768</td>
<td>-0.248338</td>
<td>0.090905</td>
<td>-0.781305</td>
<td>0.994074</td>
</tr>
</tbody>
</table>

Eigenvalues: 1.362951
Total (Cumulative) variance (%): 31.61, 25.36, 23.84 (79.80)

Data of Table 1 shows that there is a strong moderating correlation of the parameters “SEX”, “NTM” and “RESIDENCE” with the Factor 1 and this is in full agreement with the results of CA.

3.3 Cluster Analysis (K-Means Clustering)

The goal of the k-means algorithm of CA is to find the optimum “partition” for dividing a number of objects into k clusters. This procedure moves objects around from cluster to cluster with the goal of minimizing the within-cluster variance and maximizing the between-cluster variance.

CA (K-Means Clustering) for all data set was conducted to find the most significant parameters and the number of the formed clusters. The number of formed clusters and their parameters are presented below:

- 1st Cluster: “SEX”.
- 2nd Cluster: “BIOSAMPLE”, “NTM” and “RESIDENCE”.

The data above shows that 2 clusters were formed; the 1st cluster included the parameter “SEX” while the 2nd cluster included the parameters “BIOSAMPLE”, “NTM” and “RESIDENCE”.

3.4 Principal Components and Classification Analysis

Based on Cluster Analysis (K-Means Clustering) results, Principal Components and Classification Analysis for all data set were performed. All the parameters were considered active parameters. Table 2 presents the factor loadings and the eigenvalues of the extracted factors, which explained the 100.0% of total variation.

Table 2
Factor coordinates of the variables (Factor-variable correlations (factor loadings)).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
<th>Factor 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEX</td>
<td>0.730629</td>
<td>0.114005</td>
<td>0.254300</td>
<td>0.072800</td>
</tr>
<tr>
<td>BIOSAMPLE</td>
<td>-0.134500</td>
<td>-0.334582</td>
<td>0.500000</td>
<td>0.042567</td>
</tr>
<tr>
<td>NTM</td>
<td>0.548855</td>
<td>-0.304105</td>
<td>0.500000</td>
<td>0.433437</td>
</tr>
<tr>
<td>RESIDENCE</td>
<td>0.527584</td>
<td>-0.248338</td>
<td>-0.781305</td>
<td>0.994074</td>
</tr>
</tbody>
</table>

Using the data of Table 2, it could shown that Factor 1 included three variables (“SEX”, “RESIDENCE” and “NTM”) which correlated moderately to each other. The second extracted factor included the variable “BIOSAMPLE” that correlated weakly to moderately with “NTM”; the variables “RESIDENCE” and “NTM” correlated moderately to each other in third factor, and lastly, in the fourth factor they showed a moderately to weakly correlation between the variables “SEX” and “NTM”. Additionally, the contribution of each variable in the extracted factors is presented in Table 3.

Table 3
Variable contributions per extracted Factor.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
<th>Factor 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEX</td>
<td>0.375846</td>
<td>0.010465</td>
<td>0.010465</td>
<td>0.549307</td>
</tr>
<tr>
<td>RESIDENCE</td>
<td>0.225420</td>
<td>0.063786</td>
<td>0.558867</td>
<td>0.05193</td>
</tr>
<tr>
<td>BIOSAMPLE</td>
<td>0.107971</td>
<td>0.697017</td>
<td>0.003971</td>
<td>0.261942</td>
</tr>
<tr>
<td>NTM</td>
<td>0.529555</td>
<td>0.224428</td>
<td>0.819508</td>
<td>0.238959</td>
</tr>
</tbody>
</table>

Data of Table 4 shows that the variables “SEX”, “RESIDENCE” and “NTM” mainly contributed in Factor 1; the variable “BIOSAMPLE” with “NTM” in Factor 2; Factor 3 contributed from variables “RESIDENCE” and “NTM”; and the contributed variables in Factor 4 are “SEX” and “NTM”. Figure 6 shows how the studied variables of the extracted factors were located in the space.
Figure 6
Projection of the studied variables on the factor-plane.

In the above graphical presentation (Figure 6) when the variables were located near each other in space it meant that there is a correlation among them. Thus, when we see the plots of Factor 1 vs Factor 2, Factor 1 vs Factor 4 and Factor 2 vs Factor 4, the variables “NTM” and “RESIDENCE” are plotted nearly in space, showing that these variables were correlated to each other.

DISCUSSION

Only a few studies have attempted to look into the impact of nontuberculous mycobacteria on human’s health and specifically in Greece there are very few. Research studies in the city of Patra for the period 1998 to 2000 (Peloponnese) showed that the types of nontuberculous mycobacteria isolates in patients were M. flavescens, M. gordonae, M. chelonae and M. Terrae; research studies in central Greece for the period 2004 to 2006 showed that the NTM species most commonly isolated in patients from this area were M. fortuitum (51.6%) and M. peregrinum (34.5%). In the study carried out in Crete during the period from 2000 to 2009, the most common NTM species that were isolated from patients were the MAC complex (M. avium and M. intracellularare) (8.6%) and M. kansasii (3.4%). In research study carried out in Athens during the period from 2007 to 2013, M. avium (13%), M. intracellularare (10%), M. gordonae (14%) and M. fortuitum (12%) were isolated. Finally, in a study carried out in the region of Larissa, the most common NTM species that were isolated were the M. fortuitum (30.8%), M. gordonae (22.7%) and M. peregrinum (12%), while the percentage of M. avium and M. intracellularare was 2.1% and 1.8%, respectively [Vantarakis et al. 1998; Tsintzou et al. 2000; Gerogianni et al. 2008; Gitti et al. 2011; Panagiotou et al. 2014; Dovriki et al. 2016].

From the Figure 1 it can be seen that positive NTM patients are dispersed widely in the Thessaly region and this is in full agreement with the literature. Moreover, the distribution of NTM species worldwide varies by geographic regions. Recent studies in the USA, Japan and other counties worldwide show that the most frequently reported NTM species that mainly affect middle aged or elderly patients are M. avium, M. fortuitum and M. kansasii [Gerogianni et al. 2008]. These differences in results are mainly due to the different methods used for the determination and identification of NTM species as well as the different composition of the population of studied patients, and finally the regional variations in environmental conditions.

A total of 296 patients (60% male) of the University Hospital of Larissa were tested positive for NTM during the period 2010 to 2016. The permanent residence of these patients was the region of Thessaly (central Greece) and specifically 18 geographical locations there (Figure 1). The majority of the positive NTM patients lived in the cities, while the rest was distributed over the whole areas of the Thessaly region (Figure 1).

As it is shown in Figure 2, the biological samples that were studied came mainly from the pulmonary (53%) and pathological (30%) clinic of the mentioned hospital. This is in line with other studies which report that the most common source of biological samples taken to isolate NTM is the pulmonary clinic [Archibald and Jarvis 2011; Billinger et al. 2009; Epson and Winthrop 2012; Gerogianni et al. 2008; Moore et al. 2010; Iseman and Marras 2008; Lyu et al. 2011; Maekawa et al. 2011; Panagiotou et al. 2014; Primm et al. 2004; Chou et al. 2014; Dovriki et al. 2016].

Overall, for the aforementioned set term, the nontuberculous mycobacteria identified were M. fortuitum (36%), M. gordonae (21%), M. chelonae (15%); M. intracellularare (9%).
M. avium (5%) and M. abscessus (4%). It is noteworthy that M. tuberculosis was isolated in one patient and in six patients we isolated simultaneously 2 NTM species (Figure 3). According to the scientific literature, the presence of NTM is the cause for the false (positive or negative) diagnosis of tuberculosis [Gerogianni et al. 2008].

Recently, the study of possible correlation between the presence of NTM in humans and sarcoidosis disease became of great interest. Thus, there is research data arguing that a correlation between NTM and sarcoidosis may exist [Kon and Du Bois 1997; Dumouchel-Champagne et al. 2009; Mortaz et al. 2014]

In the present study, the geographical distribution of sarcoidosis cases in the Thessaly region gives us indications that there may be a correlation between NTM and sarcoidosis disease. Additionally, multivariate analysis methods have been used to identify possible correlations between the presence of NTM in patients and other parameters concerning them. For this purpose, Box & Whisker plot analysis, Cluster Analysis, and Factor Analysis were performed.

The Box & Whisker plot analysis showed that there is correlation among the parameters “BIOSAMPLE”, “NTM” and “RESIDENCE” (Figure 4). However, there is no correlation of them with the studied parameter “SEX”. When CA was conducted to all data sets, the produced dendrogram (Figure 5) showed that NTM isolated from patients correlated with the parameters “SEX” and “RESIDENCE” while there is no correlation with the parameter “BIOSAMPLE”. Both, the "Box & Whisker plot analysis" and the "Cluster analysis" proved that there is a correlation between the parameters "NTM" and "RESIDENCE". Furthermore, the FA results confirmed the CA results reinforcing their validity. Concerning FA results, when the parameters located closer to each other in space, it could be concluded that they were more closely correlated to each other. Thus, Figure 6 shows that the parameters “NTM” and “RESIDENCE” are indeed correlated to each other because these parameters were located close to each other in space.

Additionally, CA (K-Means Clustering) was applied on the whole data set and two clusters were generated. The most significant parameters for this clustering were the parameters “BIOSAMPLE”, “NTM” and “RESIDENCE”, grouped together into the second cluster, indicating that there is a correlation among them.

Subsequently, Principal Component Analysis (PCA) and Classification Analysis were performed to the whole data set. Three new factors were extracted which explained the 79.7% of the total variance (Table 2). Specifically, the factor coordinates of the parameters “NTM” and “RESIDENCE” are around 0.50, thus concluding to the fact that they are moderately interrelated. Finally, this interrelation is shown in Figure 6 where the parameters are close to each other in the space, indicating the existence of their interrelation. The results show a moderate correlation between the parameters “NTM” and “RESIDENCE” and this fully comply with the results of CA Analysis and FA Analysis.

The results of the present study support the argument that NTM are widely dispersed in the environment. Moreover, there is evidence that the presence of NTM in patients may be correlated with sarcoidosis disease. However, because of the differences observed in the results of the relevant studies, further studies are needed to determine the significance of NTM in human health in Greece and worldwide, and to find if there is a correlation of them with sarcoidosis disease.

4. CONCLUSION

Multivariate statistical methods, like Box & Whisker plot analysis, CA, and FA were used to discover possible correlations among the NTM found in patients and other parameters “SEX”, “BIOSAMPLE” and “RESIDENCE” concerning these patients. Box & Whisker plot analysis showed that there is correlation among the parameters “BIOSAMPLE”, “NTM” and “RESIDENCE”. CA results indicated that there is a correlation between the parameters “NTM” and “RESIDENCE”. Subsequently, FA was conducted on the same data set and the results of CA were confirmed by the FA results. In addition, CA (K-Means Clustering) was applied to the same data set confirming the correlation between these parameters once again. Furthermore, based on the results of CA (K-means clustering), Principal Components and Classification Analysis we were able to show showed that the two above mentioned parameters are indeed correlated.

Consequently, these three multivariate statistical methods (i.e. Box & Whisker plot analysis, Cluster Analysis and Factor Analysis) have shown that the “NTM” found in patients and the “RESIDENCE” of these patients correlate to
each other and proved that NTM are widely dispersed in the environment. In addition, based on the geographical distribution of NTM and sarcoidosis cases in the region of Thessaly, we can assume that there is evidence of a possible correlation between them.

References

Environmental Risk Factors for Pulmonary Mycobacterium avium-intracellulare Complex Disease. CHEST, 140 (3):723-729.
Non-Tuberculous Mycobacteria As Environmental Risk Factors For Human Infectious Diseases – Isolation Of Non-Tuberculous Mycobacteria From Patients In The Thessaly Region (Central Greece) And Their Correlation With Sarcoidosis Disease

Author Information

Eleni Dovriki
Department of Respiratory Medicine, Medical School, University of Thessaly
Larissa, Greece

Fotini Bardaka
Department of Respiratory Medicine, Medical School, University of Thessaly
Larissa, Greece

Irini Gerogianni
Department of Respiratory Medicine, Medical School, University of Thessaly
Larissa, Greece

Efi Petinaki
Department of Microbiology, Medical School, University of Thessaly
Larissa, Greece

Kostas I. Gourgoulianis
Department of Respiratory Medicine, Medical School, University of Thessaly
Larissa, Greece

Agelos Papaioannou
Department of Medical Laboratories, Technological Education Institution of Thessaly
Larissa, Greece